

FINAL REPORT OF ANALYSIS OF STIPEND EQUITY BY GENDER, CITIZENSHIP AND RACE/ETHNICITY

In order to assess whether graduate assistant stipend amounts are influenced by gender (Male versus Female), citizenship (Domestic versus International), or underrepresented minority (URM) status (Domestic URM versus non-URM), University leadership enlisted the assistance of statistics experts from the Eberly College of Science, Department of Statistics. This included, first and foremost, Dr. David Hunter, Professor and former Head of the Department of Statistics, who provided extensive consultation and conducted the analyses, and with input and insights from Dr. Aleksandra Slavkovic, Associate Dean for Graduate Education and Professor, Department of Statistics and Public Health Sciences, and Dr. Kirsten Eilertson, co-Director of the Statistical Consulting Center and Assistant Research Professor, Department of Statistics who also provided additional data visualization. Their collective time, effort, and expertise were invaluable.

Data for all graduate assistant appointments for Fall 2018 were included in the analysis. Because ½-time appointments constitute 97 percent of all assistantships, analysis of stipend levels were focused only on ½-time appointments (n=3,849) for comparable data. Stipend appointments are based upon grade levels, with all students appointed to a given grade level, from 9 through 26, paid the same fixed dollar amount. A grade 27 is allowed only for special circumstances with approval from the Graduate School, and is a variable rate determined by the circumstances, but above the highest fixed dollar grade (26). Because of the unique nature of the grade 27, there were only five such appointments at the University this fall, which were included in the analysis. The unit of comparison was fall/spring (10 month) stipend amount in dollars.

Because decisions around stipend offers are made at the level of academic departments/graduate programs, and non-academic units that may support graduate assistants (e.g., the University Libraries, Student Affairs, Outreach, etc.), the analysis took the approach of disaggregating the data by looking at the department/graduate program/non-academic unit level. For colleges, the analysis looked at individual majors; for non-academic units, each unit was considered as a whole; and for intercollege graduate degree programs (IGDPs) that cross college boundaries, the college effects were ignored by pooling together all students from the same IGDP major.

To this end, two separate linear regression analyses were performed for each academic unit, each non-academic unit, and the group of Intercollege Graduate Degree Programs (IGDPs) taken together: In the first analysis, the effects of primary interest were the gender-by-major interactions after accounting for the main effects of major, as well as candidacy (yes/no) and status as a 2nd-year UGF recipient (yes/no). In the second analysis, the effects of primary interest were the group-by-major interactions—where “group” is a three-level factor with levels “International student,” “Domestic underrepresented minority (URM) student,” and “Domestic non-URM student”—after accounting for the same main effects as in the first analysis. In other words, regression analysis was used to predict, for every major separately, the mean change in stipend due to membership in one gender category relative to another; similarly, a second analysis predicted the mean change in stipend due to international status or domestic URM status. These predictions corrected for overall differences due to majors as well as doctoral candidacy and 2nd-year UGF status within each unit.

Not all comparisons could be made in every program or unit, including because of the size or nature of the unit. For example, the Applied Research Laboratory (ARL) is restricted to domestic students because of the nature of the projects conducted and restrictions placed by research sponsors. Many small

programs with very limited numbers of GAs did not have students of more than one gender or citizenship, or a URM student.

Using the above regression analysis to generate t values and probability levels, the following number of effects were tested:

- A total of 113 gender effects were tested.
- A total of 109 international effects were tested.
- A total of 88 URM effects were tested.

- I. Initially, in the interest of being conservative in exploring differences, all effects that were statistically significant at a probability level of less than .1 ($P < .1$; i.e., the effect might be expected by chance alone no more than 1 time in 10) were identified, first without correcting for the large number of multiple comparisons (310 total):
 - A total of 113 gender effects were tested and 12 were found to be statistically significant at the 0.1 level; 6 in favor of females and 6 in favor of males. It is expected, based on chance alone, to have 11 of 113 effects turn up as statistically significant at the 0.1 level even if no systematic bias is taking place; furthermore, the fact that half of the significant effects favor females and the other half favor males supports that statistical significance of these effects may be due to chance alone.
 - A total of 197 international/URM effects were tested (International, Domestic non-URM, Domestic URM). These include 109 international effects, of which 9 were statistically significant at the 0.1 level (4 favoring Domestic non-URM and 5 favoring International students). It is expected, based on chance alone, to have 11 of 109 effects turn up as statistically significant at the 0.1 level even if no systematic bias is taking place; furthermore, the fact that half of the significant effects favor international students and the other half favor domestic students supports that statistical significance of these effects may be due to chance alone.
 - The 197 international/URM effects include 88 URM effects, of which 16 were statistically significant at the 0.1 level (11 in favor of URM, 5 in favor of non-URM). This is slightly more than we would expect to see based on chance alone (i.e., 16 vs. 9).
- II. To correct for the large number of multiple comparisons (113 gender and 197 International/URM effects = 310 total tests), a Bonferroni multiple comparison correction was applied such that only effects with a probability of less than $0.1/310=0.00032$ were considered significant. Applying this correction, only one effect was still significant, and upon consulting the unit, it was determined the student was given additional funds to counter a competing offer from another institution.
- III. Outcomes derived prior to the Bonferroni correction were shared with the respective colleges/units who were asked to review and reconcile numerical differences to the Graduate School (e.g., the 12 gender effects, 9 citizenship effects, and 16 URM effects). Responses confirmed that these reflected non-bias related factors as explained above (e.g., higher stipend offer to exceed a competing offer documented from another institution).

- IV. Cohort effects were not analyzed for multiple reasons. First, the effect of doctoral students passing their Qualifying Examination (previously called the Candidacy Examination) was already considered in the analyses, which contains some information about cohorts. Second, certain prestigious fellowship awards that are based on academic merit require second year support at matching (high) stipend levels, and this was also considered in the analysis (2nd-year UGF status) and contains information about cohorts. Third, further subdividing the students risks losing statistical power. Finally, there have been multiple changes to the stipend system in recent years, which means that any existing cohort effects due to factors of interest to this analysis would certainly be confounded with the effects due to those changes. Such changes have included the establishment of a minimum stipend grade; “smoothing” of stipend grades for students supported over 12 months so that the same stipend grade is provided fall, spring and summer semesters; and the addition of six higher stipends grades to the stipend tables to enable more competitive offers.

Findings:

Penn State requires a minimum level of stipend support for all graduate assistant appointments, to assure that students meet or exceed the cost-of-living for the region of campus and have sufficient financial support to enable them to focus on their education, while providing valuable professional development experience. However, beyond the required minimum, because differences in stipend offers to prospective students can reflect a variety of factors, including disciplinary differences (e.g., Engineering and Business fields are generally higher than Arts and Education fields) and efforts to meet or exceed competing offers from peer institutions, there should not necessarily be an expectation of uniform stipend levels across colleges or even among departments within colleges, and any disparities that might exist are not necessarily indicative of bias or inequity, but can be explained by other non-discrimination factors.

A comprehensive analysis of assistantship stipends across the University based upon the methodology described above did not reveal evidence to suggest bias based on gender, citizenship, or URM status. Although such evidence was not found, it was confirmed that a relatively small number of numerical differences reflected non-bias related factors as explained above (e.g., a higher offer to exceed a competing offer documented from another institution).