Combining junior and NCEA assessments to ‘prove’ impact on learning

Darcy Fawcett, darcy@sounddata.co.nz

Director, Sound Data.

Today’s resources:

- Preliminary survey https://bit.ly/4qVXBWz
- Online dashboards sounddata.co.nz/dashboard/

SOUND DATA

Removing Noise From Decision Making

How can you ‘prove’ the impact of initiatives on learning?

- You must compare assessment data of those who experienced the initiative with those who didn’t
- The initiative group must make *significantly* more *progress* than the comparison group *and* this effect must be large enough to be worth attending to.

Experimental method

- Compare the progress of two equivalent groups
  - control and experiment group
- Often very difficult to arrange and can be unethical

Quasi-experimental method

- Compare naturally occurring groups
  - Compare the progress of cohorts
- Assumptions
  - latest cohort is otherwise typical
  - The historical record is large enough to describe the population

Establishing differences in achievement is not enough,

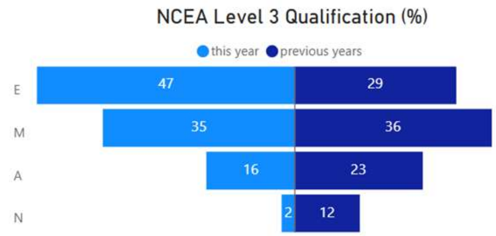
NCEA Qualifications: Demographic differences

Demographic

L3 Cohort

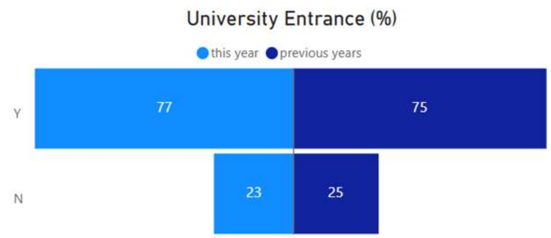
Level 3 NCEA

Demographic	Statement	n (D)	n (nD)	med (D)	med (nD)	p	ES
Asian	NO SIGNIFICANT DIFFERENCE, Asian > nonAsian	407	1669	M	M	0.275	0.02
Gender	minimal highly significant difference, Female > Male	1032	1044	M	M	0.000	0.09
L3 Cohort	small highly significant difference, this year > previous years	175	1901	M	M	0.000	0.12
Maori	NO SIGNIFICANT DIFFERENCE, nonMaori > Maori	289	1787	M	M	0.204	-0.03
MELAA	NO SIGNIFICANT DIFFERENCE, MELAA > nonMELAA	87	1989	M	M	0.975	0.00
Other	NO SIGNIFICANT DIFFERENCE, Other > nonOther	39	2037	M	M	0.208	0.03
Pacific	NO SIGNIFICANT DIFFERENCE, nonPacific > Pacific	289	1787	M	M	0.204	-0.03
Pakeha	minimal highly significant difference, Pakeha > nonPakeha	1373	703	M	M	0.000	0.08



University Entrance

Demographic	Statement	n (D)	n (nD)	med (D)	med (nD)	p	ES
Asian	NOT SIGNIFICANT, Asian > nonAsian	407	1669	Y	Y	0.75	0.02
Gender	NOT SIGNIFICANT, Female > Male	1032	1044	Y	Y	0.17	0.07
L3 Cohort	NOT SIGNIFICANT, this year > previous years	175	1901	Y	Y	0.65	0.05
Maori	NOT SIGNIFICANT, nonMaori > Maori	289	1787	Y	Y	0.51	-0.05
MELAA	NOT SIGNIFICANT, nonMELAA > MELAA	87	1989	Y	Y	0.80	-0.04
Other	NOT SIGNIFICANT, Other > nonOther	39	2037	Y	Y	0.58	0.13
Pacific	NOT SIGNIFICANT, Pacific > nonPacific	321	1755	Y	Y	0.89	0.01
Pakeha	NOT SIGNIFICANT, Pakeha > nonPakeha	1373	703	Y	Y	0.36	0.05



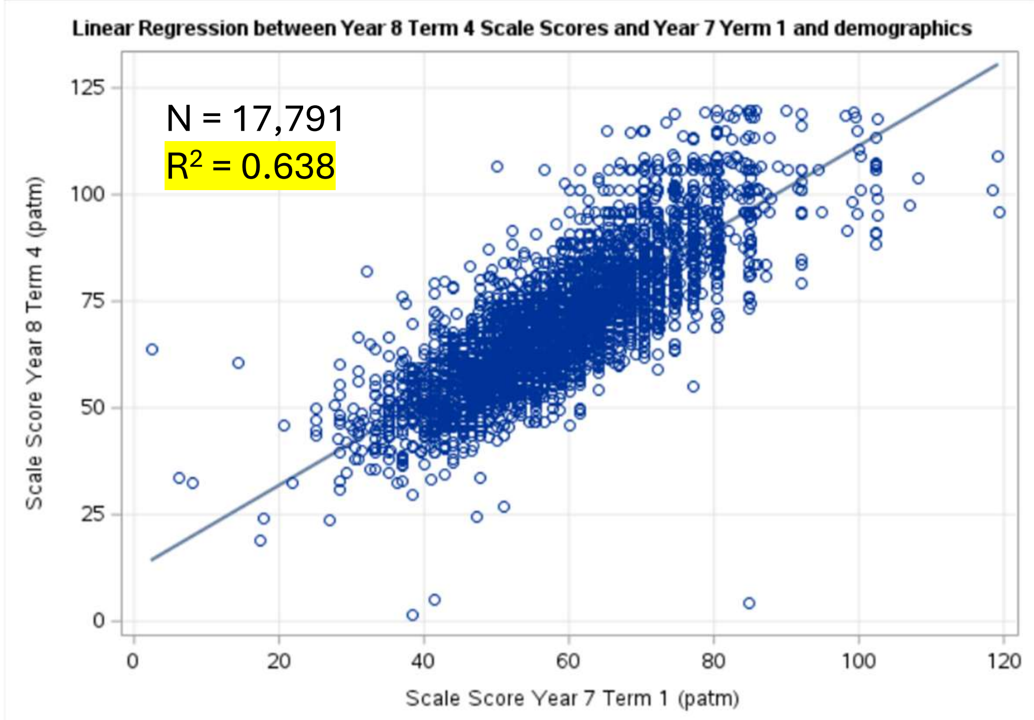
they may just have been a more able cohort!

Measuring progress requires evaluating achievement with a pre- and post-test

- If using the same assessment measure, then an independent sample or paired sample test is commonly used.
- If using different assessment measures, then some form of regression model can be used and the features of the model can describe learning.
- Think about science experiments and x-y graphs e.g., F vs a graphs

Linear regression model

- **Single or multivariate Linear Regressions**
  - **Independent variables:** 1+ continuous independent variables e.g., scale scores, credits; binary variables e.g., cohort, gender, non/Māori, etc
  - **Dependent variable**: 1 continuous variable e.g., scale scores, credits
- Explains that increases in or presence of independents causes increase in continuous dependent
- Impact of a teaching initiative would be shown by a significant cohort effect
  - e.g., this year's students gained an average of 1.30 more scale scores than in previous years
- However, NCEA Year Level Qualifications are ordinal rather than Scale Variables so a LINEAR regression cannot be used



Predictor	B	p
Intercept	14.2	< 0.0001
ScaleScore_TA	0.85	< 0.0001
Gender (Female)	1.03	< 0.0001
Cohort (this year)	1.30	< 0.0001

Evaluating Across Sector Learning.

Evaluating learning requires assessing students twice. When single pre- and post-tests use the same form of measurement, then a simple subtraction will represent progress/learning that took place. When different forms of measurement are used, more complex analysis is required. In this case where we use PAT Reading Comprehension and PAT Mathematics as our pre-tests and NCEA Qualifications for our post-tests, we analyse our data using an Ordinal Logistic Regression as illustrated on the following dashboard.

