Playing Politics? A Meta-Analysis on Active Learning in Political Science Classrooms

Emily Dunlop[a] and Christopher Way

Department of Government, Cornell University, Ithaca, New York, USA

[a] Corresponding author e-mail: emily.dunlop@cornell.edu

Paper prepared for the APSA Teaching and Learning Conference,
February 7 – 9, Alexandria, Virginia.

Abstract

Active learning (AL) has gained significant traction in political science education, often touted as an evidence-based alternative to traditional lecture methods. Despite its widespread adoption and the establishment of numerous AL initiatives and support centers at institutions across the US, the empirical evidence supporting AL's effectiveness in political science remains limited. While over 300 articles on AL have been published in political science journals since 2010, the majority of robust evidence comes from STEM fields. This paper aims to critically evaluate the existing evidence for AL effectiveness in political science. Here, we outline a first foray into a meta-analysis of Active Learning in Political Science, using a robust design and checks on internal validity of the studies under review. We also propose a nuanced approach to examining AL practices, recognizing that specific AL techniques, such as role-play and simulation commonly used in political science, may have varying impacts. Ultimately, we call for higher-quality studies and a shift from questioning whether AL works to exploring which AL practices are most effective.

1. Introduction

Active learning pedagogy has gained momentum in political science education in recent years. Although the extent of its adoption varies across institutions and instructors, it is increasingly employed as a teaching method for engaging students and fostering critical thinking. Since 2010, over 300 articles on active learning (AL) have been published in journals focused on political science instruction (Dunlop and Way 2025). Colleges have begun launching active learning initiatives to encourage the adoption of AL across disciplines, including ones at our home institution (Cornell), Indiana University, and UC Riverside. Many more colleges have set up active learning offices to support AL, including the University of Georgia, the University of Wisconsin, Harvard, Northeastern, Columbia, and others, with *Centers for Teaching and Learning* expanding across university campuses (Wright, 2023). The establishment of platforms such as the Active Learning in Political Science (ALPS) blog has helped spread the practice, and the blog's recognition with the 2024 APSA-PSA International Partnerships Award underscores its impact on the discipline.

The rapid spread of AL in political science has been driven by the belief that it is an evidence-based pedagogy that enhances understanding, retention, and applied knowledge. AL is widely believed to yield superior learning outcomes compared to the traditional lecture format, in which students are passive recipients of information presented by the instructor. By shifting the focus to students as active participants in their education, AL is thought to cause better learning outcomes, while also matching the expectations of Millennials and Gen Z students who expect more interactive and experiential learning environments and perhaps making education more inclusive and equitable (Theobald et al. 2020).

This belief has been supported by a large body of research on the effectiveness of AL in STEM fields. Several systematic reviews and meta-analyses conclude that AL improves learning outcomes by a significant amount in these fields (Freeman et al. 2014; Michael 2006; Prince 2004). Moreover, AL may have the added benefit of reducing inequities in the classroom; one study reports a 76% reduction in the probability of passing between minority and over-represented students in STEM classes (Theobald et al. 2020). However, a recent study has called into question the quality of evidence supporting these conclusions; most of the studies of AL in STEM fields do not score well on an inventory of internal validity measures (Martella et al. 2023).

Despite the apparent effectiveness of AL in STEM fields, there are three reasons to be sceptical about concluding ALPS will prove equally effective. First, according to researchers who have studied the transfer of knowledge across disciplines, there are reasons to why it is often difficult to transfer lessons learned in one field to another (Adler et al. 2018; Hill and Jordan 2021). Relevant differences across fields can be related to the nature of knowledge, types of skills, and types of learning outcomes specific to various fields of study. Second, it seems probable that the limitations of the research designs typical in STEM AL research (Martella et al. 2023) will also be common in the smaller body of ALPS research. Third, we are unaware of any quantitative meta-analyses of the effectiveness of AL in political science comparable to the Freeman et. al. paper which is the foundation of evidence for AL effectiveness in STEM fields (Freeman et al. 2014).[1]

[1] As discussed below, a recent meta-analysis has attempted to cover the social science and humanities in the same way Freeman reviewed the STEM literature (Kozanitis and Nenciovici 2023). However, this meta-analysis includes only 6 papers on political science.

