Using NCEA data and statistics to ‘prove’ impact on achievement

Darcy Fawcett, darcy@sounddata.co.nz

Director, Sound Data.

Today’s resources:

- Preliminary survey https://bit.ly/46itL7m
- Online dashboards sounddata.co.nz/dashboard/

SOUND DATA

Removing Noise From Decision Making

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

- You must compare assessment data of those who experienced the initiative with those who didn’t
- The initiative group must achieve *significantly* higher grades, credits etc than the comparison group *and* this effect must be large enough to be worth attending to.

Experimental method

- Compare the performance 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 performance of latest cohort with historical performance
- Assumptions
  - latest cohort is typical
  - The historical record is large enough to describe the population

Summary statistics

Measures of centre (averages)

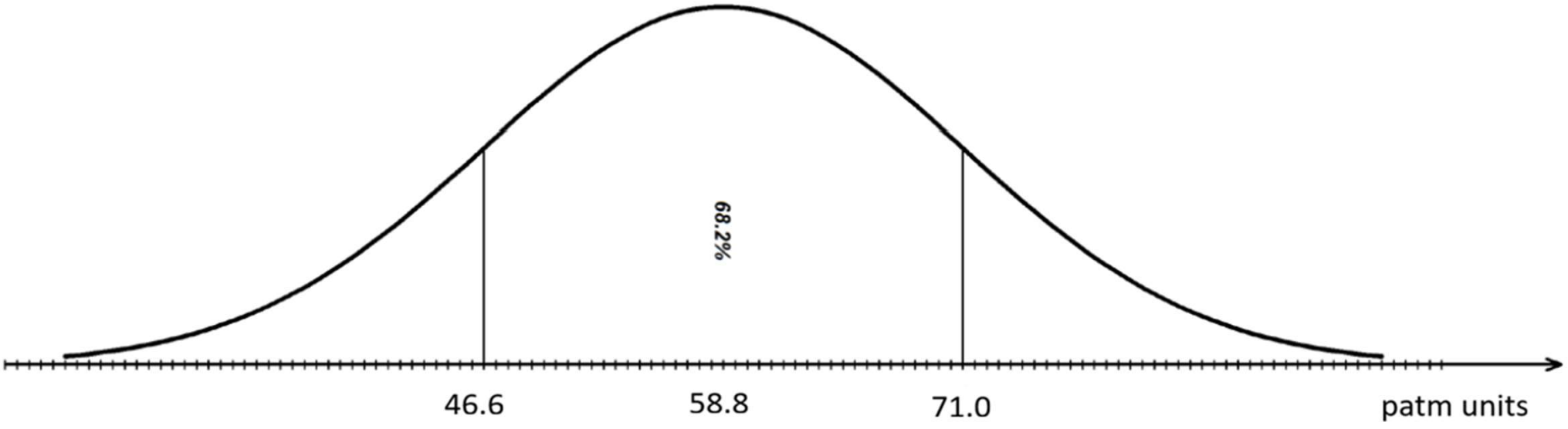
- mode
  - most frequently occurring value
- Medium
  - the middle of ranked data
- mean
  - $\mu = \bar{x} = \frac{x_1 + x_2 + x_3 + \cdots}{n}$

Measures of spread (distribution)

- range
  - $R = x_{max} - x_{min}$
- interquartile range
  - $IQR = x_{Q3} - x_{Q1}$
- standard deviation
  - $\sigma = \sqrt{\frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \cdots}{n}}$

Continuous scale data - parametric

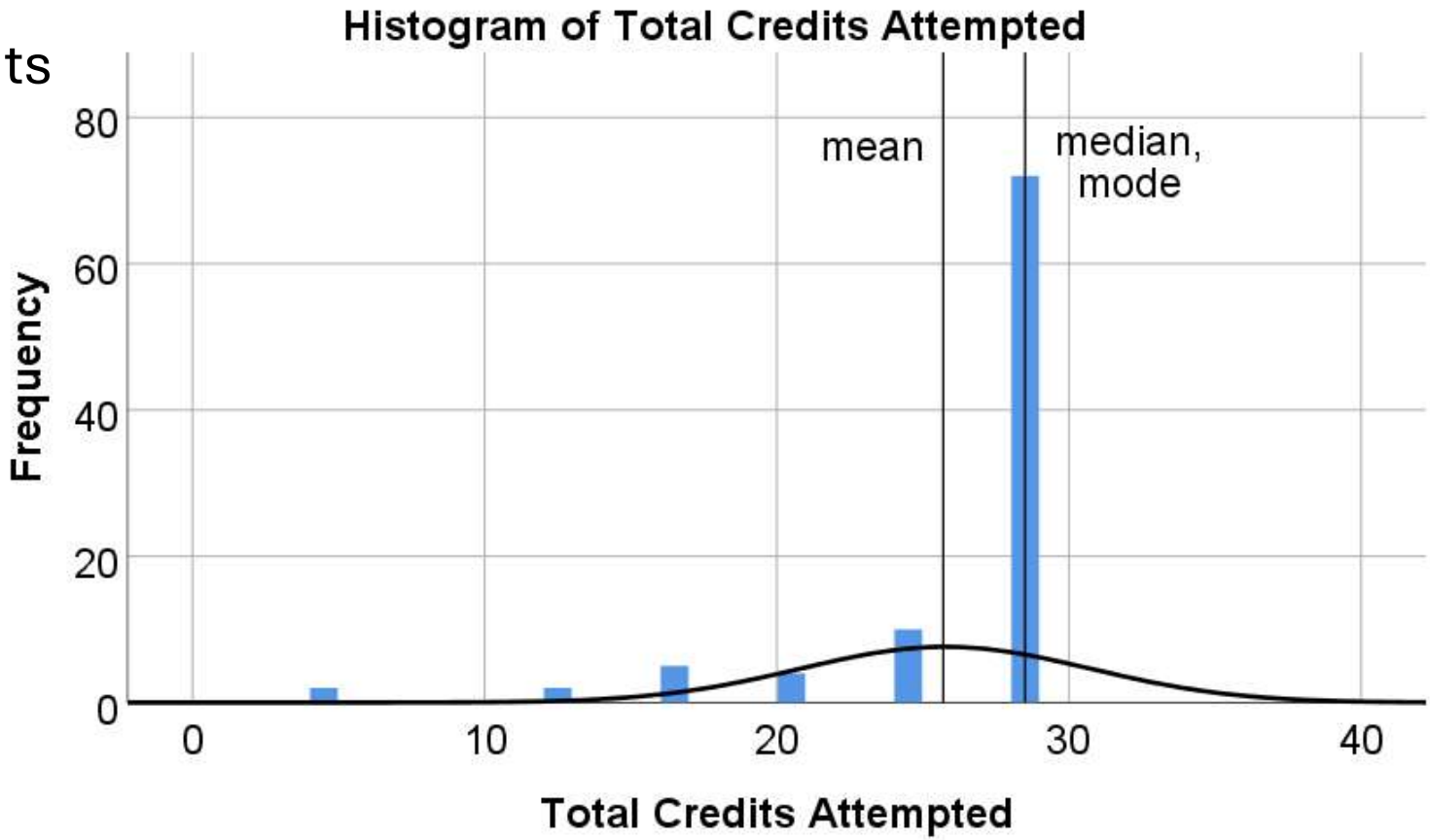
- **Continuous scale:** +1 has consistent meaning, so progress from all start points has same meaning
- **Parametric:** normally distributed (bell curve), mean = median = mode
- e.g. PAT and e-asTTle scale scores, IQ, heights