Dependent Variables: NCEA Level 3 and University Entrance Qualifications

NCEA Level 3 Qualification

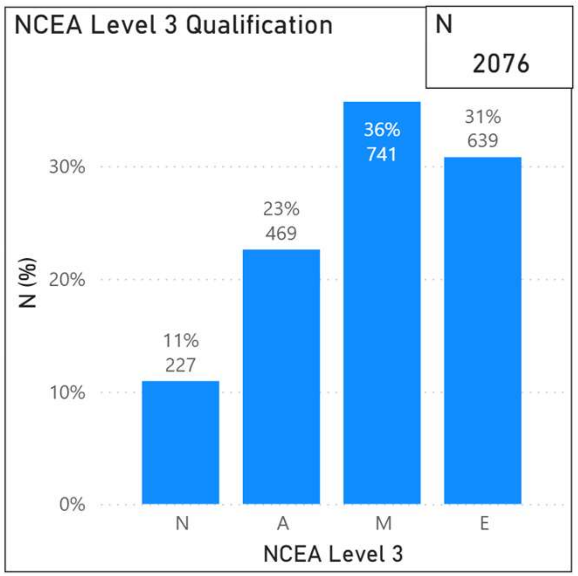
NCEA Level 3 forms ordered variables

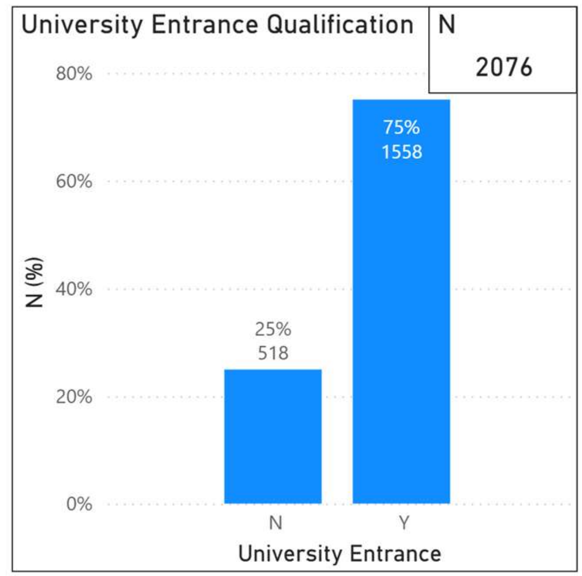
NCEA Level	Order
E	4
M	3
A	2
N	1

University Entrance

University Entrance forms a binary ordered variable

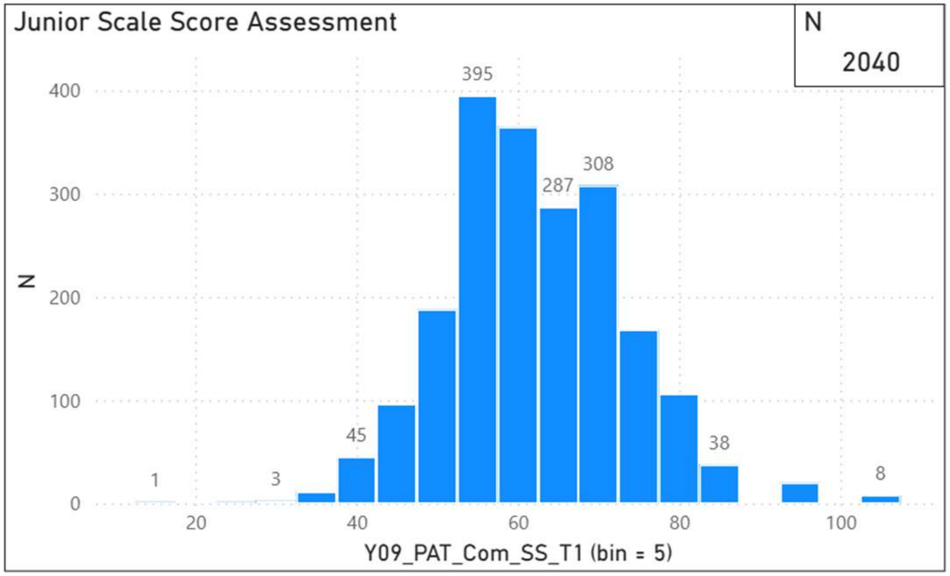
University Entrance	Order
Y	2
N	1

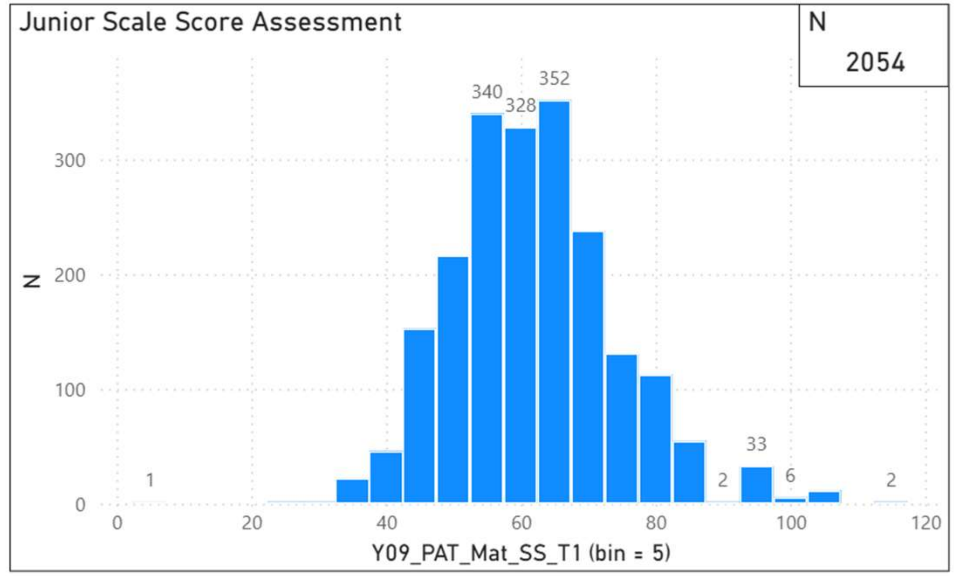




Are the chances of students gaining a higher NCEA Qualification able to be modeled and predicted by student and school attributes and activities, and thereby potentially influenced and changed?

Independent variables: Junior Assessment pre-tests, Demographics and Cohort





Continuous Variable Symbols and Names

Symbol	Name
Y09_PAT_Com_SS_T1	Year 09 PAT Reading Comprehension Scale Score Term 1
Y09_PAT_Com_ST_T1	Year 09 PAT Reading Comprehension Stanine Term 1
Y09_PAT_Mat_SS_T1	Year 09 PAT Reading Mathematics Scale Score Term 1
Y09_PAT_Mat_ST_T1	Year 09 PAT Reading Mathematics Stanine Term 1

Demographic variables

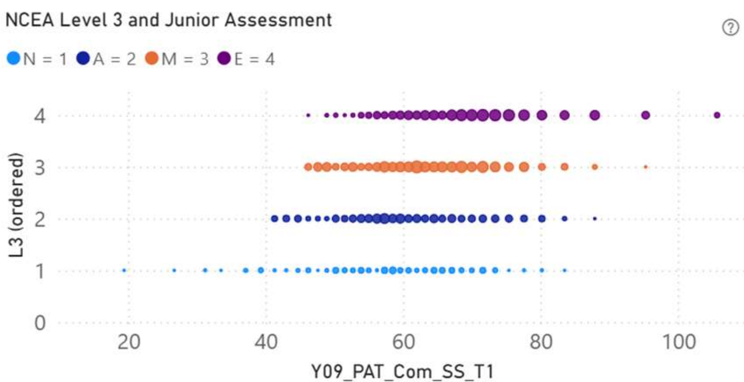
Asian	Female	Male	Maori	MELAA	Other	Pacific	Pakeha	previous years	this year
407	1032	1044	289	87	39	321	1373	1901	175

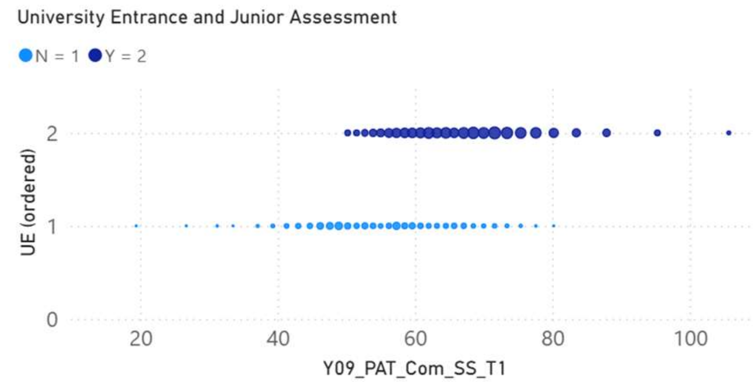
Junior assessment scale scores are continuous normally-distributed variables. Does a higher scale score increase the likelihood of a higher NCEA Qualification?
Demographic memberships are binary variables. Does being in a particular demographic (1) or not (0) increase the likelihood of a higher NCEA L3 Qualification?

