Quantitative intersectionality and student success at HSIs: two examples using administrative data

April 27, 2023

Dr. Christopher Erwin
Auckland University of Technology

Background

- 2017:
  - An interdisciplinary team of economists and sociologists was assembled to study student success at the University of New Mexico
  - The overarching goal was to develop a flexible quantitative approach to studying inequality using intersectionality and Critical Race Theory
  - Lopez *et al.* (2018) has since been cited nearly 200 times by various outlets:

American Educational Research Journal
Handbook of Critical Race Theory in Education
Handbook of Population
Handbook of Research on Science Education
Journal of Education Policy
Journal of Higher Education
Journal of Hispanic Higher Education
Journal of Policy Analysis and Management
Nature
PloS one
Review of Educational Research
Social Problems
Sociology of Education
Studying Latinx/a/o Students in Higher Education

Background

- 2022:
  - Methods from Lopez *et al.* (2018) were expanded to examine first-generation college student success at two HSIs in the American Southwest
  - This work in progress is near completion!
  - This work highlights flexibility of the methods in:
    1. capturing unobserved sources of heterogeneity (e.g., university effects, cohort effects, etc.)
    2. being applicable to many disciplines and contexts

Outline of today’s talk

- Part I: Regression-based approaches to intersectionality

- Part II: López *et al.* (2018) study of how race, ethnicity, gender, and socioeconomic status (SES) map to remedial course assignment and six-year graduation rates at a medium-sized research HSI

- Part III: Erwin *et al.* (2023) study of how race, ethnicity, gender, income, and first-generation college status map to outcomes in higher education

- Part IV: Future research: HSI collaboration/data sharing for external validity

Part I: Regression-based approaches to intersectionality

- Intersectionality refers to the idea that people experience discrimination differently depending on their overlapping identities (Crenshaw 1989)
    - E.g., a black woman is neither singularly black nor singularly a woman, but experiences discrimination based on the interaction of both characteristics
        - Her experiences with discrimination likely differ from that of a white woman or a black man, for example
- Our approach attempts to operationalize this thinking in terms of statistical/econometric models
- The fatal flaw in many regression-based studies measuring or "accounting for" discrimination is treating individual characteristics as independent rather than interdependent

Part I: Regression-based approaches to intersectionality

- Suppose we wish to measure discrimination in the OECD's definition of "low-skill" by indigeneity, SES, and gender.
  - This setup is from a working paper in New Zealand using the OECD the PIAAC Survey of Adult Skills (talk to me afterwards!)
- A "naïve" regression may take the form:

$$\Pr(LowSkill_i) = \alpha_0 + \alpha_1 Female_i + \alpha_2 Indigenous_i + \alpha_3 LowIncome_i + u_i$$

- where non-Indigenous, high-income men are the (arbitrary) reference group and the three regressors are binary indicator variables

Part I: Regression-based approaches to intersectionality

$$\Pr(LowSkill_i) = \alpha_0 + \alpha_1 Female_i + \alpha_2 Indigenous_i + \alpha_3 LowIncome_i + u_i$$

- Here coefficients have a *ceteris paribus* (everything else held constant) interpretation
  - E.g., α_1 is the measured effect of being a woman on the likelihood of being classified as “low-skill” by the OECD, everything else held constant
- Individual characteristics are modelled as independent from one another

Part I: Regression-based approaches to intersectionality

$$\Pr(LowSkill_i) = \alpha_0 + \alpha_1 Female_i + \alpha_2 Indigenous_i + \alpha_3 LowIncome_i + u_i$$

- 2 genders x 2 indigenous statuses x 2 income groups = 8 social locations
- Overall (or adjusted, if other covariates included) likelihoods for each social location by taking linear combinations:

High-income, non-Indigenous men = α_0

High-income, non-Indigenous women = $\alpha_0 + \alpha_1$

Low-income, non-indigenous men = $\alpha_0 + \alpha_3$

Low-income, non-Indigenous women = $\alpha_0 + \alpha_1 + \alpha_3$

High-income, Indigenous men = $\alpha_0 + \alpha_2$

High-income, Indigenous women = $\alpha_0 + \alpha_1 + \alpha_2$

Low-income, Indigenous men = $\alpha_0 + \alpha_2 + \alpha_3$

Low-income, Indigenous women = $\alpha_0 + \alpha_1 + \alpha_2 + \alpha_3$

Part I: Regression-based approaches to intersectionality

$$\Pr(LowSkill_i) = \alpha_0 + \alpha_1 Female_i + \alpha_2 Indigenous_i + \alpha_3 LowIncome_i + u_i$$

