Gauging the Impact of Academic Support Programs: A Quasi Experimental Design Using Propensity Scoring

Serge Herzog, PhD
Director, Institutional Analysis
Consultant, CRDA StatLab
University of Nevada
Reno, NV 89557, serge@unr.edu

What can we learn from education research?

- Questions of causality have been at the forefront of educational debates, in part because of the dissatisfaction with the quality of education research and recent federal initiatives designed to promote the accumulation of scientific evidence in education that rely on randomized controlled trials (RCTs). A common concern revolves around the design and methods used in education research, which many claim have resulted in fragmented and often unreliable findings.


Purpose of Study

- Estimate the effect of participation in living-and-learning communities (LLC) and use of math support center on:
  - First-semester academic success (GPA, credits earned)
  - Enrollment persistence into second semester
  - Second-semester academic success
  - Enrollment persistence into second year

- Control for selection bias via counter-factual analytical framework
  - Compare parametric vs. non/semi-parametric estimates

Source and Data

- New full-time freshmen at the U. of Nevada, Reno
- Data Elements (Covariates):
  - Student demographics (age, gender, race, residency, parent education)
  - Student degree goal and employment plans when in college
  - High school academic experience (GPA, AP, admission scores/test date; class rank)
  - Fall semester academic experience (English, Math, dropped credits, distance ed., major, GPA, earned credits, F/I/W grades)
  - Financial aid profile: unmet need, aid type offered/received
  - Campus services used: math/diversity centers, campus jobs
  - LLC reference groups: off campus, non-LLC on campus
  - Math support ref group: did not use math support center
Data Sample and Statistics

- Fall semester cohorts 2011 and 2012
- Excluding:
  - Students with less than 12 attempted credits (PT)
  - Students with complete credit withdrawal
  - Statistical outliers (using Cook’s, Mahalanobis’)
- Effective sample: 4,871 (math), 3,138 (LLC) students
- Computed variables
  - Precollege preparation index (GPA-test score composite)
  - Delayed college entry: months to UNR matriculation
  - Academic momentum: index (GPA, credits earned comp.)
  - Imputation (mean, predicted value) of missing values
- Analysis:
  - Linear/logistic regression, weighted-sample analysis

Predicting First-Semester Academic Momentum
Baseline Model

Significant Predictors*

Effect size = 7.15 pts or 1/3 of Std deviation over off-camp student

Ranked by Beta weight (standardized coefficient)  
*Alpha <= 0.001; *Negative; Adj R-square=.27; VIF highest = 2.08, all others < 2

Predicting Second-Semester Academic Momentum
Baseline Model

Significant Predictors*

Effect size = 4.6 pts or 1/3 of StDev over off campus student

Ranked by Beta weight (standardized coefficient)  
*Alpha <= 0.001; *Negative; Adj R-square=.27; VIF highest = 2.09, all others < 2

Predicting Spring Enrollment
At End of Fall Semester

Significant Predictors*

Ranked by Wald Significance Level  
Alpha ***<=0.001, **0.01, *0.5; *Negative; Nagelkerke R-square =.32; HL= 0.94

*Alpha <= 0.001; *Negative; Adj R-square=.27; VIF highest = 2.08, all others < 2
Predicting Spring Enrollment
At End of Fall Semester

Significant Predictors

<table>
<thead>
<tr>
<th>Factor</th>
<th>Odds Ratio</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Momentum</td>
<td>1.048</td>
<td>0.660 - 1.633</td>
</tr>
<tr>
<td>Clerk*</td>
<td>1.631</td>
<td>1.089 - 2.435</td>
</tr>
<tr>
<td>Outreach*</td>
<td>1.910</td>
<td>1.300 - 2.835</td>
</tr>
<tr>
<td>Uninvolved*</td>
<td>1.420</td>
<td>0.960 - 2.090</td>
</tr>
<tr>
<td>CampWork</td>
<td>2.215</td>
<td>1.500 - 3.250</td>
</tr>
<tr>
<td>Male</td>
<td>1.299</td>
<td>0.900 - 1.840</td>
</tr>
<tr>
<td>LLCMember*</td>
<td>1.771</td>
<td>1.070 - 2.920</td>
</tr>
</tbody>
</table>

