

MEMORANDUM

TO: Faculty Members,
Political Science Department

FROM: David R. Elkins
Chair and Associate Professor
Political Science Department

DATE: November 3, 2005

RE: International Relations Major and Political Science Major Student Assessment
Results for Spring 2005

Overview

This report summarizes the Department of Political Science's spring 2005 student assessment for Political Science Majors and International Relations Majors. In brief, the six members of the faculty that reviewed the eighteen randomly selected senior seminar papers found that three-quarters (75.6%) either met or exceeded departmentally established expectations. The mean and median score for this semester's student assessment is 3.17 and 3, respectively. A score of 3 indicates meeting Departmental expectations. The paired faculty reviewers had a modest level of agreement among their assessments ($r = .22$), which represents a continuing decline from the previous semesters.

Student Assessment Process

The Department's student assessment process required the faculty members teaching the spring 2005 senior seminars (PSC 420 American Politics and PSC 421 Comparative Politics) to submit copies of all senior seminar term papers to the student assessment coordinator. The senior seminar papers were un-graded, un-marked, and anonymous versions of the papers submitted to the instructor of record for a grade. A total of twenty-

nine papers from the spring 2005 senior seminars were submitted for student assessment review (PSC 420 = 14 and PSC 421 = 15).

Eighteen of the twenty-nine papers were randomly selected for review (62.1%). Six faculty members reviewed the papers, and the reviewers were paired. In addition, each reviewer received a student assessment packet that included six randomly assigned seminar papers, six Senior Seminar Student Assessment forms, and one Senior Seminar Student Assessment Criteria Explanation form on May 18, 2005. The due date for the reviews was August 23, 2005. An oral reminder was issued during a faculty meeting on August 24, 2005 and one written reminder was distributed on August 31, 2005. The last set of reviews was submitted in late September 2005.

Two technical issues complicated this iteration of student assessment. First, the student assessment coordinator erroneously sent out an evaluation form that included the Diction component in the Articulation and Communication Criteria. In fall 2004 this criterion was eliminated from the Department's assessment matrix in compliance with a suggestion from the Office of Assessment. Some faculty completed this criteria component and some did not thus leaving missing data. In coding this semester's assessment the Diction criteria component was eliminated from calculation of both diagnostic and assessment statistics. The second technical issue relates to more missing data. In five instances a faculty reviewer left blank an assessment criteria component.¹ In these instances, these criteria components were excluded from the calculations of both diagnostic and assessment statistics. It is unlikely that these omissions affected the results of the spring 2005 student assessment.

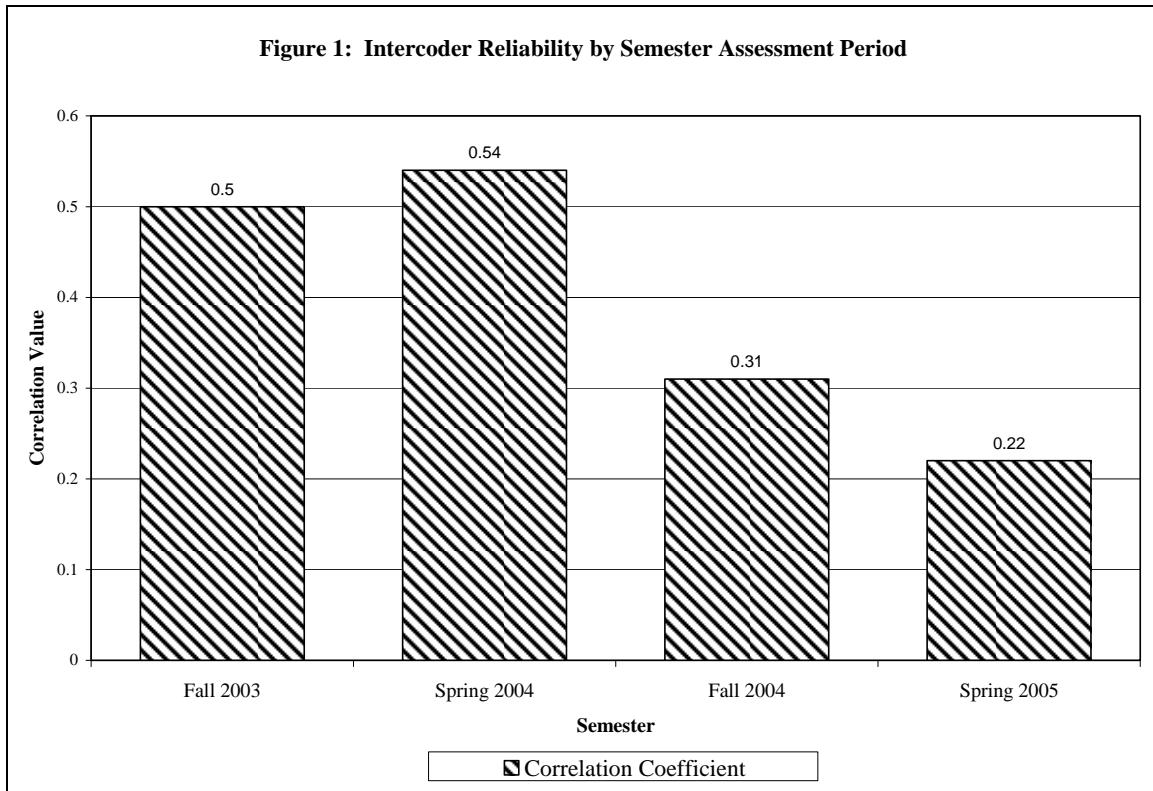
¹ Two in the Evidence component and one each in Organization, Citations, and Conclusions.

