

To: Mark Button, Department Chair, Political Science
From: Matthew Burbank, Director of Undergraduate Studies
Date: June 2017
RE: Report on initial assessment data

This report is a summary of our pilot test on assessment. The purpose of the pilot test was to try out the proposed method of collecting and evaluating assessment data for the undergraduate BA and BS degrees in Political Science.

Assessment Plan and Pilot Test

The department's assessment plan involves collecting student papers from POLS courses taught at the 5000 level during the fall and spring semester of each year and sampling approximately 100 of those papers for evaluation by the faculty serving on the Undergraduate Studies (UGS) committee each year (25 papers per faculty assuming four on the committee each year). The 5000-level POLS courses were chosen because all majors are required to take at least three 5000-level POLS courses as part of the major. The intent is to use the 5000-level courses as a basis for assessing the summative performance of political science majors. After each year's evaluation, the director of the UGS committee will draft a brief report of the assessment for that year to be submitted to the department chair. After three years of data have been accumulated, the UGS will examine the results from the yearly reports and provide recommendations to the chair and the faculty regarding changes in curriculum or instruction.

In order to see whether this assessment plan was viable, the UGS committee recommended that a pilot test be carried out using data from spring semester 2016. The report is based on the data from that pilot test.

Learning Outcomes in Political Science

The faculty had previously approved the following expecting learning outcomes for the BA and BS degrees in political science.

Students who graduate with a major in political science should:

1. demonstrate an understanding of fundamental political ideas, institutions, policies, and behavior in the United States, other countries, and internationally;
2. demonstrate an understanding of major concepts, theories, approaches to research in the study of politics;
3. be able to identify, analyze, and assess information from a variety of sources and perspectives;
4. be able to formulate an argument and express that argument clearly and cogently both orally and in writing;
5. have sufficient ability in a foreign language to enhance their knowledge of the culture and politics of nations or people outside the United States [BA only]; have sufficient

understanding to evaluate and apply numerical data in the context of social scientific analysis [BS only];

6. be prepared for entry level jobs in the public, private, or nonprofit sectors, or to undertake graduate study in an academic or professional program;
7. possess the research and communication skills necessary to understand and participate in the world of politics.

The department's assessment plan proposed that five of the seven outcomes (numbers 1, 2, 3, 4, and 7) be evaluated simultaneously using evidence from existing papers submitted in 5000-level political science classes. The plan did not attempt to assess learning outcome 5 which was intended to capture the difference between the BA and BS degree (see recommendations below). The plan also did not attempt to assess outcome 6 since other evidence (such as employment data for majors) would be more relevant to assess this outcome. In addition, this pilot test included a new learning outcome proposed by a UGS member that students "show a level of knowledge and critical thinking expected of a major."

The pilot test thus included six criteria to be evaluated using student papers as evidence:

- A. Shows understanding of political ideas, institutions, policies, or behavior in the US, other countries, or internationally;
- B. Uses major concepts, theories, or approaches to research;
- C. Identifies, analyzes, and assesses information from a variety of sources;
- D. Expresses an argument or thesis clearly in writing;
- E. Shows evidence of skills in research and communication;
- F. Shows level of knowledge and critical thinking expected of major.

The scale used for each criterion was a five point ordinal scale:

- 4 = Clear and consistent evidence of meeting criterion;
- 3 = Some evidence of meeting criterion;
- 2 = Limited or inconsistent evidence of meeting criterion;
- 1 = No evidence of meeting criterion;
- 0 = Not applicable.

Evidence Collection

During spring semester 2016, I collected 162 papers from 11 of the 13 eligible 5000 level classes taught that semester. The term "paper" is used here in a general sense, the artifacts collected were a range of end-of-term submissions including traditional research papers, essay final exams, and a variety of other writing assignments such as project proposals, legal briefs, reports on political meetings, and short reviews of a book, theory, or concept. The eclectic nature of these writing assignments was intentional since the purpose of this pilot test was to assess the types of writing that our students undertake in 5000-level courses. The papers were collected by contacting instructors of 5000-level

POLS classes and asking whether they had an end of term assignment that would meet the criteria and whether they would be willing to share it for the assessment pilot. All instructors responded positively, but I was not able to obtain papers from two of the eligible courses for logistical reasons. Since most of the papers were submitted in Canvas as part of the regular class assignment, instructors were able to provide electronic copies of papers after they were submitted for the class. This method of collecting papers, however, was overly time consuming and should be automated by developing a function in Canvas to allow papers to be gathered for this purpose.

After all the papers were assembled, I used a random number generator to create two samples of 25 papers (a total of 50 of the 162 papers were used in the evaluation). The reason for creating two samples was specific to the pilot test because I wanted to keep the number of papers that each faculty member had to evaluate at the level called for in the department's plan but also to provide a way to assess inter-rater reliability on a larger number of papers. Of the four faculty on the UGS committee, each was asked to assess 25 papers using a rubric consisting of the six ELOs and the ratings scale. Faculty were not given any training as to how to do the assessment beyond the general information on the purpose of the ratings and the rubric. The lack of "how to" instruction was intentional in order to have the pilot test be a "real world" test of how such evaluation would typically be done in the department.

