Strains and Gains

Estimating a First Year University Student Online Engagement Effect
Stata Extended Regression Model (ERM) Framework

Bill Tyler

Consultant to the Enabling Engagement Project 2017-19

Charles Darwin University

Email: *willtyler@msn.com.au*

or

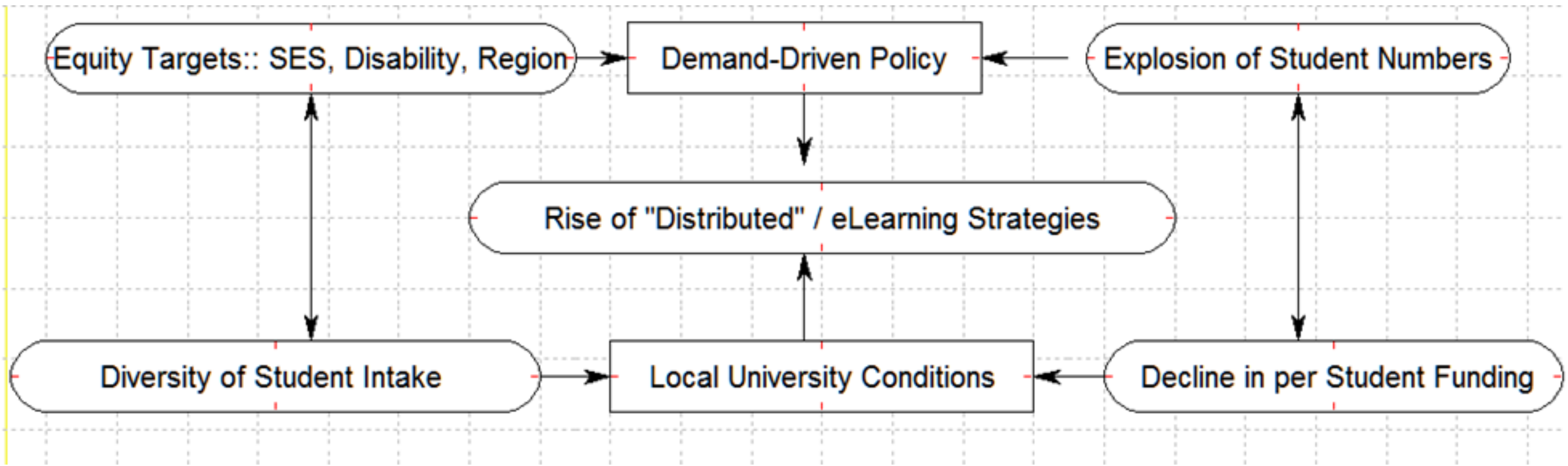
: bill.tyler@cdu.edu.au

Enabling Engagement : Contexts and Questions

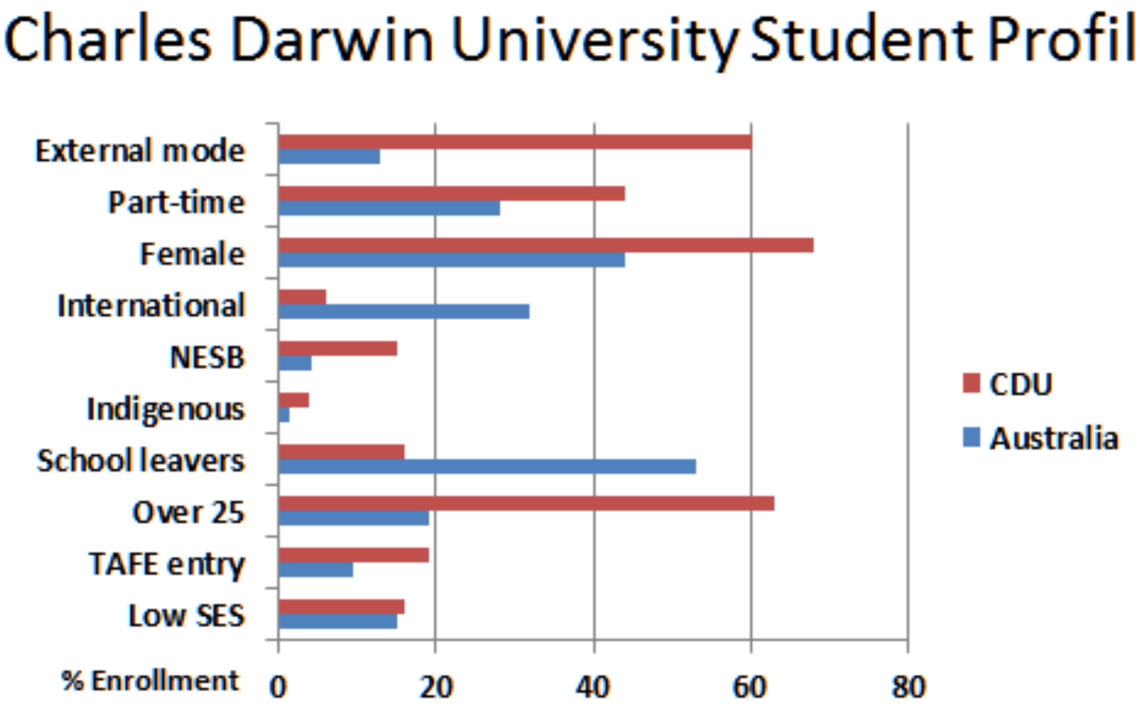
This paper addresses three main issues***:***

