DOCUMENT RESUME

ED 330 409 JC 910 198

AUTHOR Tiffany, Gerald E.; And Others
TITLE Student Learning Outcomes Assessment for English 101 at the Wenatchee Campus of Wenatchee Valley College: The Relationship of Student Outcomes and Rater Consistency in English 101, Winter 1990 and Spring 1990.
INSTITUTION Wenatchee Valley Coll., Wenatchee, Wash.
PUB DATE 5 Mar 91
NOTE 70p.
PUB TYPE Reports - Research/Technical (143) -- Tests/Evaluation Instruments (160) -- Statistical Data (110)

EDRS PRICE MF01/PC03 Plus Postage.
DESCRIPTORS Community Colleges; Comparative Analysis; *Curriculum Evaluation; *Essay Tests; Evaluation Methods; *Interrater Reliability; *Outcomes of Education; Tables (Data); Two Year Colleges; Two Year College Students; *Writing (Composition); *Writing Evaluation; Writing Skills
IDENTIFIERS Wenatchee Valley College WA

ABSTRACT

In 1991, a student learning outcomes assessment was conducted at Wenatchee Valley College, Washington. All English 101 students in the winter and spring quarters of 1990 wrote a 2-hour final exam. Winter quarter students wrote on the same topic while spring quarter students wrote on one of three randomly assigned topics. Five English 101 instructors evaluated the 262 student essays. Each essay was rated by at least two instructors. During the winter rating session, the second rater was able to see the first rater's evaluation. The purposes of the assessment were to evaluate the relationship between exam and course performance; the degree of student learning consistency between class sections; the consistency in faculty ratings of the same essays; the effect of viewing previous rater's scores on the subsequent rater's scoring; and the dimensions of English instruction where students performed the best and where performance (or instruction) needed improvement. Study results included the following: (1) course grades were not necessarily predictive of performance on the exam; (2) a high degree of consistency was found in 12 of 14 course learning objectives across all course sections; (3) students' scores and faculty ratings were not affected by having one essay topic assigned to all students, or one of three topics randomly assigned; (4) knowledge of a previous rater's score did not affect the subsequent rater's evaluation; and (5) students performed best in the area of essay organization and most poorly in the area of mechanics. Evaluative criteria, survey instruments, and detailed data tables and graphs with analysis are appended. (JMC)

STUDENT LEARNING OUTCOMES ASSESSMENT FOR ENGLISH 101 AT THE WENATCHEE CAMPUS OF WENATCHEE VALLEY COLLEGE: The Relationship of Student Outcomes and Rater Consistency in English 101, Winter 1990 and Spring 1990

By

Gerald E. Tiffany
Edward Chenevert
Nancy D. Howard

March 5, 1991

Table of Contents

OVERVIEW

The WVC Composition 101 instructors undertook a student learning outcomes assessment (SLOA) project that began its planning stage in November of 1989 and finished data collection in late June of 1990. Analysis of these data commenced in November of 1990 and continues through the present.

Our interest in student outcomes lies in what current students here at our college are taking with them from English 101. Thus, our methodology was not so much based upon the experimental paradigms of pure research and formal hypothesis casting as it was a fishing expedition in which we hoped to capture and measure student writing.

Two hundred and sixty-two students served as subjects, roughly half in each quarter of Winter 1990 and Spring 1990. Five English 101 instructors, comprised of full-time and part-time faculty, rated essays the students wrote during a two-hour finals period.

This SLOA project produced results in a variety of areas:

- o Students' grades in English 101 do not necessarily relate to students' scores on a Departmental Exam.
- o Overall, we have excellent consistency in English 101 objectives. Twelve of 14 sections scored very similarly.
- o The English faculty was able to agree upon a definition of student competency as described in an evaluation criteria scoring sheet.
- o Faculty scoring of student writing in a Departmental exam is extremely consistent.
- o Students in the Winter and Spring terms scored substantially the same in both administrations of the exam.
- o Students' scores and faculty's ratings are apparently not affected by having one topic for a group at large or three topics randomly assigned.
- o Knowledge of a previous rater's scoring of an essay did not effect a rater's subsequent evaluation.
- o A departmental exam can offer writing students another source of feedback regarding their writing skills.
- o Students write best (or teachers teach best) in the dimension of organization.
- o Students perform worst (or teachers teach worst) in the dimension of mechanics.

HISTORY AND RATIONALE

During the Fall and Winter terms of 1989-1990 the writing instructors from several divisions (Humanities, Developmental Education, High School Completion, and Adult Basic Education) met to compare their course syllabi and discuss mutual concerns. First, we wanted a sense of unity and identity as instructors working in the same pursuit, albeit at different locations on campus and under the headings of varying divisions. Second, we felt that a more complete understanding on our parts of content and objectives from diverse courses would help us better to serve students fed from several sources into the two-tiered general composition requirements.

Most students in academic and vocational programs are required to take English 101. Students working toward a transferable Associate's Degree in Arts and Sciences take one of three two-hundred level courses: English 202 Composition (Essay), English 203 Composition (Literary Analysis), or English 204 Composition (Research). The students choose the second tier of composition based upon the requirements of the institution to which they intend to transfer and their proposed major coursework.

During our final meeting in that academic year we discussed exit exams in English 101. This was a rather heated discussion. Some instructors vigorously opposed such an exam, on the reasonable basis that students taught to revise and refine essays in two week blocks could not perform adequately in a two-hour period. Others voiced concerns that an exit exam forced teaching to a test at the cost of skimming over or rushing through important matters of style. Some present at that meeting favored the prospect of checking curricular consistency and instructional effectiveness: Were students studying under different faculty members leaving their classes at the end of the term with similar and sufficient levels of writing skill?

Some present at that meeting were concerned with consistency of faculty evaluation: Would different faculty members here at WVC judge student writing similarly? Or would faculty members judge essays from disparate evaluative criteria? More important, could this group even agree on a set of evaluative criteria? Still others present at the meeting were simply interested in data that would reflect what must assuredly be minimum student competencies, given that an examination at the end of the course would only be two hours in length.

The point here is important: while our SLOA project examined student writing outcomes, it evolved from a very complex set of human agendas, opinions, and interactions. Hence, this SLOA project also explored FIPAR (Faculty Interactive Processes and Relationships). In abstract terms, SLOA and FIPAR cannot help but intersect at some point. For several of us, therefore, the final value of this project was based as much upon what we learned about fellow faculty members and our varying philosophies of English instruction as it was based upon what we learned about assessing student learning outcomes.

Any large SLOA project requires tremendous communication and mutual assistance between the project coordinators.

We decided to run a Departmental Exam, as opposed to an exit exam. Very large colleges, offering dozens of composition sections, must often rely upon exit exams to ensure that students entering the more advanced compositions courses possess the skills one might expect of a graduate of English 101. Since we are a tiny department (two and two-thirds full-time and two part-time faculty members), this was not a concern. Thus, our project would study our students' writing rather than police our teachers' competencies. Further, we decided that instructors would count each student's score on this departmental final in whatever form best suited each instructor. In some sections, the student's score was worth as much as one-sixth of the course grade; in other sections the score counted as less than one-tenth of the course grade.

Our first step was a formal proposal submitted to our associate academic dean. This proposal outlined several research questions and a few ideas regarding SLOA project procedures. Later in the SLOA process we came upon more questions our methodology and data might answer, and so the task of analysis became (for poor, inumerate English teachers) increasingly more complex as time went by. The proposal was soon accepted . Our second step was to receive the permission of our Instructional Council to change the final examination schedule so that almost all of the English 101 students would write their exams at the same time and (as much as possible) the same location. For the sake of procedural consistency we thought it best to test all of our students in equivalent environments and conditions. Our Instructional Council was amenable to the changes we proposed, and our planning continued. We held faculty meetings, formed schedules, held a norming session, tested students in the Winter, made two important procedural changes, and tested students again in the Spring. We debriefed one another, coded our data, and elicited the assistance of a statistician.

SLOA PROJECT QUESTIONS

Instruction and Evaluation

1. Will student success over the quarter predict student success on the Departmental Final? In other words, is there correlation between students' scores on the test and students' grades in the English 101 sections? Looked at another way, this question might read: Do teachers evaluate in regular assigned essays the same criteria with the same rigor that they evaluate on a Departmental Exam?

2. Do we have departmental consistency in our objectives? If we teach from identical objectives, then there should be no great differences between the scores for different sections. While we share a common WVC syllabus and objectives, instructors take students along varying paths to reach the common department goal of student writing competencies. Are the writing sections similar enough in their scores to indicate that several roads lead successfully to Rome?

3. Can the Composition faculty agree to a definition of student competency? Can the faculty become consistent in their evaluations of student writing? Overall, on which dimension (Organization-O, Development-D, Mechanics-M), do faculty most agree in their ratings? On which dimension do they least agree? Each essay was rated by two evaluators. In those cases when our scores differed by four or more points, a third reader rated the essay. Do we have a good level of agreement? Will we most agree among ourselves in the Winter or in the Spring? In other words, will we gain or lose consistency over time? If we disagree significantly, what will cause that disagreement?

4. In the Winter quarter all of the students wrote on the same topic. In the Spring quarter students wrote on one of three topics randomly assigned. Does this different procedure affect students' writing or faculty's evaluation? Do Departmental Exams need to have only one topic to be consistent in their measures and fair to the students?

5. During the Winter rating session, the second rater was able to see the first rater's evaluation. During the Spring session, we masked the first rater's score in order to reduce the evaluation effect. Will that masking create a result? Will there be more or less consistency in raters' evaluations?

Student Learning Outcomes

6. This concern is best phrased as a statement rather than a question. We would like to assure students of some objectivity in evaluation by having their writings read by more than one instructor. We hope that students will then be more at ease with their own learning outcomes. We want them to know that the skills they develop are not just the whims or fantasies of some idiosyncratic instructor of English. We want them to know that their developing writing skills reflect the way that people in the real world (or at least other English 101 sections) really do write. The student will learn that more than one teacher really does perceive that he has strong skills in one area or lacks skill in another.

7. Are student outcomes in writing different in different academic terms? Do students write better in one quarter than another? This study collects data in two quarters. We are interested to see if there are differences.

8. On which dimension (Organization, Development, Mechanics) do students at WVC best perform? On which do they not perform as well? This is an important matter. Armed with answers to these questions, we can better serve our students. This study should provide data to show us both strengths and weaknesses in student learning outcomes, and indicate the directions we teachers need to take in our English 101 classes.

PROCEDURE

Students

All English composition students in the Winter and Spring quarters of 1990 wrote a two hour exam during a scheduled English 101 final exam period. The 262 students are for the most part local to the Wenatchee service district. WVC averages 97% of its student body from "feeder" highschools--but of varying graduation dates. Some students attend WVC in order to take advantage of particular academic, vocational, and extra-curricular programs. Our students are, in large part, native English speakers. Like students in most community colleges, they are a heterogeneous lot, representing diverse ages, personal histories, and educational goals.

Entrance to English 101 is limited to those students who have scored sufficiently high on the ASSET examination or have passed English 81, Basic Writing Techniques. The ASSET measures students' skills in grammar and syntax, reading, and math. The English portion of ASSET is a timed, objective-type grammar-based test, using test forms and machine scoring. Currently the English department of WVC does not use diagnostic essay exams to determine students' writing readiness . English 81 prepares students for success in English 101. English 81 emphasizes grammar, the paragraph, topic sentence, and development. On the whole, students who pass English 81 are quite well prepared to succeed in the requirements of English 101.

The 262 students tested--14 English 101 sections--received copies of the evaluative instrument some weeks prior to the examination date.

Evaluators

Five English 101 instructors evaluated student essays written in the Departmental Exam. As previously mentioned, three of the five are full-time instructors. Composition sections comprise at least two-thirds of the full-timers' teaching loads, and all of the part-timers' teaching loads. Their backgrounds vary: high school instruction and administration, literature, composition theory, communication research, fine arts, rhetoric, and general humanities. Four of the five evaluators have at least ten years of English teaching experience. Two were in favor of the project, and three were not, but agreed to see what the results would be in one study.

Evaluative Criteria

The evaluators in this WVC project employed evaluative criteria developed by the English Department of Spokane Community College in District 17 (See Appendix A). Advanced Composition instructors at SCC had often voiced concerns that students with average and good English 101 grades entered advanced composition courses with inadequate skill levels. That department mandated exit exams in its English 101 classes. Students had to achieve an arbitrarily imposed score on the criteria's scale to receive a passing grade in the course, regardless of the students' grades on regular English 101 writing assignments.

While we at WVC had no intention of using students' scores on the evaluative criteria to determine students' worthiness to advance or pass, we did admire the instrument that SCC faculty had developed through several years of discussion and refinement. We agreed with their standards and wording and were not willing to reinvent the wheel.