Findings

This report describes two issues. It describes the diagnostics of the spring 2005 assessment methods and it describes the results from this iteration of student assessment.

Diagnostics. Assessment diagnostics explains the level of agreement and disagreement that were recorded in the spring 2005 assessment process. The inter-coder reliability was disappointing ($r = .22$). This represents an ongoing decline from previous iterations of student assessment. Figure 1 illustrates the levels of inter-coder reliability as measured by a Pearson's r correlation coefficient for three previous semesters and the spring 2005 student assessment. This decline represents a worrisome pattern that must be addressed. Fluctuation in personnel, additions and losses of faculty members, coupled with the episodic participation of some faculty reviewers suggests a need to revisit faculty understanding of assessment guidelines and measurements. At the end of last academic year the Department decided that it would need to convene a specific meeting to review assessment, particularly focusing on crafting further guidelines to improve inter-coder reliability (see attached memorandum). The source of this decline in inter-coder reliability is directly related to disagreements among faculty reviewers during assessment.

Disagreements frequently occur in this student assessment process. Assessment disagreement is defined in several ways. First, there are major and minor disagreements. A minor disagreement is where one of the paired reviewers differs by one point on a dimensional component with his or her reviewing pair. For instance, if one reviewer scores a dimension a 3 while his or her pair scores it a 2, this is a minor disagreement. A major



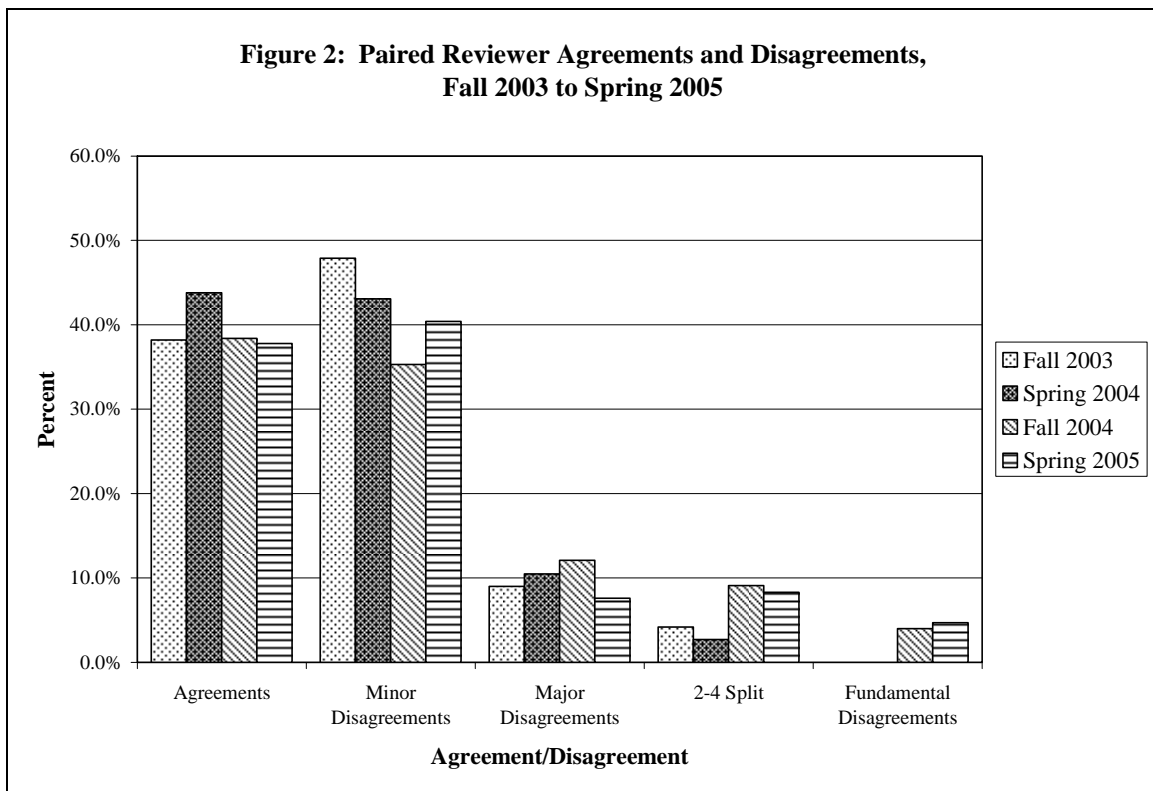
disagreement is where the split between the paired reviewers is greater than one. For instance, one reviewer scores a dimensional component a 3 while his or her pair scores the same dimensional component a 5. In addition, I divide disagreements into high and low categories. A high category disagreement is when at least one reviewer indicated that a dimensional component exceeded expectations. By contrast, a low category disagreement indicates that at least one paired reviewer found that a dimensional component did not meet expectations. Consequently, in the first example above, where one reviewer found a dimensional component a 3 while his or her counterpart gave it a 2 would be defined as a Minor Low Disagreement. The other example, where one reviewer found a dimensional component met expectations (a score of 3) and his or her pair scored that dimensional component a 5 would be defined as a Major High Disagreement.

Finally, there are two additional classes of disagreements that pose particularly difficult problems with inter-coder reliability and I treated differently, though they met the definition above. The first is the 2-4 Split. A 2-4 Split disagreement is, by definition, a major disagreement. However, it is one where the reviewers split on whether a dimensional component exceeded expectations, a score of 4, and did not meet expectations, a score of 2. The other category is even more problematic. In this category, what I call Fundamental Split Disagreements, the split is by three or more points and indicates that a fundamental disagreement exists between the paired reviewers about a dimensional component. For instance, one reviewer scoring a component with a 5 while his or her pair scores it 1 is the prime example of a fundamental split.