The distribution of Year 8 students' PAT: Mathematics scores (Term 4)
(Lawes, E, 2016) mean = 58.8, sd = 12.2 patm

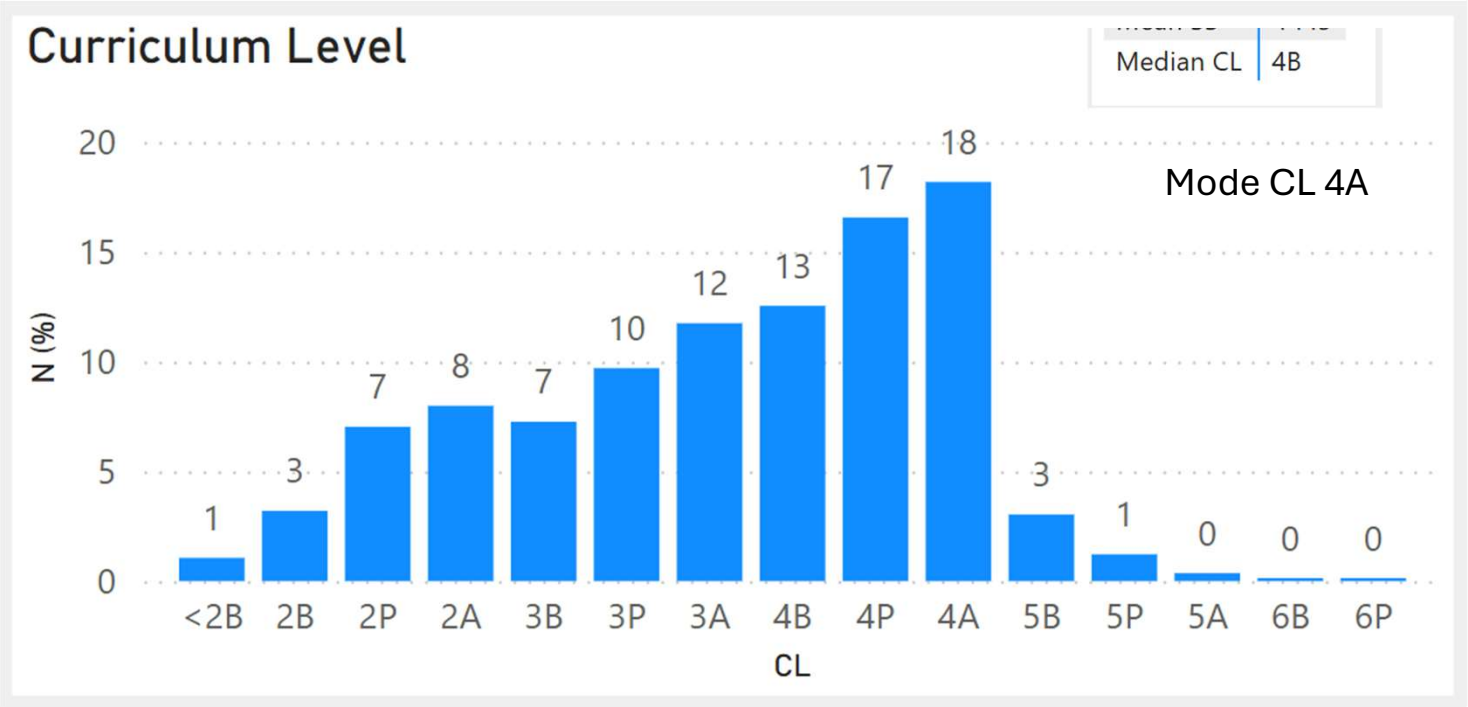
Continuous/scale data – non-parametric

- **not** normally distributed
- E.g. NCEA credits



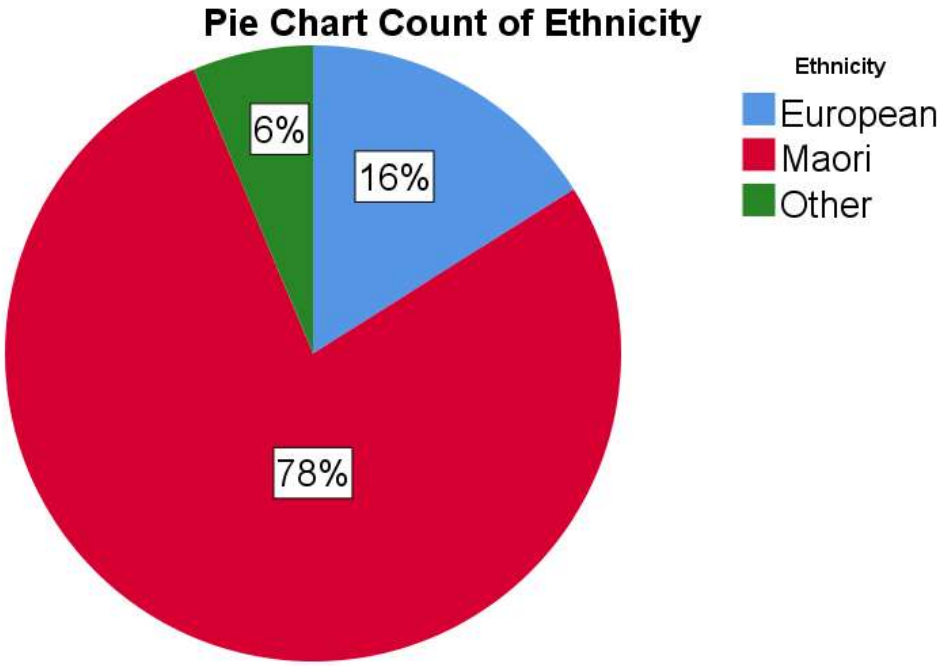
Categorical Ordinal data

- Categories that have order
- e.g. curriculum levels, achievement standards …



Categorical nominal data

- Categories that have no order
  - Count/frequency or percentage
- Summary: mode
- e.g. Gender, ethnicity, cohor