- *Do reported levels of online engagement of First Year enrolments predict the Numerical Grades Awarded?*
- *Can an ERM "wash out" the endogenous effects of covariate bias, sample selection in estimation of an "engagement effect?*
- *What are the strengths and weaknesses of the Extended Regression Modelling (ERM) framework in evaluation research for HE innovation?*

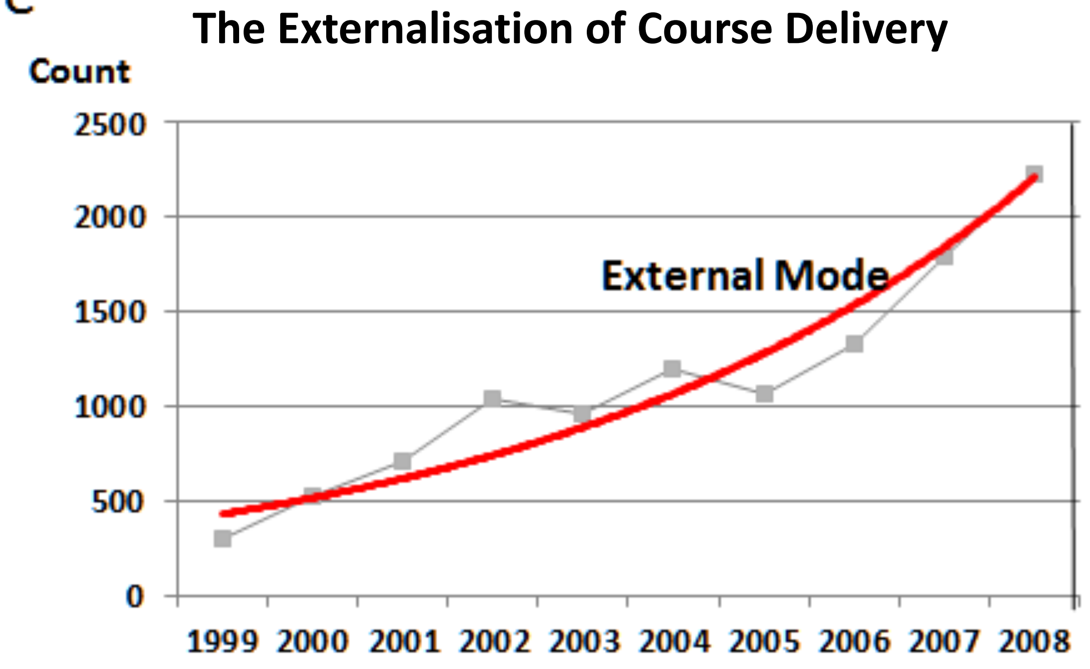
Policy and Performance
The Rise of "Distributed Learning"



Diversity and Delivery: the CDU Context



*Data based on James et al (2009), Bradley Report (2008) and CDU (2010).



* All Common Unit Enrolments (n=21, 615)