To reduce the uncertainty about the effectiveness of AL in political science, we are conducting the first meta-analysis of the active learning in political science (ALPS) literature. This is an essential step if we are committed to teaching based on evidence rather than tradition. Our preliminary searches lead us to believe enough ALPS research comparing the results of traditional lecturing with AL interventions exists to support a quantitative meta-analysis. We propose doing an analysis analogous to the Freeman et.al. (2014) paper in the STEM fields, as well as coding all the articles for internal validity following the procedures of Martella et.al. (2023) and re-doing the analysis with only the papers featuring the higher-leverage research designs.

In the rest of this paper, we first review the evidence produced by meta-analyses for the effectiveness of AL in the STEM fields. We then turn to the social sciences before proposing our own meta-analysis. We conclude with some suggestions for future research directions in ALPS.

2. Evidence for the effectiveness of active learning in STEM fields

The evidence for the effectiveness of active learning in STEM fields is impressive. Several systematic reviews and meta-analyses conclude that AL improves learning outcomes by a significant amount in STEM fields (Freeman et al. 2014; Michael 2006; Prince 2004). In particular, a large body of research was cataloged and analysed in the meta-analysis by Freeman et. al. (Freeman et al. 2014), which has become the most-cited paper in this area.[2] Freedman et. al. identified 642 papers on AL that aimed to compare learning outcomes in traditional lecture courses

[2] According to Google Scholar, the paper has been cited 11,913 times.

with those featuring AL interventions in the fields of astronomy, biology, chemistry, computer science, engineering, geology, mathematics, natural resources/environmental science, nutrition/food science, physics, psychology, and statistics. Restricting coverage to those that featured efforts to ensure assessment equivalence, student equivalence, instructor equivalence, and that employed a quantitative assessment indicator yielded a total of 225 studies comparing AL approaches to traditional lecturing.

The marquee finding from the meta-analysis is that AL interventions improved student scores on exams and concept inventories by an average of about 0.47 standard deviations, with course failure rates falling from 33.8% in traditional formats to 21.8% with active learning. Students showed greater improvement on concept inventories than on instructor-written exams. Importantly, the study found no significant differences across disciplines for which there were more than 10 studies.[3] AL was found to yield the largest benefits in smaller classes (with less than 50 students), although students in medium and large classes benefited as well. The authors conclude that adopting AL instruction could increase student retention in STEM majors, reduce failure rates, and improve conceptual understanding.

Building on Freedman et. al., a recent meta-analysis focusses on one particular AL pedagogy in STEM fields: flipped classrooms (Gong, Cai, and Cheng 2024). Applying similar inclusion criteria to those of Freedman et. al., the authors found 53 studies between 2010 and 2023 studying the effects of flipped classrooms in science, technology, engineering, and mathematics fields. The meta-analysis yielded a more modest effect than in the Freedman et. al. study, with a

[3] However, in two fields with a small number of studies – geology and computer science – AL was not discernibly more effective than traditional teaching approaches.

mean effect size of 0.26 standard deviations (this difference was significant) and considerable heterogeneity across studies. Once again, effects were largely comparable across fields, at least for those with more than a handful of studies: medical science, mathematics, and technology. Notably, the effectiveness of flipped classroom activities was greater when they were paired with pre- and post-class activities; for example, including quizzes or learning checks either at the start of class or after class. Although the estimated effects are more modest, the results reinforce those of the Freedman et. al. analysis.

Taken together, these two meta-analyses, along with numerous articles on AL in STEM fields that have been published since 2014, seem to present over-whelming evidence for the superiority of AL over traditional instructional methods. However, a recent review of the evidence casts doubt on this conclusion (Martella et al. 2023). Covering all the papers in the Freeman et. al. meta-analysis (Freeman et al. 2014) as well as 260 additional articles published between 2015 and 2022, Martella et. al. focussed on the methodological rigor of the papers with an emphasis on internal validity (Martella et al. 2023). Drawing on well-established standards from the medicine/health sciences field, the authors catalog 11 internal validity controls. For our purposes, the most important of these focus on whether the study is a randomized controlled trial design and whether evidence of equivalence between treatment and control groups was provided.[4]

The authors then assessed whether each article had met each of the 11 internal validity standards. Of the nearly 450 papers, only 1.9% of the articles met all 11 of the standards. The