Statistical tests for significance

p: the probability that your observations are a random occurrence

Test result	probability
similar to / no significant difference	$p > 0.10$
notable difference	$0.10 \geq p > 0.05$
significant difference	$0.05 \geq p > 0.01$
highly significant difference	$0.01 \geq p$

The p-value acts as a filter for changes that are statistically similar to the normal vaiations that occur from year to year

Associated measure for strength of association or effect size

Effect size	ES
None/minimal	$0.0 \leq \lvert ES \rvert < 0.1$
Weak/small	$0.1 \leq \lvert ES \rvert < 0.3$
Moderate	$0.3 \leq \lvert ES \rvert < 0.5$
Strong/large	$0.5 \leq \lvert ES \rvert$

- Large populations can show significant relationships for very small shifts. Use related tests to show the strength of the association or effect size
  - The relative size of the skew on the graph
- Which measure you use depends on the nature of the test and variables E.g.,
  - Comparing two sample with scale (e.g. Rosenthal's R) or ordinal variables (e.g., Gamma); comparing pass rate to expected (Phi).
- These trigger points are context dependent and should be varied depending on data properties and theoretical assumptions (Cohen 1969; Kraft 2020).

Tests for significant differences and associated effect sizes

Variables		Tests for significance (and effect size)	
Independent	Dependent = 1	Independent samples	Related samples
0 variables - Theoretical	**Categorical, Ordinal** National rates, averages, distributions	**Cat2: One sample binomial test (Rosenthal's *R*)** - e.g., Year Level Achieved, School vs National % **Cat3-4: One sample chi-sq test (Phi)** - e.g., Level Qualifications, School vs National distribution **Ordinal 5+ and non-parametric scale: One sample median (Rosenthal's *R*)** - e.g., Credits gained, School vs National average ; **Parametric scale: One sample t-test (Rosenthal's *R*)** - e.g., PAT Scale Scores, School vs National average ;	
1 variable **2 levels** - Cohort - non/Māori - Gender **3+ levels** - Prioritised Ethnicity - Year - Room	**Ordinal, Ord ≤ 3** Course Endorsement, UE, Below/At+	**Chi-Square Linear-by-Linear Association test (Gamma)** e.g., UE; this year vs previous years.	**McNemar's test (Phi, AR)** - e.g., Below/At+; T1 vs T4
	Ordinal, Ord ≥ 4 **Y**ear **L**evel **Q**ual. **A**chievement **S**tds **C**urriculum **L**evels **Non-Parametric Scale** Credits, SMART	1IV2L: **Wilcoxon-Mann-Whitney U rank test (R)** - e.g., YLQ; male vs female. 1IV3+L: **Kruskal-Wallis Anova by ranks (Eta)** - e.g., AS; Prioritised Ethnicity. 1IV3+L: **Anom of transformed ranks.** - e.g., progress on CL; School vs all rooms.	1IV2L: **Wilcoxon signed ranks test (Cohen's D)** - e.g., Curriculum Levels; T1 vs T4
	Parametric scale PAT scale scores e-asTTle s. scores PaCT s. scores	1IV2L: **Independent samples t-test (D or R)** - e.g., Scale score, Māori vs non-Māori. 1IV3+L: **Analysis of Variance by means (Eta).** - e.g., progress on s.sc; School vs Room1. 1IV3+L: **Analysis of means, Anom.** - e.g., progress on s.sc; School vs all rooms.	1IV2L: **Paired samples t-test (D)** - e.g., Scale score; T1 vs T4

https://stats.oarc.ucla.edu/other/mult-pkg/whatstat/

Wilcoxon-Mann-Whitney U rank test

- 1 independent variable with 2 levels: ty vs py
- 1 non-parametric scale variable: NCEA credits

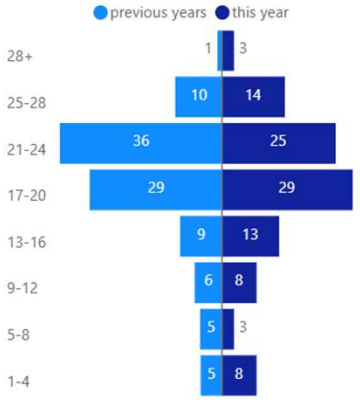
Total credits

Level 2 English

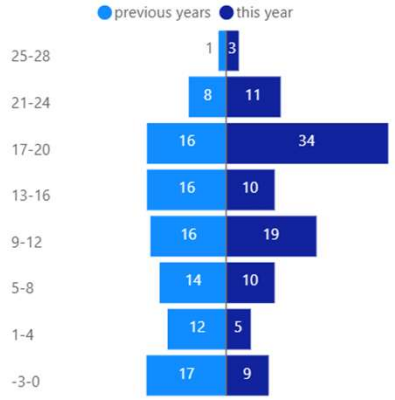
Gender: All | Ethnicity: All | Teacher: All | Year: All

Median total credits

Total credits attempted (%)



Total credits earned (%)



14+ credits (%)

Cohort	Yes
previous years	37
this year	56

Cohort	n
previous years	835
this year	80

Cohort	NSN	Student Name	TA	TE
previous years	5002	Student, 5002	25	7
previous years	5003	Student, 5003	18	6
previous years	5005	Student, 5005	21	0
previous years	5008	Student, 5008	22	10
previous years	5009	Student, 5009	22	18
previous years	5011	Student, 5011	18	4
previous years	5012	Student, 5012	28	6
previous years	5014	Student, 5014	14	10
previous years	5015	Student, 5015	22	8