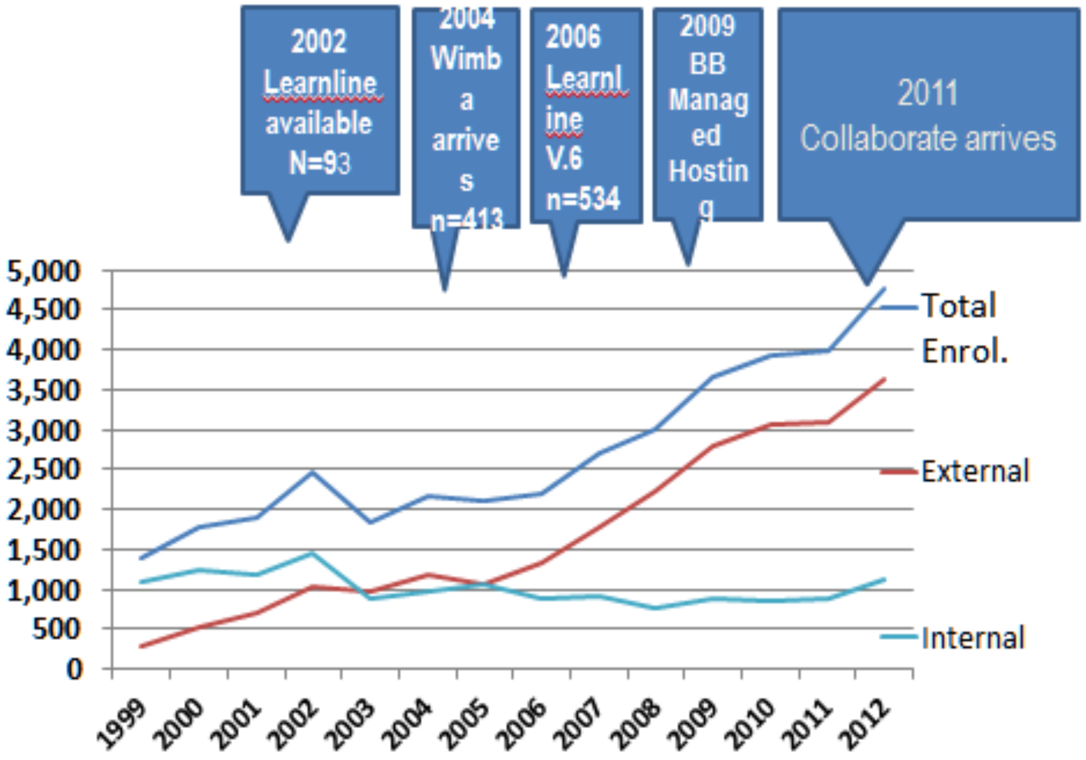
The Flexible Learning Response: Phasing in Online Delivery at CDU

Flexible Learning: the CDU Response: 2002-2016

- All units and courses now available online (offered to Open University Australia, plus 3 MOOCS)

- Learnline Management System and Virtual classroom (Collaborate) widely available in both attendance modes (with mobile access)

- Online portal for full student services plus social media platform (with ShareStream for video)



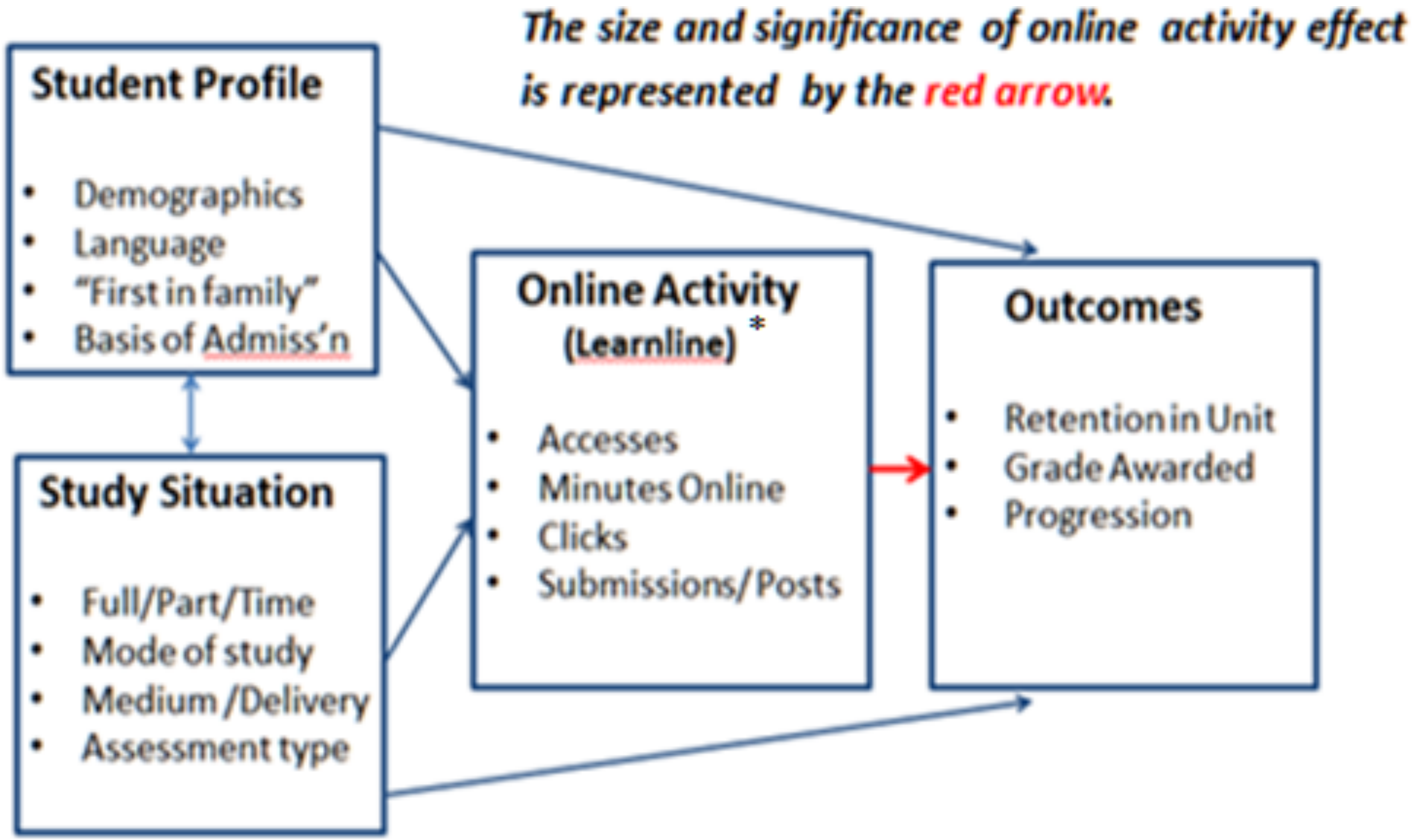
The Research Questions: Online Activity and Student Success