Table 1 depicts the frequency of agreements and disagreements of paired assessment reviewers for the spring 2005 assessment cycle. Nearly three times out of eight (37.8%, n = 73), the paired reviewers agree in their assessment of discrete components. Agreement is highest among reviewers in the Articulation and Communication Criteria and lowest among the Critical and Analytical Criteria, particularly the Hypothesis component. Despite the number of agreements, reviewers are more likely to disagree while assessing students' senior seminar papers. Nearly two-fifths (40.4%, n = 77) of the disagreements are minor with a slight tendency for reviewers to disagree on whether the component met or exceeded expectations. While nearly four out of five reviewers either agreed or disagreed in a minor fashion, one in five have a significant disagreement of some sort or another (Major, Split, or Fundamental). The majority of these forms of disagreements were in the Critical and Analytical Criteria. Over a quarter of reviewers (28.9%, n = 20) have non-minor forms of disagreements in this criteria whereas less than one in five have non-minor disagreements in

either the Research Criteria (18.9%, n = 10) or Articulate and Communicate (15.5%, n = 11). Removing these significant disagreements from the calculation of inter-coder reliability renders an $r = .67$. The problem is that a fraction of reviewer disagreements, particularly in the Critical and Analytical Criteria, is having a non-trivial impact on inter-coder reliability.²

Figure 2 depicts the levels of agreement or disagreement over four assessment periods. The fact is that a relatively small shift in the number of significant disagreements over time has created problems with inter-coder reliability. For instance, as noted in Figure 1, the assessment periods fall 2003 and spring 2004 had correlations greater than or equal to .5 while fall 2004 and spring 2005 have had correlations less than .35. In this same division



² The problems of inter-coder reliability associated with disagreements exist across the three teams of paired reviewers. The three teams have correlations of .23, .19, and .23.

of time, the proportion of agreements and minor disagreements for the 2003/2004 period was around 85% (86.1%, fall 2003 and 86.9%, spring 2004). By contrast, the proportion of agreements and minor disagreements in the 2004/2005 period dropped to less than 80% (73.7%, fall 2004 and 78.2%, spring 2005). This relatively small shift, approximately 10%, has created a decline in the inter-coder reliability. By moving a handful of non-minor disagreements into the minor disagreement or agreement categories, the Department can significantly improve the inter-coder reliability of its student assessment procedures.

Findings. Table 2 depicts the results of the spring 2005 student assessment for International Relations Majors and Political Science Majors. The reviewers, in an unpaired frequency analysis, found that over three-quarters (75.6%, n = 292) of criteria components either met or exceeded the Department's expectations. In a paired analysis using the means of two reviewers to represent component scores, the papers' mean score is 3.17 with a median score of 3.

The reviewers were most satisfied with the Articulate and Communicate Criteria with well over four-fifths (83.8%, n = 119) of the papers either meeting or exceeding the Department's expectations. Still, the reviewers found that over half (51.9%, n = 55) of the papers exceeded the Department's expectations in the Research Criteria. However, a consistent finding among these student assessments has been relative disappointment in the Critical and Analytical Criteria. In spring 2005 assessment over a third (34.1%, n = 47) of papers did not meet the Department's expectations.

The findings for the Critical and Analytical Criteria indicate mixed results. The reviewers were very satisfied with the Thesis component with nearly six out of every seventh

Table 2: Frequency Distribution of Scores and Average of Scores for Assessment Papers, Spring 2005

Dimension	<i>Frequency of Scores</i> ¹			Average ²
	Exceeds Expectations (5 or 4)	Meets Expectations (3)	Does Not Meet Expectations (2 or 1)	
<i>Critical and Analytical Criteria</i>				
Thesis	16	15	5	3.42
Hypothesis	13	9	14	3.03
Evidence	6	11	15	2.56
Conclusions	9	12	13	2.79
<i>Criteria Subtotal</i>	<i>31.9%</i> <i>(44)</i>	<i>34.1%</i> <i>(47)</i>	<i>34.1%</i> <i>(47)</i>	<i>2.94</i>
<i>Research Criteria</i>				
Sources	20	12	4	3.52
Citations	13	10	11	3.09
Bibliography	22	5	9	3.5
<i>Criteria Subtotal</i>	<i>51.9%</i> <i>(55)</i>	<i>25.5%</i> <i>(27)</i>	<i>22.6%</i> <i>(24)</i>	<i>3.38</i>
<i>Articulate and Communicate Criteria</i>				
Organization	11	17	6	3.26
Paragraphs	12	22	2	3.33
Sentence Structure	12	18	6	3.22
Grammar	14	13	9	3.11
<i>Criteria Subtotal</i>	<i>34.5%</i> <i>(49)</i>	<i>49.3%</i> <i>(70)</i>	<i>16.2%</i> <i>(23)</i>	<i>3.23</i>
TOTAL	38.3% (148)	37.3% (144)	24.4% (94)	3.17

N = 386 (11 Dimensions × (2 Reviewers × 18 Papers)) – 10 Missing. The missing data comes from the five dimensions that were deleted due to incomplete response of at least one of paired reviewers.