- The *ceteris paribus* interpretation is akin to Crenshaw's (1989) critique of the "conceptual limitations of ... single-issue analyses"
- In our example, a Māori woman living in poverty in New Zealand may face discrimination through three separate avenues: being Indigenous, being low-income, and being a woman
  - However, these effects are assumed to be independent (and additive) rather than interdependent (and multiplicative)
- Our proposed solution to such "conceptual limitations" is to assume interdependency of individual characteristics (i.e., multiplicative rather than additive effects)

Part I: Regression-based approaches to intersectionality

- A more realistic model can be expressed as:

$$\Pr(LowSkill_i)$$
$$= \beta_0 + \beta_1 Female_i + \beta_2 Indigenous_i + \beta_3 LowIncome_i + \beta_4 Female_i * Indigenous_i + \beta_5 Female_i$$
$$* LowIncome_i + \beta_6 Indigenous_i * LowIncome_i + \beta_7 Female_i * Indigenous_i * LowIncome_i + \varepsilon_i$$

- We call this a "saturated" model in that it includes level (or main) effects and all possible interaction effects

- This model allows for additional sources of discrimination from jointly belonging to multiple groups
  - "Overlap" effects

Part I: Regression-based approaches to intersectionality

$$\Pr(LowSkill_i)$$
$$= \beta_0 + \beta_1 Female_i + \beta_2 Indigenous_i + \beta_3 LowIncome_i + \beta_4 Female_i * Indigenous_i + \beta_5 Female_i$$
$$* LowIncome_i + \beta_6 Indigenous_i * LowIncome_i + \beta_7 Female_i * Indigenous_i * LowIncome_i + \varepsilon_i$$

- As before, one can take linear combinations of coefficients to arrive at predicted likelihoods for each of the 8 social locations:

High-income, non-Indigenous men = β_0

High-income, non-Indigenous women = $\beta_0 + \beta_1$

Low-income, non-indigenous men = $\beta_0 + \beta_3$

Low-income, non-Indigenous women = $\beta_0 + \beta_1 + \beta_3 + \beta_5$

High-income, Indigenous men = $\beta_0 + \beta_2$

High-income, Indigenous women = $\beta_0 + \beta_1 + \beta_2 + \beta_4$

Low-income, Indigenous men = $\beta_0 + \beta_2 + \beta_3 + \beta_6$

Low-income, Indigenous women = $\beta_0 + \beta_1 + \beta_2 + \beta_3 + \beta_4 + \beta_5 + \beta_6 + \beta_7$

Part I: Regression-based approaches to intersectionality

$$\Pr(LowSkill_i) = \beta_0 + \beta_1 Female_i + \beta_2 Indigenous_i + \beta_3 LowIncome_i + \beta_4 Female_i * Indigenous_i + \beta_5 Female_i * LowIncome_i + \beta_6 Indigenous_i * LowIncome_i + \beta_7 Female_i * Indigenous_i * LowIncome_i + \varepsilon_i$$

- Statistical significance for linear combinations is tested using the delta method (i.e., via first-order Taylor approximation)
- We estimate the model in two steps, each equally insightful:

1. Estimate the more realistic model, take marginal effects if necessary, and examine the main and interaction effects
2. Take the appropriate linear combinations to calculate predicted likelihoods for each social location

Part I: Regression-based approaches to intersectionality

1. Estimate the more realistic model, take marginal effects if necessary, and examine the main and interaction effects

   Coefficients on main and interaction effects tell us which sources of discrimination are driving predicted likelihoods for social locations

2. Take the appropriate linear combinations to produce predicted likelihoods for each social location

   Likelihoods for each social location are a) easy to compare and b) reveal complex landscapes of inequality that are often unseen when individual-level characteristics are assumed to be independent of one another

Part I: Regression-based approaches to intersectionality

- Challenges encountered (so far):

1. Small cell sizes and empty cells

   Some social locations include no subjects!

2. No variation in outcomes for some cells

   Some social locations perfectly predict success or failure!

3. Right-hand side variables increase quickly when additional individual characteristics are added

   *Pr(college graduate) = f(first-gen., low-income, female, Hispanic, **race**)*
   gives 2 x 2 x 2 x 2 x 5 = 80 social locations!