The scale measures skill levels in three important areas: Organization, Development, and Mechanics. Under each heading the instrument describes six levels of achievement, from polished excellence to unacceptable failure. Hence, the highest possible score is 18 (perfect six in all three areas), and the lowest is three (one point in each of the three areas). The strength of these evaluative criteria resides in their behavioral descriptors. While these descriptors may seem vague to people unschooled in composition, English instructors find them immediately comprehensible and "norm-able."

Norming

Several weeks prior to the administration of the first departmental examination, the evaluators met to discuss the evaluative criteria and undergo norming. In timed settings the evaluators read handwritten photocopies of several essays and scored them using the criteria's descriptors in Organization, Development, and Mechanics. After scoring several essays, the evaluators agreed among themselves that they had arrived at acceptable levels of correlation. We were interested to find that agreement on operational definitions within the criteria's descriptors is more crucial than agreement on the qualities of the essay under scrutiny. Unfortunately, we did not record the scores from that norming session, and thus cannot comment on the success of the norming session in terms of statistical correlation.

Test Administration

The Departmental Examination was administered twice, during finals week of Winter 1990 and finals week of Spring 1990. These two testing situations had many commonalties and one important difference.

In both tests we attempted to create a double blind procedure. Students seated themselves in our two largest lecture halls (one of 110 seats and one of 72 seats) on the same date at the same time to enhance treatment similarity.

The instruments (See Appendix B) were numerically coded. On the top of the instrument package we stapled a slip that held the student's name and the code number. The next page contained the code number, space for three sets of scores and the writing prompt. The prompt read as follows: "DIRECTIONS: Write an essay on the following topic. You may use a dictionary and thesaurus, but no other aid." To this page were attached lined pages for the students to write upon. Thus, through this rather elaborate anonymity, we hoped no student knew which evaluators would score his essay, and the evaluators would not know which student had written any essay, the student's section, or the student's instructor. The students had no prior knowledge of the topic.

Prior to the administration of the final exam we scrambled the consecutively numbered tests so that they were no longer in numerical sequence. We noted that the students tended of their own volition to sit together according to section, and so we scrambled the instruments once again after the students submitted their essays and before evaluators scored them.

The students were allowed two hours to write their essays. Two proctors who were instructors of English 101 were present in each testing room and administered the examination. The students handed in their essays when they felt they were finished.

The proctors immediately detached the identifying cover slips when the students submitted their examinations.

In both administrations students in night classes were given the option of writing their exams during the scheduled two-hour Departmental Examination period or during their last scheduled class period, also two hours. Make-up exams were also administered during this last night class. To protect the test topic security we used two sets of topics: one topic for the night and make-up period, and another for the large group setting. The make-up topics were "The Future" in the Winter exam and "Responsibility" in the Spring exam. The Winter large group prompt was "Decisions."

There was one major procedural revision between the Winter administration and the Spring administration. The evaluators felt that having to read so many essays on the same topic was a numbing experience at best, and a devastating experience at worst. Evaluators feared that they lost necessary mental freshness and excitement as well as physical vigor because so many essays on the same topic tended to blend together into one long and rather unpleasant discourse. To alleviate this we we used three different prompts in the Spring large group departmental exam. They were "Learning," "Entertainment," and "Human Relationships."

Evaluation

We attempted to make both evaluation sessions as similar as possible. We used the same evaluators for both administrations. Each essay was read and scored by two evaluators. A work study student was present in each evaluation session to time us and keep us on schedule and on task. We allowed three minutes per essay for this holistic reading. In the event that scores assigned by evaluators differed by four or more points, we asked a third evaluator to read and score the essay. All evaluators took part in each aspect of the evaluation

process. Both sessions started in the morning, broke for lunch, and continued into the afternoon.

Several instructors voiced concerns that the first reader's score might influence the following evaluation. During the first administration the second evaluator could see the score that the first evaluator had written on the cover of the instrument, and the third evaluator could see the scores written by the first and second evaluators. During the second administration of the exam we used Post-It Notes to cover our scores so that subsequent readers would not be influenced by the opinions of others.

Coding

After we completed our evaluations, the need for student anonymity was over. Teachers needed the scores in their grade books. We put the essays back into a numerical sequence and reattached the identifying cover slips. The instructors listed the students in their sections by identification number and scores for each reading. Thus we had for each student an identification code number, and at least two (in some cases three) sets of scores: two (sometimes three) each for Organization, Development, and Mechanics. Further, the students' scores were grouped into their respective English 101 sections.

These data were entered into an Excell computer spreadsheet (See Appendix C). We next scrambled the order of sections in the data without destroying any section's group integrity. Further, we renamed each section with letters from the end of the alphabet. Thus, while we can discuss the differences between sections, among sections, between administrations of the exam, and within the entire fourteen sections that took part in this SLOA study, we cannot say which instructor taught any given class section from the data we have. We felt that a SLOA project had no business placing any instructor under jeopardizing scrutiny.

RESULTS

Instruction and Evaluation

1. *Will student success over the quarter predict student success on the Departmental Final? In other words, is there correlation between students' scores on the test and students' grades in the English 101 sections? Looked at another way, this question might read: Do teachers evaluate the same criteria with the same rigor in regular assigned essays that they evaluate on a Departmental Exam?*

Our results here are based partially upon anecdotal evidence and partially upon matching test scores with class grades in a very informal, non-statistical comparison. Anecdotally, instructors (evaluators) report that those students who did well in class also did well in their Departmental Final.

Data do exist for 126 students of the population tested. Their final exam scores were converted by instructors to regular assignment equivalencies. In that particular subset of the population, 24 scored higher on the final than in the course. 46 received Departmental Exam grades that matched course grades. 65 received Departmental Exam grades that were lower than their course grades. In those cases for which we have data, 36.5% of the students received Departmental Exam grades that match course grades, but 46% of the students received course grades lower than Departmental Exam grades. Our data suggest that student success over the academic quarter will not predict student success on a Departmental Exam.

Discussion:

First, why is there a discrepancy between the anecdotal report and the data? Perhaps those instructors who did not supply matched scores and grades indeed found strong correlation between scores and grades. More probably, the instructors reported what they believed to have been true. Six months had passed between the Spring administration and our request for this information.

Second, why do test scores and course grades parallel one another only one-third of the time? We would have predicted that in the event the scores and grades did not match, then test scores would be lower than course grades, because students had only two hours to focus, organize, draft, and revise their essays, but in a normal class assignment students have one or two weeks for the same process. We like to believe that our students understand that the final product is worth great effort, and if nothing else, we teach that writing is a process, and revision is the greatest part of that process.

Perhaps we forgive the errors of our students when we compel them to write under the gun. We realized during the evaluation sessions that we were

scoring minimum competencies, and gave greater allowances for disfluencies, fuzzy thinking, and minor surface errors than we would in normal classroom assignments. Another factor may be grading criteria unrelated to the classroom assignments, or only marginally related. Some instructors deduct points if students do not attend regularly, on the grounds that English 101 is a performance course, and students must be present to learn. Additionally, students sometimes submit late essays, or fail entirely to submit an assignment. Thus, sanctions leveled against students' final grades may cause students' final grades to reflect a lower performance than the Departmental Exam.

2. *Do we have departmental consistency in our objectives? If we teach from identical objectives, then there should be no great differences between the scores for different sections. While we share a common WVC syllabus and objectives, instructors take students along varying paths to reach the common department goal of student writing competencies. Are the writing sections similar enough in their scores to indicate that several roads lead successfully to Rome?*

A two-tail test of probability between the fourteen combined sections of Winter and Spring indicates 12 of the 14 sections correlate at or above the .05 level of confidence (See Appendix D). Two of the sections varied at or above the .05 level: one section scored significantly lower than other sections, and one section scored significantly higher. Thus, we have strong departmental consistency on this exam in twelve out of fourteen sections. 86% of the sections are very similar in student outcomes. In one out of 14 registrations, a student will find that he is on the express road to Rome, and the ride is a smooth one. In one out of 14 registrations, a student will find that the Appian Way has detours and is somewhat rocky.

Discussion:

We are quite pleased with the consistency of scores between sections. Teachers are confident of their expertise, and our results tend to show that no matter the procedural differences in the classrooms, our students exit the course with a coherent and unified skill level.

Did the instructors "teach to the exam"? Perhaps they did to a small degree, but since instructor anonymity was assured prior to the project, we probably have data that reflect a fairly normal teaching environment. Teachers knew their students would be compared with other students, but they also knew that the comparison would not be used to endanger any particular instructor.

In one case a section of students scored significantly high above the pack. The instructor of that section taught other sections that scored within the normal range of the entire population. Conversely, in one case a section of students scored significantly far below the pack. Again, that instructor taught other sections that scored within the normal range. What could have caused these two sections to fall outside of the norm? Teachers report that 8:00 AM sections are tougher to teach than other sections, because students don't want to be in class at such an early hour. Teachers also complain that 11:00 AM sections are

restless due to hunger, and 12:00 PM sections are dopey due to fatigue. The results may be the effect of teaching skill or may not

3. *Can the Composition faculty agree to a definition of student competency? Can the faculty become consistent in their evaluations of student writing? Overall, on which dimension (O, D, M), do faculty most agree in their ratings? On which dimension do they least agree? Each essay was rated by two evaluators. In those cases when our scores differed by four or more points, a third reader rated the essay. Do we have a good level of agreement? Will we most agree among ourselves in the Winter or in the Spring? In other words, will we gain or lose consistency over time? If we disagree, what will cause that disagreement?*

We did achieve our goal of arriving at a definition of student competency. While we left the matter of "pass" and "fail" to individual instructors, we managed to agree in faculty meetings to use the evaluative criteria discussed in "Procedure." We normed with that yardstick and employed it to capture student outcomes in a testing situation. Also, the evaluators left the evaluation sessions having agreed at an almost incredible level of consistency in our evaluatior

Random analysis shows no difference in evaluation one and evaluation two (See Appendix E). Evaluation three (when evaluators disagreed by four or more points) was done so seldom it could not be included in the analysis. Thus, our evaluations were consistent enough that it really did not matter who evaluated an essay.

We suspected when we posed this question that we would disagree to some discussable extent, and we thought that such disagreements might have commonalties. Combining the Winter and Spring evaluations, we differed by four or more on our evaluations' total scores only twelve times. Six essays required three readings in the Winter, and six essays required three readings in the Spring. This is not a significant level of disagreement.

Discussion:

While acknowledging this insignificance, it is interesting to note that on those rare occasions when we did disagree by four or more out of a total of 18 possible points, the total scores on the essay were rather high. We tend not to disagree in our evaluations. When we do disagree, the disagreement seems to be based on our differing definitions of excellent writing. The number of disagreements is too small for breakdowns into the O, D, M dimensions.

Our data will not tell us if we gained or lost agreement in the second Departmental Exam scoring session, because the identity of the evaluators for any particular essay is unknown. Over all, the difference between the two evaluations of the essays in the Spring was as insignificant as the difference between the two evaluations in the Winter. In both Winter and Spring, the two evaluations of each essay were virtually equivalent.

We met to norm only once in the two academic terms, and that meeting was only one hour in length. We would suggest, then, that our consistency in evaluation results from highly specific evaluative criteria. Those criteria are

descriptive and somewhat behavioral in their wording. They often do not leave a great deal of room for indecision and disagreement. For example, the discriptor for "two" in the Mechanics dimension leaves little room for doubt: "Major errors that distract the reader (frag, comma splice, run-on, spelling).

4. *In the Winter quarter all of the students wrote on the same topic. In the Spring quarter students wrote on one of three topics randomly assigned. Does this different procedure affect students' writing or faculty's evaluation? Do Departmental Exams need to have only one topic to be consistent in their measures and fair to the students?*

The data do not indicate big differences between the evaluations of student outcomes in Winter and Spring. Apparently, three topics randomly assigned to a group of students produces writing by students and evaluations by raters very similar to a situation in which one topic is assigned to a group of students.

Mean Scores on Dimensions

	Organization	*Development*	*Mechanics*	*Total*
Winter	**3.6844**	**3.2092**	**2.9894**	**9.8830**
Spring	**3.5178**	**3.1976**	**2.8538**	**9.5692**

Our data indicate that a group of students, overall, writes as well when they have more than one topic as they do when the entire group writes on one topic. Further, as previously reported, evaluators score as consistently with three topics as with one topic. This suggests that Departmental Exams do not need to have only one topic to be consistent and to be fair. Sadly, we do not have the data necessary to determine if any of the three topics in the Spring administration produced essays more likely to receive high scores.