I. Do increased levels of online activity exert a uniform and positive effect on grade levels, after “confounding “ variables (student background and admission entry categories are controlled?

II. Does an effect (sign, size, significance) also depend on learning context- External Mode, Part-time Status or Unit Type (Common Unit or Core Unit)?

III. How might we infer a **causal effect** for exposure to and participation in online participation on student grades?

Estimating Online Effect: From Regression to Causal Inference



Can we reconfigure this path model emulate an experimental (RCT) model?

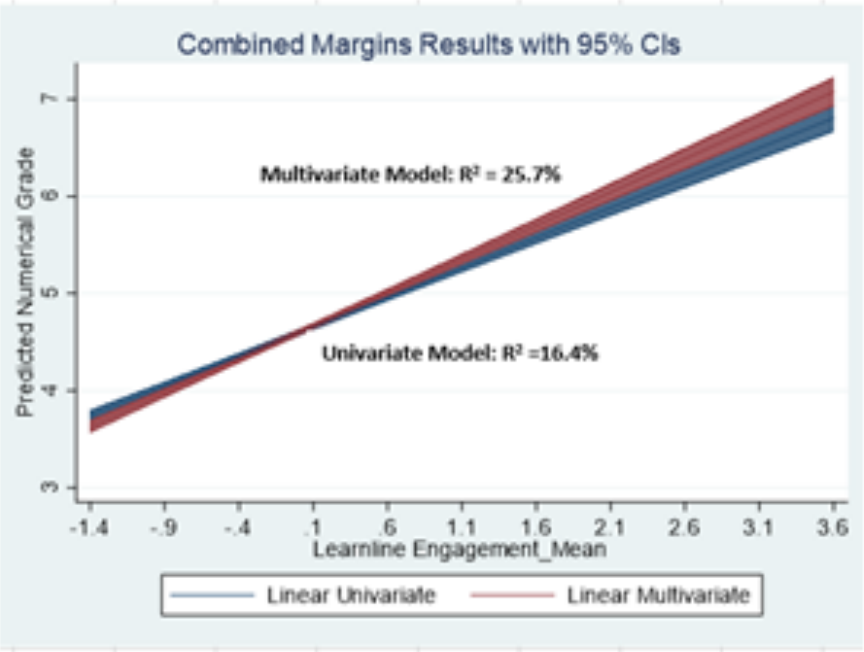
The Sample: Outcome, "Treatment" & Confounders

Variable	*S2 2017*			*S1 2018*			*Combined*		
Outcome	Obs	Mean	Std. Dev.	Obs	Mean	Std. Dev	Obs	Mean	Std. Dev.
Numerical Grade (2-7)	2,398	4.52	1.34	2,581	4.66	1.42	4,979	4.60	1.39
Learnline Engagement									
Number of Unitaccess	2,398	60.00	41.43	2,581	73.29	47.03	4,979	66.89	44.91
TotalClicks/Interactions	2,398	398.68	302.22	2,581	528.05	389.71	4,979	465.74	356.19
Totalminutes online	2,398	1286.29	1193.59	2,581	1698.95	1421.66	4,979	1500.20	1332.68
Learning Situation									
Common Unit	2,398	0.48	0.50	2,581	49.32%	50.01%	4,979	48.83%	49.99%
External_Mode	2,398	59.55%	49.09%	2,581	65.05%	47.69%	4,979	62.40%	48.44%
Part-time Status	2,398	31.65%	46.52%	2,581	28.86%	45.32%	4,979	30.21%	45.92%
Student Demographics									
Male	2,397	38.55%	48.68%	2,581	26.39%	44.08%	4,978	32.24%	46.74%
NESB	2,398	31.40%	46.42%	2,581	23.01%	42.10%	4,979	27.05%	44.43%
Indigenous_Status	2,398	5.67%	23.13%	2,581	6.47%	24.60%	4,979	6.09%	23.91%
Age in Years	2,398	28.05	9.63	2,581	28.66	9.52	4,979	28.36	9.58
TER_Present	2,398	18.27%	38.65%	2,581	10.23%	30.31%	4,979	14.10%	34.80%