Part I: Regression-based approaches to intersectionality

- Basic extensions of the model:
  - Fixed effects
    - removing unobserved, time-invariant heterogeneity from estimates
    - e.g., institution and cohort effects in Erwin *et al.* (2023)
  - Multilevel modelling (models become “mixed effects”)
    - Allowing for correlated outcomes when subjects are naturally clustered in groups
    - e.g., feeder high schools in López *et al.* (2018) and Erwin *et al.* (2023), local government districts in New Zealand working paper

Part II: López *et al*. (2018)

López, N., Erwin, C., Binder, M., & Chavez, M. J. (2018). Making the invisible visible: Advancing quantitative methods in higher education using critical race theory and intersectionality. *Race Ethnicity and Education*, *21*(2), 180-207.

- **Data:** Administrative data from a medium-sized research HSI in the American Southwest
- **Outcomes:** Graduation within 6 years; remedial English placement; remedial mathematics placement
- **Methods:** Saturated mixed-effects logistic models

Part II: López *et al.* (2018)

- **Research questions:**

1. What patterns of educational inequalities remain invisible when we treat race, gender, and class as independent?

2. How do estimated achievement gaps change when we recognize that such characteristics are dependent on one another?

3. How is the simultaneity of race/structural racism, settler colonialism, gender relations/patriarchy and class/capitalism experienced differently by students according to their location in intersecting systems of power, privilege, oppression and resistance in a given context?

Part II: López *et al.* (2018)

Table 1. Descriptive statistics, graduation, and remediation models.

Variable	2000–2008	2000–2015
Graduated within 6 Years	.406	–
Developmental English	.294	.268
Developmental Mathematics	.326	.301
Any developmental	.431	.397
Female	.582	.577
White	.406	.371
Black	.030	.024
Hispanic	.444	.499
American Indian	.069	.058
Asian	.050	.047
Low-income	.539	.498
Observations	6427	13953

Note: Graduation models include cohorts from 2000 to 2008. Remediation models include cohorts from 2000 to 2015. Low-income is defined by being in the bottom quartile of the income distribution. High-income is defined as being in the top quartile of the income distribution. Students in the middle quartile range are not included in the analysis. The sample is limited to students that graduated high school in the same state as the university. Six-year graduation rates for students in the developmental models are not reported as they have not had sufficient time to graduate.

Part II: López *et al.* (2018)

- **Empirical model:**

$$y_{ij}^* = \alpha_0 + \boldsymbol{X\beta} + \boldsymbol{Z\gamma} + \boldsymbol{W\delta} + \zeta_j + \varepsilon_{ij} \qquad (1)$$

$$\zeta_j \sim N(0, \psi) \qquad (2)$$

where i denotes the student, j denotes the high school, and y denotes one of the three outcomes described above. Idiosyncratic errors, ε_{ij}, are assumed to have a standard logistic distribution with variance ϕ. The model assumes that ζ_j are independent across high schools and independent of main and interaction effects for student i. $\boldsymbol{X}$ is a vector of main effects, $\boldsymbol{Z}$ is a vector of interaction effects, and $\boldsymbol{W}$ is a vector of cohort effects. Cohort effects are included to capture differences in incoming students over time. Binary outcomes, y_{ij}^*, are determined by latent continuous responses via a threshold model

$$y_{ij} = \begin{cases} 1 & \text{if } y_{ij}^* > 0 \\ 0 & \text{otherwise} \end{cases}. \qquad (3)$$

Table 2. Multilevel logistic estimates of probability of six-year graduation by race, ethnicity, gender, and class, 2000–2008.

Variable	Marginal effect		Standard error
Black	−.226	***	.069
Hispanic	−.033		.026
American Indian	−.093	*	.055
Asian	.0009		.071
Low-income	−.142	***	.026
Male	−.137	***	.025
Black × Low-income	.183	**	.091
Hispanic × Low-income	−.051		.036
American Indian × Low-income	−.161	**	.074
Asian × Low-income	.004		.085
Male × Low-income	−.009		.040
Black × Male	.058		.144
Hispanic × Male	−.002		.039
American Indian × Male	−.140		.091
Asian × Male	−.075		.099
Black × Low-income × Male	.050		.175
Hispanic × Low-income × Male	.133	**	.056
American Indian × Low-income × Male	.230	*	.123
Asian × Low-income × Male	.141		.124
Likelihood ratio statistic			48.39
Residual intraclass correlation			.026
Observations			6427

Note: Marginal effects from a saturated logistic model are reported. The baseline group is high-income, white women. *, **, and *** Denote statistical significance at the 10, 5, and 1 percent levels, respectively. Robust standard errors are reported. Cohort fixed effects are included in the model. Low-income is defined as being in the bottom income quartile in the sample. Students from middle income quartiles are not included in the analysis.