Discussion:

The Spring students did receive slightly lower scores in each dimension of O, D, M, but that difference is not great. Possibly the use of three topics rather than one topic can account for the slightly lower scores. Just as possible, the changing seasons may account for the differences.

We feel confident that we can use more than one topic if we should ever wish to run the Departmental Exam again.

5. *During the Winter rating session, the second rater was able to see the first rater's evaluation. During the Spring session, we masked the first rater's score in order to reduce the evaluation effect. Will that masking create a result? Will there be more or less consistency in raters' evaluations?*

As previously noted, the difference between the two evaluations was equally insignificant in Winter and in Spring. Both ratings of an essay in the Winter were very similar, and both ratings of an essay in the Spring were very similar. Our data suggests that a rater's knowledge of a previous rater's scoring of an essay does not influence subsequent scoring. It doesn't matter if the second reader sees the first reader's score. Just as interesting, it doesn't matter if the second reader doesn't see the first reader's score.

Discussion:

We are pleased that our evaluators have sufficient self-confidence to score an essay on their best judgement regardless of a previous rater's opinion. Once more, this self-confidence is most likely the result of our rigorous evaluative criteria. Also, the evaluators apparently understood the substantive terms in the criteria and agreed on those terms' application.

Student Learning Outcomes

6. *This concern is best phrased as a statement rather than a question. We would like to assure students of some objectivity in evaluation by having their writing read by more than one instructor.*

We accomplished this goal. Each student received feedback from two, and sometimes three, instructors who did not know the student's identity. Thus, the feedback was not biased for or against any student or group of students. Also, the high level of agreement among the raters would be welcome news to the students, if they were given that information.

Discussion:

We must admit that this was the first concern that we formalized in the planning process, yet it seems to have gotten lost. We did not relay the results to the students as efficiently as we might have done. The students wrote their exams after the last regular class meeting. Hence, only those students sufficiently motivated by curiosity or fear sought out their Departmental Exam scores after they had finished the course.

7. *Are student outcomes in writing different in different academic terms? Do students write better in one quarter than another? This study collects data in two quarters. We are interested to see if there are differences.*

Again, the data revealed no difference in evaluators' consistency or student outcome in Winter and Spring. Students write as well or as poorly in Spring as they do in Winter.

Discussion:

None needed.

8. *On which dimension (O, D, M) do students at WVC best perform? On which do they not perform as well? This is an important matter. Armed with answers to these questions, we can better serve our students. This study should provide data to show us both strengths and weaknesses in student learning outcomes, and indicate the directions we need to take in our English 101 classes.*

As the following table indicates, students scored highest on Organization and lowest on Mechanics.

Percent of Scores in Each Dimension for Each Value

Value	*Organization*	*Development*	*Mechanics*
1	2.1	2.9	7.1
2	11.3	20.6	25.8
3	29.4	44.8	41.5
4	41.3	21.2	20.2
5	13.1	7.9	4.6
6	2.7	2.7	.8

"One" is the lowest score on the evaluative criteria, and "six" is an excellent score. 57.1% of evaluations on Organization scored four or higher. 31.8% of evaluations on Development were four or higher. Only 25.6% of evaluations on Mechanics scored four or higher. Evaluators were twice as pleased with Organization as they were with Mechanics. A cross correlation of each scale variable to all other scale variables shows variances between dimensions at a .001 level of significance.

A frequency distribution indicates a mode of 9 (the total criteria's actual midpoint is 10.5) with some positive skewness and a leptokurtic distribution. The total group of students is homogeneous in the writing traits measured.

Frequency Distribution on Total Score in All Dimensions

Value	Frequency	Percent	Cum. Percent
3	7	1.3	1.3
4	3	.6	1.9
5	9	1.7	3.7
6	26	5.0	8.7
7	48	9.2	17.9
8	71	13.7	31.5
9	93	17.9	49.4
10	83	16.0	65.4
11	65	12.5	77.9
12	54	10.4	88.3
13	21	4.0	92.3
14	17	3.3	95.6
15	10	1.9	97.5
16	7	1.3	98.8
17	3	.6	99.4
18	3	.6	100.0

For this distribution, 7 falls on the 16th percentile, 10 falls on the 50th percentile, and 12 falls on the 84th percentile.

An unexpected result derives from a comparison of central tendencies between sections. While twelve of the 14 sections scored very similarly to one another, two sections differed in kurtosis; that is, two sections' scores distribute themselves into a narrow, leptokurtic curve, and other sections' scores distributed themselves into a wider, platykurtic curve:

Total Means and Their Standard Deviations

Section	Total Mean	Standard Deviation
K	9.3889	3.017
L	10.2000	2.902
M	9.6579	2.172
N	9.7500	3.349
O	9.2045	2.226
P	9.9744	2.109
Q	10.8158	1.971
R	9.6429	3.122
S	8.8000	1.436
T	10.0000	3.094
U	9.5526	1.927
V	9.7879	2.012
W	10.2727	3.042
X	9.1389	2.758

The range of the standard deviations reflects the various degrees of kurtosis. The students in Section S and Section Q, for example, all wrote at a rather homogeneous, uniform level, while the students in Section T wrote at levels that ranged across the evaluative criteria, and their scores formed a flatter, platykurtic curve.

Finally, the mean scores for each dimension and each section demonstrate once again that, on the whole, the sections scored similarly, and that overall they share the same strengths (Organization) and weaknesses (Mechanics).

Mean Scores for Sections in Each Dimension

Section	Organization	Development	Mechanics
K	3.4444	3.1111	2.8333
L	3.8250	3.5250	2.8500
M	3.6842	3.0789	2.8947
N	3.4000	3.0750	3.2750
O	3.5227	2.7045	2.9773
P	3.7436	3.2821	2.9487
Q	4.1579	3.6053	4.1579
R	3.5000	3.1667	2.9762
S	3.1500	2.8750	2.7750
T	3.7059	3.1765	3.1176
U	3.6316	3.2895	2.6316
V	3.6667	3.4545	2.6667
W	3.8182	3.2273	3.2273
X	3.2778	3.1667	2.6944

Discussion:

Why do we find such a dramatic difference between Organizational skills and Mechanical skills? First, Organization quite possibly is an easier task in English 101 than is mechanics. For a high score in Organization, students needed to demonstrate "Interesting lead-in; clear and focused thesis stated as an opinion; rest of organization is immediately apparent; coherence: smooth transitions, fluid style, definite and effective conclusion." While this may sound daunting to those who don't teach English, is translates roughly to "Good coffee tables have four legs and a flat surface. Each leg should be attached firmly. Coffee drinkers should be able to use the table without too much trouble." The basic theory of building a good, easy-to-use five-paragraph essay can be taught in two to three hours, and students usually learn the skill of organization quickly.

On the other hand, a good Mechanics score certainly sounds easier to obtain, but is actually far more difficult: "No apparent errors: unusually good mechanics--what we would expect from ourselves." This descriptor has fewer words, certainly, than the good organization discriptor, but the attempt of it is a far different universe of discourse, and a huger universe at that. Any survivor of English 101 can tell an interested party that the more one learns about grammar and punctuation, the more complex they become. Definitions of grammatical

terms are themselves full of grammatical terms that must be defined. For students in their first two years of college, few undertakings are more harrowing, more frustrating, than the road to grammatical excellence.

At WVC English 101 stresses the correcting of errors in students' essays rather than diagramming of sentences and naming of parts for the intellectual excitement of it. Nevertheless, our data indicate that even at the end of English 101 our students are left with a hard row to hoe: "Major errors that distract the reader (frag, comma splice, run-on, spelling)."

Second, lower Mechanics scores may be the result of the two-hour exam situation. Students can knock together a coffee table in a hurry, but can they sand off all the splinters? Can they apply three layers of varnish? Teachers of English like to think that their students spend two weeks revising (re-wording, re-thinking, re-writing) their essays. Perhaps students spend two weeks editing (correcting) their essays.

Third, Mechanical flaws may irritate raters more than organizational or developmental flaws irritate raters. There are degrees of "focus," and degrees of "completeness," but there are no degrees of "frag." Teachers of English 101 grow amazingly adept at spotting and attacking mistakes of grammar and punctuation. Even though teachers grow increasingly weary of these errors, every new crop of English 101 students continues to make those errors. Although students score sufficiently high on ASSET and then spend 55 hours in the composition classroom, they still, in Mechanics, average 2.6 on a scale of six.

Development may be easier for students because it is more a matter of the ear, more a matter of fluency. Development does not require memorization and application of endless, arcane rules; rather it demands taste and judgement--an easier and more humane requirement than excellence in the voodoo of Mechanics.

Still, our students find Development tougher than Organization. Organization requires students to learn and follow a simple, linear blueprint, to distinguish between the general and the specific, to remember that a coffee table can't stand upright on just two legs. Development (as described in our evaluative criteria) requires students to sound, in their essays, like reasonable people with important messages, and that is no easy task. Of the three dimensions, Development is the hardest to pin down.

Finally, all the scores may seem too low: Ninety percent of the scores fell on the lower two-thirds of the scales. Once again, we must return to the criteria against which we evaluate our students' writing. We could make the scale a warmer, fuzzier yardstick: "Ten frags: 8; twenty frags: 5." We must also remember that the dimensions are truly different from one another and not allow ourselves to confuse apples with oranges.

CONCLUSIONS

Recommendations for English 101 at Wenatchee Valley College

1. Make no changes as regards the teaching of Organization and Development.

2. Emphasize the teaching of Mechanics in English 101. Meet as a department to discuss methods, ideas, what works, what doesn't work. For instance, it may be helpful for the students if English teachers type up copies of errors from an assignment and students work together in peer-teaching groups to correct those errors. If necessary, revise the English 101 syllabus so that it reflects a greater concern for Mechanical correctness in students' writing.

3. Encourage instructors to exchange batches of papers now and again to allow students to receive feedback from other evaluators. Our instructors see the same strengths and weaknesses, and so those exchanges would be fair.

4. Encourage new instructors to meet with currently teaching instructors for informal norming sessions of short duration. This will help us to continue the highly similar evaluation that we have established.

5. Encourage instructors to develop similar grading policies for regular assignments in English 101. Some may consider this dictatorial, but if the grading criteria are developed in concert by all English 101 faculty, we will find that we currently grade on the same bases using different terminology.

6. Encourage instructors to meet and refine or redefine instructional goals.

7. Encourage instructors of advanced composition courses to share with English 101 instructors their expectations of entering students' skills . Do we need to develop a common vocabulary of rhetoric to share with our students? Do advanced composition instructors have concerns about their entering students' skills?

8. Encourage administrators to pay part-time instructors for their efforts in large SLOA projects. Such projects are outside the realm of normal teaching and class preparation for any community college instructor. Part-time instructors especially need to be paid, praised, and promoted when their efforts on behalf of students are above and beyond the call of duty.

We do not recommend further Departmental Finals for the sake of judging instructors' competency or curricular consistency. We do recommend further Departmental Finals so that we may establish a relationship between other variables in the policies and processes of English instruction at WVC.

Recommendations for Further Research

We recommend that if faculty and administrators agree to run more Departmental Finals, their procedures include the gathering and coding of additional data. We can argue that our procedure works, that it has some degree of validity. What we don't know is how our evaluations and how students' learning outcomes in Organization, Development, and Mechanics relate to ASSET, placement, grades, and learning.

Briefly, we should run both a Departmental Pre-test and Post-test using our same procedures. In addition, we should gather and factor in the students' ASSET scores and final course grades in our analyses. Thus, we will be able to answer the following questions.

1. In which dimension (O,D,M) do entering students in English 101 need the most help? Do students share any common strengths when they enter the course?

2. In which dimensions (O,D,M) do students make the most and least progress?

3. Which outcome best relates with student success in English 101: ASSET or pre-tests?

4. Do ASSET and pre-test predict better in one dimension (O, D, M) than in another?

The answers to these questions could have powerful influences upon the current policies and procedures in English 101 at Wenatchee Valley College.

Appendix A., Evaluative Criteria

Evaluative Criteria for English Department Final

ORGANIZATION

6. Interesting lead-in; clear and focused thesis stated as an opinion; rest of organization is immediately apparent; coherence: smooth transitions, fluid style, definite and effective conclusion.

5. Relevant lead-in; focused thesis; clear topic sentences; coherence: helpful transitions, definite conclusion.

4. Obvious and focused thesis statement; appropriate topic sentences with relevant support; recognizable conclusion.

3. Unfocused thesis statement; erratic topic sentences and support; forced conclusion.

2. Unclear thesis statement; general absence of topic sentences; lack of support; weak conclusion.

1. No apparent organization; thesis statement, topic sentences, support and conclusion vague or non-existent.

DEVELOPMENT

6. Excellent unity, coherence, completeness through paragraphs. Provocative topic sentences, sophisticated transitions, memorable examples and details. Excellent diction.