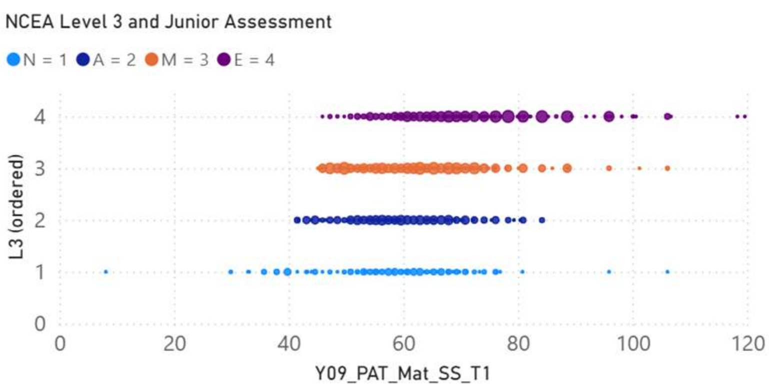
Correlations with ordinal variables:

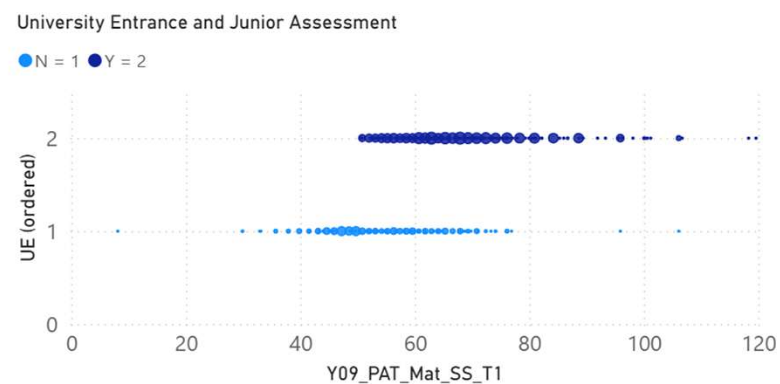
As the scale score increases, the **probability** of gaining a higher grade increases

Correlations between NCEA Qualification Ordinal Variables and Year 9 Scale Scores for Term 1









Spearman Correlations between NCEA Qualifications and Junior Assessments

Dep Var	Ind Var	Statement	N	R	p
L3_Endorsement_Ord	Y09_PAT_Com_SS_T1	significant correlation	2040	0.37	0.000
L3_Endorsement_Ord	Y09_PAT_Mat_SS_T1	significant correlation	2054	0.39	0.000

Spearman Correlations between University Entrance and Junior Assessments

Dep	Ind Var	Statement	N	R	p
UE_Ord	Y09_PAT_Com_SS_T1	significant correlation	2040	0.47	0.000
UE_Ord	Y09_PAT_Mat_SS_T1	significant correlation	2054	0.50	0.000

Logistic ordinal regression

- Logistic ordinal regressions investigate the relationship between ordered responses and a set of explanatory variables.
  - Ordinal outcome variables: NCEA Level 3 and University Entrance Qualifications
  - A range of predictor variable types: parametric (PAT and e-asTTle scale scores), ordinal (effort scores) and binary (cohort, gender, ethnic heritages, ...)
- Increase in or presence of independents cause increase in the probability of higher ordinal outcomes (n levels)

$\hat{Y}_i = \frac{e^u}{1-e^u}$ where $\hat{Y}_i$ is the cumulative probability of Y greater than or equal to an ordinal category (i = 1, 2, n-1)

where $u = B_{i0} + B_1X_1 + B_2X_2 + \ ... + B_kX_k$ with constant B_0, coefficients B_j , and predictors X_j , for *k* predictors (*j = 1, 2, 3, … k*).

The linear regression equation is the natural logarithm of the odds ratio to the predictors

$$ln\left(\frac{\hat{Y}_i}{1-\hat{Y}_i}\right) = B_{i0} + \sum B_jX_{ij}$$

where the goal is to find the *best* linear combination of predictors to maximise the likelihood of obtaining the observed outcome frequencies of the ordinal variable.

(Tabachnick & Fidell, 2014)

Assumptions and conditions

- **Absence of Multicollinearity:** independent variables not *too* strongly correlated ($R < 0.9$)
- **Optimisation technique**: Fisher's Scoring (backwards selection, $p \leq 0.2$)
- **Model Convergence Status (relative gradient convergence criterion)**: the maximum likelihood algorithm has converged
- **Score Test for the Proportional Odds Assumption:** the same coefficients for all ordinal contrasts
- **Model Fit Statistics (Akaike Information Criterion**) i.e., the backwards selection signifiantly improves the model
- **Likelihood ratio**: i.e., at least one coefficient is not equal to 0.
- **Adjusted / Pseudo R^2 (Nagelkerke):** describes the goodness of fit of the model to the data (weak ≤ 0.2, $0.2 \leq$ moderate < 0.4, $0.4 \leq$ strong)

Key features for teachers

Model name, dependent variable, number of students, p-value of model

First independent variable name, p-value, Odds ratio average, lower confidence limit, upper confidence limit

Ordinal Logistic Regression - Backwards Selection

Model	Dep	N	p	Convergence	Adj_R_sq	B1	B1_p	B1_Odds	B1_LCL	B1_UCL	B2	B2_p	B2_Odds	B2_LCL	B2_UCL
UE_M09_T1b	UE	2054	0.000	satisfied	0.41	Y09_PAT_Mat_SS_T1	0.000	1.18	1.16	1.20	Female_Ord	0.030	1.35	1.03	1.77

Convergence of model, Adjusted R^2 (goodness of fit)

Second independent variable name, p-value, Odds ratio average, lower confidence limit, upper confidence limit

The Odds Ratio is most important. For each one-unit increase in a scale score predictor, the **odds** of gaining a higher qualification (versus lower ones) change by a factor of the Odds Ratio, assuming all other variables remain constant. For categorical predictors, the odds ratio represents the ratio of the odds for a specific group compared to a reference group. It describes how much more (or less) likely members of that group are to reach a higher outcome level compared to the baseline group.

The Ordinal Logistic Regression Model linking **University Entrance** to Year 9 Term 1 assessments and demographics (version a)

Actual Qualification vs Most Likely Predicted Qualification

UE	N	Y	Total
N	10.9%	12.7%	23.6%
Y	3.4%	72.9%	76.4%
Total	14.3%	85.7%	100.0%

Correct predictions

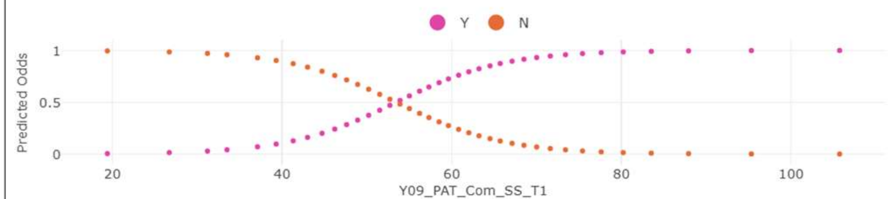
Summary of Odds

Predicted Grade	Min	Mean	Max
N	0.00	0.24	1.00
Y	0.00	0.76	1.00

Number of students

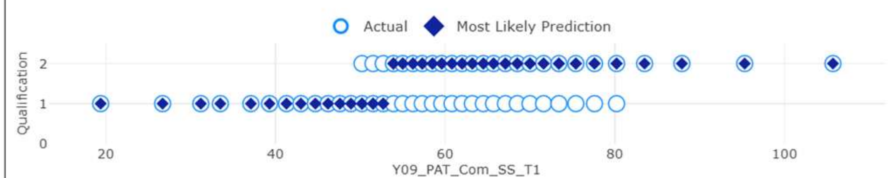
previous years	this year	Total
1867	173	**2040**

Predicted Odds for University Entrance Qualification Achievement by Junior Assessment Scale Score



Actual University Entrance Qualification and Most Likely Prediction by Junior Assessment Scale Score

N = 1, Y = 2



Asian: All
Maori: All
MELAA: All
Other: All
Pacific: All
Pakeha: All
Gender: All
L3 Cohort: All
Y09_PAT_Com_ST_T1: All

Ordinal Logistic Regression - Direct

Model	Dep	N	p	Convergence	Adj_R_sq	B1	B1_p	B2	B2_p	B3	B3_p	B4	B4_p	B5	B5_p	B6	B6_p	B7	B7_p
UE_M09_T1a	UE	2040	0.000	satisfied	0.37	Y09_PAT_Com_SS_T1	0.000	L3_Cohort_Ord	0.684	Female_Ord	0.349	Maori_Ord	0.723	Pacific_Ord	0.224	Asian_Ord	0.521	MELAA_Ord	0.713

B8	B8_p	B9	B9_p	B10	B10_p	B11	B11_p	B12	B12_p	linear_combination
Other_Ord	0.898	Pakeha_Ord	0.486							

Ordinal Logistic Regression - Backwards Selection

Model	Dep	N	p	Convergence	Adj_R_sq	B1	B1_p	B1_Odds	B1_LCL	B1_UCL	B2	B2_p	B2_Odds	B2_LCL	B2_UCL	B3	B3_p	B3_Odds	B3_LCL	B3_UCL
UE_M09_T1a	UE	2040	0.000	satisfied	0.37	Y09_PAT_Com_SS_T1	0.000	1.17	1.15	1.19										

B4	B4_p	B4_Odds	B4_LCL	B4_UCL	B5	B5_p	B5_Odds	B5_LCL	B5_UCL	B6	B6_p	B6_Odds	B6_LCL	B6_UCL

The Odds Ratio is the most important feature!