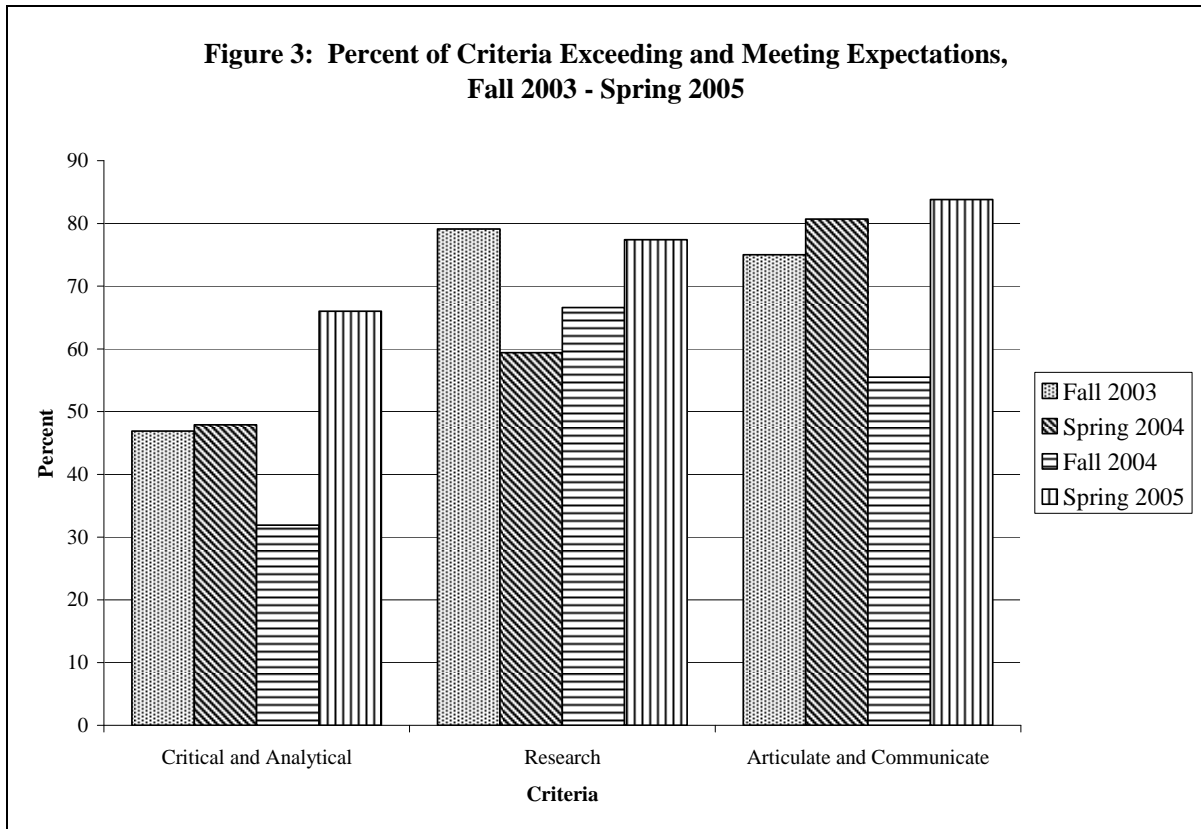
1 = The frequency of scores columns represent the rankings that each faculty members gave to a paper. In this portion of the analysis the scores are treated as discrete and not paired measures.

2 The arithmetic average was derived by establishing a mean for each dimension for each paper. An average was created for a total criteria.

(86.1%, n = 31) paper either meeting or exceeding the Department's expectations. By contrast, reviewers were slightly split on the Hypothesis component. It is as if the students either exceeded reviewer's expectations or did not meet them. Still, the majority (61.1%, n = 22) of the reviewers judged the papers either met or exceeded expectations for the Hypothesis component. When it turns to the Evidence component, the reviewers clearly indicate that the students are not making a convincing case for either their thesis or hypothesis. The modal (n = 15) response in this component is in the Does Not Meet Expectations category. This is followed by the Conclusions component and it suggests a similar representation of the empirical evidence with its modal category (n = 13) falling in the Does Not Meet Expectations column. Despite the fact that the overall results for to Critical and Analytical Criteria are defined as either meeting or exceeding Department's expectation, there remains a sense of disappointment. It appears that the reviewers are finding that students are able to form coherent theses and, the majority of time, form appropriate and acceptable hypotheses, but they are not able to test them convincingly or to draw appropriate conclusions from their apparently muddled evidence.

Despite the frustrations with the Critical and Analytical Criteria, the results of this iteration of student assessment demonstrate improvement. Figure 3 illustrates four semesters of student assessment results for the three departmentally defined criteria.³ It

³ Spring 2003 student assessment results are omitted from the depiction in Figure 3 because they are not directly comparable. The spring 2003 student assessment was the first attempt at this model of student assessment, and the measurement instrument of assessment was altered after spring 2003 from a three category to the current five category measurement procedure.



is premature to state whether a pattern of improvement is emerging, but the results are encouraging. The fall 2004 results indicate a retrenchment from some forward movement and thus is a limiting factor in determining whether improvement is taking place. It is also unclear, if improvement is occurring, as to why improvement is occurring. For Research and Articulate and Communicate Criteria, there has been relative satisfaction of performance. By contrast, there has been consistent frustration with the Critical and Analytical Criteria. The spring 2005 semester's improvement in this criterion suggests two explanations. One is students are getting better at meeting our expectations and the other is faculty members are getting better communicating our expectations to them both in senior seminars and in other courses we teach. I suspect a little of both is happening. Still, the results are encouraging.