[4] Specifically, the 11 are: 1) research design; 2) group equivalence; 3) attrition equivalence; 4) matched sample size; 5) matched instructor; 6) operational definition of variables; 7) dosage of variables; 8) matched implementation in time frame; 9) matched content; 10) equivalent dependent measures; 11) implementation fidelity.

majority of articles (62.7%) met half or fewer of the control standards, and only 22.7% met 7 to 10 of the 11 standards. Of particular interest, only 9.6% of studies used the "gold standard" of randomized controlled trials, whereas 88.1% were quasi experiments.[5] Perhaps most troubling is the failure of the majority of studies to adequately document the equivalence of control and treatment groups (often referred to in political science research as "balance" across groups). Here, we can consider two dimensions of equivalence. The first refers to knowledge and skills directly related to the learning outcome being assessed; essentially, familiarity with the material. This can be assessed via pre-tests, past exam scores on similar materials, or content-specific assessments. The second refers to measures that are indirectly related to likely performance in the class: high school/college GPAs, prerequisite classes taken, SAT scores, etc. If the groups are not balanced on characteristics relevant to performance in the class, selection on confounders undercuts our ability to make valid comparisons. On this point, typical practices are not reassuring. A majority of articles (67%) failed to provide any information about equivalence directly related to the outcome of interest, and 53.5% did not include any information about group equivalence on factors indirectly related to the assessed outcomes.

Weaknesses in these two categories of internal validity carry disproportionate weight among the 11 standards assessed. RCTs are the "gold standard" in research for good reasons: only randomized assignment to treatment can control for both known and unknown confounders and achieve balance across treatment and control groups with a high degree of confidence. However, the vast majority of AL research is quasi-experimental for good reasons: it is difficult to design

[5] 1.9% were within-subjects designs.

and gain approval for a true RCT in typical instructional environments. In a quasi-experimental design, ensuring group equivalence is vitally important, preferably on a measure related to the learning outcome. If the groups are not equal on variables related to the outcome (either directly or indirectly), valid comparisons cannot be made. For this reason, it is particularly troubling that a majority of quasi-experimental papers did not establish that the groups were equal, not even on factors indirectly related to the assessed outcomes. It is not surprising that the vast majority of research in this area is quasi-experimental; it is surprising that the majority fails to establish that the groups are comparable on key dimensions.

With these caveats in mind, the authors' coding of outcomes finds that 64.6% of the articles showed that AL was more effective than traditional lecture formats (however, only 33% showed a statistically significant difference). Most of these studies were in the fields of engineering, chemistry, physics, mathematics, and biology.

3. Evidence for the effectiveness of active learning in political science

Rapidly growing interest in AL in STEM fields has led to a large body of research into its effectiveness, even if recent research has questioned the evidentiary foundations of the conclusion that AL outperforms traditional instructional techniques. Despite a large number of papers describing AL practice in the social sciences and political science (Ishiyama 2010; 2013; ADD OTHERS; Dunlop and Way 2025), until recently systematic analysis of the evidence for the effectiveness of AL in the social sciences and humanities has been thin compared to that for STEM fields. However, a recent paper seeks to close this gap by providing a meta-analysis for the social

sciences modeled closely on the Freeman et.al. paper (Kozanitis and Nenciovici 2023). Kozanitis and Nenciovici (2023) searched databases for articles on the effectiveness of AL in the social sciences and humanities, screening the results down to 519 possible papers. They then applied five criteria (similar to those of Freeman et.al. 2014) to these 519 papers, which winnowed the number to 104.[6] Of the 104 studies included, the fields of psychology, sociology, economics, education, and language instruction accounted most of the studies; the authors located only 6 studies focused on political science meeting their inclusion criteria.