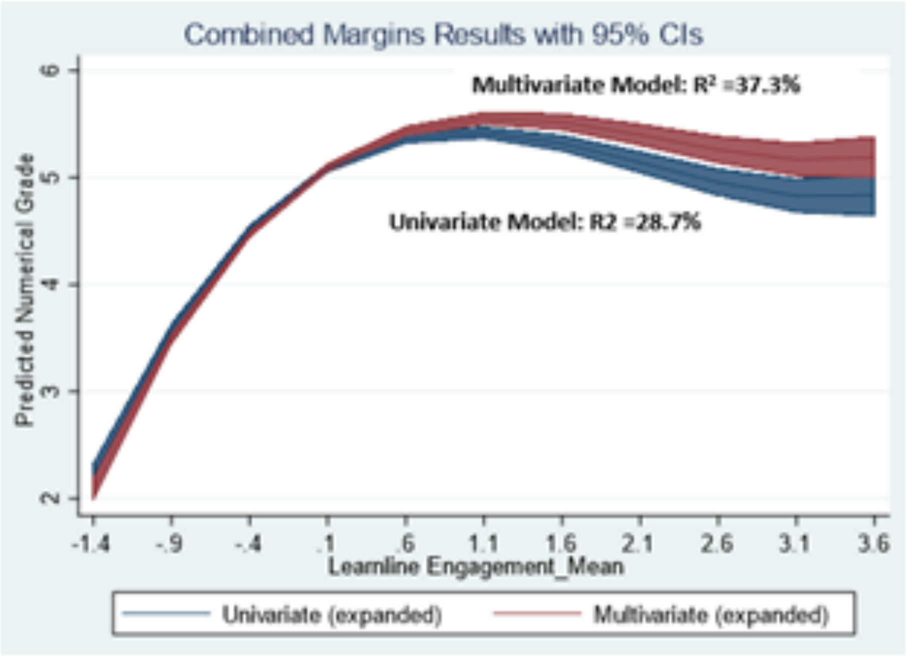
Regression Results: Linear and Non-Linear

Fig. 5 Marginsplot Comparisons of Engagement Effect in Univariate and Multivariate Regressions