5. Very good unity, coherence, completeness through paragraphs. Effective topic sentences, smooth transitions, strong examples and details. Good diction.

4. Above average unity, coherence, completeness through paragraphs. Clear topic sentences, useful transitions, appropriate examples and details.

3. Adequate unity, coherence and completeness through paragraphs. Use of topic sentences and transitions; details and examples are adequate.

2. Lack of any of the following: unity, coherence or completeness.

1. Lack of two or more of the following: unity, coherence or completeness.

MECHANICS

6. No apparent errors: unusually good mechanics--what we would expect from ourselves.

5. Competent, no apparent errors: mechanics enhance essay.

4. Adequate mechanics: they neither enhance nor detract from the writing.

3. Occasional mechanical errors.

2. Major errors that distract the reader (frag, comma splice, run-on, spelling).

1. Major errors that fail the essay.

Appendix B, Instruments

Student Code Number:_____

English 101 Final Examination

Winter, 1990

Evening Classes

DIRECTIONS: Write an essay on the following topic:

Choices

You may use a dictionary and thesaurus, but no other aid.

Organization	Organization	Organization
Development	Development	Development
Mechanics	Mechanics	Mechanics
Total	Total	Total

Student Code Number:_____

English 101 Final Examination

Winter, 1990

Make-up

DIRECTIONS: Write an essay on the following topic:

The Future

You may use a dictionary and thesaurus, but no other aid.

Organization	Organization	Organization
Development	Development	Development
Mechanics	Mechanics	Mechanics
Total	Total	Total

English 101 Final Examination

Spring, 1990

Student Code Number: _____

DIRECTIONS: Write an essay on the following topic:

Learning

You may use a dictionary and thesaurus, but no other aid.

Organization	Organization	Organization
Development	Development	Development
Mechanics	Mechanics	Mechanics
Total	Total	Total

English 101 Final Examination

Spring, 1990

Student Code Number: _____

DIRECTIONS: Write an essay on the following topic:

Entertainment

You may use a dictionary and thesaurus, but no other aid.

Organization	Organization	Organization
Development	Development	Development
Mechanics	Mechanics	Mechanics
Total	Total	Total

English 101 Final Examination

Spring, 1990

Student Code Number: _____

DIRECTIONS: Write an essay on the following topic:

Human Relationships

You may use a dictionary and thesaurus, but no other aid.

Organization	Organization	Organization
Development	Development	Development
Mechanics	Mechanics	Mechanics
Total	Total	Total

English 101 Final Examination

Spring, 1990

Make-up

Student Code Number: _____

DIRECTIONS: Write an essay on the following topic:

Responsibility

You may use a dictionary and thesaurus, but no other aid.

Organization	Organization	Organization
Development	Development	Development
Mechanics	Mechanics	Mechanics
Total	Total	Total

Appendix C, Excell Spreadsheet Coded Data

COMPOSITION 101 FINALS WINTER 1990

Section	Code Nmbr	Organize	Develop	Mechan	Total
K	8	4	3	3	10
K	8	4	4	3	11
K	9	4	4	4	12
K	9	4	3	2	9
K	10	4	3	4	11
K	10	4	3	2	9
K	13	4	3	2	9
K	13	4	3	4	11
K	25	1	1	1	3
K	25	1	1	1	3
K	27	5	5	5	15
K	27	4	3	3	10
K	27	4	5	4	13
K	35	1	1	1	3
K	35	1	1	1	3
K	52	4	3	3	10
K	52	4	3	3	10
K	84	5	5	3	13
K	84	5	5	3	13
K	85	4	4	3	11
K	85	4	4	3	11
K	86	4	4	3	11
K	86	4	4	3	11
K	91	4	4	2	10
K	91	3	3	2	8
K	112	4	3	4	11
K	112	3	2	4	9
K	113	3	3	2	8
K	113	2	2	3	7
K	115	4	3	2	9
K	115	4	3	2	9
K	131	5	5	5	15
K	131	4	4	4	12
K	137	3	2	3	8
K	137	3	4	3	10
K	140	1	2	3	6
K	140	2	2	3	7
L	7	5	5	4	14
L	7	5	5	4	14
L	26	4	4	4	12
L	26	4	3	3	10
L	41	4	3	3	10
L	41	4	3	3	10
L	42	4	3	2	9
L	42	4	4	2	10
L	55	6	6	5	17

L	55	3	3	4	10
L	55	5	5	4	14
L	56	2	3	2	7
L	56	2	2	1	5
L	72	4	4	4	12
L	72	4	4	4	12
L	74	4	3	3	10
L	74	4	3	3	10
L	87	4	3	3	10
L	87	3	2	3	8
L	88	3	3	2	8
L	88	3	2	3	8
L	89	5	5	3	13
L	89	4	5	3	12
L	90	3	3	3	9
L	90	3	3	2	8
L	92	5	4	2	11
L	92	5	4	3	12
L	98	4	3	2	9
L	98	4	3	1	8
L	99	5	4	4	13
L	99	6	5	3	14
L	104	5	5	4	14
L	104	6	6	5	17
L	114	4	3	3	10
L	114	4	4	3	11
L	116	2	3	1	6
L	116	2	3	2	7
L	119	3	3	2	8
L	119	2	2	2	6
L	120	3	3	3	9
L	120	2	2	1	5
L	120	2	2	2	6
M	15	4	3	2	9
M	15	4	3	3	10
M	16	3	3	3	9
M	16	4	3	3	10
M	19	4	3	3	10
M	19	4	4	4	12
M	28	5	4	4	13
M	28	3	4	3	10
M	29	4	3	3	10
M	29	4	3	3	10
M	36	5	4	4	13
M	36	4	3	3	10
M	37	3	2	3	8
M	37	4	3	3	10
M	38	4	3	2	9
M	38	4	3	3	10

M	51	2	2	3	7
M	51	3	3	1	7
M	57	4	3	3	10
M	57	3	2	2	7
M	58	4	4	5	13
M	58	4	3	4	11
M	62	4	3	4	11
M	62	4	3	3	10
M	73	3	4	4	11
M	73	6	6	3	15
M	73	3	4	3	10
M	108	4	4	2	10
M	108	4	3	3	10
M	117	2	2	2	6
M	117	2	2	2	6
M	132	4	3	2	9
M	132	3	3	1	7
M	135	3	2	2	7
M	135	2	2	1	5
M	136	3	3	3	9
M	136	3	3	3	9
M	150	5	3	4	12
M	150	5	3	4	14
N	17	1	1	1	3
N	17	1	1	1	3
N	78	4	4	3	11
N	78	4	2	3	9
N	173	5	3	4	12
N	173	4	4	3	11
N	174	3	3	3	9
N	174	3	2	3	8
N	175	4	3	3	10
N	175	3	3	4	10
N	176	3	2	4	9
N	176	4	3	5	12
N	177	3	3	3	9
N	177	3	3	3	9
N	178	3	2	2	7
N	178	4	3	3	10
N	179	5	4	3	12
N	179	5	5	3	13
N	180	6	6	6	18
N	180	6	6	5	17
N	181	3	2	3	8
N	181	3	3	3	9
N	182	2	2	3	7
N	182	3	3	3	9
N	183	2	1	3	6
N	183	2	2	3	7

N	186	4	3	4	11
N	186	3	3	3	9
N	187	4	3	4	11
N	187	4	4	4	12
N	188	2	2	3	7
N	188	2	3	3	8
N	67	5	5	4	14
N	67	5	5	4	14
N	128	5	5	5	15
N	128	5	5	4	13
N	184	2	2	4	8
N	184	2	3	4	8
N	185	2	2	1	5
N	185	2	2	1	5
O	107	3	2	1	6
O	107	3	2	1	6
O	44	4	3	3	10
O	44	4	3	4	11
O	22	4	2	5	11
O	22	4	3	2	9
O	14	4	2	4	10
O	14	4	3	4	11
O	21	3	3	3	9
O	21	3	3	2	8
O	3	5	5	4	14
O	45	4	3	2	9
O	45	3	3	2	8
O	127	2	3	3	8
O	127	2	2	3	7
O	60	1	1	1	3
O	60	2	1	3	6
O	111	4	4	4	12
O	111	4	4	4	12
O	43	3	2	2	7
O	43	3	2	2	7
O	75	4	2	4	10
O	75	3	2	3	8
O	49	4	1	4	9
O	49	4	1	4	9
O	5	4	3	3	10
O	5	4	2	2	8
O	68	3	3	3	9
O	68	5	4	3	12
O	48	4	2	4	10
O	48	4	2	4	10
O	34	4	3	2	9
O	34	4	3	2	9
O	109	3	2	2	7

O	109	3	2	2	7
O	23	4	3	3	10
O	23	3	3	2	8
O	118	4	3	4	11
O	118	3	2	3	8
O	71	4	3	4	11
O	71	2	4	2	8
O	20	5	6	6	17
O	20	5	4	4	13
O	20	3	4	4	11
P	2	4	3	4	11
P	2	3	2	2	7
P	2	3	2	3	8
P	39	3	4	4	11
P	39	4	3	4	11
P	40	2	4	3	9
P	40	4	4	3	11
P	53	4	3	2	9
P	53	4	3	2	9
P	54	3	3	1	7
P	54	3	3	1	7
P	59	5	3	4	12
P	59	3	2	3	8
P	59	4	4	3	11
P	61	4	3	2	9
P	61	4	4	3	11
P	65	4	3	3	10
P	65	4	3	3	10
P	66	4	4	4	12
P	66	4	4	4	12
P	70	4	4	4	12
P	70	5	4	3	12
P	76	3	3	2	8
P	76	4	3	3	10
P	77	4	3	2	9
P	77	4	3	3	10
P	110	4	4	3	11
P	110	4	4	2	10
P	121	4	3	2	9
P	121	3	2	2	7
P	122	3	3	3	9
P	122	3	3	1	7
P	124	2	2	2	6
P	124	3	3	3	9
P	125	4	3	4	11
P	125	5	2	4	11
P	126	3	4	4	11
P	126	3	3	3	9
P	129	6	5	5	16

W90EFforEd

P	129	5	5	5	15
Q	1	4	4	2	10
Q	1	6	5	3	14
Q	1	5	4	2	11
Q	6	4	4	4	12
Q	6	4	4	4	12
Q	11	3	2	1	6
Q	11	3	2	2	7
Q	12	5	4	4	3
Q	12	4	4	3	11
Q	18	4	3	2	9
Q	18	4	3	2	9
Q	24	5	4	5	14
Q	24	4	4	4	12
Q	30	4	4	3	11
Q	30	4	3	3	10
Q	31	4	4	4	12
Q	31	4	4	3	11
Q	32	4	3	3	10
Q	32	4	1	3	8
Q	33	5	3	4	12
Q	33	4	3	3	10
Q	46	3	5	1	9
Q	46	4	4	1	9
Q	50	4	5	2	11
Q	50	4	3	4	11
Q	63	5	5	5	15
Q	63	4	4	4	12
Q	64	4	4	3	11
Q	64	4	4	3	11
Q	69	4	4	4	12
Q	69	5	4	4	13
Q	79	4	3	3	10
Q	79	4	2	2	8
Q	105	4	4	3	11
Q	105	5	4	3	12
Q	106	5	4	3	12
Q	106	5	5	3	13
Q	123	4	3	3	10
Q	123	3	2	3	8
	Averages	3.65799257	3.20921986	2.9893617	9.84751773

40

std dev org	std dev dev	std dev mech	std dev tot
1.02445428	1.05819296	1.02455282	2.66547937

COMPOSITION 101 FINALS WINTER 1990

Section	Code Nmbr	Organize	Develop	Mechan	Total
R	1	2	2	1	5
R	1	3	2	3	8
R	96	4	3	3	10
R	96	4	3	3	10
R	99	6	5	4	15
R	99	5	6	5	16
R	100	4	3	2	9
R	100	2	2	2	6
R	101	4	4	4	12
R	101	4	3	4	11
R	102	3	2	3	8
R	102	5	6	5	16
R	102	5	5	4	14
R	103	3	2	3	8
R	103	4	4	3	11
R	108	2	2	3	7
R	108	3	3	2	8
R	109	4	5	4	13
R	109	1	1	2	4
R	109	3	3	3	9
R	110	3	3	2	8
R	110	2	2	2	6
R	116	3	2	2	7
R	116	2	2	2	6
R	117	4	4	4	12
R	117	4	4	4	12
R	118	3	3	3	9
R	118	4	3	4	11
R	120	5	4	3	12
R	120	3	3	4	10
R	125	5	5	4	14
R	125	4	4	4	12
R	126	4	3	4	11
R	126	4	4	4	12
R	127	4	3	3	10
R	127	4	4	3	11
R	128	2	1	1	4
R	128	2	1	1	4
R	129	2	2	2	6
R	129	3	3	2	8
R	131	5	4	3	12
R	131	4	4	3	11
R	140	3	3	2	8
R	140	5	4	3	12
S	49	4	3	3	10
S	49	3	3	2	8