Odds Ratio per unit change in IV

*Negative; Nagelkerke R-square = .32; HL= 0.94

Findings

• Impact on academic momentum (GPA, credits earned)
  – On average, LLC students are ~ 10 percentile points higher in GPA and credits earned for both fall and spring semesters compared to off-campus students, controlling for all other covariates
  – On average, LLC students are ~ 5 percentile points higher in GPA and credits earned for both fall and spring semesters compared to non-LLC on-campus students, controlling for all other covariates
  – Thus, compared to off-campus students, students living on campus are more likely to earn a higher GPA and more credits in their first year, an advantage that is magnified with participation in a living-and-learning community

• Impact on enrollment persistence
  – On average, LLC students are 9.5% more likely to persist than off-campus students \[\left(\frac{\text{base} - \text{p}}{1 - \text{base} - \text{p}} \right)\text{LLC-OR} = \text{pp}; \left(\frac{\text{pp}}{1 + \text{pp}}\right) = 0.95\]
  – On average, non-LLC on-campus students are 9.4% more likely to persist than off-campus students \[\left(\frac{0.922}{1 - 0.922}\right)\text{1.346} = 15.9; \left(\frac{15.9}{16.9}\right) = 0.94\], however that result did not meet statistical significance (alpha <=0.05)
  – The LLC participation benefit accrues net off all other covariates!!

Caveats

• Findings estimate average effect using parameter data from all students, both ‘treated’ and ‘untreated’ (e.g., LLC and non-LLC students)
• Linearity assumption in regression fails if parameters in model (Xs) are highly nonlinear with outcome, distribution of Xs differs between groups of interest (e.g. LLC students vs others)
• Parameter models typically fail to answer to counterfactual H: The outcome for the ‘treated’ had they not received the treatment
• Probability of student selections/choices often correlate with outcomes of interest (selection/endogeneity bias)

Causal Inference Estimation

• Randomized Control Trials (RCT)
  – Costly, ethical issues, operational difficulties, disruption of natural (campus) setting
  – Replication not possible with RCT
• Nonexperimental designs in observational studies
  – Regression discontinuity
  – Instrumental variable (IV) techniques
  – Econometric models to adjust for selection bias
  – Propensity score methods
    • Inverse probability of treatment weighting (IPTW)
    • Subclass stratification
    • Propensity score matching (PSM)
    • Regression covariate adjustment (linear/nonlinear)
Causal Inference Estimation

• Propensity Scoring: Rubin Causal Model
  – Two potential outcomes: \( Y_i(1) \) and \( Y_i(0) \) and let \( Z \) denote LLC (1) or No LLC (0)
  – But only one is observed: \( Y_i = Y_i(1) + (1 - Z)Y_i(0) \)
  – Thus, \( Y_i(1) - Y_i(0) = \text{effect of LLC and } E[Y_i(1) - Y_i(0)] = \text{average LLC effect (ATE)} \)
  – Average LLC effect (ATE) for \( Y_i = 1 \) is defined as \( E[Y_i(1) - Y_i(0)|Z = 1] \) where \( X_i = \text{pre-Z covariates} \)

• Propensity score (PS) accounts for covariates (characteristics) that predict treatment selection (participation in LLC or using math center)
• PS captures the conditional probability of treatment selection given observed (measurable) covariates
• Matching treated (LLC) with non-treated (no LLC) students on their PS enables “counterfactual” analysis, i.e. the expected outcome values with and without LLC participation for those who actually participated!