Conclusions

The result of the spring 2005 student assessment for Political Science Majors and International Relations Majors demonstrates some continuing problems but promising results. The on-going problems relate to the challenges we continue to confront with inter-coder reliability. The decline of agreement among paired reviewers suggests a response. Late last spring, following a discussion of the fall 2004 student assessment results, the Department proposed that we convene a meeting, perhaps in the form of a retreat, to examine the continuing problems confronted with the Critical and Analytical Criteria of student assessment (see attached memorandum). Specifically, the idea is to focus on improving collective understanding of what we believe to be appropriate indicators of meeting components of this criterion.

Despite this continuing challenge there were hopeful results in this student assessment. The overall assessment scores suggest that we are finding results more in line with our expectations. Indeed, even with the challenges we continue to face in the Critical and Analytical Criteria, the results show promise. Most Political Science Majors and International Relations Major are meeting our expectations for establishing a thesis and many are meeting or exceeding our expectations for the creation of a hypothesis. Many students remain less successful in validating their hypothesis to a satisfactory degree through the evidence they present and they are not drawing conclusions we believe meet our expectations. Still, headway is being made.

OUTCOME ASSESSMENT REPORT

Fall 2005/Spring 2006

Department of Political Science

Political Science Majors and International Relations Majors

Introduction: The Political Science Department conducts outcome assessment for two majors: Political Science and International Relations. The Department conducts its assessment through a careful examination of consensually determined criteria and in a methodologically appropriate manner. The Department frequently discusses the results of the results of its assessments during regularly scheduled faculty meetings. The Department has made modifications to both its methodology and its two majors, particularly the International Relations major.

Goals and Outcomes: The goals for the Political Science Department's Political Science Major and International Relations Major assessment are to have students demonstrate critical and analytic thinking, to engage in proper research, and effective communication. The outcomes of these goals are as follows:

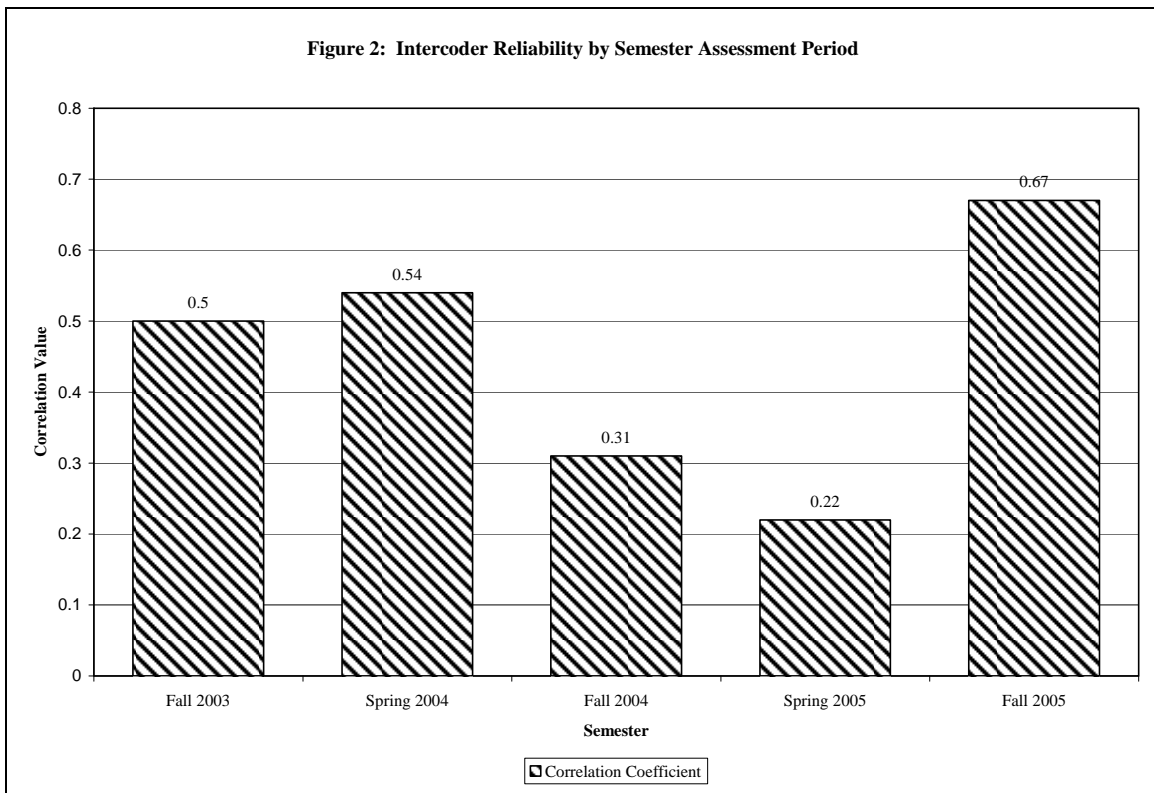
Table 1: Assessment Goals for International Relations Majors and Political Science Majors	
DIMENSION	EXPECTATIONS
<i>Critical and Analytical Criteria Goals</i>	
Thesis Component	Clearly articulate thesis
Hypothesis Component	Research question or hypothesis is clearly formulated.
Evidence Component	Evidence is generally appropriate.
Conclusions Component	Draws appropriate conclusions.
<i>Research Criteria Goals</i>	
Sources Component	Five to ten scholarly sources cited or combination of scholarly resources, government documents, interviews, foreign news sources, and articles from newspapers of record.*
Citations Component	Appropriate citations (footnotes, endnotes, embedded)
Bibliography Component	Properly organized bibliography.
<i>Communication Criteria Goals</i>	
Organization Component	Good organization.
Paragraphs Component	Consistently developed paragraphs.
Sentence Structure Component	Concise sentence structure.
Grammar Component	Infrequent grammatical errors.
* = Non-scholarly sources (government documents, interviews, specialized news sources, and article from newspapers have a multiplier of five. This means that to "count" for a scholarly source, a student must provide five non-scholarly sources. For example, one article from the <i>American Political Science Review</i> will have the equivalent weight of five newspaper articles from <i>The New York Times</i> .	