Linear Regression



Cubic Regression



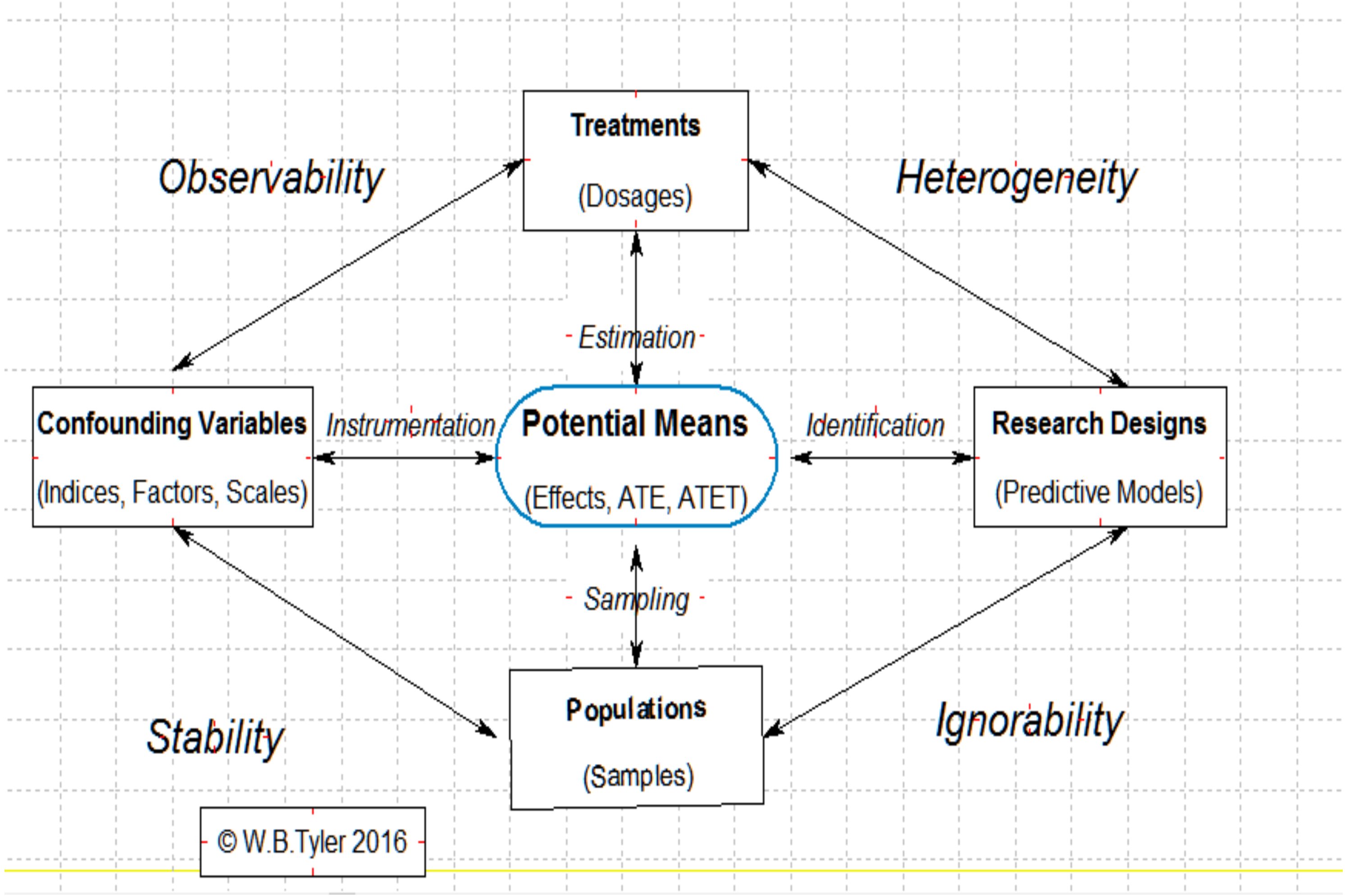
Developing an Extended Regression Model*

The term "endogenous" is used most frequently to encompass the distorting confounding or "non-ignorable" effects of :

- **endogenous covariates**: where a background variable which may have a confounding effect on response to a treatment. These need to be included in the estimation (cf Analysis of Covariance).
- **sample selection**: where the participants in such a trial were overwhelmingly drawn from a non-representative group of the target population (e.g. in the weightloss example, from a group that had a history of chronic eating disorders);
- **treatment assignment**: where those who were assigned to the treatment group rather than the 'control' or non-treatment group were unbalanced across one or more critical dimensions (e.g. on ethnicity, age or gender).

*Users are referred to similar ERM model for estimating an intervention effect for a "Fictional University" in Chuck Huber's presentation at this Conference, available at https://tinyurl.com/2019CausalInference

Estimating Online Effect: A Generic Framework



Building an ERM of Online Engagement Effect

These three sources of endogeneity will be addressed within an Extended Regression Modelling framework. Each is followed by a research question of practical interest:

- **Endogenous covariates** In this model two covariates, Part-time Status (defined as an EFTSL score below .375 or one or two units associated with each enrolment per semester) and External Mode of Attendance, are identified as endogenous.
- **Sample Selection Bias** The status of the lowest scoring group (FNS/DNS)* in the scale of Grade Awarded outcomes raises an important issue of endogeneity that precedes that of treatment assignment or levels of engagement. These were treated by a Heckman-type selection model (similar to Chuck Huber's use of the same approach for missing data).
- **Endogeneity in Assignment to Treatment** - in a self-selection design, recognises that :
  1. more motivated and committed students will be more likely to have higher activity scores than others, even after adjustment self select to a level of online engagement;
  2. conversely, lower ability students who are more at risk of attrition or failure may be more likely to rely on the resources and support offered by Blackboard and other systems.