The takeaway from the resulting analysis is extremely similar to that of Freeman et.al.: AL practices, as compared to traditional lecturing, yielded a 0.49 standard deviation increase in learning outcomes. AL seems to increase learning outcomes by about half a standard deviation in both STEM fields and social sciences/humanities. However, the Kozanitis and Nenciovici analysis reports considerably more heterogeneity than Freeman et.al. found (Kozanitis and Nenciovici 2023). There was no evidence for greater effectiveness in five fields: philosophy, history, law, library instruction, and inter-disciplinary classes. In two other fields, political science and management, the 95% confidence intervals were positive, but the lower bounds were very close to zero, indicating a high level of uncertainty about the magnitude of effects. AL appears to be more effective in smaller classes than larger ones, with effect sizes decreasing in stair-step fashion as class size increased. However, positive effects were greater in upper division classes compared to introductory classes, suggesting that perhaps AL is more effective when students are engaging with

[6] The selected papers had to: 1) describe experimental or quasi-experimental research; 2) focus on the social sciences or humanities; 3) focus on post-secondary education; 4) measure learning outcomes with identical assessments across two groups; 5) provide quantitative measures of learning outcomes.

higher-order cognitive skills. Moreover, there was no evidence for the effectiveness of several AL interventions, including project-based learning, case analyses, role-playing simulations, and combination activities. The non-result on role-playing is particularly notable given that simulations are by far the most popular AL intervention employed in political science, a tension between popularity and evidence of effectiveness that has been noted before (Ishiyama 2010). Notably, most of the non-results are for categories with a very small n, suggesting the possibility that a lack of power is causing an inability to detect effects and highlighting the need for more research.

Given that Kozanitis and Nenciovici modeled their procedures on those of Freeman et. al., it seems likely that most of the papers included would feature many of the limitations documented in Martella et. al.'s evaluation. The latter found that most papers on AL in STEM field satisfied only from four to six of 11 internal validity standards. We plan to apply the Martella et. al. criteria to the 6 papers included in Kozanitis and Nenciovici's meta-analysis; however, we have not yet received a list of papers included in the analysis from the authors. As a preliminary probe, we used several of the search terms employed by Kozenitis and Nenciovici to search for political science articles that may have been included in their analysis. One article that matches their search criteria employs a quasi-experimental research design to explore the effects of adding active learning into an introductory course in peace and conflict studies (Sjöstedt 2015). This paper meets all five criteria for inclusion outlined by Kozenitis and Nenciovici (Kozanitis and Nenciovici 2023). As a baseline, this study used classes taught one year using traditional methods as the control group. The investigators then introduced a mixture of AL teaching strategies in the same courses in two subsequent years for the treatment group. The main finding is that failure rates dropped roughly in

half, whereas the percentage of students awarded the highest grade increased by about 60% (Sjöstedt 2015).

These are large effects and would constitute strong evidence for the effectiveness of ALPS if they proved to be robust. However, we then scored the paper using Martella et.al's 11 standards for internal validity (Martella et al. 2023). First, of course, the paper is not an RCT, in keeping with the finding that more than 90% of papers on AL rely on quasi-experimental research designs. Given that, establishing group equivalence is of paramount importance in ensuring research design quality. Here, the paper falls a bit short. No attempt is made to establish equivalence on knowledge or skills related to course assessments: no pre-tests were administered, no past exam scores or grades on similar material were recorded, and no content-specific assessments were used. On measures indirectly related to likely performance in the classes, only rough demographic data was discussed (gender and age). No measures that could be expected to be strongly related to likely performance in the class were collected: no prior grades, test course, related classes taken, etc. Thus, we have no basis for strong confidence that the groups were balanced across treatment and control groups. Overall, of the 11 criteria for internal validity employed by Martella et.al., we assess that five were met by this study.

Another political science paper turned up by the Kozanitis and Nenciovici search procedure focuses on the under-studied outcome of student reactions to AL (although course performance was also recorded) in a large introduction to international relations course (Lambach, Kärger, and Goerres 2017). This paper focuses specifically on flipped-classroom instruction as an innovation offered in a 2014/15 course offering (the treatment), which was compared with results from a traditional lecture version of the same class taught in 2013/14 (the control). In terms of learning

outcomes, the study found no significant difference in student performance across the two versions of the course; in fact, the average scores were slightly worse under the flipped classroom treatment. In terms of student satisfaction, students rated the course more highly than the traditional format of the same class, and higher than the averages for political science courses in general at the same university. However, there was no clear evidence that the students felt that they had learned more. Overall, the study concludes that students seem to *like* a flipped classroom format more, but the authors failed to find evidence that students actually *learned* more or believed that they had learned more.