Variable	Demographic	Statement	n (py)	n (ty)	med (py)	med (ty)	p	ES
TotalAttempted	All	NOT SIGNIFICANT reduction	835	80	18.0	18.0	0.420	-0.03
TotalEarned	All	small highly significant improvement	835	80	10.0	14.0	0.000	0.12
TotalEarned	Asian	invalid test, n < 20 improvement	127	8	10.0	14.5	0.425	0.07
TotalAttempted	Asian	invalid test, n < 20 reduction	127	8	18.0	18.0	0.706	-0.03
TotalAttempted	Female	NOT SIGNIFICANT reduction	441	47	18.0	18.0	0.102	-0.07
TotalEarned	Female	small highly significant improvement	441	47	10.0	17.0	0.000	0.19
TotalAttempted	Male	NOT SIGNIFICANT improvement	394	33	18.0	18.0	0.551	0.03
TotalEarned	Male	NOT SIGNIFICANT improvement	394	33	10.0	12.0	0.345	0.05
TotalAttempted	Maori	NOT SIGNIFICANT reduction	230	20	18.0	17.5	0.392	-0.05
TotalEarned	Maori	small significant improvement	230	20	7.5	10.5	0.016	0.15
TotalAttempted	MELAA	no test, missing cohort	4	0	19.5			
TotalEarned	MELAA	no test, missing cohort	4	0	13.5			
TotalAttempted	Other	no test, missing cohort	11	0	18.0			
TotalEarned	Other	no test, missing cohort	11	0	10.0			
TotalEarned	Pacifica	invalid test, n < 20 improvement	99	12	10.0	11.0	0.760	0.03
TotalAttempted	Pacifica	invalid test, n < 20 reduction	99	12	21.0	18.0	0.502	-0.06

Are students gaining enough credits in this course? Is this year similar to previous years? Do any subgroups stand out?

Year Level and demographic
(All, Y7, Y8, M, F, Māori, Pacifica, Asian, MELAA, Pakeha)

number of students
(previous years or this year)

median credits
(previous years or this year)

Variable	Demographic	Statement	n (py)	n (ty)	med (py)	med (ty)	p	ES
TotalAttempted	All	NOT SIGNIFICANT reduction	835	80	18.0	18.0	0.420	-0.03
TotalEarned	All	small highly significant improvement	835	80	10.0	14.0	0.000	0.12

Variable
(Total Credists Attempted or Earned)

Computer generated statistical conclusion
(all you need if you don't like maths!)

the p-value

The effect size

Wilcoxon-Mann-Whitney U rank test

- 1 independent variable with 2 levels: ty vs py
- 1 ordinal dependent variable (4+): NCEA grades

Total grades produced

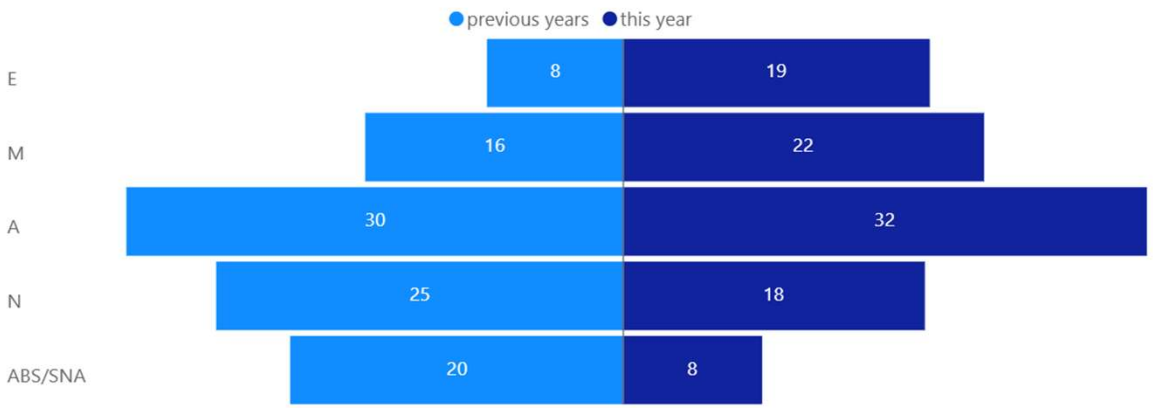
Level 2 English

Gender: All | Ethnicity: All | Type: All | Teacher: All | Year: All

Cohort	n (grades)	n (students)
previous years	3517	835
this year	330	80

Year	NSN	Student Name	n (grades)
2025	5020	Student, 5020	5
2025	5098	Student, 5098	7
2025	5101	Student, 5101	5
2025	5110	Student, 5110	1
2025	5209	Student, 5209	4
2025	5234	Student, 5234	4
2025	5270	Student, 5270	3
2025	5312	Student, 5312	6
2025	5320	Student, 5320	5
2025	5380	Student, 5380	1

Total grades (%)



Type	Demographic	Statement	n (py)	n (ty)	med (py)	med (ty)	p	ES
All	All	small highly significant improvement	3517	330	A	A	0.000	0.13
All	Asian	NOT SIGNIFICANT improvement	520	33	A	A	0.121	0.07
All	Female	moderate highly significant improvement	1896	190	A	M	0.000	0.21
All	Male	NOT SIGNIFICANT improvement	1621	140	A	A	0.448	0.02
All	Maori	small highly significant improvement	937	77	N	A	0.000	0.13
All	MELAA	no test, missing cohort	17	0	A			
All	Other	no test, missing cohort	43	0	N			
All	Pacifica	NOT SIGNIFICANT improvement	434	46	A	A	0.091	0.08
All	Pakeha	small highly significant improvement	2145	250	A	A	0.000	0.14
EX	All	small highly significant improvement	1889	170	N	A	0.000	0.14
EX	Asian	NOT SIGNIFICANT improvement	267	20	N	A	0.240	0.07
EX	Female	moderate highly significant improvement	1001	96	N	A	0.000	0.21
EX	Male	NOT SIGNIFICANT improvement	888	74	N	A	0.201	0.04
EX	Maori	small highly significant improvement	537	38	N	A	0.006	0.11
EX	MELAA	no test, missing cohort	9	0	N			
EX	Other	no test, missing cohort	24	0	N			

Have the students succeeded this year? Is this year similar to previous years? Do any sub-groups stand out?