• IPTW with focus on (ATT) reweights ps of non-treated students with \( w(Y_i(0)) = \frac{P_i(X)}{1-P_i(X)} \)
• Stratification matches treated with non-treated on the propensity score within 6 subclasses (using support range of treated)
• PSM uses 7 algorithms with LLC/math-defined support
  – NN1:1 and K:1 nearest neighbor matching, with/no replacement, random, within 0.2 stddev of ps (caliper), equal weights of controls
  – Kernel1:1 seeks smallest ps distance (difference), no replacement
  – Kernel3:1 seeks smallest ps distance (difference), with up to 3 control (non-LLC/math) cases, weights based on ps distance
  – Full subclass weighted seeks minimal ps distance within max classes
  – Genetic seeks set of weights for each covariate to optimize balance, with replacement to result in 1:1 match, varying optimization iteration
• Regression adjustment uses matched data in linear/log models with LLC/math status and post-selection covariates
Average Outcomes and Naïve Estimator

<table>
<thead>
<tr>
<th>On-Campus Students</th>
<th>Non-LLC</th>
<th>LLC</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fall momentum</td>
<td>79</td>
<td>88</td>
<td>9</td>
</tr>
<tr>
<td>Spring retention</td>
<td>93%</td>
<td>96%</td>
<td>3%</td>
</tr>
<tr>
<td>Spring momentum</td>
<td>85</td>
<td>91</td>
<td>6</td>
</tr>
<tr>
<td>Second yr retention</td>
<td>79%</td>
<td>90%</td>
<td>11%</td>
</tr>
</tbody>
</table>

Estimation of LLC Participation Effect in Weighted Regression Covariate Adjustment

<table>
<thead>
<tr>
<th>Unmatched</th>
<th>IPTW Subclass</th>
<th>NN1:1</th>
<th>NNK:1</th>
<th>Kernel1:1</th>
<th>Kernel3:1</th>
<th>Genetic</th>
<th>Avg Diff to Naïve Est</th>
</tr>
</thead>
<tbody>
<tr>
<td>(N=3138)</td>
<td>(N=3138)</td>
<td>(N=1201)</td>
<td>(N=1105)</td>
<td>(N=1214)</td>
<td>(N=1214)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fall momentum</td>
<td>4.2</td>
<td>4.1</td>
<td>4.4</td>
<td>4.4</td>
<td>3.3</td>
<td>4.2</td>
<td>4.9</td>
</tr>
<tr>
<td>Spring retention</td>
<td>2.2</td>
<td>1.1</td>
<td>1.0</td>
<td>0.6</td>
<td>1.2</td>
<td>1.2</td>
<td>2.0</td>
</tr>
<tr>
<td>Spring momentum</td>
<td>2.4</td>
<td>2.5</td>
<td>2.4</td>
<td>3.0</td>
<td>2.3</td>
<td>2.5</td>
<td>2.8</td>
</tr>
<tr>
<td>Second yr retention</td>
<td>6.9%</td>
<td>5.2%</td>
<td>6.0%</td>
<td>5.1%</td>
<td>6.0%</td>
<td>6.4%</td>
<td>6.0%</td>
</tr>
</tbody>
</table>

Note: Bolded = sig ≤ 0.05 alpha; SE bootstrapped 1000 replications (not listed)
Fall momentum post-treatment covariates: Acad cts, div cts, no Engl, no Math, online course
Other outcomes post-treatment covariates: as above plus fall unearned credits, flags for (W/U/F) grades
Findings on LLC Effect
• Impact on academic momentum (GPA, credits earned)
  – On average, inverse weighting, stratification, and matching reduce the LLC effect size by ~ 50%, 1:1 matching with optimization weights (genetic) showing the greatest reduction (60 to 70%)
  – Results are largely consistent for both fall and spring momentum
  – Weighted regression adjustment produces almost the same results after controlling for post-treatment selection covariates
• Impact on enrollment persistence
  – On average, inverse weighting, stratification, and matching reduce the LLC effect size by ~ 35 to 40%
  – Weighted regression adjustment reduces the effect size only for second-year retention (spring-fall) ~ 7 to 27% after controlling for post-treatment selection covariates
  – The LLC impact grows over time, suggesting varying time effects independent of measurement technique
• Asymptotically all PS estimators should yield the same results
• Sensitivity analysis of findings: Comparing LLC-eligible vs. ineligible (see slides 11, 15) indicates unconfoundedness is more plausible

