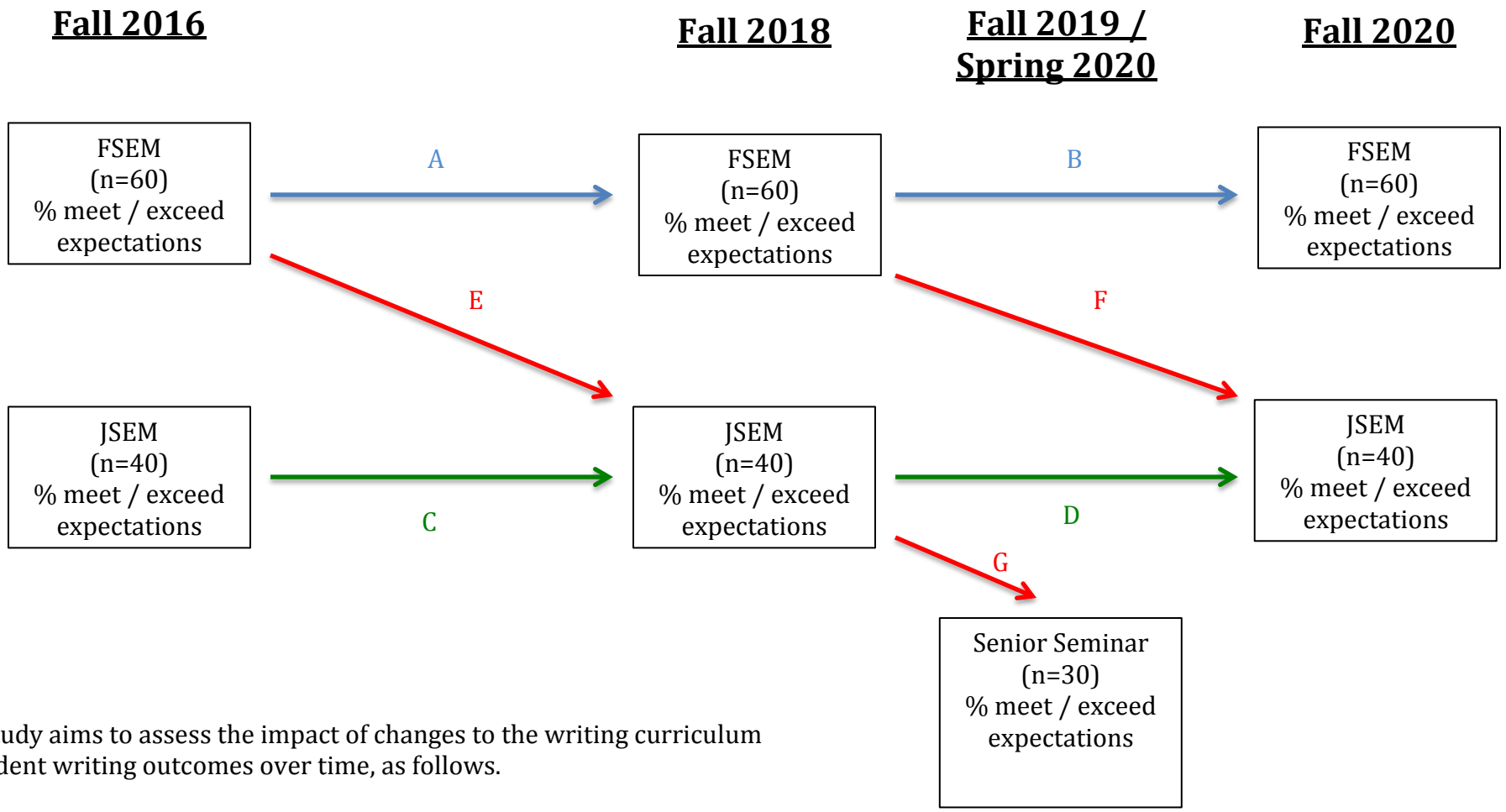
Study Design



This study aims to assess the impact of changes to the writing curriculum on student writing outcomes over time, as follows.

A,B,: Change in FSEM over time (**Δ%** meet/exceed expectations)
C,D: Change in JSEM over time (**Δ%** meet/exceed expectations)
E,F,G: Change within cohorts over time (**Δ%** meet/exceed expectations)

Potential confounds/sources of measurement error
Number of Writing Intensive Courses students completed
Differences in academic abilities of incoming students (among individual cohorts)
Changes to writing assessment rubrics / rater calibration
Differences among colleges/discipline-specific writing expectations for Senior Seminars
Sampling variability

Sampling strategy and rationale

The University Registrar will randomly sample individual students from FSEM (n=60) and JSEM (n=40) in fall semester of 2016. Study coordinators will then contact FSEM and JSEM instructors to request copies of each selected student's *best* writing sample from the semester for review. In the fall of 2018, writing samples will be collected in JSEMs (n=40) for the same cohort of students who were selected in FSEMs (n=60) in the fall of 2016, which enables assessment of changes in writing quality over time for a single cohort of students. Differences in sample sizes between JSEMs and FSEMs are designed to accommodate attrition over time. If feasible, we will also collect and evaluate discipline specific writing samples in senior seminars (n=30) of the 2020 graduating class, which would enable an evaluation of student writing as a developmental process from matriculation to graduation. Similarly, in the fall of 2020, writing samples will be collected in JSEMs for the cohort of students who were selected in FSEMs in the fall of 2018 to allow assessments of changes in writing quality over time for the graduating class of 2022.

Rating samples of student writing

Each writing sample will be reviewed by a minimum of two independent faculty raters (including one rater trained in writing assessment) using a rubric developed by General Education Assessment Committee (Appendix X). A calibration session will ensure reasonable consistency of rating using the rubric. Consistency among raters will be defined as percent agreement among raters and Cohen's κ will be estimated to control for agreement by chance. Then, differences among raters will be resolved through group deliberation before proceeding to statistical analysis. If the distribution of data approximates a normal distribution, then ICC(3,1) will be used to assess inter-rater reliability.

Assessing change in writing quality over time

To assess changes in curricular effectiveness of FSEMs over time, we will compare the percentage of student writing samples that meet or exceed expectations in the fall semester of 2016, 2018, and 2020. The same comparisons will be used to assess changes in curricular effectiveness of JSEMs over time. Differences in the proportion of student writing samples that meet or exceed expectations will be tested for statistical significance using X^2. No statistically significant differences are expected given the limited size of the sample, and the same limitations in sample size prevent correction for α-inflation. If the distribution of ratings approximates a normal distribution, then a mixed effects linear model will also be constructed to recover statistical power, and differences among marginal means will be tested for statistical significance. Residual errors will be compared and tested for independence, normality, and homoskedasticity as indicators of model robustness.