Results

The results revealed that student papers generally showed evidence of meeting our ELOs. Table 1 shows the result for each of the four faculty raters and the mean rating for papers given a rating of 3 ("some evidence of meeting criterion") or 4 ("clear and consistent evidence of meeting criterion"). These results indicate that papers from our 5000-level courses generally show a high level of understanding of political ideas, institutions, policies, or behavior (outcome A), evidence of research skill (E), and critical thinking (F). These papers showed somewhat less evidence of using a variety of sources (C) and using major concepts, theories, and approaches to research (B).

One point that was apparent to the faculty doing the assessment was that the variety of paper assignments made it more difficult to evaluate some types of papers on each learning outcome. For papers that were standard research papers, it was relatively easy to evaluate the paper on all of the outcomes whatever the substance of the paper. For papers that were tailored more specifically to a particular course, however, it was often more difficult to evaluate all the learning outcomes especially the outcome on identifying, analyzing, and assessing information from a variety of sources (outcome C). For example, some of the papers being evaluated asked students to critique a book or a particular theory. In contrast to a typical research paper, these assignments did not require or encourage students to use and/or evaluate a variety of sources and, as a result, led to differences in scoring. Some evaluators simply gave low scores for such papers on outcome C, while others scored this outcome as not applicable.

Table 1, Percentage of papers receiving a score of 3 or 4 by outcome and rater

Outcome	R1	R2	R3	R4	Mean
A. Shows understanding	92	96	72	84	86
B. Major concepts	64	84	68	72	72
C. Variety of sources	60	88	52	72	68
D. Writing	84	84	60	60	72
E. Research skill	84	76	76	68	76
F. Critical thinking	80	88	68	84	80

Note: Rater 1 and rater 2 evaluated the same 25 papers (sample 1) and rater 3 and rater 4 evaluated the same 25 papers (sample 2).

Table 2, Two measures of inter-rater reliability for each criterion

Criterion	<u>Spearman's r</u>		<u>Krippendorff's alpha</u>	
	Sample 1	Sample 2	Sample 1	Sample 2
A	.52	.41	.15	.31
B	.17	.47	.08	.47
C	.36	.84	.31	.63
D	.33	.39	.31	.40
E	.43	.75	.30	.73
F	.21	.47	.08	.25

Note: Reliabilities are based on two independent ratings of 25 papers in each sample.

A goal specific to the pilot test was to evaluate inter-rater reliability. The results from our pilot test suggest that the level of inter-rater reliability was not high. Table 2 shows the inter-rater reliabilities for the two samples for each criterion using both Spearman's r as a measure of correlation for ordinal data and Krippendorff's alpha for ordinal data as a measure of inter-rater reliability. Since reliability should be .80 or above, these results suggest that only a couple of results approached a high level of reliability. The reliabilities would likely be improved by modifying how the department identifies or evaluates student writing assignments or by providing faculty with specific instructions as to how to apply the learning outcomes to different types of papers.

Actions and Recommendations

Based on the results from the pilot test, the UGS committee made several recommendations to the political science faculty. First, and most importantly, the UGS recommended that the department continue with its proposed assessment plan. The UGS committee regarded the plan as feasible and likely to provide the department with worthwhile assessment data. In particular, the pilot test indicated that the yearly assessment based on a sample of approximately 100 papers would not be too onerous and

that three years' worth of data would provide a reasonable amount of information from which to make recommendations regarding expected learning outcomes and any suggestions for changes to the major or curriculum.

The UGS committee also suggested several minor modifications to the department's expected learning outcomes and process for evaluating student papers. Based on the pilot test, the UGS made the following two recommendations to the faculty:

1. Make three minor modifications to the department's current expected learning outcomes:
 - a. Make the BA and BS learning outcomes the same by eliminating the learning outcome that refers to language skill (for BA) or quantitative skill (for BS).
 - b. Eliminate the phrase "both orally and" from our fourth learning outcome as it cannot be assessed with this method and it would be difficult to establish a means for assessment.
 - c. Add a new expected learning outcome that states all students should "Show a level of knowledge and critical thinking expected of major."
2. Modify the assessment procedure to better match student papers to the learning outcomes in order to improve inter-rater reliabilities. Specifically:

The proposed assessment rubric includes six outcomes: (A) Understanding of political ideas, institutions, policies, or behavior; (B) Understanding of major concepts, theories or approaches to research; (C) Identifies, analyzes, and assesses information from a variety of sources, (D) Expresses an argument or thesis clearly in writing; (E) Shows level of knowledge and critical thinking expected of a major; and (F) Shows evidence of skills in research and communication.

After student papers are collected and sampled, each paper will first be classified as either a research paper or an argumentation paper. Research papers will be defined as those papers of ten or more pages that investigate a single topic using a range of sources. Argumentation papers are defined as papers of a generally shorter length that are intended as specific, directed projects for students. Research papers would be assessed using all learning outcomes while argumentation papers would be assessed on all criteria except for outcome C because such writing assignments often do not require information from a range of sources.

The faculty discussed these recommendations at a meeting in March 2017 and voted to approve the department's assessment plan with these modifications to the department's expected learning outcomes and assessment procedures.