Table 3. Multilevel logistic estimates of probability of six-year graduation by race, ethnicity, gender, and class, 2000–2008.

Variable	Marginal effect		Standard error	Cell size
White, high-income women (base)	–	–	–	869
White, low-income women	−.142	***	.026	594
White, high-income men	−.137	***	.025	705
White, low-income men	−.288	***	.031	440
Black, high-income women	−.226	***	.069	57
Black, low-income women	−.185	***	.059	76
Black, high-income men	−.305	**	.126	18
Black, low-income men	−.223	***	.077	45
Hispanic, high-income women	−.033		.026	599
Hispanic, low-income women	−.225	***	.024	1094
Hispanic, high-income men	−.172	***	.029	462
Hispanic, low-income men	−.240	***	.027	699
American Indian, high-income women	−.093	*	.055	85
American Indian, low-income women	−.396	***	.050	186
American Indian, high-income men	−.371	***	.072	66
American Indian, low-income men	−.453	***	.066	108
Asian, high-income women	.0009		.071	50
Asian, low-income women	−.137	***	.046	128
Asian, high-income men	−.211	***	.069	54
Asian, low-income men	−.217	***	.055	92
Likelihood ratio statistic				48.23
Residual intraclass correlation				.025
Observations				6427

Note: Probabilities for groups based on linear combinations of marginal effects from a saturated logistic model. The baseline group is high-income, white women. *, **, and *** Denote statistical significance at the 10, 5, and 1 percent levels, respectively. Robust standard errors are reported. Cohort fixed effects are included in the model. Low-income is defined as being in the bottom income quartile in the sample. Students from middle income quartiles are not included in the analysis.

Table 4. Multilevel logistic estimates of probability of developmental English by race, ethnicity, gender, and class, 2000–2015.

Variable	Marginal effect		Standard error
Black	.188	***	.047
Hispanic	.142	***	.019
American Indian	.152	***	.041
Asian	.129	***	.049
Low-income	.085	***	.022
Male	.032		.021
Black × Low-income	.017		.061
Hispanic × Low-income	.065	**	.026
American Indian × Low-income	.163	***	.049
Asian × Low-income	.129	**	.057
Male × Low-income	.015		.031
Black × Male	.020		.081
Hispanic × Male	.004		.027
American Indian × Male	.074		.056
Asian × Male	−.075		.072
Black × Low-income × Male	−.062		.102
Hispanic × Low-income × Male	−.031		.038
American Indian × Low-income × Male	−.179	**	.070
Asian × Low-income × Male	.039		.085
Likelihood ratio test statistic			372.37
Residual intraclass correlation			.075
Observations			13,953

Note: Marginal effects from a saturated logistic model are reported. The baseline group is high-income, white women. ** and *** Denote statistical significance at the 5 and 1 percent levels, respectively. Robust standard errors are reported. Cohort fixed effects are included in the model. Low-income is defined as being in the bottom income quartile in the sample. Students from middle income quartiles are not included in the analysis.

Table 5. Multilevel logistic estimates of probability of developmental English by race, ethnicity, gender, and class, 2000–2015.

Variable	Marginal effect		Standard error	Cell size
White, high-income women (base)	–	–	–	1843
White, low-income women	.085	***	.022	1043
White, high-income men	.032		.021	1578
White, low-income men	.133	***	.023	718
Black, high-income women	.188	***	.047	97
Black, low-income women	.291	***	.040	118
Black, high-income men	.240	***	.066	45
Black, low-income men	.295	***	.048	75
Hispanic, high-income women	.142	***	.019	1665
Hispanic, low-income women	.292	***	.018	2455
Hispanic, high-income men	.178	***	.020	1260
Hispanic, low-income men	.312	***	.019	1588
American Indian, high-income women	.152	***	.041	153
American Indian, low-income women	.400	***	.029	331
American Indian, high-income men	.258	***	.040	126
American Indian, low-income men	.342	***	.033	203
Asian, high-income women	.129	***	.049	118
Asian, low-income women	.343	***	.031	233
Asian, high-income men	.086		.053	117
Asian, low-income men	.354	***	.033	187
Likelihood ratio test statistic				372.37
Residual intraclass correlation				.075
Observations				13,953

Note: Probabilities for groups based on linear combinations of marginal effects from a saturated logistic model. The baseline group is high-income, white women. *** Denotes statistical significance at the 1 percent level. Robust standard errors are reported. Cohort fixed effects are included in the model. Low-income is defined as being in the bottom income quartile in the sample. Students from middle income quartiles are not included in the analysis.