The goals and outcomes were developed in a subcommittee consisting of Dr. Schulz, Dr. Charlick and Dr. Elkins during the spring of 2003. The goals and outcomes were presented to the entire Department for approval. The goals and outcomes were refined by the full department and approved by the department in late spring of 2003. The goals and have been modified once since the approval. The 2004 Assessment Report indicated “We question how you are measuring “diction” using a written paper” (Page 6). The department agreed with the Assessment Team’s suggestion and removed the criteria as a goal.

Research Methods: The method of assessment is based on students demonstrating outcomes as indicated by their final papers in the Department’s senior seminars. The faculty members teaching a senior seminar submit unmarked and anonymous final papers from each senior seminar to the chair. The student assessment coordinator randomly selects from these papers a representative sample to distribute to paired teams of reviewers. The reviewers are fulltime faculty members in the Political Science Department that did not teach senior seminars (faculty members that did teach senior seminars are excluded from the pool of reviewers). The reviewers are randomly assigned to two-person review teams. The reviewers assess each paper using an instrument measuring the outcomes of the discrete components of the goals (see Appendix A Electronic Attachment). The measurement instrument was modified between the spring and the fall of 2003 to increase the level of inter-coder reliability by increasing the number of measurement categories:

Figure 1: Illustration of Spring 2003 and Fall 2003 to Present Student Assessment Instruments				
Spring 2003 Measurement Instrument				
Exceeds Expectations		Meets Expectations		Does Not Meet Expectations
3		2		1
Fall 2003 to Current Measurement Instrument				
Exceeds Expectation		Meets Expectations	Does Not Meet Expectations	
5	4	3	2	1

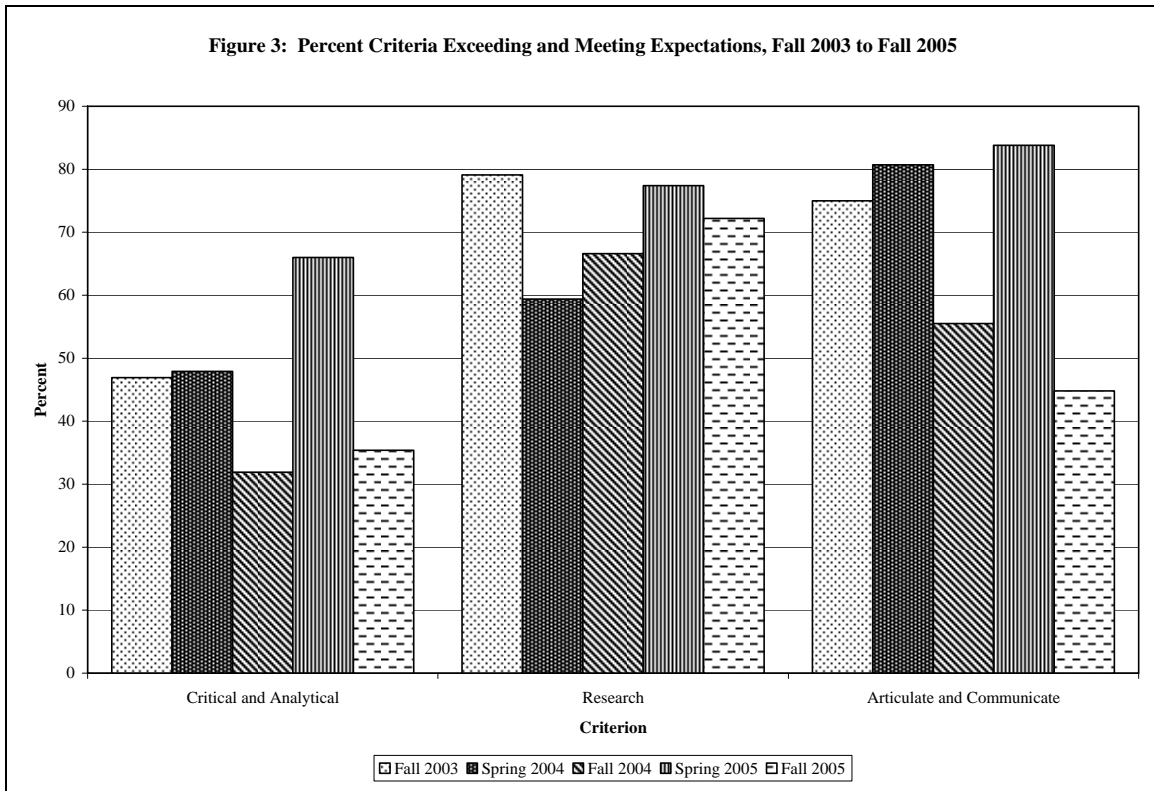
The Review of the Department’s assessment instrument by the Office of Assessment indicated that the rating instrument should have only three categories (Research 8. “Consider having three rates for each paper.”) The Department has only three rates for each paper, as indicated above. However, it allows faculty member reviewers to select scoring categories that indicated a feature of an assessment component that may not completely meet its target expectation, while granting that it either excelled or was deficient in some degree of this feature. As figure 2 illustrates, the result has been a modest, but inconsistent, improvement in the inter-coder reliability scores for the teams of reviewers



As noted above, the department altered its measurement instrument in the fall of 2005 to correspond to suggestions made by the Office of Assessment. The Office of Assessment noted, “We question how you are measuring “diction” using a written paper.” The Department eliminated this feature from the rating form.