S'90EFforEd

S	33	3	3	3	9
S	33	3	2	3	8
S	34	3	2	3	8
S	34	4	3	3	10
S	32	4	3	2	9
S	32	4	2	3	9
S	51	3	3	2	8
S	51	3	3	3	9
S	54	4	4	3	11
S	54	3	3	2	8
S	55	4	5	2	11
S	55	4	3	2	9
S	26	4	3	3	10
S	26	4	5	3	12
S	7	4	3	3	10
S	7	2	2	4	8
S	13	2	2	2	6
S	13	3	3	2	8
S	87	3	3	3	9
S	87	3	3	4	10
S	86	4	3	3	10
S	86	4	3	3	10
S	84	4	3	3	10
S	84	3	3	2	8
S	83	3	3	2	8
S	83	1	2	3	6
S	72	2	2	3	7
S	72	3	2	3	8
S	71	4	3	3	10
S	71	3	4	4	11
S	69	3	3	3	9
S	69	2	3	3	8
S	67	3	2	3	8
S	67	3	3	3	9
S	58	2	2	2	6
S	58	3	3	3	9
S	56	3	3	3	9
S	56	2	2	2	6
T	2	4	5	4	13
T	2	5	4	3	12
T	3	6	4	6	16
T	3	6	6	6	18
T	4	4	4	3	11
T	4	4	4	4	12
T	5	4	4	4	12
T	5	4	5	5	14
T	6	3	2	2	7
T	6	2	2	1	5
T	7	4	3	3	10

T	7	3	3	2	8
T	9	4	3	3	10
T	9	3	3	3	9
T	10	2	2	3	7
T	10	4	3	3	10
T	11	2	2	2	6
T	11	2	2	2	6
T	12	3	2	3	8
T	12	3	3	3	9
T	19	5	4	5	14
T	19	6	5	5	16
T	15	4	3	3	10
T	15	4	3	3	10
T	92	3	3	2	8
T	92	3	3	2	8
T	111	4	2	2	8
T	111	3	2	2	7
T	136	4	4	4	12
T	136	4	3	4	11
T	137	3	2	3	8
T	137	3	2	2	7
T	139	4	3	2	9
T	139	4	3	2	9
U	28	3	3	2	8
U	28	3	2	2	7
U	70	4	3	2	9
U	70	3	3	2	8
U	73	3	3	3	9
U	73	3	3	4	10
U	85	4	5	3	12
U	85	3	4	2	9
U	59	5	6	3	14
U	59	4	5	4	13
U	66	4	3	2	9
U	66	3	2	1	6
U	52	4	3	2	9
U	52	3	3	3	9
U	50	5	4	3	12
U	50	4	3	3	10
U	48	2	2	2	6
U	48	4	2	2	8
U	40	3	3	2	8
U	40	3	3	1	7
U	14	4	3	3	10
U	14	3	3	3	9
U	27	4	3	4	11
U	27	5	4	2	11
U	25	3	3	3	9
U	25	4	4	2	10

U	39	5	3	3	11
U	39	5	5	4	14
U	29	4	4	2	10
U	29	4	3	3	10
U	30	3	4	3	10
U	30	3	3	3	9
U	31	4	3	2	9
U	31	4	3	3	10
U	24	2	2	3	7
U	24	3	3	2	8
U	35	4	3	3	10
U	35	4	4	4	12
V	74	4	3	2	9
V	74	4	3	2	9
V	75	5	5	3	13
V	75	5	5	3	13
V	81	4	4	2	10
V	81	4	3	2	9
V	80	2	2	3	7
V	80	3	6	3	12
V	80	3	4	3	10
V	79	4	4	2	10
V	79	3	3	1	7
V	78	4	3	3	10
V	78	3	3	2	8
V	77	5	5	3	13
V	77	3	3	3	9
V	77	3	2	4	9
V	76	3	2	2	7
V	76	3	3	2	8
V	82	4	3	2	9
V	82	3	2	3	8
V	57	5	4	2	11
V	57	4	3	2	9
V	61	3	3	3	9
V	61	1	2	2	5
V	61	4	4	2	10
V	60	5	4	4	13
V	60	5	4	3	12
V	68	4	4	3	11
V	68	4	4	3	11
V	500	3	3	3	9
V	500	4	4	4	12
V	36	4	4	3	11
V	36	4	4	3	11
V	37	3	4	3	10
V	37	3	2	3	8
V	38	4	3	4	11
V	38	3	3	3	9

W	300	4	3	3	10
W	300	3	2	2	7
W	301	3	2	2	7
W	301	3	3	3	9
W	302	4	3	2	9
W	302	3	2	2	7
W	303	3	4	5	12
W	303	5	3	3	11
W	304	4	3	2	9
W	304	3	3	2	8
W	305	5	6	5	16
W	305	5	4	4	13
W	306	4	3	4	11
W	306	3	3	2	8
W	307	3	2	3	8
W	307	2	3	3	8
W	308	5	3	3	11
W	308	5	2	5	12
W	309	6	6	6	18
W	309	5	6	4	15
W	310	3	2	3	8
W	310	3	3	3	9
X	311	4	4	2	10
X	311	3	3	3	9
X	312	3	3	3	9
X	312	2	2	3	7
X	313	2	2	1	5
X	313	2	3	2	7
X	314	2	4	3	9
X	314	2	2	2	6
X	315	5	4	3	12
X	315	2	2	3	7
X	315	5	5	2	12
X	316	2	3	3	8
X	316	3	3	3	9
X	317	3	3	3	9
X	317	4	3	3	10
X	318	3	3	2	8
X	318	2	2	2	6
X	319	3	3	3	9
X	319	2	3	2	7
X	320	4	3	1	8
X	320	3	3	2	8
X	321	5	6	4	15
X	321	5	5	5	15
X	322	4	3	4	11
X	322	3	3	4	10
X	323	5	5	3	13
X	323	6	5	5	16

X	324	3	2	4	9
X	324	4	4	4	12
X	325	2	3	1	6
X	325	4	3	1	8
X	326	3	3	1	7
X	326	3	3	1	7
X	327	5	4	5	14
X	327	4	3	2	9
X	327	3	3	3	9
X	328	3	2	2	7
X	328	3	2	2	7
	Averages	3.51778656	3.19762846	2.85375494	9.56916996

47

std dev org	std dev dev	std dev mech	std dev tot
1.01987896	0.96689023	2.51979428	#DIV/0!

Appendix D, Two-Tailed Probability t-Test

COMPARISON OF WINTER AND SPRING GROUPS

T-TEST /GROUPS QUARTER ('WINTER','SPRING') /VARIABLES ORGANIZA DEVELOPM
MECHANIC TOT.

t-test for: ORGANIZA

	Number of Cases	Mean	Standard Deviation	Standard Error
WINTER	282	3.6844	1.024	.061
SPRING	253	3.5178	.986	.062

Pooled Variance Estimate				Separate Variance Estimate			
F Value	2-Tail Prob.	t Value	Degrees of Freedom	2-Tail Prob.	t Value	Degrees of Freedom	2-Tail Prob.
1.08	.539	1.91	533	.056	1.92	530.34	.056

t-test for: DEVELOPM

	Number of Cases	Mean	Standard Deviation	Standard Error
WINTER	282	3.2092	1.058	.063
SPRING	253	3.1976	1.020	.064

Pooled Variance Estimate				Separate Variance Estimate			
F Value	2-Tail Prob.	t Value	Degrees of Freedom	2-Tail Prob.	t Value	Degrees of Freedom	2-Tail Prob.
1.08	.550	.13	533	.898	.13	530.26	.897

t-test for: MECHANIC

	Number of Cases	Mean	Standard Deviation	Standard Error
WINTER	282	2.9894	1.025	.061
SPRING	253	2.8538	.967	.061

Pooled Variance Estimate				Separate Variance Estimate			
F Value	2-Tail Prob.	t Value	Degrees of Freedom	2-Tail Prob.	t Value	Degrees of Freedom	2-Tail Prob.
1.12	.347	1.57	533	.117	1.57	531.63	.116

t-test for: TOT

	Number of Cases	Mean	Standard Deviation	Standard Error
WINTER	282	9.8830	2.635	.157
SPRING	253	9.5692	2.520	.158

	Pooled Variance Estimate			Separate Variance Estimate		
F Value	2-Tail Prob.	t Value	Degrees of Freedom	2-Tail Prob.	t Value	Degrees of Freedom	2-Tail Prob.
1.09	.468	1.40	533	.161	1.41	530.82	.160

Table of t-test Probability for 2-tail test
between
Groups K through X

ORGANIZATION

	K	L	M	N	O	P	Q	R	S	T	U	V	W	X
K	-	.16	.34	.88	.74	.21	**.00**	.83	.20	.34	.43	.40	.24	.54
L	.16	-	.55	.12	.17	.72	.12	.19	**.00**	.64	.39	.52	.98	**.04**
M	.34	.55	-	.27	.41	.76	**.01**	.41	**.01**	.93	.79	.94	.61	.09
N	.88	.12	.27	-	.61	.16	**.00**	.71	.30	.28	.35	.33	.20	.66
O	.74	.17	.41	.61	-	.24	**.00**	.92	**.04**	.40	.56	.49	.23	.27
P	.21	.72	.76	.16	.24	-	**.02**	.27	**.00**	.86	.54	.17	.76	**.04**
Q	**.00**	.12	**.01**	**.00**	**.00**	**.02**	-	**.00**	**.00**	**.03**	**.00**	**.01**	.12	**.00**
R	.83	.19	.41	.71	.92	.27	**.00**	-	.10	.42	.55	.49	.27	.38
S	.20	**.00**	**.01**	.30	**.04**	**.00**	**.00**	.10	-	**.01**	**.01**	**.01**	**.01**	.56
T	.34	.64	.93	.28	.40	.86	**.03**	.42	**.01**	-	.74	.87	.70	.10
U	.43	.39	.79	.35	.56	.54	**.00**	.55	**.01**	.74	-	.86	.44	.12
V	.40	.52	.94	.33	.49	.71	**.01**	.49	**.01**	.87	.86	-	.58	.12
W	.24	.98	.61	.20	.23	.76	.12	.27	**.01**	.70	.44	.58	-	.07
X	.54	**.04**	.09	.66	.27	**.04**	**.00**	.38	.56	.10	.12	.12	.07	-

Note: Significant at or above the .05 level indicated in **bold**.

Table of t-test Probability for 2-tail test
between
Groups K through X

DEVELOPMENT

	K	L	M	N	O	P	Q	R	S	T	U	V	W	X
K	-	.11	.89	.90	.09	.44	**.05**	.84	.28	.81	.46	.19	.72	.83
L	.11	-	**.04**	.09	**.00**	.25	.73	.16	**.00**	.16	.30	.77	.33	.13
M	.89	**.04**	-	.99	.06	.25	**.01**	.71	.24	.66	.28	.08	.58	.67
N	.90	.09	.99	-	.14	.39	**.04**	.74	.40	.72	.40	.17	.66	.73
O	.09	**.00**	.06	.14	-	**.00**	**.00**	.06	.37	**.05**	**.01**	**.00**	.07	**.04**
P	.44	.25	.25	.39	**.00**	-	[illegible]	.62	**.02**	.62	.97	.40	.83	.57
Q	**.05**	.73	**.01**	**.04**	**.00**	.10	[illegible]	.08	**.00**	.07	.14	.51	.19	**.05**
R	.84	.16	.71	.74	.06	.62	.08	-	.20	.97	.61	.27	.85	1.0
S	.28	**.00**	.24	.40	.37	**.02**	**.00**	.20	-	.15	**.03**	**.05**	.17	.14
T	.81	.16	.66	.72	**.05**	.62	.07	.97	.15	-	.63	.27	.87	.97
U	.46	.30	.28	.40	**.01**	.97	.14	.61	**.03**	.63	-	.46	.83	.57
V	.19	.77	.08	.17	**.00**	.40	.51	.27	**.05**	.27	.46	-	.46	.22
W	.72	.33	.58	.66	.07	.83	.19	.85	.17	.87	.83	.46	-	.84
X	.83	.13	.67	.73	**.04**	.57	**.05**	1.0	.14	.97	.57	.22	.84	-

Note: Significant at or above the .05 level indicated in **bold**.