For every 1 increase in Com scale scores, there is a _____ increase in the chances of gaining University Entrance

This model is making correct predictions of N _____% of the time and Y _____% of the time, for a total of _____% correct predictions

At what scale score does it become more likely to gain UE than not ?

The Ordinal Logistic Regression Model linking **University Entrance** to Year 9 Term 1 assessments and demographics (version b)

Actual Qualification vs Most Likely Predicted Qualification

UE	N	Y	**Total**
N	13.1%	11.1%	**24.2%**
Y	3.4%	72.4%	**75.8%**
Total	**16.5%**	**83.5%**	**100.0%**

Correct predictions

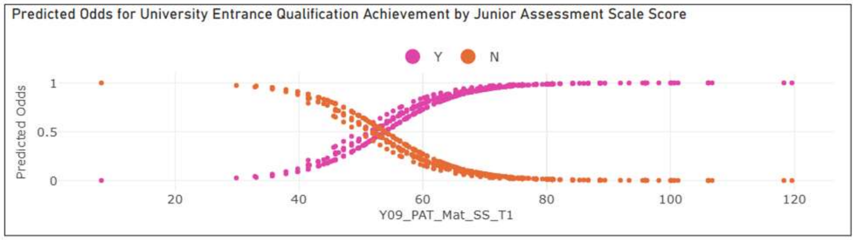
N 14% | Y 86%

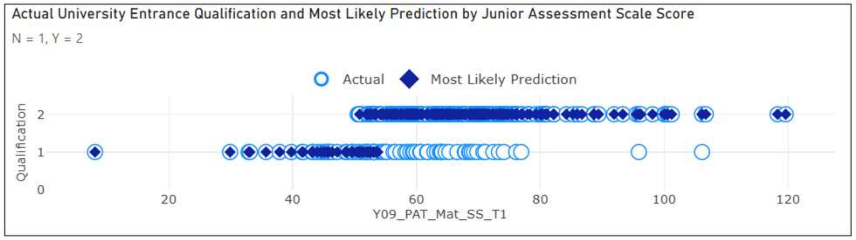
Summary of Odds

Predicted Grade	Min	Mean	Max
N	0.00	0.24	1.00
Y	0.00	0.76	1.00

Number of students

previous years	this year	**Total**
1881	173	**2054**





Asian: All
Maori: All
MELAA: All
Other: All
Pacific: All
Pakeha: All
Gender: All
L3 Cohort: All
Y09_PAT_Mat_ST_T1: All

Ordinal Logistic Regression - Direct

Model	Dep	N	p	Convergence	Adj_R_sq	B1	B1_p	B2	B2_p	B3	B3_p	B4	B4_p	B5	B5_p	B6	B6_p	B7	B7_p
UE_M09_T1b	UE	2054	0.000	satisfied	0.41	Y09_PAT_Mat_SS_T1	0.000	L3_Cohort_Ord	0.084	Female_Ord	0.068	Maori_Ord	0.735	Pacific_Ord	0.332	Asian_Ord	0.490	MELAA_Ord	0.821

B8	B8_p	B9	B9_p	B10	B10_p	B11	B11_p	B12	B12_p	linear_combination
Other_Ord	0.842	Pakeha_Ord	0.862							

Ordinal Logistic Regression - Backwards Selection

Model	Dep	N	p	Convergence	Adj_R_sq	B1	B1_p	B1_Odds	B1_LCL	B1_UCL	B2	B2_p	B2_Odds	B2_LCL	B2_UCL	B3	B3_p	B3_Odds	B3_LCL	B3_UCL
UE_M09_T1b	UE	2054	0.000	satisfied	0.41	Y09_PAT_Mat_SS_T1	0.000	1.18	1.16	1.20	Female_Ord	0.030	1.35	1.03	1.77	Pacific_Ord	0.089	0.72	0.49	1.05

B4	B4_p	B4_Odds	B4_LCL	B4_UCL	B5	B5_p	B5_Odds	B5_LCL	B5_UCL	B6	B6_p	B6_Odds	B6_LCL	B6_UCL
L3_Cohort_Ord	0.090	1.46	0.94	2.27										

For every 1 increase in Mat scale scores, there is a _____ increase in the chances of gaining University Entrance

If female, these odds go up by a factor of _________, for a total odds of _________

The odds ratios for the Cohort_ord variable range between ________ and ________? How is this related to the p-value of _______? Did this year's students really gain UE at a higher rate when their Mat Scale scores are taken into account?

Why are their 4 lines of Excellence in the "Predicted Odds for University Entrance …"?

The Ordinal Logistic Regression Model linking **University Entrance** to Year 9 Term 1 assessments and demographics (version c)

Actual Qualification vs Most Likely Predicted Qualification

UE	N	Y	Total
N	14.2%	8.6%	22.8%
Y	4.0%	73.1%	77.2%
Total	18.2%	81.8%	100.0%

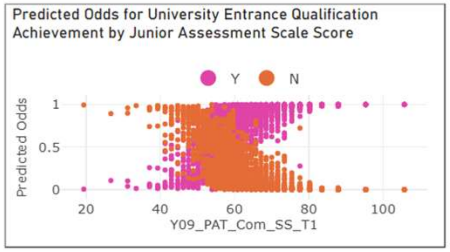
Correct predictions

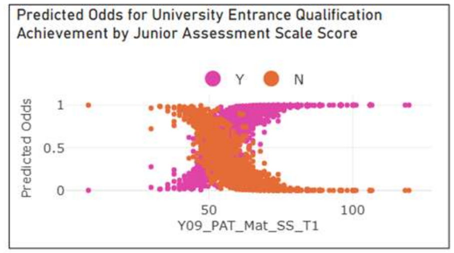
Summary of Odds

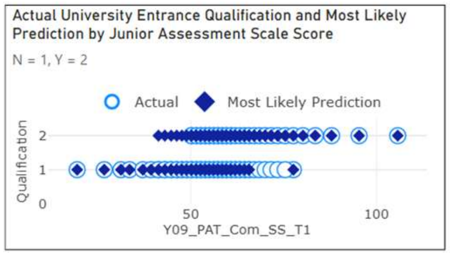
Predicted Grade	Min	Mean	Max
N	0.00	0.23	1.00
Y	0.00	0.77	1.00

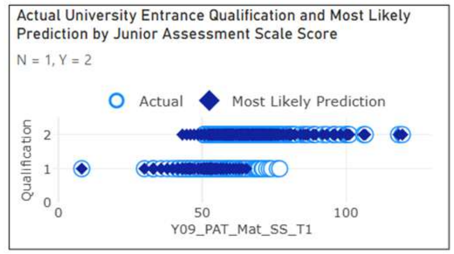
Number of students

previous years	this year	Total
1847	171	**2018**









Asian	All
Maori	All
MELAA	All
Other	All
Pacific	All
Pakeha	All
Gender	All
L3 Cohort	All
Y09_PAT_Com_ST_T1	All
Y09_PAT_Mat_ST_T1	All

Ordinal Logistic Regression - Direct

Model	Dep	N	p	Convergence	Adj_R_sq	B1	B1_p	B2	B2_p	B3	B3_p	B4	B4_p	B5	B5_p	B6	B6_p	B7	B7_p
UE_M09_T1c	UE	2018	0.000	satisfied	0.55	Y09_PAT_Mat_SS_T1	0.000	Y09_PAT_Com_SS_T1	0.000	L3_Cohort_Ord	0.239	Female_Ord	0.139	Maori_Ord	0.910	Pacific_Ord	0.214	Asian_Ord	0.941

B8	B8_p	B9	B9_p	B10	B10_p	B11	B11_p	B12	B12_p	linear_combination
MELAA_Ord	0.823	Other_Ord	0.639	Pakeha_Ord	0.728					

Ordinal Logistic Regression - Backwards Selection

Model	Dep	N	p	Convergence	Adj_R_sq	B1	B1_p	B1_Odds	B1_LCL	B1_UCL	B2	B2_p	B2_Odds	B2_LCL	B2_UCL	B3	B3_p	B3_Odds	B3_LCL	B3_UCL
UE_M09_T1c	UE	2018	0.000	satisfied	0.55	Y09_PAT_Mat_SS_T1	0.000	1.16	1.14	1.18	Y09_PAT_Com_SS_T1	0.000	1.12	1.10	1.14	Female_Ord	0.065	1.34	0.98	1.82

B4	B4_p	B4_Odds	B4_LCL	B4_UCL	B5	B5_p	B5_Odds	B5_LCL	B5_UCL	B6	B6_p	B6_Odds	B6_LCL	B6_UCL
Pacific_Ord	0.071	0.67	0.43	1.04										

For every 1 increase in Com scale scores, there is a _____ increase in the chances of gaining University Entrance

For every 1 increase in Mat scale scores, there is a _____ increase in the chances of gaining University Entrance.