Findings: The empirical data are compiled by the assessment coordinator in a numeric format based on the results evaluations provided by the assessment review teams. The assessment coordinator writes an assessment report outlining the methods, diagnostics of the analysis, and the results of the assessment. The Political Science Department has produced five reports based on its current assessment process.¹ Figure 3 illustrates the findings of five iterations of student assessment and, in general, indicates that the majority of Political Science Majors and International Relations Majors are meeting or exceeding departmentally established goals. The department is currently conducting its review of senior seminar final papers for the spring 2006 semester. The report will be distributed in the fall of 2006 and discussed at the first departmental faculty meeting.

¹ Spring 2005 and Fall 2005 Assessment Reports were submitted along with this document. Earlier Assessment Reports are available upon request to the Department of Political Science. The reports can only be provided in hardcopy due to the loss of digital versions.

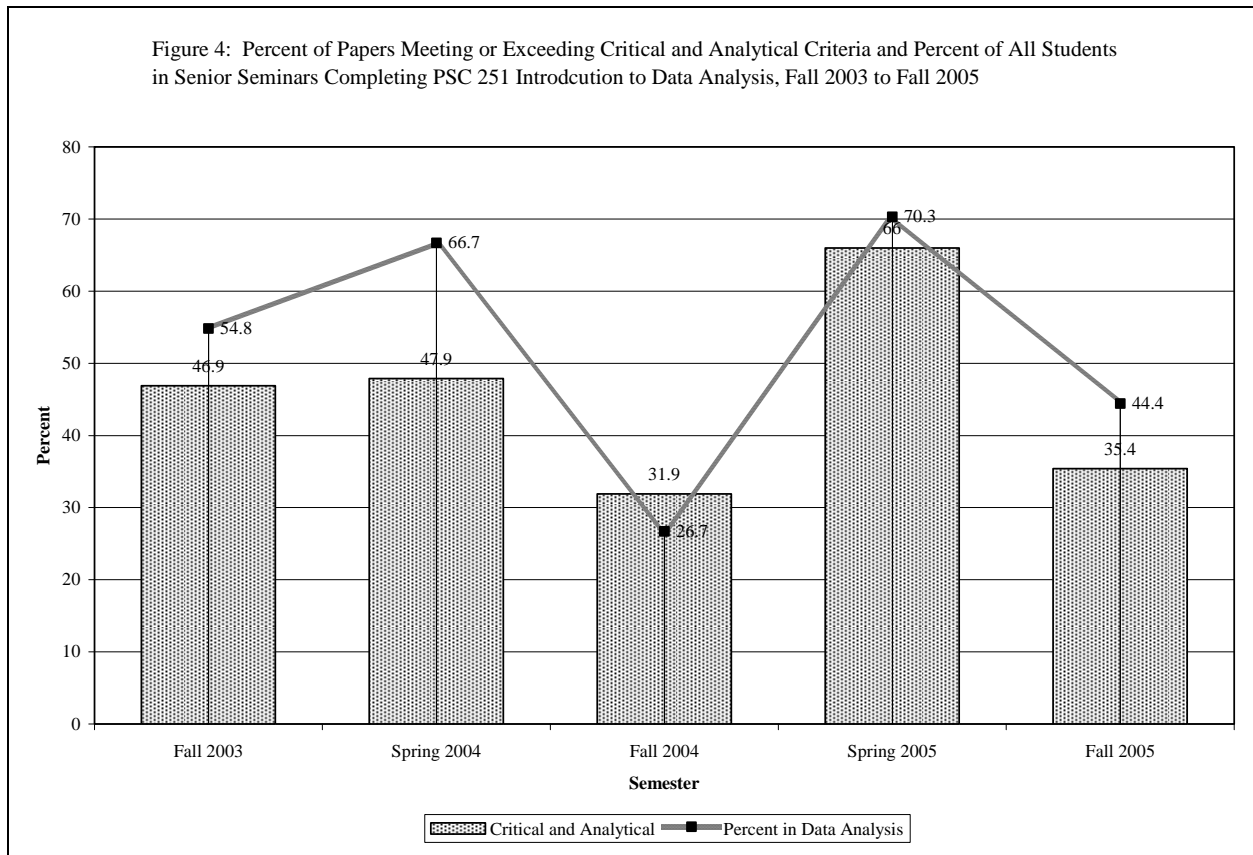


Review: Students are involved in two general ways. First, students are involved in the review process by their submission of senior seminar papers to the instructor of record. Second, faculty members teaching senior seminars frequently inform students of the nature of student assessment and the department’s goals. All fulltime faculty members are involved in the review process either as instructors in senior seminars or as reviewers for the purpose of student assessment. Reports are distributed to faculty members and discussed in subsequent department faculty meetings.

Action: The Department has determined that the guidelines for the International Relations majors do not clearly enough indicate the proper sequencing of the senior seminar. International Relations majors may take the seminar according to current guidelines, as if it were a regular course. The idea, however, is that the senior seminar is a capstone course. The Department has directed the chair to take steps to articulate clearly that the senior seminar is a capstone course and should only be taken near the completion of the degree, specifically after the student has completed the core and track requirements. The chair informs International Relations Majors during the advising process that senior seminars are capstone courses. In addition, all students, both Political Science Majors and International Relations Majors, must have authorization to enroll in a senior seminar. As part of the senior seminar enrollment process, the department reviews a student’s file and determines whether she has met core requirements for her major before allowing her to enroll.