Table of t-test Probability for 2-tail test
between
Groups K through X

MECHANICS

	K	L	M	N	O	P	Q	R	S	T	U	V	W	X
K	-	.94	.79	.07	.53	.63	.35	.54	.76	.30	.35	.44	.19	.59
L	.94	-	.84	.08	.57	.67	.38	.58	.69	.31	.30	.39	.20	.54
M	.79	.84	-	.10	.70	.81	.47	.71	.49	.38	.19	.25	.23	.41
N	.07	.08	.10	-	.20	.17	.35	.20	**.01**	.56	**.00**	**.01**	.87	**.03**
O	.53	.57	.70	.20	-	.90	.74	1.0	.27	.58	.09	.14	.38	.52
P	.63	.67	.81	.17	.90	-	.65	.90	.35	.52	.13	.18	.34	.32
Q	.35	.38	.47	.35	.74	.65	-	.74	.13	.80	**.04**	.06	.54	.15
R	.54	.58	.71	.20	1.0	.90	.74	-	.28	.59	.10	.14	.38	.26
S	.76	.69	.49	**.01**	.27	.35	.13	.28	-	.12	.36	.47	**.05**	.70
T	.30	.31	.38	.56	.58	.52	.80	.59	.12	-	**.05**	.07	.74	.14
U	.35	.30	.19	**.00**	.09	.13	**.04**	.10	.36	**.05**	-	.84	**.02**	.79
V	.44	.39	.25	**.01**	.14	.18	.06	.14	.47	.07	.84	-	**.03**	.91
W	.19	.20	.23	.87	.38	.34	.54	.38	**.05**	.74	**.02**	**.03**	-	.10
X	.59	.54	.41	**.03**	.52	.32	.15	.26	.70	.14	.79	.91	.10	-

Note: Significant at or above the .05 level indicated in **bold**.

Table of t-test Probability for 2-tail test
between
Groups K through X

TOTAL

	K	L	M	N	O	P	Q	R	S	T	U	V	W	X
K	-	.24	.66	.64	.75	.33	**.02**	.72	.27	.41	.78	.54	.29	.72
L	.24	-	.36	.52	.08	.69	.28	.41	**.01**	.78	.25	.49	.93	.11
M	.66	.36	-	.89	.36	.52	**.02**	.98	**.04**	.59	.84	.80	.38	.37
N	.64	.52	.89	-	.38	.72	.09	.88	.10	.74	.75	.96	.55	.39
O	.75	.08	.36	.38	-	.11	**.00**	.45	.33	.19	.46	.24	.11	.91
P	.33	.69	.52	.72	.11	-	.08	.58	**.01**	.97	.36	.70	.65	.14
Q	**.02**	.28	**.02**	.09	**.00**	.08	-	**.05**	**.00**	.18	**.01**	**.03**	.41	**.00**
R	.72	.41	.98	.88	.45	.58	**.05**	-	.12	.62	.88	.82	.44	.46
S	.27	**.01**	**.04**	.10	.33	**.01**	**.00**	.12	-	**.03**	**.05**	**.02**	**.01**	.50
T	.41	.78	.59	.74	.19	.97	.18	.62	**.03**	-	.46	.74	.75	.22
U	.78	.25	.84	.75	.46	.36	**.01**	.88	**.05**	.46	-	.62	.27	.46
V	.54	.49	.80	.96	.24	.70	**.03**	.82	**.02**	.74	.62	-	.48	.27
W	.29	.93	.38	.55	.11	.65	.41	.44	**.01**	.75	.27	.48	-	.15
X	.72	.11	.37	.39	.91	.14	**.00**	.46	.50	.22	.46	.27	.15	-

Note: Significant at or above the .05 level indicated in **bold**.

Appendix E, Random Analysis of Raters' Consistency

Frequency Distributions by Evaluator
for
Combined Winter and Spring 90 Groups

ORGANIZA	BEAR	FOX	GOAT	(Total)
1	4	0	7	11
2	29	1	30	60
3	73	6	80	159
4	108	3	107	218
5	39	5	29	73
6	6	0	8	14
	259	15	261	535

DEVELOPM	BEAR	FOX	GOAT	(Total)
1	7	0	8	15
2	52	3	55	110
3	116	3	117	236
4	56	4	54	114
5	22	4	19	45
6	6	1	8	15
	259	15	261	535

MECHANIC	BEAR	FOX	GOAT	(Total)
1	16	0	21	37
2	68	4	66	138
3	109	6	107	222
4	51	4	54	109
5	12	0	12	24
6	3	1	1	5
	259	15	261	535

TOT	BEAR	FOX	GOAT	(Total)
3	3	0	4	7
4	1	0	2	3
5	3	0	6	9
6	14	1	12	27
7	21	0	27	48
8	37	1	34	72
9	50	3	43	96
10	37	4	46	87
11	34	1	31	66
12	26	1	28	55
13	12	1	9	22
14	9	2	8	19
15	6	0	4	10
16	3	0	4	7
17	1	1	2	4
18	2	0	1	3
	259	15	261	535

Computation of Central Tendencies for Bear

ORGANIZA	Count =	259
	Rows =	535
	Minimum =	1
	Maximum =	6
	Sum =	944
	Average =	4
	Std Dev =	0
	Variance =	0
DEVELOPM	Count =	259
	Rows =	535
	Minimum =	1
	Maximum =	6
	Sum =	829
	Average =	3
	Std Dev =	1
	Variance =	1
MECHANIC	Count =	259
	Rows =	535
	Minimum =	1
	Maximum =	6
	Sum =	761
	Average =	3
	Std Dev =	0
	Variance =	0

Total	Count =	259
	Rows =	535
	Minimum =	3
	Maximum =	18
	Sum =	2534
	Average =	10
	Std Dev =	2
	Variance =	6

Computation of Central Tendencies for Goat

ORGANIZA	Count =	261
	Rows =	535
	Minimum =	1
	Maximum =	6
	Sum =	928
	Average =	4
	Std Dev =	1
	Variance =	1
DEVELOPM	Count =	261
	Rows =	535
	Minimum =	1
	Maximum =	6
	Sum =	828
	Average =	3
	Std Dev =	1
	Variance =	1
MECHANIC	Count =	261
	Rows =	535
	Minimum =	1
	Maximum =	6
	Sum =	756
	Average =	3
	Std Dev =	0
	Variance =	0
TOT	Count =	261
	Rows =	535
	Minimum =	3
	Maximum =	18
	Sum =	2512
	Average =	10
	Std Dev =	2
	Variance =	6

CROSS CORRELATION OF EACH SCALE VARIABLE
TO
ALL OTHER SCALE VARIABLES

Correlations:	ORGANIZA	DEVELOPM	MECHANIC	TOTAL
ORGANIZA	1.0000	.7087**	.5359**	.8828**
DEVELOPM	.7087**	1.0000	.4907**	.8675**
MECHANIC	.5359**	.4907**	1.0000	.7934**
TOTAL	.8828**	.8675**	.7934**	1.0000

N of cases: 520 1-tailed Signif: * - .01 ** - .001

" . " is printed if a coefficient cannot be computed

t-test for: ORGANIZA

	Number of Cases	Mean	Standard Deviation	Standard Error
Section K	36	3.4444	1.206	.201
Section L	40	3.8250	1.130	.179
Section M	38	3.6842	.904	.147
Section N	40	3.4000	1.297	.205
Section O	44	3.5227	.876	.132
Section P	39	3.7436	.818	.131
Section Q	38	4.1579	.638	.103
Section R	42	3.5000	1.110	.171
Section S	40	3.1500	.770	.122
Section T	34	3.7059	1.060	.182
Section U	38	3.6316	.786	.127
Section V	33	3.6667	.924	.161
Section W	22	3.8182	1.053	.224
Section X	36	3.2778	1.111	.185

t-test for: DEVELOPM

	Number of Cases	Mean	Standard Deviation	Standard Error
Section K	36	3.1111	1.141	.190
Section L	40	3.5250	1.062	.168
Section M	38	3.0789	.784	.127
Section N	40	3.0750	1.289	.204
Section O	44	2.7045	.978	.147
Section P	39	3.2821	.759	.122
Section Q	38	3.6053	.946	.153
Section R	42	3.1667	1.228	.189
Section S	40	2.8750	.723	.114
Section T	34	3.1765	1.058	.181
Section U	38	3.2895	.898	.146
Section V	33	3.4545	.971	.169
Section W	22	3.2273	1.270	.271
Section X	36	3.1667	.971	.162