Part II: López *et al.* (2018)

- **Conclusions:**

- Assuming independence of race, gender, and class oversimplifies the complex nature of achievement gaps in higher education
    - Statistical significance of interaction effects is evidence of interdependence
    - Statistical significance of main effects reveals they also have their own measureable effects on success in college as well

- Our paper offers a new method of assessing the complex nature of inequality along multiple interdependent individual-level characteristics

Part III: Erwin *et al.* (2023)

Erwin, C., López, N., Wise, C., Torres-Velasquez, V., Zerai, A., Jenrette, M., & Martinez, V. (2023). Inequity in Graduation Rates at HSIs: An Intersectional Analysis of Outcomes by Race, Gender and First-Generation College Status. Working paper.

- **Data:** Administrative data from two research (R1 and R2) HSIs within the same state in the American Southwest
- **Outcomes:** Graduation within 4 years; developmental English placement; development mathematics placement
- **Methods:** Saturated mixed-effects logistic models (including university fixed effect and cohort effects)

Part III: Erwin *et al*. (2023)

- **Research questions:**

1. What patterns of educational inequalities remain invisible when we treat race, gender, and first-generation college status as independent?

2. How do estimated achievement gaps change when we recognize that such characteristics are dependent on one another?

Table 1. Descriptive statistics for incoming first-time, full-time freshmen resident students, Southwest University (SU) and Borderland University (BU), 2014 to 2020 cohorts

Variable	(1) Southwestern Public University	(2) Borderlands University	
First Generation College Student	.275	.376	***
College Graduation:			
Within 4 Years	.082	.041	***
Within 5 Years	.093	.070	***
Within 6 Years	.094	.078	***
Remediation:			
Mathematics Required	.238	.105	***
English Required	.056	.131	***
Female	.575	.562	*
Ethnicity:			
Hispanic	.591	.635	***
Race:			
White	.310	.317	
American Indian	.036	.027	***
Asian	.046	.011	***
Black	.018	.010	***
Observations	12,269	6,354	

Table 2. Mixed effects logistic models of 4-year completion rates, marginal effects

Variable	Coefficient (Standard Error)
Male	-.024*** (.007)
First-Generation	-.015** (.007)
Hispanic	-.013*** (.005)
American Indian	-.041*** (.014)
Black	-.033* (.018)
Asian	-.013 (.010)
Male x Hispanic	.002 (.007)
Male x Black	.008 (.029)
Male x Asian	.018 (.014)
Male x American Indian	.035* (.020)
Male x First-Generation	.009 (.010)
First-Generation x Hispanic	.009 (.008)
First-Generation x Black	.035 (.035)
First-Generation x Asian	.019 (.020)
First-Generation x American Indian	.019 (.019)
Male x First-Generation x Hispanic	-.008 (.014)
Male x First-Generation x Black	-
Male x First-Generation x Asian	-.056 (.040)
Male x First-Generation x American Indian	-.060 (.049)
BU	-.032*** (.006)
Cohort Fixed Effects	YES
ρ	.017
LR $\chi^2(1)$	9.05***
Observations	13,949

Table 3. Logistic probabilities of graduation within four years, by social location

Social Location	Coefficient (Standard Error)	Cell Size
Continuing-Generation White Women	(Reference)	2,044
First-Generation White Women	-.015** (.007)	431
Continuing-Generation White Men	-.024*** (.007)	1,740
First-Generation White Men	-.030*** (.011)	276
Continuing-Generation Hispanic Women	-.013*** (.004)	2,914
First-Generation Hispanic Women	-.019*** (.006)	1,884
Continuing-Generation Hispanic Men	-.035*** (.006)	2,214
First-Generation Hispanic Men	-.040*** (.008)	1,326
Continuing-Generation American Indian Women	-.041*** (.014)	178
First-Generation American Indian Women	-.037** (.019)	94
Continuing-Generation American Indian Men	-.031** (.015)	134
First-Generation American Indian Men	-.078* (.042)	45
Continuing-Generation Asian Women	-.013 (.010)	188
First-Generation Asian Women	-.009 (.017)	76
Continuing-Generation Asian Men	-.019 (.015)	140
First-Generation Asian Men	-.062** (.027)	62
Continuing-Generation Black Women	-.033* (.018)	99
First-Generation Black Women	-.013 (.032)	23
Continuing-Generation Black Men	-.049** (.022)	81
First-Generation Black Men	-	0
Observations	13,949	

Source: Offices of Institutional Analytics at Southwest University (SPU) and Borderlands University (BU). †First generation college student status was based on voluntary information on the Free Application for Federal Student Aid (FASFA). *, **, and *** denote statistical differences at the ten five, and one percent levels respectively.