The Mantel Haenszel linear-by-linear chi-square test

- 1 independent variable with 2 levels: ty vs py
- 1 ordinal dependent variable (3-): Course Endorsement

Course endorsements

Level 2 English

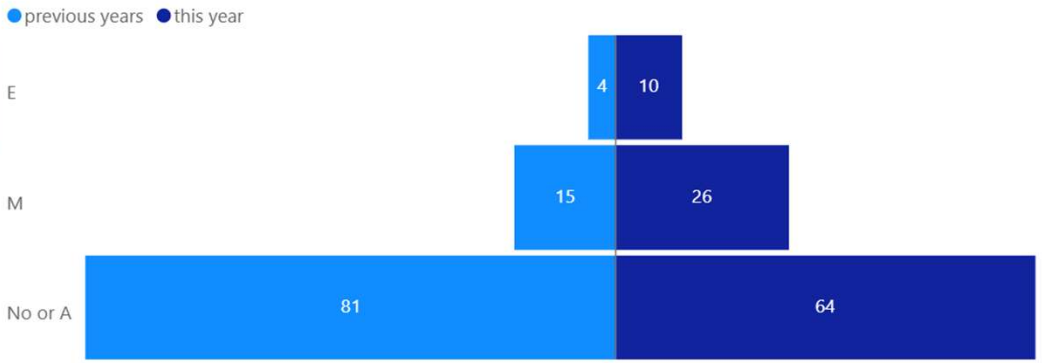
Gender: All

Ethnicity: All

Teacher: All

Year: All

Course endorsement (%)



Cohort	n
previous years	835
this year	80

Year	NSN	Student Name	Endorsement
2020	5002	Student, 5002	NO
2013	5003	Student, 5003	NO
2013	5005	Student, 5005	NO
2019	5008	Student, 5008	M2
2013	5009	Student, 5009	M2
2014	5011	Student, 5011	NO
2021	5012	Student, 5012	NO
2017	5014	Student, 5014	NO
2015	5015	Student, 5015	NO

Demographic	Statement	n (py)	n (ty)	med (py)	med (ty)	p	ES
All	large highly significant improvement	835	80	No or A	No or A	0.001	0.39
Asian	invalid test, n < 20 reduction	127	8	No or A	No or A	0.540	-0.44
Female	large highly significant improvement	441	47	No or A	No or A	0.009	0.37
Male	large significant improvement	394	33	No or A	No or A	0.028	0.41
Maori	very large significant improvement	230	20	No or A	No or A	0.022	0.53
MELAA	no test, missing cohort	4	0	No or A			
Other	no test, missing cohort	11	0	No or A			
Pacifica	invalid test, n < 20 improvement	99	12	No or A	No or A	0.153	0.50
Pakeha	large highly significant improvement	504	58	No or A	No or A	0.000	0.42

Have the students succeeded this year? Is this year similar to previous years? Do any sub-groups stand out?

University Entrance Longitudinal

Year Level

13

Gender: Female

Ethnicity: All

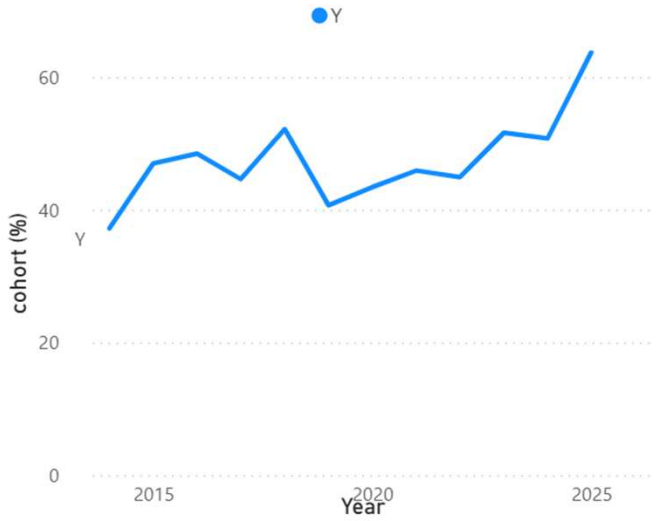
Year: All

Cohort	N
previous years	706
this year	69

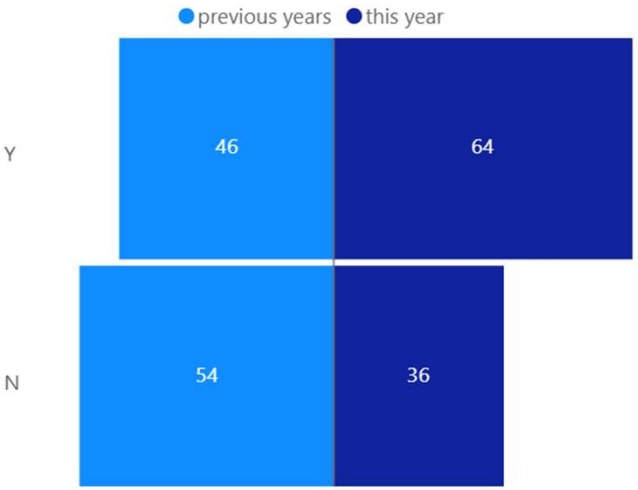
Search

Year	NSN	Student Name	UE
2025	10196	Student, R10196	Y
2025	10263	Student, R10263	Y
2025	10271	Student, R10271	Y
2025	10312	Student, R10312	N
2025	10325	Student, R10325	Y
2025	10438	Student, R10438	Y
2025	10507	Student, R10507	N
2025	10944	Student, R10944	Y
2025	11054	Student, R11054	Y
2025	11089	Student, R11089	N

Level Qualification by Year



Level Qualifications (%)



Demographic	Statement	n (py)	n (ty)	med (py)	med (ty)	p	ES
All	NOT SIGNIFICANT, improvement	1963	198	Y	Y	0.371	0.070
Asian	NOT SIGNIFICANT, improvement	566	68	N	Y	0.072	0.239
Female	large highly significant, improvement	706	69	N	Y	0.005	0.352
Male	NOT SIGNIFICANT, reduction	1257	129	Y	N	0.356	-0.088
Maori	NOT SIGNIFICANT, improvement	590	64	Y	Y	0.361	0.122
MELAA	invalid test, n < 20, improvement	21	5	Y	Y	1.000	0.154
Other	invalid test, n < 20, reduction	26	3	N	N	1.000	-0.029
Pacific	NOT SIGNIFICANT, improvement	633	58	N	Y	0.493	0.108
Pakeha	NOT SIGNIFICANT, improvement	1196	118	N	Y	0.289	0.110

Have the students succeeded this year? Is this year similar to previous years? Do any sub-groups stand out?