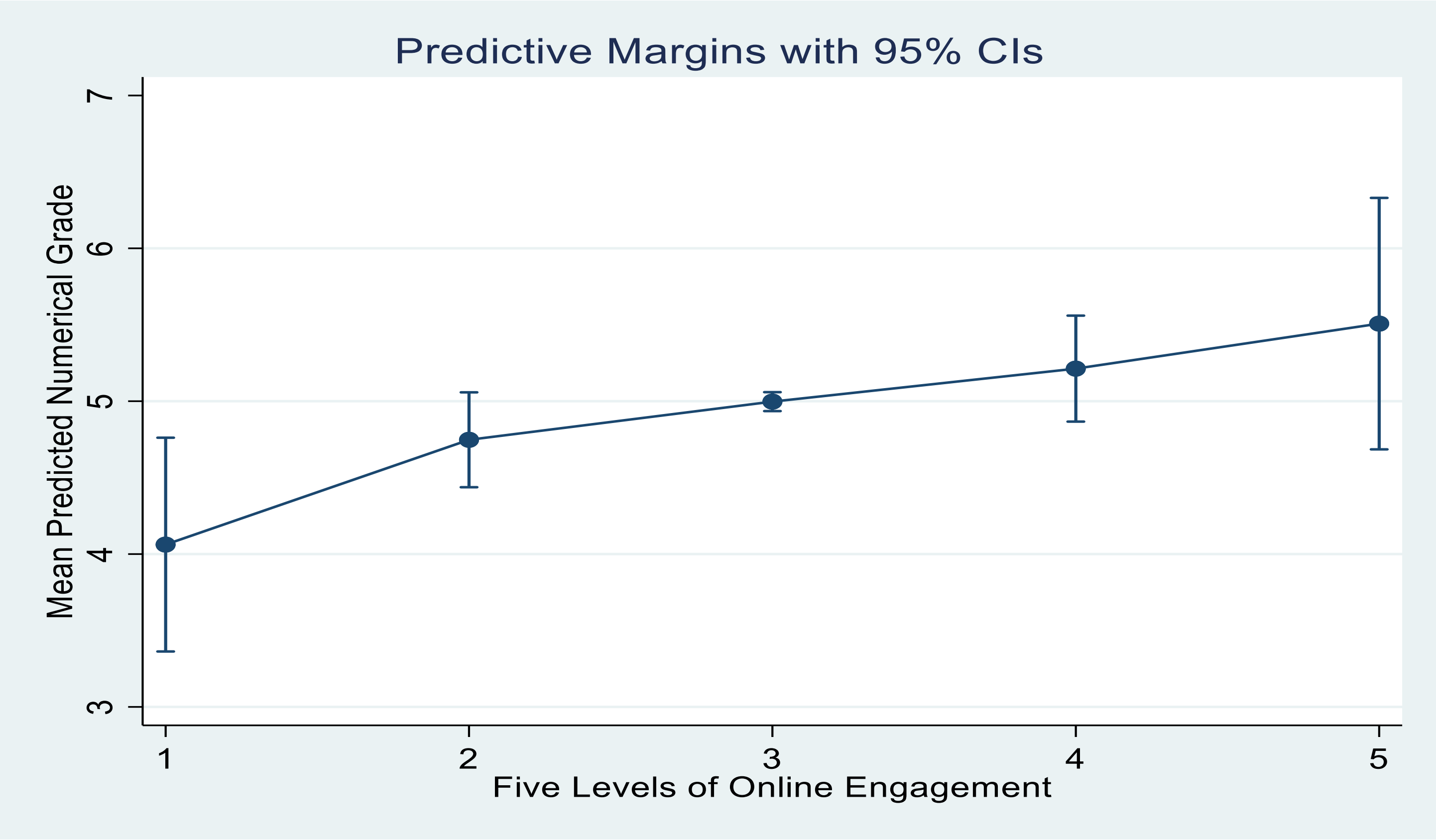
Sourcing Endogeneity : Covariate, Engagement and Sample Selection*

Source of Endogeneity	*Correl. Coeff.*	*Standard Error*	*Z*	*P>z*	*[95% Conf. Level]*	
corr(e.graded_3plus,e.gradescale2)	0.312	0.088	3.570	0.000	0.132	0.473
corr(e.engage_strat5a,e.gradescale2)	0.574	0.076	7.560	0.000	0.407	0.704
corr(e.Part_time,e.gradescale2)	0.544	0.149	3.650	0.000	0.193	0.772
corr(e.External_Mode,e.gradescale2)	0.058	0.015	3.790	0.000	0.028	0.088
corr(e.engage_strat5a,e.graded_3plus)	**0.743**	**0.018**	**41.990**	**0.000**	**0.706**	**0.775**
corr(e.Part_time,e.graded_3plus)	-0.038	0.025	-1.510	0.131	-0.088	0.011
corr(e.External_Mode,e.graded_3plus)	-0.076	0.025	-3.000	0.003	-0.125	-0.026
corr(e.Part_time,e.engage_strat5a)	-0.007	0.018	-0.380	0.705	-0.043	0.029
corr(e.External_Mode,e.engage_strat5a	0.048	0.018	2.680	0.007	0.013	0.082
corr(e.External_Mode,e.Part_time)	0.119	0.016	7.350	0.000	0.087	0.151

* ”engage_strat5a” is multivalued “treatment”variable defined as a five-level grouping of of the means of three Learnline activity zscores.
“gradescale2” and “graded_3plus” are the dependent variables for the full sample (includes the DNS/FNS grades)and the “selected” sample (excludes the DNS/FNS) respectively.