At the most recent faculty meeting (May 3, 2005) the Department decided that it must study further ways to improve its student assessment for its two majors (Political Science Major and International Relations Major). After successive iterations of student assessment a pattern is emerging about the proper course of action to address this problem. The problem appears to be

that International Relations Majors are not required to take a research methods and data analysis course as part of their curriculum. Figure 3 illustrates the percent of randomly selected papers reviewed as part of student assessment that either met or exceeded departmental goals (the bar in the chart) and the percent of the total number of students in those senior seminars that had taken the department's data analysis course (the trend line in the chart), PSC 251 Introduction to Data Analysis.



Action Plan: The department wants to hold a faculty retreat to discuss various options to address this issue. Among initially proposed options are: (1) eliminate the Critical and Analytical criteria, (2) revise criteria measurement, (3) increase research methods component in 300-level courses, (4) assign a research paper writing text in seminars, (4) require methods course for International Relations Majors. The department chair will attempt to arrange such a retreat if time and expenses allow.

APPENDIX A
SENIOR SEMINAR STUDENT ASSESSMENT CRITERIA EXPLANATION AND STUDENT ASSESSMENT FORM

<i>DIMENSION</i>	<i>EXCEEDS EXPECTATIONS</i>	<i>MEETS EXPECTATIONS</i>	<i>DOES NOT MEET EXPECTATIONS</i>
Critical and Analytical Criteria			
<i>Thesis Component</i>	Clearly articulated thesis linked to either theory or cogent statement of the nature of the problem being investigated	Clearly articulated thesis.	Unclear or nonexistent thesis
<i>Hypothesis Component</i>	Theoretically appropriate hypothesis or explanation of a significant political problem	Research question or hypothesis is clearly formulated.	Research question or hypothesis is unclear or nonexistent.
<i>Evidence Component</i>	Superior evidence appropriate for the hypothesis or problem being investigated.	Evidence is generally appropriate.	Evidence does not test hypothesis or answer research question.
<i>Conclusions Component</i>	Conclusions linked to theory or problem being investigated.	Draws appropriate conclusions.	Unclear or no conclusions.
Research Criteria			
<i>Sources Component</i>	Multiple (10 or more) scholarly sources cited or combination of scholarly sources, government documents, interviews, foreign news sources, and articles from newspapers of record.*	Five to ten scholarly sources cited or combination of scholarly sources, government documents, interviews, foreign news sources, and articles from newspapers of record.*	Less than five scholarly sources cited or combination of scholarly sources, government documents, interviews, foreign news sources, and articles from newspapers of record.*
<i>Citations Component</i>	Consistent use of appropriate citations (footnotes, endnotes, or embedded).	Appropriate citations (footnotes, endnotes, or embedded).	Poor, inconsistent use of style or no use of formal style.
<i>Bibliography Component</i>	Consistent, complete, and properly organized bibliography.	Properly organized bibliography.	

* In general, five of government documents, interviews, foreign news sources, and articles from newspapers of record should be considered the equivalent of one scholarly article.

APPENDIX A
SENIOR SEMINAR STUDENT ASSESSMENT CRITERIA EXPLANATION AND STUDENT ASSESSMENT FORM

<i>DIMENSION</i>	<i>EXCEEDS EXPECTATIONS</i>	<i>MEETS EXPECTATIONS</i>	<i>DOES NOT MEET EXPECTATIONS</i>
Articulate and Communicate Component			
<i>Organization Component</i>	Excellent Organization	Good Organization	Uneven and ineffective organization
<i>Paragraphs Component</i>	Thoroughly developed paragraphs with clean transitional phrases.	Consistently developed paragraphs.	Incompletely developed paragraphs and inconsistent transitional phrases.
<i>Sentence Structure Component</i>	Interlinked and concise sentence structures.	Concise sentence structures	Coherent but vague sentence structures
<i>Grammar Component</i>	No grammatical errors	Infrequent grammatical errors	Isolated serious mechanical errors and occasional minor mechanical errors.

APPENDIX A
SENIOR SEMINAR STUDENT ASSESSMENT CRITERIA EXPLANATION AND STUDENT ASSESSMENT FORM

Paper Identification Number: _____			Evaluator's Identification: _____			
<i>Directions:</i> Faculty evaluator should circle the number corresponding to his or her evaluation. The faculty evaluator should provide any comment about the paper or the evaluation in the space provided (if needed, write on back of the form).						
Dimension	Evaluator's Comments	Exceeds Expectations	Meets Expectations	Doest Not Meet Expectations		
Critical and Analytical Criteria						
<i>Thesis</i>		5	4	3	2	1
<i>Hypothesis</i>		5	4	3	2	1
<i>Evidence</i>		5	4	3	2	1
<i>Conclusions</i>		5	4	3	2	1
Research Criteria						
<i>Sources</i>		5	4	3	2	1
<i>Citations</i>		5	4	3	2	1
<i>Bibliography</i>		5	4	3	2	1
Articulate and Communicate Criteria						
<i>Organization</i>		5	4	3	2	1
<i>Paragraphs</i>		5	4	3	2	1
<i>Sentence Structure</i>		5	4	3	2	1
<i>Grammar</i>		5	4	3	2	1