Gender

Level 2 English

Total grades

cohort	Statement	n (F)	n (M)	med (F)	med (M)	p	ES
previous years	minimal highly significant difference, Male > Female	1896	1621	A	A	0.002	-0.05
this year	large highly significant difference, Female > Male	190	140	M	A	0.000	0.31

Endorsement

Cohort	Statement	n (F)	n (M)	med (F)	med (M)	p	ES
previous years	moderate highly significant difference Female > Male	441	394	No or A	No or A	0.002	0.25
this year	NOT SIGNIFICANT Female > Male	47	33	No or A	No or A	0.315	0.23

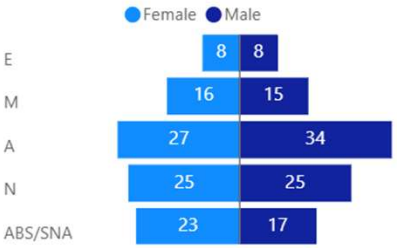
Total credits

Cohort	Variable	Statement	n (F)	n (M)	med (F)	med (M)	p	ES
previous years	TotalAttempted	NO SIGNIFICANT DIFFERENCE, Female > Male	441	394	18.0	18.0	0.410	0.03
this year	TotalAttempted	NO SIGNIFICANT DIFFERENCE, Male > Female	47	33	18.0	18.0	0.305	-0.11
this year	TotalEarned	NO SIGNIFICANT DIFFERENCE, Female > Male	47	33	17.0	12.0	0.164	0.16
previous years	TotalEarned	NO SIGNIFICANT DIFFERENCE, Male > Female	441	394	10.0	10.0	0.099	-0.06

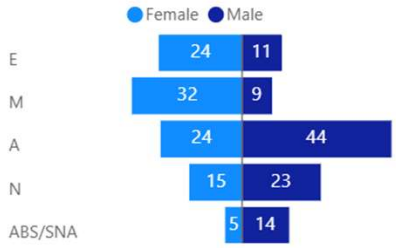
How does the achievement of females compare to males? Has this changed?

Teacher: All

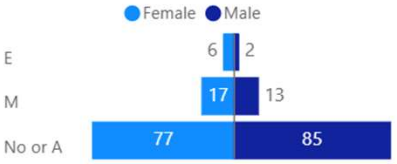
Total grades (%) previous years



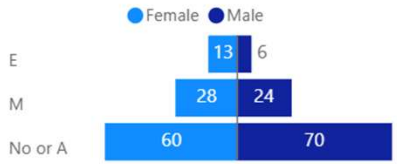
Total grades (%) this year



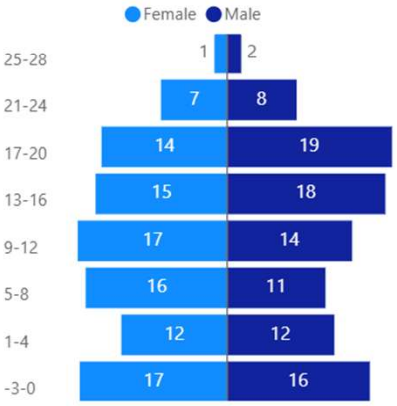
Endorsement (%) previous years



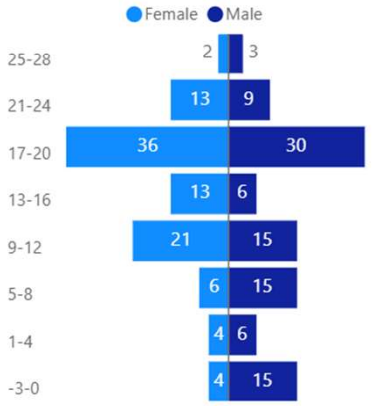
Endorsement (%) this year



Total earned (%) previous years



Total earned (%) this year



One sample binomial test

- 0 independent variable: theoretical comparison
- 1 ordinal binary variable: Qualification Pass rates or UE

Qualification Pass Rate National comparison

Year Level

13

Cohort: All

Dem: All

EQI: All

Cohort	Dem	EQI	Statement	N	School	National	p	ES
Previous Years	All	Equity Band: Above Average	NO SIGNIFICANT DIFFERENCE, Nat > Sch	1963	65.7%	66.5%	0.430	0.00
Previous Years	All	Equity Group: Moderate	small highly significant difference, Nat > Sch	1963	65.7%	71.3%	0.000	-0.12
Previous Years	All	National: All Schools	small highly significant difference, Nat > Sch	1963	65.7%	70.9%	0.000	-0.11
Previous Years	Asian	National: All Schools	small highly significant difference, Nat > Sch	566	67.1%	73.9%	0.000	-0.14
Previous Years	Female	National: All Schools	moderate highly significant difference, Nat > Sch	706	63.0%	74.9%	0.000	-0.26
Previous Years	Male	National: All Schools	NO SIGNIFICANT DIFFERENCE, Sch > Nat	1257	67.1%	66.6%	0.676	0.01
Previous Years	Maori	National: All Schools	small highly significant difference, Sch > Nat	590	65.9%	58.6%	0.000	0.14
Previous Years	MELAA	National: All Schools	NO SIGNIFICANT DIFFERENCE, Sch > Nat	21	66.7%	66.3%	1.000	0.00
Previous Years	Other	National: All Schools	NO SIGNIFICANT DIFFERENCE, Nat > Sch	26	53.8%	67.5%	0.146	-0.21
Previous Years	Pacific	National: All Schools	NO SIGNIFICANT DIFFERENCE, Sch > Nat	633	63.3%	63.0%	0.869	0.04
Previous Years	Pakeha	National: All Schools	moderate highly significant difference, Nat > Sch	1196	64.8%	76.3%	0.000	-0.26
This Year	All	Equity Band: Above Average	NO SIGNIFICANT DIFFERENCE, Nat > Sch	198	68.2%	72.5%	0.176	-0.07
This Year	All	Equity Group: Moderate	NO SIGNIFICANT DIFFERENCE, Nat > Sch	198	68.2%	74.2%	0.061	-0.11
This Year	All	National: All Schools	small significant difference, Nat > Sch	198	68.2%	74.6%	0.041	-0.12
This Year	Asian	National: All Schools	NO SIGNIFICANT DIFFERENCE, Nat > Sch	68	69.1%	75.1%	0.263	-0.08
This Year	Female	National: All Schools	NO SIGNIFICANT DIFFERENCE, Sch > Nat	69	79.7%	77.7%	0.774	0.09
This Year	Male	National: All Schools	small significant difference, Nat > Sch	129	62.0%	71.4%	0.025	-0.17
This Year	Maori	National: All Schools	NO SIGNIFICANT DIFFERENCE, Sch > Nat	64	68.8%	66.2%	0.792	0.10
This Year	MELAA	National: All Schools	invalid test, n < 20, Sch > Nat	5	100.0%	67.5%	0.181	0.41
This Year	Other	National: All Schools	invalid test, n < 20, Nat > Sch	3	33.3%	75.6%	0.150	-0.60
This Year	Pacific	National: All Schools	NO SIGNIFICANT DIFFERENCE, Nat > Sch	58	60.3%	69.7%	0.152	-0.14
This Year	Pakeha	National: All Schools	moderate highly significant difference, Nat > Sch	118	69.5%	79.6%	0.008	-0.22