We then applied the standards of Martella et.al. to this paper. Once again, the paper is clearly not an RCT, so is in keeping with the greater than 90% of papers in broader AL research that are also quasi-experimental. The focus then turns to establishing equivalence of the two group of students. The paper does not discuss group equivalence directly related to the outcome variables. No measures of students' prior knowledge of international relations are provided, meaning there is no way to determine if one group entered the class with a stronger background than the other. The study also comes up short in establishing equivalence between groups on measures indirectly related to learning outcomes, such as test scores or grades in prior coursework. Thus, we cannot be confident the differences (or lack thereof) are due to teaching methods or pre-existing disparities. In applying all 11 internal validity standards, we score five as met. The scoring for these two ALPS papers – 5 of 11 standards met – are very comparable to the modal scores assigned to STEM AL papers by Martella et.al.

Again, we do not (yet) have a list of the six specific political science papers included in a recent meta-analysis of AL research in the social sciences (Kozanitis and Nenciovici 2023). However,

given their inclusion criteria and search terms, these two papers are plausible proxies for the types of research they reviewed. Applying the standards suggested by a recent evaluation of AL research in STEM fields (Martella et al. 2023) to them suggests that papers included in the meta-analysis of social science research will feature similar methodological limitations.

4. A meta-analysis of the ALPS literature

To fill this gap in the ALPS literature, we are conducting a meta-analysis of research in political science. In this brief preview, we discuss some of the assessment types to be covered, AL pedagogies featured, and search terms employed. Our overview of the literature (Dunlop and Way 2025) is used to help identify search terms for identifying candidate studies. Search terms will be combined to cover three areas: 1) active learning pedagogies; 2) disciplinary focus on political science; and 3) types of assessment employed.

For the first category (active learning pedagogy), candidate search terms include:

- Active learning
- Flipped classroom
- Role-playing
- Simulation
- Problem-based learning
- Experiential learning
- Peer-based learning
- Debates
- Project-based learning
- Gallery walk

These search terms are paired with disciplinary keywords:

- Political science education
- Teaching political science

- Political science pedagogy
- International relations
- Comparative politics
- American politics
- Political theory

Finally, we add key-words to indicate aspects of research design and assessment:

- Experiment
- Quasi-experiment
- Learning outcomes
- Assessments

Some preliminary searches indicate that these terms should yield a sufficient number of papers for a meta-analysis. For example, a search combining "international relations," "active learning," "experiment," and "learning outcomes" yields 726 results. A quick perusal of the results indicates that the vast majority of these papers are reviewing the literature or describing teaching practices rather than describing a study in sufficient detail to qualify for a meta-analysis but nonetheless seems like to yield several dozen candidate papers. Similarly, a search for "flipped classroom," "political science," and "experiment," and "learning outcome" yields 185 results.

We hope the results will provide political science instructors with research-based evidence for them to choose their teaching methods. A possible contribution from our analysis is not just to provide an overall estimate of the effectiveness of the rather amorphous set of teaching practices lumped under the "active learning" label, but to evaluate the evidence for specific pedagogies (eg, simulations) and to suggest areas where further research would be most valuable. For example, thus far most AL research seems to focus on a "AL vs. lecture" dichotomy, with little attention being paid to the relative effectiveness of different AL practices.

5. Conclusion

ALPS offers an exciting opportunity for instructors to support student success in university classrooms, with some suggesting benefits beyond specific knowledge and understanding, including critical thinking and other skills that are transferable to the workplace. However, while much has been written about ALPS, including hundreds of articles in the last decade and a half, much of the scholarship in the field is descriptive (detailing simulation or other activities and providing resources to instructors – an important goal in itself). When evidence is presented of its effectiveness, the studies often rely on quasi-experimental results. Of course, there have been several literature reviewers summarizing the current state of the field (see: Craig, 2014; Dunlop and Way 2025; Ishiyama, 2010; 2013; Krammerer and Higashi, 2021). However, our proposed meta-analysis will provide a robust understanding of the effectiveness of active learning and also may potentially mitigate the expectations of such educational practices in the long run. It is important, perhaps now more than ever, that the scholarship of teaching and learning provides robust evidence for our practices in classrooms. Our proposed meta-analysis will provide a first foray into these practices.