To what extent is this model *better* than the previous two models? Justify your answer in terms of correct predictions and the goodness of fit (adjusted R^2). Describe the differecnes in the two types of graphs.

The Ordinal Logistic Regression Model linking NCEA Level 3 to Year 9 Term 1 assessments and demographics (version a)

Actual Qualification vs Most Likely Predicted Qualification

Actual \ Predicted	N	A	M	E	Total
N	0.6%	2.3%	5.7%	0.7%	**9.4%**
A		4.7%	14.3%	4.0%	**23.0%**
M		5.0%	22.9%	8.4%	**36.3%**
E		0.4%	15.7%	15.2%	**31.3%**
Total	**0.6%**	**12.4%**	**58.7%**	**28.3%**	**100.0%**

Correct predictions

N 57% Y 43%

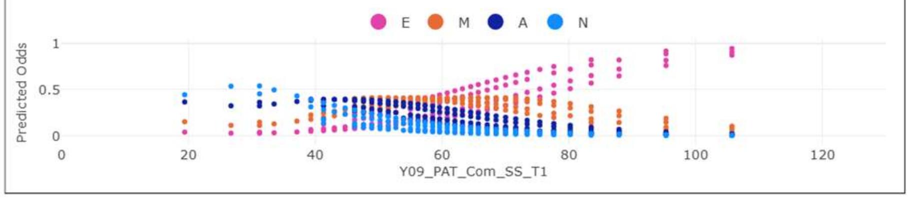
Summary of Odds

Predicted Grade	Min	Mean	Max
N	0.00	0.09	0.54
A	0.01	0.23	0.40
M	0.05	0.36	0.41
E	0.03	0.32	0.94

Number of students

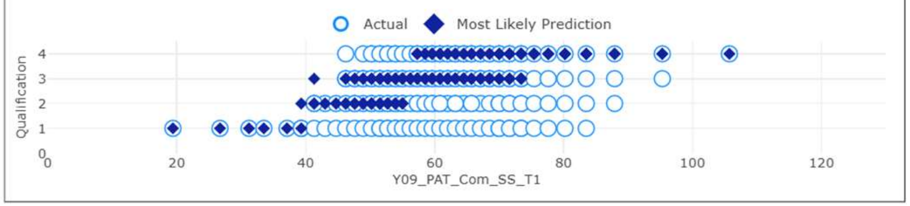
previous years	this year	Total
1867	173	**2040**

Predicted Odds for each type of NCEA L3 Qualification Grade by Junior Assessment Scale Score



Actual NCEA L3 Qualification and Most Likely Prediction by Junior Assessment Scale Score

N = 1, A = 2, M = 3, E = 4



Asian	Maori
All	All

MELAA	Other
All	All

Pacific	Pakeha
All	All

Gender	L3 Cohort
All	All

Y09_PAT_Com_ST_T1
All

Ordinal Logistic Regression - Direct

Model	Dep	N	p	Convergence	Adj_R_sq	B1	B1_p	B2	B2_p	B3	B3_p	B4	B4_p	B5	B5_p	B6	B6_p	B7	B7_p
L3_M09_T1a	NCEA L3	2040	0.000	satisfied	0.19	Y09_PAT_Com_SS_T1	0.000	L3_Cohort_Ord	0.000	Female_Ord	0.024	Maori_Ord	0.661	Pacific_Ord	0.867	Asian_Ord	0.335	MELAA_Ord	0.395

B8	B8_p	B9	B9_p	B10	B10_p	B11	B11_p	B12	B12_p	linear_combination
Other_Ord	0.746	Pakeha_Ord	0.184							

Ordinal Logistic Regression - Backwards Selection

Model	Dep	N	p	Convergence	Adj_R_sq	B1	B1_p	B1_Odds	B1_LCL	B1_UCL	B2	B2_p	B2_Odds	B2_LCL	B2_UCL	B3	B3_p	B3_Odds	B3_LCL	B3_UCL
L3_M09_T1a	NCEA L3	2040	0.000	satisfied	0.19	Y09_PAT_Com_SS_T1	0.000	1.08	1.07	1.09	L3_Cohort_Ord	0.000	2.48	1.84	3.36	Female_Ord	0.000	1.41	1.20	1.66

B4	B4_p	B4_Odds	B4_LCL	B4_UCL	B5	B5_p	B5_Odds	B5_LCL	B5_UCL	B6	B6_p	B6_Odds	B6_LCL	B6_UCL

For every 1 increase in Com scale scores, there is a _____ increase in the chances of gaining a higher Level 3 Qualification

If in this cohort and a girl, these odds go up by a factor of _________ and _________, for a total predicted odds ratio of ______

If in Stanine 9, the chance of getting an Excellence are between __________ and __________

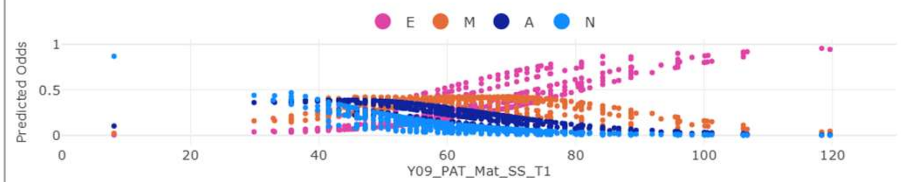
This goodness of fit of the model is weak/moderate/strong.

The Ordinal Logistic Regression Model linking NCEA Level 3 to Year 9 Term 1 assessments and demographics (version b)

Actual Qualification vs Most Likely Predicted Qualification

Actual \ Predicted	N	A	M	E	Total
N	0.8%	2.2%	6.3%	0.7%	**10.0%**
A		5.2%	14.6%	3.0%	**22.8%**
M		5.5%	22.6%	8.0%	**36.1%**
E		0.7%	14.9%	15.4%	**31.1%**
Total	**0.8%**	**13.6%**	**58.4%**	**27.2%**	**100.0%**

Predicted Odds for each type of NCEA L3 Qualification Grade by Junior Assessment Scale Score



Asian: All

Maori: All

MELAA: All

Other: All

Pacific: All

Pakeha: All

Gender: All

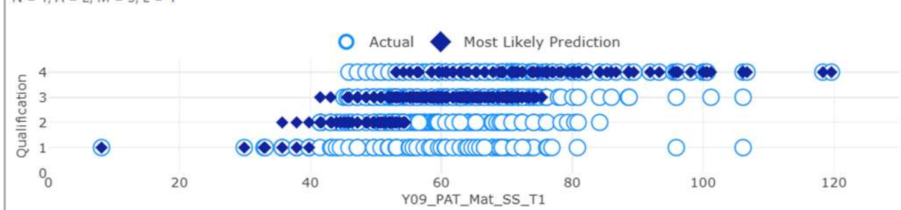
L3 Cohort: All

Correct predictions

N 56% | Y 44%

Actual NCEA L3 Qualification and Most Likely Prediction by Junior Assessment Scale Score

N = 1, A = 2, M = 3, E = 4



Summary of Odds

Predicted Grade	Min	Mean	Max
N	0.00	0.10	0.87
A	0.01	0.22	0.39
M	0.02	0.36	0.42
E	0.00	0.31	0.95

Y09_PAT_Mat_ST_T1: All

Number of students

previous years	this year	Total
1881	173	**2054**

Ordinal Logistic Regression - Direct

Model	Dep	N	p	Convergence	Adj_R_sq	B1	B1_p	B2	B2_p	B3	B3_p	B4	B4_p	B5	B5_p	B6	B6_p	B7	B7_p	B8	B8_p
L3_M09_T1b	NCEA L3	2054	0.000	satisfied	0.22	Y09_PAT_Mat_SS_T1	0.000	L3_Cohort_Ord	0.000	Female_Ord	0.002	Maori_Ord	0.355	Pacific_Ord	0.663	Asian_Ord	0.081	MELAA_Ord	0.108	Other_Ord	0.788

B9	B9_p	B10	B10_p	B11	B11_p	B12	B12_p	linear_combination
Pakeha_Ord	0.147							

Ordinal Logistic Regression - Backwards Selection

Model	Dep	N	p	Convergence	Adj_R_sq	B1	B1_p	B1_Odds	B1_LCL	B1_UCL	B2	B2_p	B2_Odds	B2_LCL	B2_UCL	B3	B3_p	B3_Odds	B3_LCL	B3_UCL
L3_M09_T1b	NCEA L3	2054	0.000	satisfied	0.22	Y09_PAT_Mat_SS_T1	0.000	1.08	1.07	1.08	L3_Cohort_Ord	0.000	3.00	2.21	4.07	Female_Ord	0.000	1.71	1.40	2.08

B4	B4_p	B4_Odds	B4_LCL	B4_UCL	B5	B5_p	B5_Odds	B5_LCL	B5_UCL	B6	B6_p	B6_Odds	B6_LCL	B6_UCL
Asian_Ord	0.086	0.81	0.63	1.03	MELAA_Ord	0.096	0.70	0.45	1.07					

For every 1 increase in Mat scale scores, there is a _____ increase in the chances of gaining a higher Level 3 Qualification

If in this cohort and a girl, these odds go up by a factor of _________ and _________, for a total predicted odds ratio of ______

If in Stanine 5 and above, your chaces of getting an Excellence are between __________ and __________

This goodness of fit of the model is weak/moderate/strong.