Causal Inference Estimation

### Average Outcomes and Naïve Estimator

<table>
<thead>
<tr>
<th>Math Support Center</th>
<th>Didn’t Use</th>
<th>Used</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fall momentum</td>
<td>78</td>
<td>83</td>
<td>5</td>
</tr>
<tr>
<td>Spring retention</td>
<td>92%</td>
<td>96%</td>
<td>4%</td>
</tr>
<tr>
<td>Spring momentum</td>
<td>84</td>
<td>86</td>
<td>2</td>
</tr>
<tr>
<td>Second yr retention</td>
<td>80%</td>
<td>86%</td>
<td>6%</td>
</tr>
</tbody>
</table>

### Causal Inference Estimation

Estimation of Math Center Use (ATT) Using Stratification, Matching, and Reweighting

<table>
<thead>
<tr>
<th>Treated</th>
<th>OLS/Logit Regr</th>
<th>IPTW</th>
<th>NN5:1 (^\text{*})</th>
<th>Kernel1:1</th>
<th>Kernel3:1</th>
<th>Full:Opt Subcl W</th>
<th>Genetic</th>
<th>Gen PSU</th>
<th>Avg Diff</th>
</tr>
</thead>
<tbody>
<tr>
<td>Untreated</td>
<td>(N=4107)</td>
<td>(N=4107)</td>
<td>(N=780)</td>
<td>(N=780)</td>
<td>(N=780)</td>
<td>(N=780)</td>
<td>(N=780)</td>
<td>(N=780)</td>
<td>(N=780)</td>
</tr>
<tr>
<td>Fall momentum</td>
<td>4.1 (0.673)</td>
<td>4.0 (1.044)</td>
<td>3.8 (1.308)</td>
<td>3.8 (0.749)</td>
<td>3.8 (1.144)</td>
<td>3.8 (1.144)</td>
<td>7.7 (1.144)</td>
<td>7.7 (1.144)</td>
<td>4.0 (1.144)</td>
</tr>
<tr>
<td>Spring retention</td>
<td>4.7 (0.673)</td>
<td>3.6 (1.044)</td>
<td>3.2 (1.308)</td>
<td>3.7 (0.749)</td>
<td>3.9 (1.144)</td>
<td>3.9 (1.144)</td>
<td>6.1 (1.144)</td>
<td>6.1 (1.144)</td>
<td>4.6 (1.144)</td>
</tr>
<tr>
<td>Spring momentum</td>
<td>2.1 (0.673)</td>
<td>1.9 (1.044)</td>
<td>2.0 (1.308)</td>
<td>1.9 (0.749)</td>
<td>2.1 (1.144)</td>
<td>2.1 (1.144)</td>
<td>2.6 (1.144)</td>
<td>2.6 (1.144)</td>
<td>4.0 (1.144)</td>
</tr>
<tr>
<td>Second yr retention</td>
<td>7.1 (0.673)</td>
<td>6.1 (1.044)</td>
<td>2.3 (1.308)</td>
<td>6.4 (0.749)</td>
<td>7.1 (1.144)</td>
<td>7.1 (1.144)</td>
<td>11.0 (1.144)</td>
<td>11.0 (1.144)</td>
<td>10.0 (1.144)</td>
</tr>
</tbody>
</table>

Bolding at 0.05 alpha; \(^\text{*}\) Weighted N=780; "Caliper 0.2 SD of TU; only retained students; SE are bootstrapped (1K replications)