Table 6. Nonlinear models of developmental course English placement, marginal effects

Variable	Coefficient (Standard Error)
Male	-.004 (.009)
First-Generation	.039*** (.012)
Hispanic	.046*** (.008)
American Indian	.106*** (.013)
Black	.071*** (.019)
Asian	.055** (.022)
Male x Hispanic	.008 (.010)
Male x Black	-.004 (.030)
Male x Asian	-.021 (.028)
Male x American Indian	-.007 (.019)
Male x First-Generation	-.031 (.019)
First-Generation x Hispanic	-.002 (.014)
First-Generation x Black	.025 (.031)
First-Generation x Asian	-.006 (.032)
First-Generation x American Indian	-.027 (.018)
Male x First-Generation x Hispanic	.016 (.021)
Male x First-Generation x Black	.056 (.042)
Male x First-Generation x Asian	.044 (.048)
Male x First-Generation x American Indian	.056* (.029)
BU	.040*** (.005)
Cohort Fixed Effects	YES
ρ	.172
LR $\chi^2(1)$	298.64***
Observations	18,623

Table 7. Logistic probabilities of developmental course English placement, by social location

Social Location	Coefficient (Standard Error)	Cell Size
Continuing-Generation White Women	(Reference)	2,372
First-Generation White Women	.039*** (.012)	545
Continuing-Generation White Men	-.004 (.009)	1,979
First-Generation White Men	.016 (.016)	368
Continuing-Generation Hispanic Women	.046*** (.008)	3,666
First-Generation Hispanic Women	.083*** (.009)	2,362
Continuing-Generation Hispanic Men	.051*** (.008)	2,701
First-Generation Hispanic Men	.072*** (.009)	1,640
Continuing-Generation American Indian Women	.106*** (.013)	220
First-Generation American Indian Women	.117*** (.014)	128
Continuing-Generation American Indian Men	.096*** (.016)	158
First-Generation American Indian Men	.132*** (.018)	65
Continuing-Generation Asian Women	.055** (.022)	242
First-Generation Asian Women	.088*** (.023)	92
Continuing-Generation Asian Men	.030 (.024)	170
First-Generation Asian Men	.076*** (.017)	65
Continuing-Generation Black Women	.071*** (.019)	113
First-Generation Black Women	.135*** (.024)	29
Continuing-Generation Black Men	.063** (.021)	91
First-Generation Black Men	.152*** (.028)	16
Observations	18,623	

Source: Offices of Institutional Analytics at Southwest University (SPU) and Borderlands University (BU). †First generation college student status was based on voluntary information on the Free Application for Federal Student Aid (FASFA). *, **, and *** denote statistical differences at the ten five, and one percent levels respectively.

Part III: Erwin *et al.* (2023)

- **Conclusions:**

- Although we find evidence of some nuanced inequities in higher education along lines of first-generation status, gender, and race-ethnicity, differences across social locations are mostly driven by main effects.

- Limited statistical significance of interaction effects provides weak evidence of interdependence between race-ethnicity, gender, and first-generation status

- Our paper expands on previous quantitative intersectionality models by pooling data across institutions and removing time-invariant institution-level heterogeneity.

Part IV: Future research:

- **Thoughts on future research…**

- HSI collaboration/data sharing system would:
  - increase our sample and cell sizes and improve external validity
  - institution fixed effects could be interesting in their own right
  - Allow us to estimate more complex and realistic models

- Data collection upon matriculation is key
  - FAFSA often not filed for high-income students
  - What about LGBTQIA students?
  - Uniformity in how race and ethnicity are recorded is key to seamless collaboration across the broader HSI community

- Many individual characteristics worth considering for such models

Part IV: Future research:

- Thank you for your time!
- Please feel free to contact me with any questions, comments, or suggestions

Dr. Christopher Erwin

christopher.erwin@aut.ac.nz

+64 027 545 7774

www.christophererwin.com