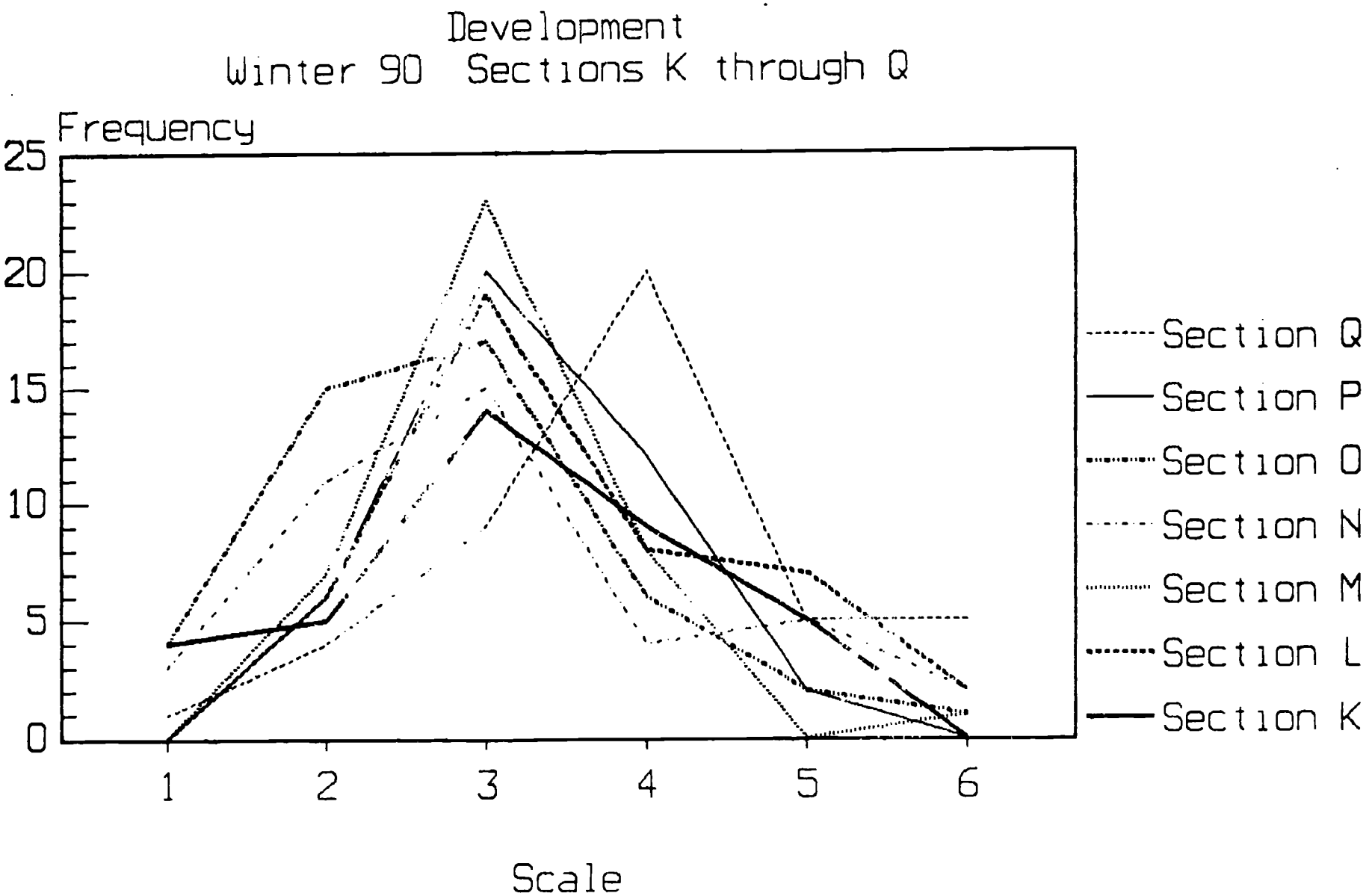
t-test for: MECHANIC

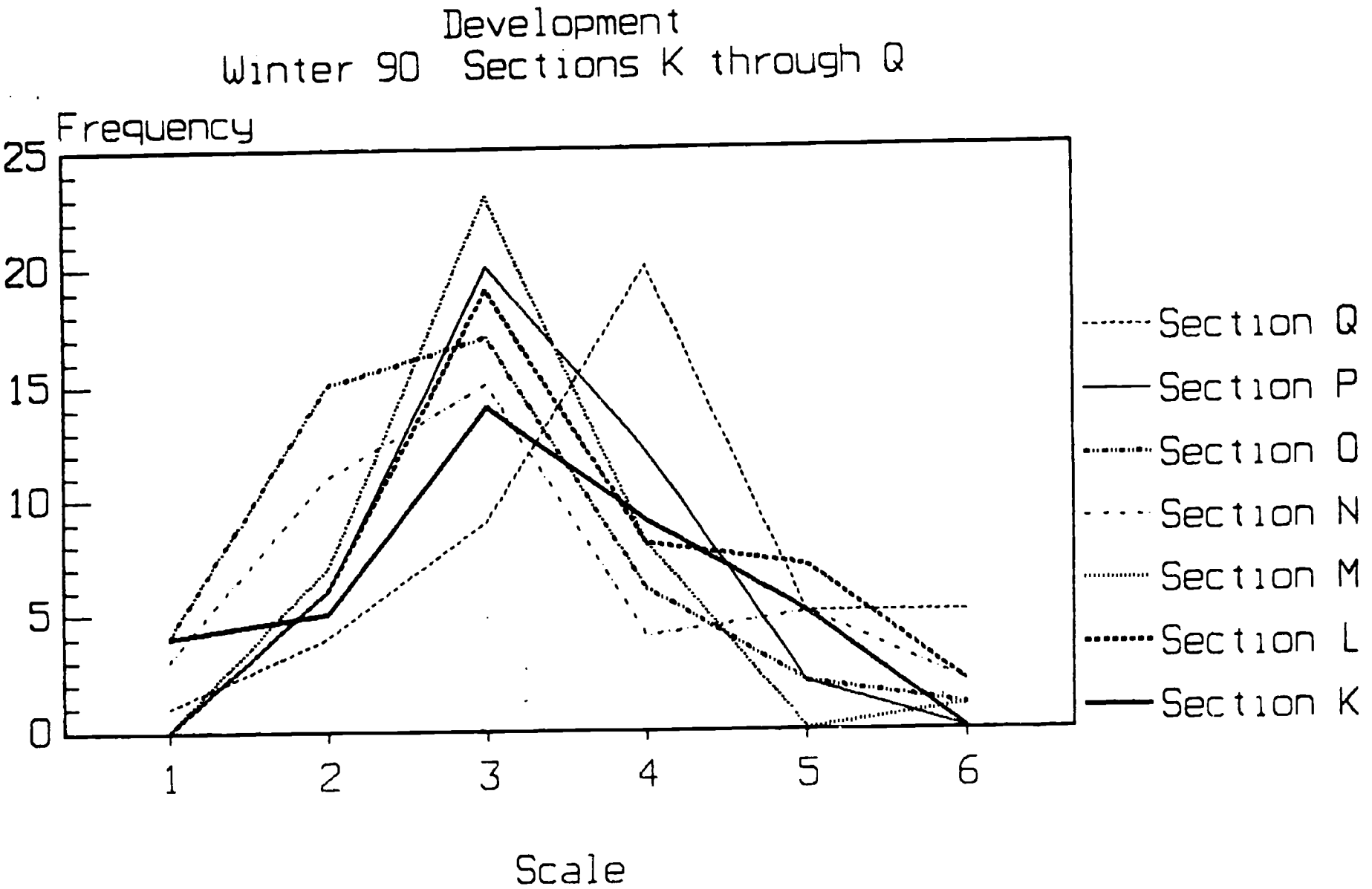
	Number of Cases	Mean	Standard Deviation	Standard Error
Section K	36	2.8333	1.028	.171
Section L	40	2.8500	1.027	.162
Section M	38	2.8947	.924	.150
Section N	40	3.2750	1.086	.172
Section O	44	2.9773	1.023	.154
Section P	39	2.9487	1.025	.164
Section Q	38	4.1579	.638	.103
Section R	42	2.9762	1.024	.158
Section S	40	2.7750	.577	.091
Section T	34	3.1176	1.225	.210
Section U	38	2.6316	.786	.127
Section V	33	2.6667	.692	.120
Section W	22	3.2273	1.193	.254
Section X	36	2.6944	1.167	.194

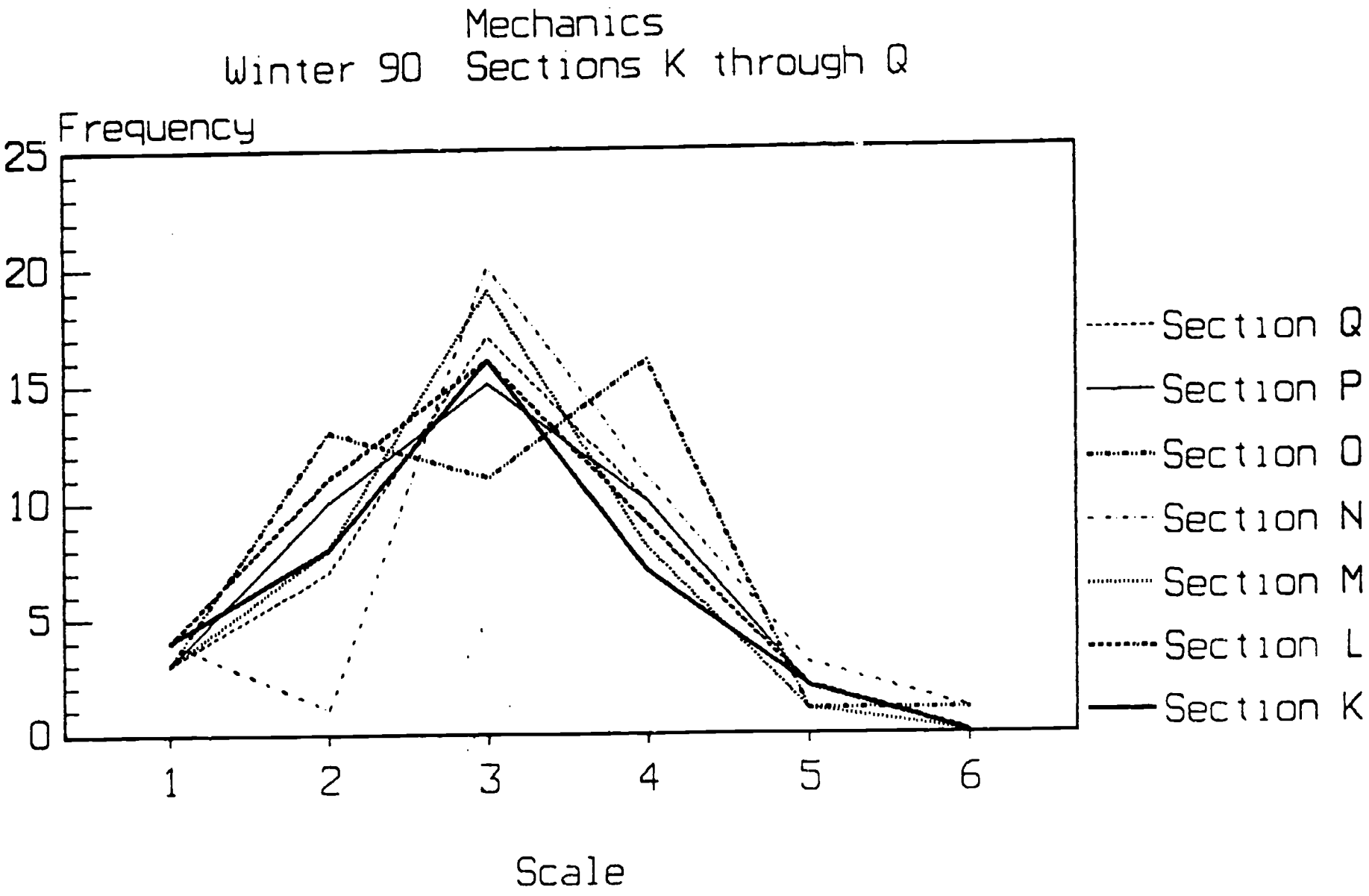
t-test for: TOTAL

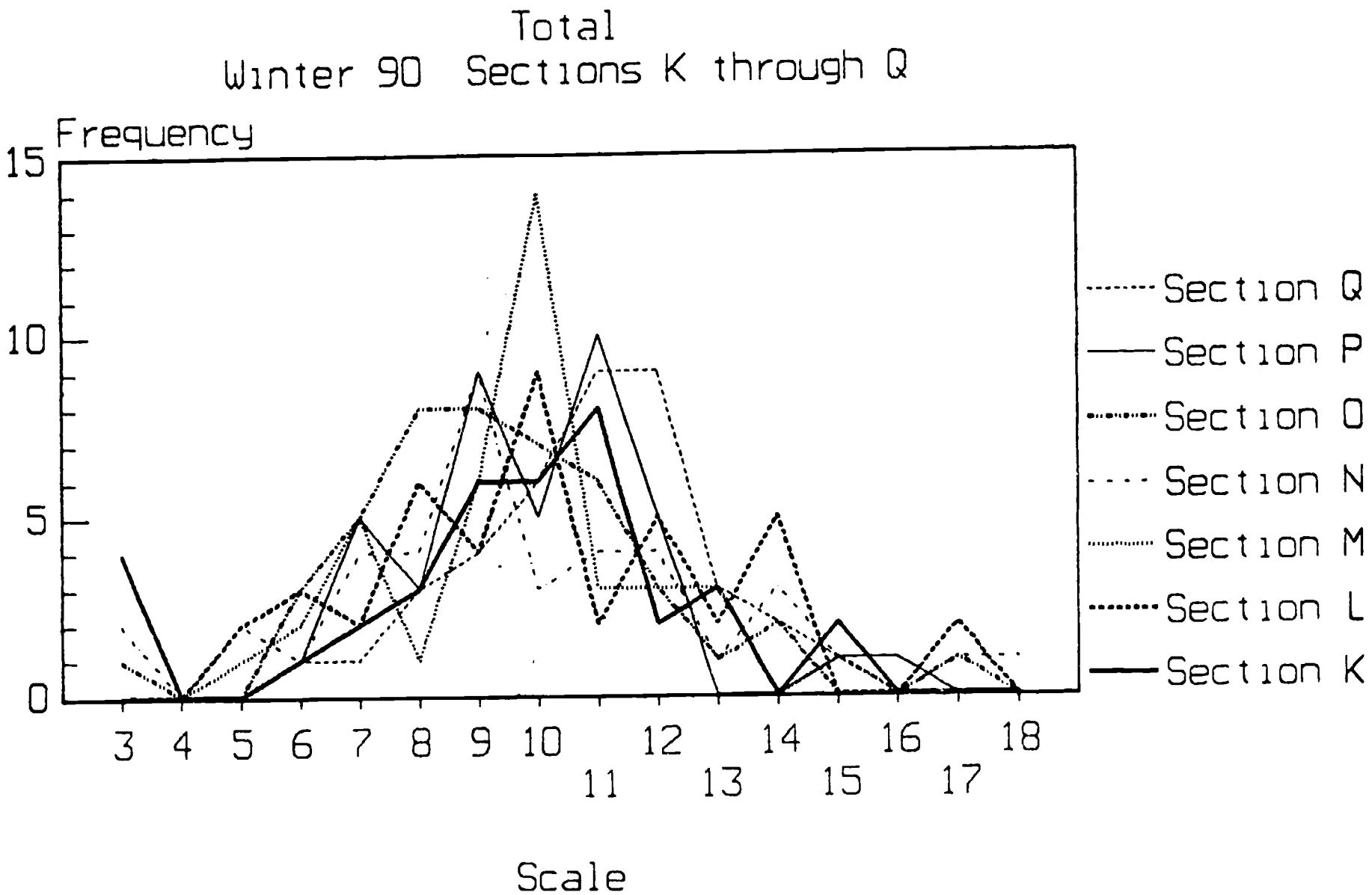
	Number of Cases	Mean	Standard Deviation	Standard Error
Section K	36	9.3889	3.017	.503
Section L	40	10.2000	2.902	.459
Section M	38	9.6579	2.172	.352
Section N	40	9.7500	3.349	.530
Section O	44	9.2045	2.226	.336
Section P	39	9.9744	2.109	.338
Section Q	38	10.8158	1.971	.320
Section R	42	9.6429	3.122	.482
Section S	40	8.8000	1.436	.227
Section T	34	10.0000	3.094	.531
Section U	38	9.5526	1.927	.313
Section V	33	9.7879	2.012	.350
Section W	22	10.2727	3.042	.649
Section X	36	9.1389	2.758	.460