Potential Grades at Five Levels of Online Engagement*

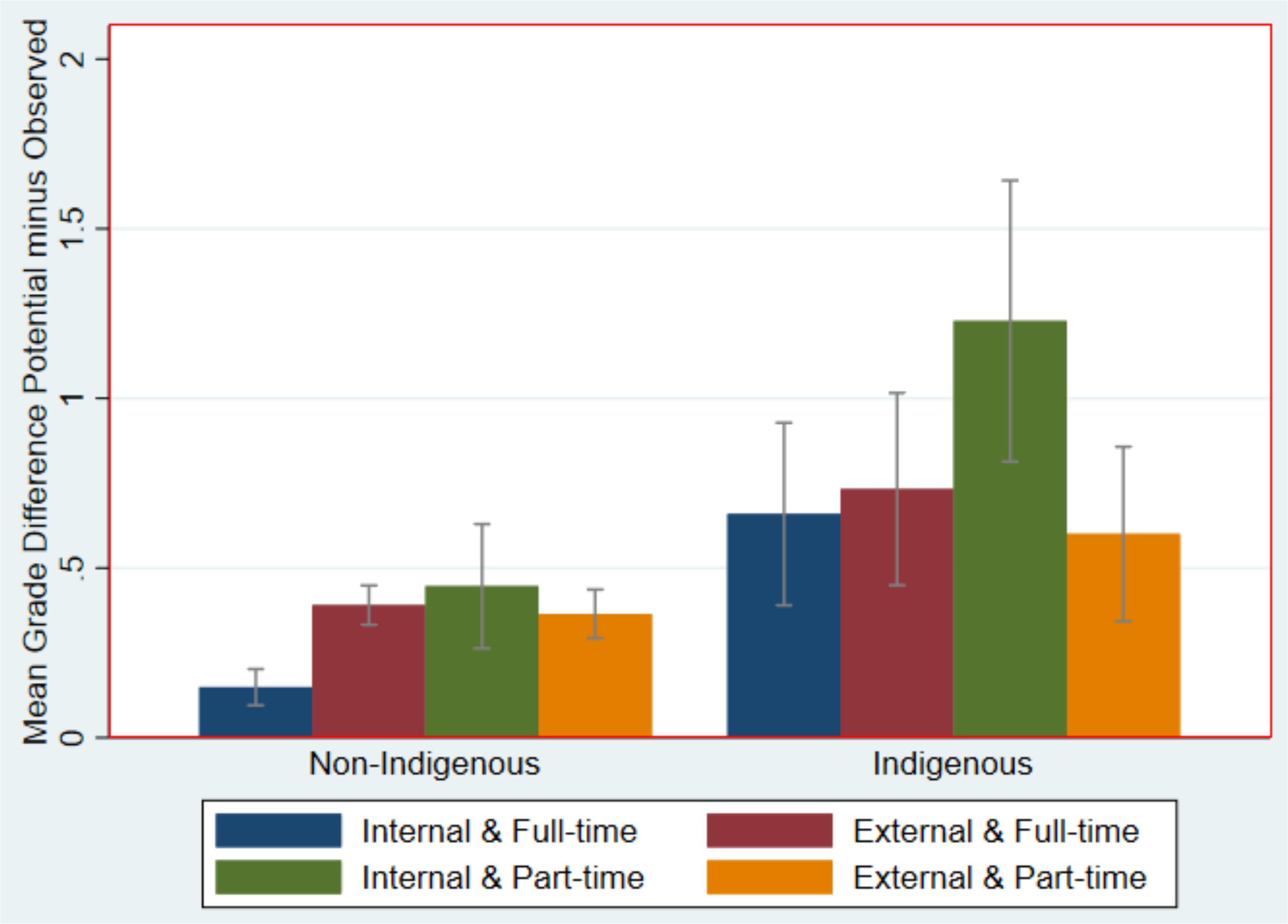
n=4,978 observations (standard error adjusted for 3,192 clusters)



*Blackboard Learnline Activity scores – Learnline is a compulsory learning system for all enrolments

Potential vs Observed Outcomes

Potential Gains: Indigenous Enrolments by Attendance Status



“Strains and Gains”: a Summary

Strains

- Causal attribution requires more sensitive discriminators for exposure vs participation when “treatment” (level of online engagemnt) is either compulsory or universal.
- Multivalued treatment scoring may complicate estimates of marginals and contrasts.
- Lack of multiway vce (cluster) restricts levels of “nested” effects estimation.

Gains

- ERM Release 15 provides consistent estimators in a complex Higher Education valuation research.
- Combined auxilliary equations (with eregress) can reproduce the non-linear fit of an OLS cubic expansion.
- Positive treatment effects of online engagement are unevenly distributed, with highest potential “gains” at the lower end of observed grade distribution.