The Ordinal Logistic Regression Model linking NCEA Level 3 to Year 9 Term 1 assessments and demographics (version c)

Actual Qualification vs Most Likely Predicted Qualification

Actual \ Predicted	N	A	M	E	Total
N	0.1%	2.4%	3.2%	0.5%	6.2%
A		2.8%	13.9%	4.4%	21.1%
M		6.0%	22.5%	9.5%	37.9%
E		0.2%	12.7%	21.8%	34.8%
Total	**0.1%**	**11.4%**	**52.4%**	**36.2%**	**100.0%**

Correct predictions

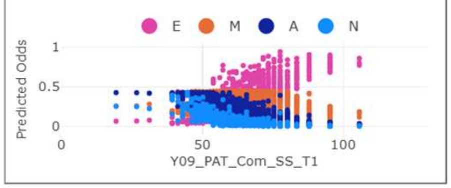
Summary of Odds

Predicted Grade	Min	Mean	Max
N	0.00	0.07	0.43
A	0.01	0.20	0.43
M	0.05	0.37	0.44
E	0.03	0.36	0.94

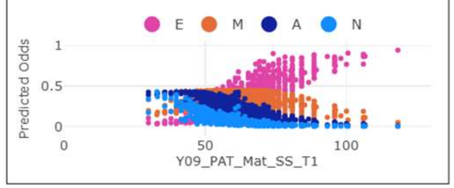
Number of students

previous years	this year	Total
919	85	**1004**

Predicted Odds for each type of NCEA L3 Qualification Grade by Junior Assessment Scale Score

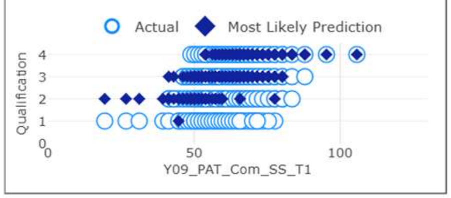


Predicted Odds for each type of NCEA L3 Qualification Grade by Junior Assessment Scale Score



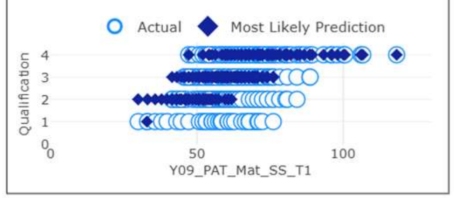
Actual NCEA L3 Qualification and Most Likely Prediction by Junior Assessment Scale Score

N = 1, A = 2, M = 3, E = 4



Actual NCEA L3 Qualification and Most Likely Prediction by Ju Assessment Scale Score

N = 1, A = 2, M = 3, E = 4



Asian	All
Maori	All
MELAA	All
Other	All
Pacific	All
Pakeha	All
Gender	Female
L3 Cohort	All
Y09_PAT_Com_ST_T1	All
Y09_PAT_Mat_ST_T1	All

Ordinal Logistic Regression - Direct

Model	Dep	N	p	Convergence	Adj_R_sq	B1	B1_p	B2	B2_p	B3	B3_p	B4	B4_p	B5	B5_p	B6	B6_p	B7	B7_p	B8	B8_p
L3_M09_T1c	NCEA L3	2018	0.000	satisfied	0.27	Y09_PAT_Mat_SS_T1	0.000	Y09_PAT_Com_SS_T1	0.000	L3_Cohort_Ord	0.000	Female_Ord	0.005	Maori_Ord	0.443	Pacific_Ord	0.738	Asian_Ord	0.165	MELAA_Ord	0.225

B9	B9_p	B10	B10_p	B11	B11_p	B12	B12_p	linear_combination
Other_Ord	0.580	Pakeha_Ord	0.158					

Ordinal Logistic Regression - Backwards Selection

Model	Dep	N	p	Convergence	Adj_R_sq	B1	B1_p	B1_Odds	B1_LCL	B1_UCL	B2	B2_p	B2_Odds	B2_LCL	B2_UCL	B3	B3_p	B3_Odds	B3_LCL	B3_UCL
L3_M09_T1c	NCEA L3	2018	0.000	satisfied	0.27	Y09_PAT_Mat_SS_T1	0.000	1.06	1.05	1.07	Y09_PAT_Com_SS_T1	0.000	1.05	1.04	1.06	L3_Cohort_Ord	0.000	3.00	2.19	4.09

B4	B4_p	B4_Odds	B4_LCL	B4_UCL	B5	B5_p	B5_Odds	B5_LCL	B5_UCL	B6	B6_p	B6_Odds	B6_LCL	B6_UCL
Female_Ord	0.000	1.50	1.27	1.77										

To what extent has the teaching initiative experienced by this year's cohort enhanced their learning? Were these gains, if any, experienced the same by all demographics?

Feedback data collected from 23 middle leaders

Dashboards are useful and easy to use!

How confident are you in your ability to use and interpret the DASHBOARD:	Confident or very confident
Independently?	61%
With coaching?	100%

How useful is the data in the DASHBOARD for:	Useful or very useful
understanding student achievement and progress?	91%
evaluating teaching/learning initiatives?	78%
informing future actions?	83%

QUESTION: Does a Wānanga day enhance student outcomes?

HERES WHAT HAPPENED IN 2019:

Wānanga Day and the MCAT
91027 Apply algebraic procedures in solving problems

Wānanga?
In essence students were given a day of supervised preparation for the 2019 MCAT. The MCAT has had a lot of media coverage over its difficultly through students and teachers being underprepared for the external.

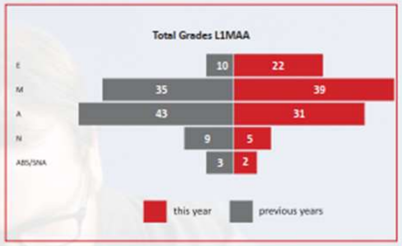
Teachers and NCEA students outraged over difficult Level 1 MCAT algebra exam – ***Stuff, 2016***
7 hacks for surviving NZ's toughest NCEA exam – ***Scoop, 2017***

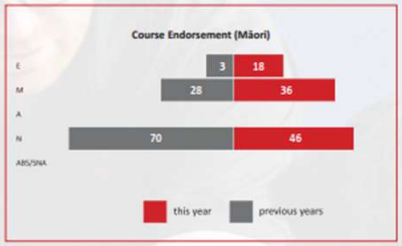
Methodology
After hearing positive results from the Art and Māori departments in instituting Wānanga days for 2019 the Mathematics department decided for all students sitting the external MCAT to have a day of organised preparation.

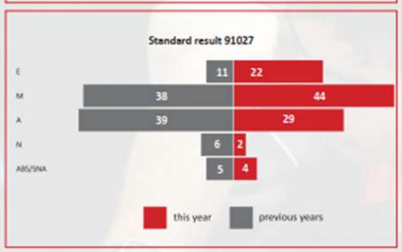
- used the library and ran various presentations and activities for students the day prior to the external.

- Students where supplied lunch during a break.

Results Summary
- L1MAA students gained better overall grades in all topics. This difference was highly significant ($p = <0.0001$).
- Course endorsement improved overall and this improvement was due to Māori students. This was large and significant ($p = 0.002$, $G = 0.475$).
- MCAT showed significant improvement of grades earned including double the percentage of students gaining excellence ($p = 0.016$).







Conclusion
- Using a Wānanga day has enhanced student outcomes.

It is an opportunity to establish a learning focused relationship between students, and teachers. It is a time to fully focus on a standard without external distractions.
Although a significant advantage was established it was only for one standard, the results indicate that this practice should be promoted within the department and school. It would require a major change to the timetable.

GOAL: Was to significantly lift the rate of higher grades and endorsements at Level 1.

HERES WHAT HAPPENED IN 2019:

Highlights
- 90914 ***(IN)*** - Historically 19% Non achieved.

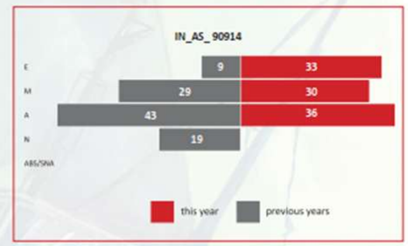
2019 - Large, highly significant improvement.