One sample chi-squared test

- 0 independent variable: theoretical comparison with national rates
- 1 **categorical** dependent variable: NAME grade distribution

Qualifications Distribution National Comparison

Year Level

13

Cohort: All

Dem: All

EQI: All

Cohort	Dem	EQI	Statement	N	N (Sch)	A (Sch)	M (Sch)	E (Sch)	N (Nat)	A (Nat)	M (Nat)	E (Nat)	N (rd)	A (rd)	M (rd)	E (rd)	p	ES
Previous Years	All	National: All Schools	small highly significant difference	1963	34.3%	37.9%	18.5%	9.3%	29.1%	41.4%	18.8%	10.6%	0.18	-0.09	-0.02	-0.12	0.000	0.12
This Year	All	National: All Schools	NO SIGNIFICANT DIFFERENCE	198	31.8%	35.9%	19.2%	13.1%	25.4%	43.5%	20.1%	11.0%	0.25	-0.18	-0.05	0.19	0.080	0.18
Previous Years	All	Equity Group: Moderate	small highly significant difference	1963	34.3%	37.9%	18.5%	9.3%	28.7%	46.0%	16.3%	9.0%	0.20	-0.18	0.13	0.04	0.000	0.17
This Year	All	Equity Group: Moderate	moderate highly significant difference	198	31.8%	35.9%	19.2%	13.1%	25.8%	47.6%	17.6%	9.0%	0.24	-0.25	0.09	0.45	0.006	0.25
Previous Years	All	Equity Band: Above Average	moderate highly significant difference	1963	34.3%	37.9%	18.5%	9.3%	33.5%	47.1%	12.7%	6.7%	0.03	-0.20	0.46	0.39	0.000	0.24
This Year	All	Equity Band: Above Average	large highly significant difference	198	31.8%	35.9%	19.2%	13.1%	27.5%	51.1%	14.1%	7.3%	0.16	-0.30	0.36	0.80	0.000	0.34
Previous Years	Asian	National: All Schools	moderate highly significant difference	566	32.9%	37.3%	20.7%	9.2%	26.1%	35.5%	23.0%	15.4%	0.26	0.05	-0.10	-0.40	0.000	0.21
This Year	Asian	National: All Schools	NO SIGNIFICANT DIFFERENCE	68	30.9%	33.8%	25.0%	10.3%	24.9%	36.5%	23.0%	15.5%	0.24	-0.07	0.09	-0.34	0.491	0.19
Previous Years	Female	National: All Schools	moderate highly significant difference	706	37.0%	34.8%	19.1%	9.1%	25.1%	39.2%	22.4%	13.3%	0.47	-0.11	-0.14	-0.32	0.000	0.28
This Year	Female	National: All Schools	NO SIGNIFICANT DIFFERENCE	69	20.3%	37.7%	23.2%	18.8%	22.3%	40.1%	24.1%	13.6%	-0.09	-0.06	-0.04	0.39	0.649	0.15
Previous Years	Male	National: All Schools	small highly significant difference	1257	32.9%	39.5%	18.1%	9.5%	33.4%	44.0%	15.0%	7.6%	-0.02	-0.10	0.21	0.24	0.000	0.13
This Year	Male	National: All Schools	moderate significant difference	129	38.0%	34.9%	17.1%	10.1%	28.6%	47.3%	15.9%	8.2%	0.33	-0.26	0.07	0.23	0.032	0.26
Previous Years	Maori	National: All Schools	large highly significant difference	590	34.1%	38.0%	18.3%	9.7%	41.4%	43.0%	10.8%	4.8%	-0.18	-0.12	0.69	1.02	0.000	0.35
This Year	Maori	National: All Schools	NO SIGNIFICANT DIFFERENCE	64	31.3%	40.6%	17.2%	10.9%	33.8%	46.5%	12.8%	6.9%	-0.08	-0.13	0.34	0.59	0.381	0.22
Previous Years	MELAA	National: All Schools	NO SIGNIFICANT DIFFERENCE	21	33.3%	42.9%	19.0%	4.8%	33.7%	40.1%	17.9%	8.3%	-0.01	0.07	0.06	-0.42	0.947	0.13
This Year	MELAA	National: All Schools	invalid test, n < 20	5	0.0%	60.0%	40.0%	0.0%	32.5%	38.9%	18.6%	10.0%	-1.00	0.54	1.15	-1.00	0.269	0.89
Previous Years	Other	National: All Schools	NO SIGNIFICANT DIFFERENCE	26	46.2%	30.8%	11.5%	11.5%	32.5%	38.2%	18.9%	10.3%	0.42	-0.19	-0.39	0.12	0.449	0.32
This Year	Other	National: All Schools	invalid test, n < 20	3	66.7%	33.3%	0.0%	0.0%	24.4%	42.3%	21.2%	12.1%	1.73	-0.21	-1.00	-1.00	0.354	1.04
Previous Years	Pacific	National: All Schools	large highly significant difference	633	36.7%	36.0%	19.4%	7.9%	37.0%	50.6%	9.4%	3.0%	-0.01	-0.29	1.07	1.60	0.000	0.48
This Year	Pacific	National: All Schools	very large highly significant difference	58	39.7%	27.6%	19.0%	13.8%	30.3%	53.9%	12.4%	3.4%	0.31	-0.49	0.53	3.06	0.000	0.71
Previous Years	Pakeha	National: All Schools	moderate highly significant difference	1196	35.2%	36.5%	18.3%	10.0%	23.7%	41.1%	22.3%	12.8%	0.48	-0.11	-0.18	-0.22	0.000	0.27
This Year	Pakeha	National: All Schools	moderate significant difference	118	30.5%	37.3%	18.6%	13.6%	20.4%	43.4%	23.8%	12.4%	0.50	-0.14	-0.22	0.09	0.039	0.27

How does the Qualifications Distribution of the school compare to national statistics? Has this changed? Do any groups stand out?