References

Adler, Carolina, Gertrude Hirsch Hadorn, Thomas Breu, Urs Wiesmann, and Christian Pohl. 2018. “Conceptualizing the Transfer of Knowledge across Cases in Transdisciplinary Research.” *Sustainability Science* 13 (1): 179–90. https://doi.org/10.1007/s11625-017-0444-2.

Craig, John. 2014. What Have We Been Writing About? Patterns and Trends in the Scholarship of Teaching and Learning in Political Science. *Journal of Political Science Education, 10*(1): 23-36.

Dunlop, Emily, and Christopher Way. 2025. “From Chalkboards to ChatGPT: Trends and the Future of Active Learning in Political Science Classrooms.”

Freeman, Scott, Sarah L. Eddy, Miles McDonough, Michelle K. Smith, Nnadozie Okoroafor, Hannah Jordt, and Mary Pat Wenderoth. 2014. “Active Learning Increases Student Performance in Science, Engineering, and Mathematics.” *Proceedings of the National Academy of Sciences* 111 (23): 8410–15. https://doi.org/10.1073/pnas.1319030111.

Gong, Jin, Sanfa Cai, and Ming Cheng. 2024. “Exploring the Effectiveness of Flipped Classroom on STEM Student Achievement: A Meta-Analysis.” *Technology, Knowledge and Learning* 29 (2): 1129–50. https://doi.org/10.1007/s10758-023-09700-7.

Hill, Joshua, and Lynda Jordan. 2021. “Instructional Strategies.” In *Design for Learning: Principals, Processes, and Praxis*. EdTech Books. https://edtechbooks.org/id/instructional_strate.

Ishiyama, John. 2010. “What Is the Impact of in Class Active Learning Techniques? A Meta Analysis of the Existing Literature.” In .

———. 2013. “Frequently Used Active Learning Techniques and Their Impact: A Critical Review of Existing Journal Literature in the United States.” *European Political Science* 12 (1): 116–26. https://doi.org/10.1057/eps.2012.3.

Kozanitis, Anastassis, and Lucian Nenciovici. 2023. “Effect of Active Learning versus Traditional Lecturing on the Learning Achievement of College Students in Humanities and Social Sciences: A Meta-Analysis.” *Higher Education* 86 (6): 1377–94. https://doi.org/10.1007/s10734-022-00977-8.

Lambach, Daniel, Caroline Kärger, and Achim Goerres. 2017. “Inverting the Large Lecture Class: Active Learning in an Introductory International Relations Course.” *European Political Science* 16 (4): 553–69. https://doi.org/10.1057/s41304-016-0078-3.

Martella, Amedee Marchand, Ronald C. Martella, Jane K. Yatcilla, Alexandra Newson, Eric N. Shannon, and Charissa Voorhis. 2023. “How Rigorous Is Active Learning Research in STEM Education? An Examination of Key Internal Validity Controls in Intervention Studies.” *Educational Psychology Review* 35 (4): 107. https://doi.org/10.1007/s10648-023-09826-1.

Michael, Joel. 2006. “Where’s the Evidence That Active Learning Works?” *Advances in Physiology Education* 30 (4): 159–67. https://doi.org/10.1152/advan.00053.2006.

Prince, Michael. 2004. “Does Active Learning Work? A Review of the Research.” *Journal of Engineering Education* 93 (3): 223–31. https://doi.org/10.1002/j.2168-9830.2004.tb00809.x.

Sjöstedt, Roxanna. 2015. “Assessing a Broad Teaching Approach: The Impact of Combining Active Learning Methods on Student Performance in Undergraduate Peace and Conflict Studies.” *Journal of Political Science Education* 11 (2): 204–20. https://doi.org/10.1080/15512169.2015.1030502.

Theobald, Elli J., Mariah J. Hill, Elisa Tran, Sweta Agrawal, E. Nicole Arroyo, Shawn Behling, Nyasha Chambwe, et al. 2020. “Active Learning Narrows Achievement Gaps for Underrepresented Students in Undergraduate Science, Technology, Engineering, and Math.” *Proceedings of the National Academy of Sciences* 117 (12): 6476–83. https://doi.org/10.1073/pnas.1916903117.

Wright, Mary C. 2023. *Centers for Teaching and Learning: The New Landscape of Higher Education.* Johns Hopkins University Press.