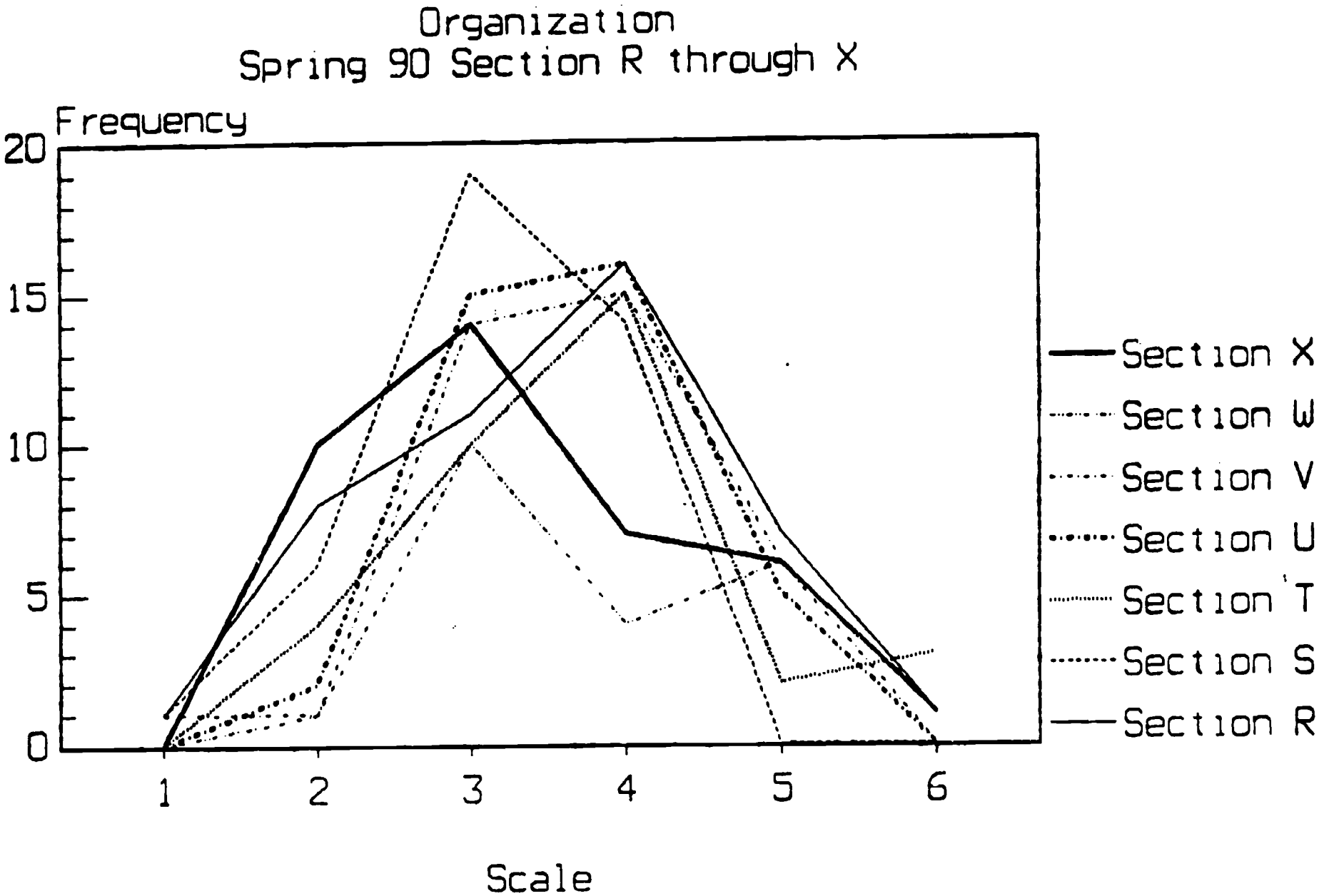
Appendix F, Student Outcomes Frequency Polygons by Section and Dimension

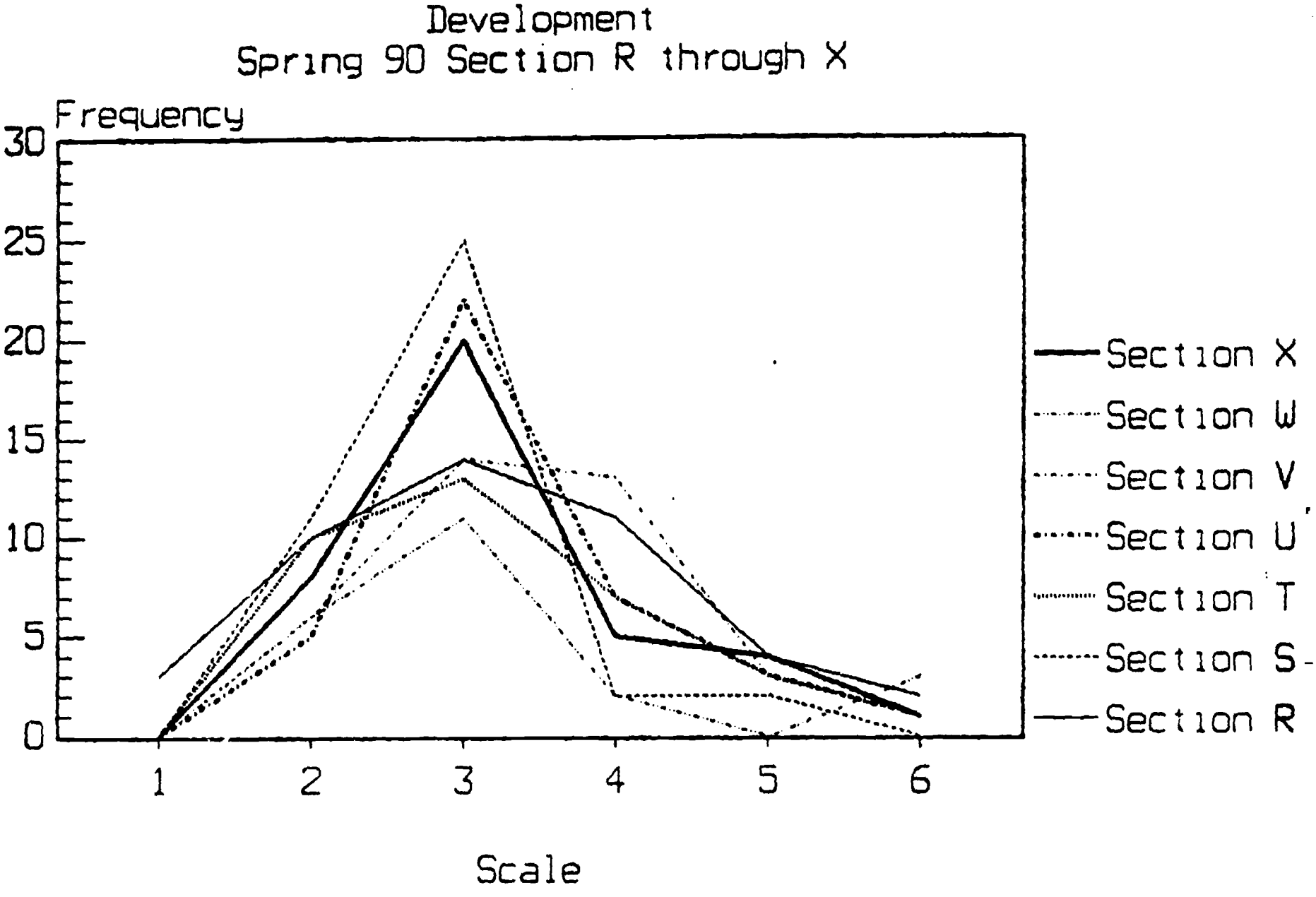

Development
Winter 90 Sections K through Q
Frequency
25
20
15
10
5
0
1
2
3
4
5
6
Scale
Section Q
Section P
Section O
Section N
Section M
Section L
Section K

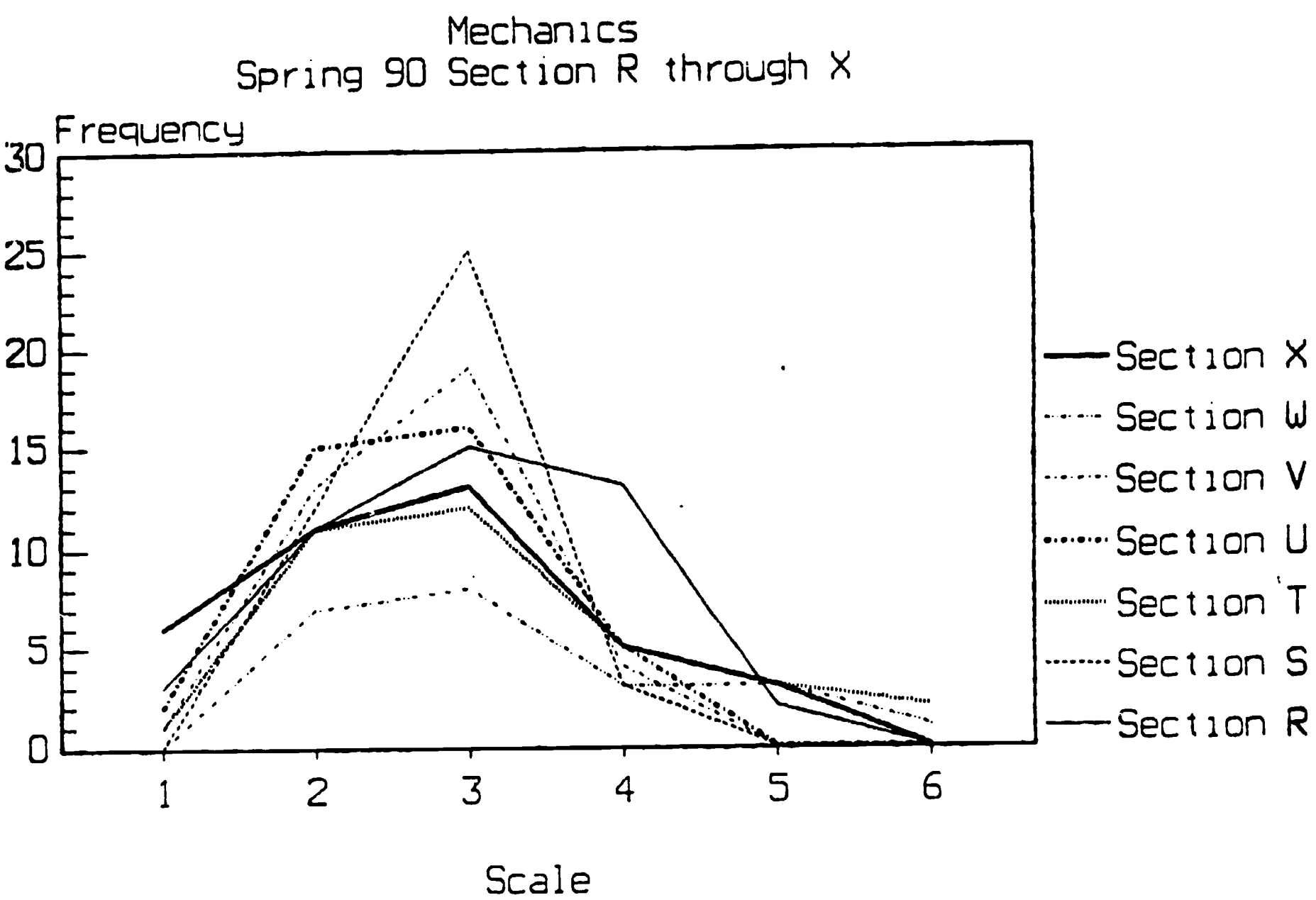

Development
Winter 90 Sections K through Q
Frequency
25
20
15
10
5
0
1
2
3
4
5
6
Scale
Section Q
Section P
Section O
Section N
Section M
Section L
Section K


Mechanics
Winter 90 Sections K through Q
Frequency
25
20
15
10
5
0
1
2
3
4
5
6
Scale
Section Q
Section P
Section O
Section N
Section M
Section L
Section K


Total
Winter 90 Sections K through Q
Frequency
15
10
5
0
3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18
Scale
Section Q
Section P
Section O
Section N
Section M
Section L
Section K


Organization
Spring 90 Section R through X
Frequency
20
15
10
5
0
1
2
3
4
5
6
Scale
Section X
Section W
Section V
Section U
Section T
Section S
Section R


Development
Spring 90 Section R through X
Frequency
30
25
20
15
10
5
0
1
2
3
4
5
6
Scale
Section X
Section W
Section V
Section U
Section T
Section S
Section R


Mechanics
Spring 90 Section R through X
Frequency
30
25
20
15
10
5
0
1
2
3
4
5
6
Scale
Section X
Section W
Section V
Section U
Section T
Section S
Section R