Note: Retention logit coefficients are converted to Delta-p percentage points
Causal Inference Estimation

| Treatment | IPTW | Kernel1 | Kernel5:1 | Kernel3:1 | Full-Subclass W | Genetic | Gen-PS | Avg Diff to OLS
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>N=780</td>
<td>N=4148</td>
<td>N=780</td>
<td>N=780</td>
<td>N=780</td>
<td>N=780</td>
<td>N=780</td>
<td>N=780</td>
<td>N=780</td>
</tr>
<tr>
<td>Full-momentum</td>
<td>3.4</td>
<td>3.3</td>
<td>3.5</td>
<td>3.5</td>
<td>3.5</td>
<td>3.5</td>
<td>3.5</td>
<td>3.5</td>
</tr>
<tr>
<td>(0.670)</td>
<td>(1.35)</td>
<td>(0.670)</td>
<td>(1.35)</td>
<td>(0.670)</td>
<td>(1.35)</td>
<td>(0.670)</td>
<td>(1.35)</td>
<td>(0.670)</td>
</tr>
<tr>
<td>Spring retention</td>
<td>3.8</td>
<td>3.2</td>
<td>3.2</td>
<td>4.0</td>
<td>3.1</td>
<td>3.1</td>
<td>3.1</td>
<td>3.1</td>
</tr>
<tr>
<td>(0.844)</td>
<td>(1.23)</td>
<td>(0.844)</td>
<td>(1.23)</td>
<td>(0.844)</td>
<td>(1.23)</td>
<td>(0.844)</td>
<td>(1.23)</td>
<td>(0.844)</td>
</tr>
<tr>
<td>Second retention</td>
<td>4.8</td>
<td>4.2</td>
<td>4.2</td>
<td>5.1</td>
<td>4.7</td>
<td>4.7</td>
<td>4.7</td>
<td>4.7</td>
</tr>
<tr>
<td>(2.1)</td>
<td>(2.1)</td>
<td>(2.1)</td>
<td>(2.1)</td>
<td>(2.1)</td>
<td>(2.1)</td>
<td>(2.1)</td>
<td>(2.1)</td>
<td>(2.1)</td>
</tr>
</tbody>
</table>

Note: Retention logit coefficients are converted to Delta-p percentage points.

Findings on Math Support Effect

- **Impact on academic momentum (GPA, credits earned)**
  - On average, PS matching increases the math support effect size in first semester by ~27%, full matching with optimal subclasses and kernel (1:1) showing the greatest rise in effect size (88 to 110%).
  - In contrast, the average PS-based effect size estimate does not differ much from standard OLS/logit estimates for second semester.
  - Weighted regression adjustment produces very similar results, both full subclass and kernel (1:1) doubling the math support effect size.

- **Impact on enrollment persistence**
  - On average, PS-based results show a greater effect size after controlling for post-treatment selection covariates, especially using kernel (1:1) and full subclass matching.
  - Given bias-variance tradeoff, weighted-distance matching (kernel, full subclass) may offer best estimate, assuming unconfoundedness holds.
  - Asymptotically all PS estimators should yield the same results.

Summary of Findings

- **Living-and-learning community effect**
  - Naïve estimators and non-weighted regression *overestimate* vis-à-vis propensity-score non/semi-parametric approaches
  - Assuming unconfoundedness, self selection may account for 50-70% of effect size on GPA/credits earned, up to 40% of effect size on enrollment persistence with non-weighted data.
  - Unconfoundedness is more plausible given results on off-campus students (treatment ‘ineligible’).

- **Math support center effect**
  - *Underestimation* vis-à-vis propensity-score weighted data
  - Self selection may account for over 50% of effect on GPA/credits earned, ~25% of effect on persistence with non-weighted data
  - Bias-variance tradeoff in PS-weighted analysis suggests average PS-weighted results are lower-bound estimates.

Further Research


- Foundational literature

Link to presentation: http://www.unr.edu/ia/research