- 90916 ***(EX)*** - Historically 23% - SNA, 9% NA.

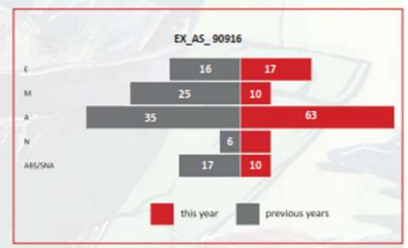
2019 - Similar to previous years.

What did we do?
- The Wānanga day was a huge success. 100% of the students achieved their internal allowing us to move as a majority onto the external submissions (closure/peace of mind/admin done).
- Proposition had a clear direction. ***Unpack Arrival***
- Collaboration - Two classes - SM/WO

Work on ...
- We achieved our goal from 2019 for the internals by significantly improving both earned and quality of grade. However there was a shift in the externals both up and down.



Externals
SNA - DOWN 7%, N/A - DOWN 6%,
Achieved - UP 28%, Merit - DOWN 15%
Excellence - UP 1%



- As with the internals from 2019, the externals show a ***large portion*** of students gaining achievement. We have not yet fixed this issue!

- Externals are a ***school concern*** and our subject is no different.

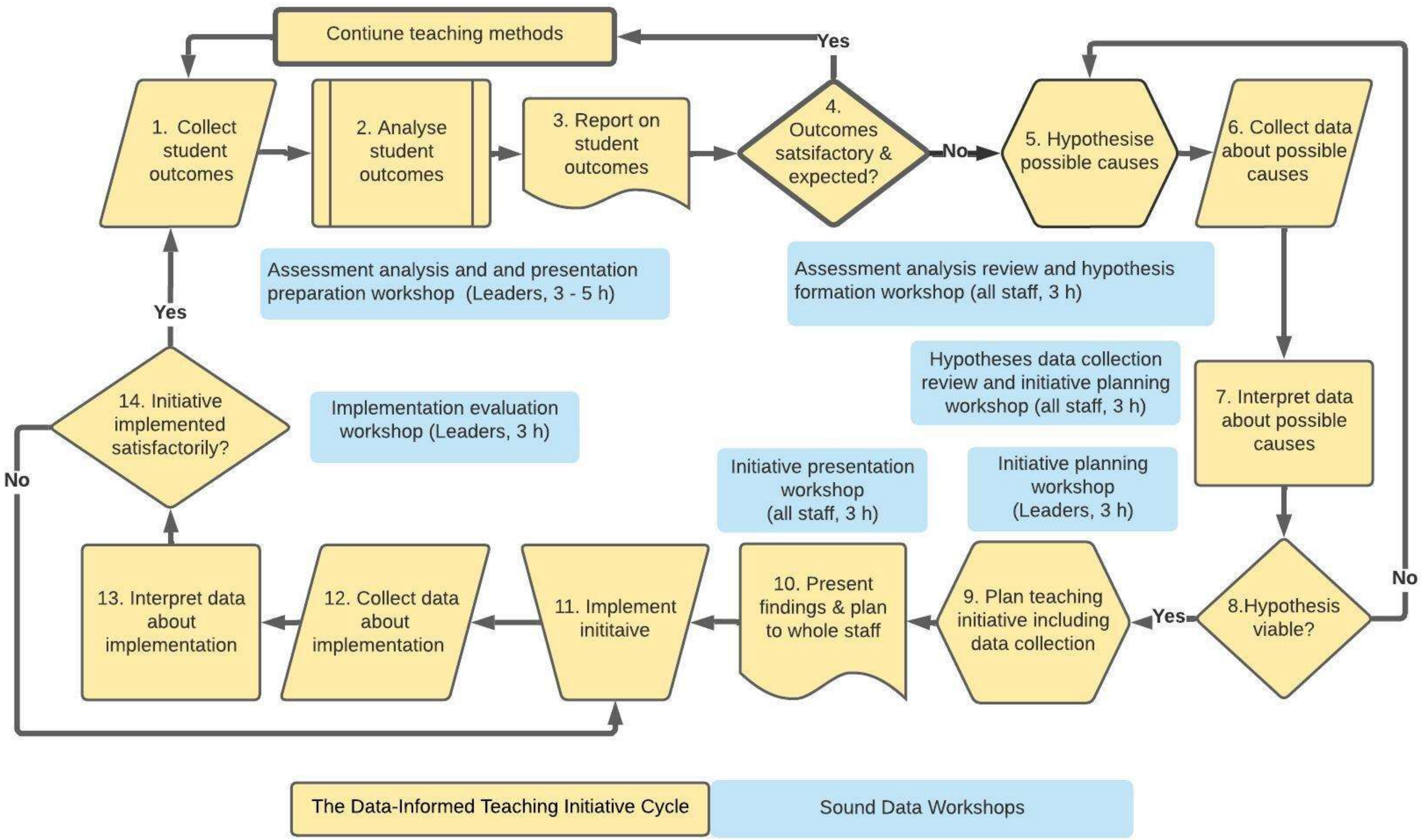
Data presentations are useful and have an impact!

How useful were the poster and presentations for:	Useful or very useful
Clarifying your departmental challenges, goals and plans?	78%
Sharing your departmental challenges, goals and plans?	83%
Understanding the challenges, goals and plans of the other departments?	83%

Survey question	Impact or large impact
Do you think your POSTER PRESENTATION had an impact on your teaching colleagues?	61%

Leaders are confident and value data-informed teaching initiatives

Survey Question	x or very x
How **confident** are you that your evidence-based initiative will lead to improvement?	78%
Were the outcomes of this process **worth** the effort you had to put in?	78%
How **likely** are you to recommend this process to colleagues at other schools?	78%
How **keen** are you to engage in this process again next year?	78%



(Lai & Schildkamp 2013, Mandinach & Gummer, 2015, Schildkamp et al 2018)

Data Driven Decison Making (DDDM)

- DDDM is a cyclic process of purposeful knowledge construction and decision/action (Lai & Schildkamp, 2013).
- *Evaluation* data use: positive or negative implications for user (King & Pechman, 1984);
- How teachers make sense of, respond to, and use *evaluation data* depends upon:
  - the data:
    - high/low stakes (Lewis & Holloway, 2019);
    - the data use technology (Fjørtoft & Lai, 2020);
  - the individual user:
    - beliefs and attiudes (Prenger & Schildkamp, 2018);
    - data literacy (Mandinach et al., 2015);
    - implications of data for teacher identity (Lasater, Bengtson, et al., 2020);
      - attribute the causes of the data:
        - to instruction, student understanding, nature of the test, student charatersitics (Bertrand & Marsh, 2015);
        - to factors outside the school, school policy, students/classroom, teacher functioning (Schildkamp, Poortman, Handelzalts, 2016);
  - social interactions:
    - leaders (Datnow & Park, 2019);
    - the depth, formality and transparency of data use (Vanlommel & Schildkamp, 2019);
    - the school data culture (Lasater, Albiladi, et al., 2020);
    - the external accountability system (Hargreaves et al., 2013).

Using data to improve teaching, learning and accountability

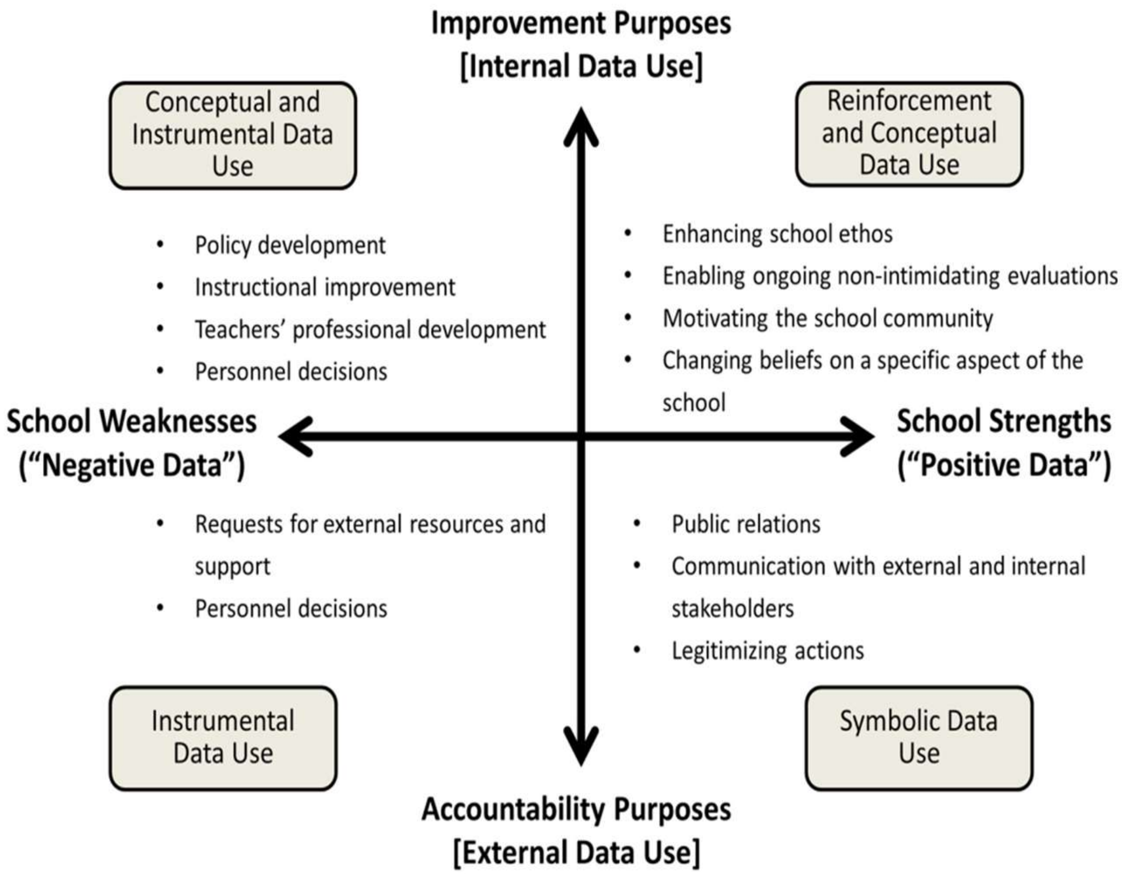
Data-informed teaching initiatives can improve teaching, learning *and* accountability when:

- They use a wide variety of *valid* data (input/process/output);
- They use postive data to celebrate *strengths* and successes as well as negative data to indicate *weakness* for improvement;
- Leaders and teachers are coached and resourced for *collaborative* data use;
- Leaders and teachers having *agency* in their data use in a *safe data-culture*.

(Marsh, Bertrand et al., 2015; Schildkamp & Poortman, 2015; Poortman & Schildkamp, 2016; Aderet-German & Ben-Peretz, 2020; Lai, McNaughton et al 2020, Lasater, Albiladi, et al., 2020)

"Accountability without improvement is empty rhetoric, and improvement without accountability is whimsical action without direction" (Earl and Katz, 2006).

The desirable uses of data in schools



- instrumental use – to directly influence actions and practice;
- conceptual use – to influence mental models and activity over time;
- persuasive or symbolic use – for a predetermined agenda
- reinforcement – to consciously highlight individual and school strengths.

(Aderet-German & Ben-Peretz, 2020)

The undesirable uses of data in schools

- non-use – ignoring or cursory data use i.e., ticking the box;
- interpreting data incorrectly – e.g., insufficient data literacy leading to inappropriate decisions or actions;
- narrowing the curriculum or teaching to the test;
- falsifying data – e.g., giving inappropriate assistance before or during assessments;
- educational triage and reshaping the data pool – focusing only on those on the threshold of passing and transferring less able students to non-examined classes;
- bullying or shaming teachers, leaders and schools

Data is misused when:

- Data use misrepresents reality to stakeholders;
- Data use has negative consequences for students;
- Data use removes the opportunity for the improvement of teaching.

(Booher-Jennings, 2005; Buly & Valencia, 2002; Ehren & Swanborn, 2012; Schildkamp & Ehren, 2013; Ford, 2018; Hardy & Lewis, 2017; Susan, 2016; Datnow & Park, 2018; Volante et al., 2020; Lockton et al., 2020; Bertrand & Marsh, 2021)

References

Aderet-German, T., & Ben-Peretz, M. (2020). Using data on school strengths and weaknesses for school improvement. *Studies in Educational Evaluation*, *64*, 100831. Retrieved 2020/03/01/, from http://www.sciencedirect.com/science/article/pii/S0191491X19300343

Bertrand, M., & Marsh, J. (2021). How data-driven reform can drive deficit thinking. *Phi Delta Kappan*, *102*(8), 35-39. https://doi.org/10.1177/00317217211013936

Booher-Jennings, J. (2005). Below the Bubble: "Educational Triage" and the Texas Accountability System. *American Educational Research Journal*, *42*(2), 231-268. http://www.jstor.org.ezproxy.waikato.ac.nz/stable/3699376

Buly, M. R., & Valencia, S. W. (2002). Below the Bar: Profiles of Students Who Fail State Reading Assessments. *Educational Evaluation and Policy Analysis*, *24*(3), 219-239. https://doi.org/10.3102/01623737024003219

Datnow, A., & Park, V. (2018). Opening or Closing Doors for Students? Equity and Data Use in Schools. *Journal of Educational Change*, *19*(2), 131-152. https://doi.org/10.1007/s10833-018-9323-6

Earl, L. M., & Katz, S. (2006). *Leading schools in a data-rich world: Harnessing data for school improvement.* Corwin Press.

Ehren, M. C. M., & Swanborn, M. S. L. (2012). Strategic data use of schools in accountability systems. *School Effectiveness and School Improvement: Special Issue: Data-Based Decision Making around the World: From Policy to Practice to Results*, *23*(2), 257-280. https://doi.org/10.1080/09243453.2011.652127

Ford, T. G. (2018). Pointing teachers in the wrong direction: understanding Louisiana elementary teachers' use of Compass high-stakes teacher evaluation data. *Educational Assessment, Evaluation and Accountability*, *30*(3), 251-283. https://doi.org/10.1007/s11092-018-9280-x

Hardy, I., & Lewis, S. (2017). The ‘doublethink’ of data: educational performativity and the field of schooling practices. *British Journal of Sociology of Education*, *38*(5), 671-685. https://doi.org/10.1080/01425692.2016.1150155

Lai, M. K., McNaughton, S., Jesson, R., & Wilson, A. (2020). *Research-practice Partnerships for School Improvement: The Learning Schools Model*. Emerald Group Publishing.

Lai, M. K., & Schildkamp, K. (2016). In-service teacher professional learning: Use of assessment in data-based decision-making. *Handbook of human and social conditions in assessment*, 77-94.

Lasater, K., Albiladi, W. S., Davis, W. S., & Bengtson, E. (2020). The Data Culture Continuum: An Examination of School Data Cultures. *Educational Administration Quarterly*, *56*(4), 533-569. https://doi.org/10.1177/0013161X19873034

Lockton, M., Weddle, H., & Datnow, A. (2020). When data don’t drive: teacher agency in data use efforts in low-performing schools. *School Effectiveness and School Improvement*, *31*(2), 1-23. https://doi.org/10.1080/09243453.2019.1647442

Mandinach, E. B., & Gummer, E. S. (2015). The Conceptual Framework for Data Literacy for Teachers. In M. C. Linn (Ed.), *Data Literacy for Educators : Making It Count in Teacher Preparation and Practice* (pp. 38-56). Teachers College Press. http://ebookcentral.proquest.com/lib/waikato/detail.action?docID=4513520

Marsh, J. A., Bertrand, M., & Huguet, A. (2015). Using data to alter instructional practice: The mediating role of coaches and professional learning communities. *Teachers College Record*, *117*(4), 1-40.

Poortman, C. L., & Schildkamp, K. (2016). Solving student achievement problems with a data use intervention for teachers. *Teaching and Teacher Education*, *60*, 425-433. https://doi.org/10.1016/j.tate.2016.06.010

Schildkamp, K., & Ehren, M. (2013). From “Intuition”- to “Data”-based Decision Making in Dutch Secondary Schools? In (2013 ed., Vol. 17, pp. 49-67). Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-94-007-4816-3_4

Schildkamp, K., Handelzalts, A., Poortman, C. L., Leusink, H., Meerdink, M., Smit, M., Ebbeler, J., & Hubers, M. D. (2018). *The Data Team™ Procedure: A Systematic Approach to School Improvement*. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-58853-7

Schildkamp, K., & Lai, M. K. (2013). Conclusions and a Data Use Framework. In Data-based Decision Making in Education: Challenges and Opportunities (pp. 177-191). Springer Netherlands. https://doi.org/10.1007/978-94-007-4816-3_10

Schildkamp, K., & Poortman, C. L. (2015). Factors Influencing the Functioning of Data Teams. *Teachers College record (1970)*, *117*(4).

Schildkamp, K., Poortman, C. L., & Handelzalts, A. (2016). Data teams for school improvement. *School Effectiveness and School Improvement*, *27*(2), 228-254. https://doi.org/10.1080/09243453.2015.1056192

Tabachnick, B. G., & Fidell, L. S. (2014). Using multivariate statistics (Vol. 6). Pearson.

Volante, L., DeLuca, C., Adie, L., Baker, E., Harju-Luukkainen, H., Heritage, M., Schneider, C., Stobart, G., Tan, K., & Wyatt-Smith, C. (2020). Synergy and Tension between Large-Scale and Classroom Assessment: International Trends. *Educational measurement, issues and practice*.