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.

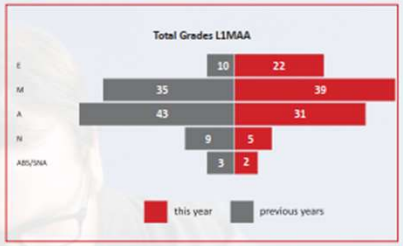
Teachers and NCEA students outraged over difficult Level 1 MCAT algebra exam – ***Stuff, 2016***
7 hacks for surviving NZ's toughest NCEA exam – ***Scoop, 2017***

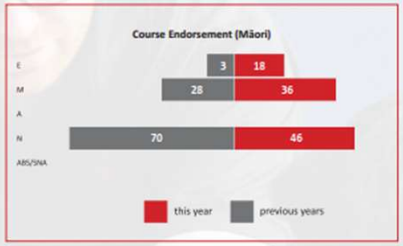
Methodology
After hearing positive results from the Art and Māori departments in instituting Wānanga days for 2019 the Mathematics department decided for all students sitting the external MCAT to have a day of organised preparation.

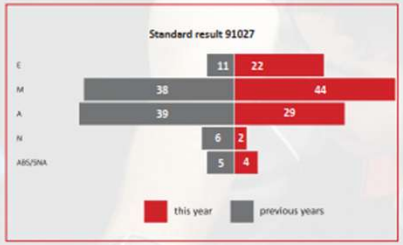
- [redacted]used the library and ran various presentations and activities for students the day prior to the external.

- 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.

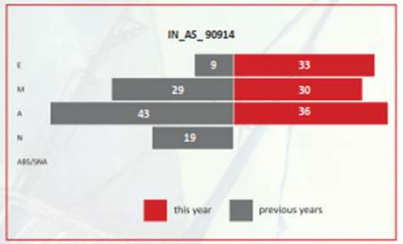
What did we do?
- The Wānanga day was a huge success. 100% of the students achieved their internal allowing us to move as a majority onto the external submissions (closure/peace of mind/admin done).
- Proposition had a clear direction. ***Unpack Arrival***
- Collaboration - Two classes - SM/WO

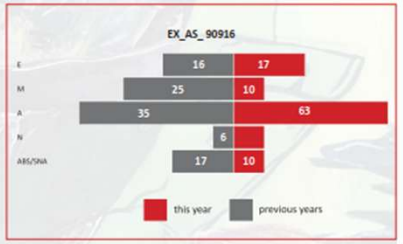
Work on ...
- We achieved our goal from 2019 for the internals by significantly improving both earned and quality of grade. However there was a shift in the externals both up and down.

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.





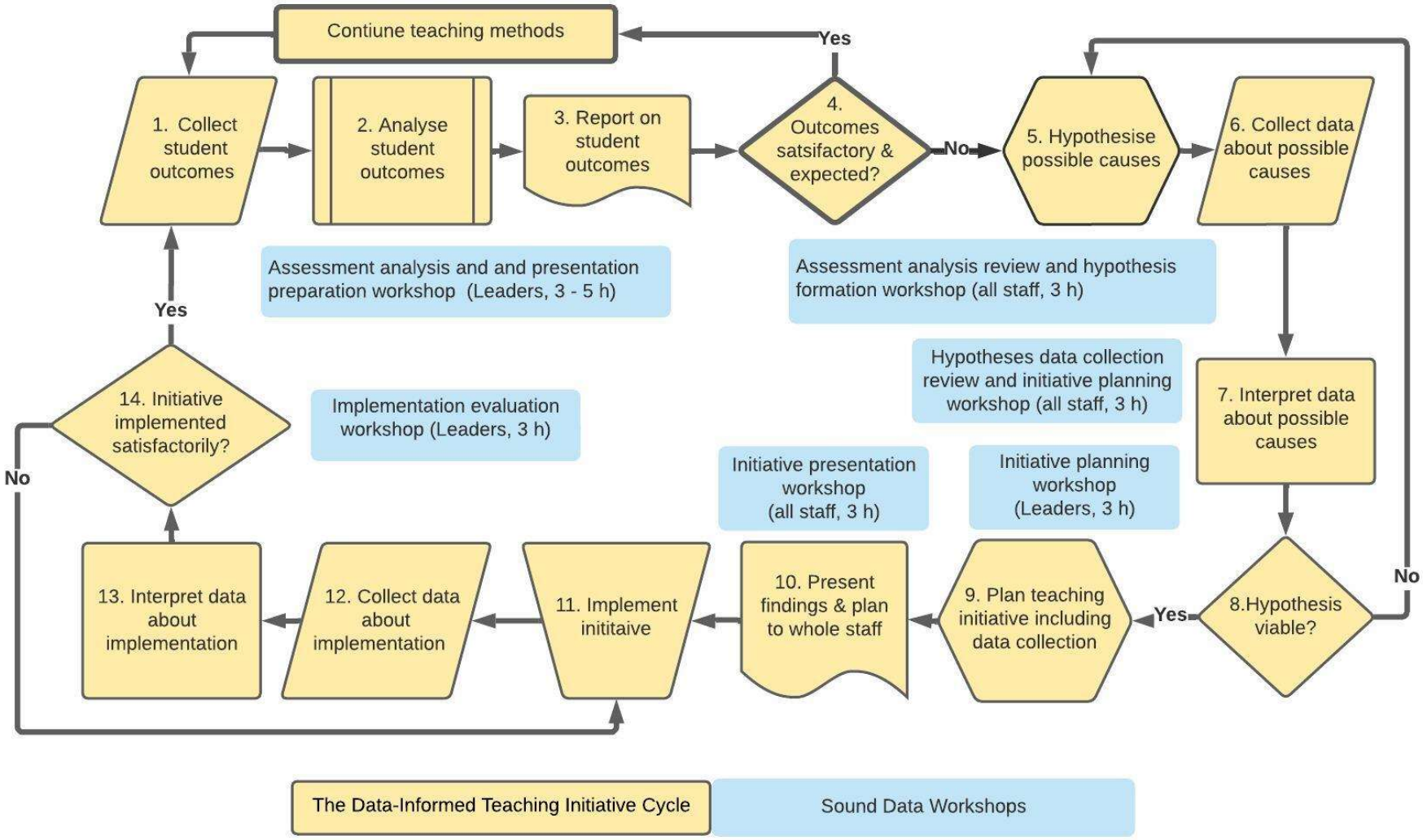
Data presentations are useful and have an impact!

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

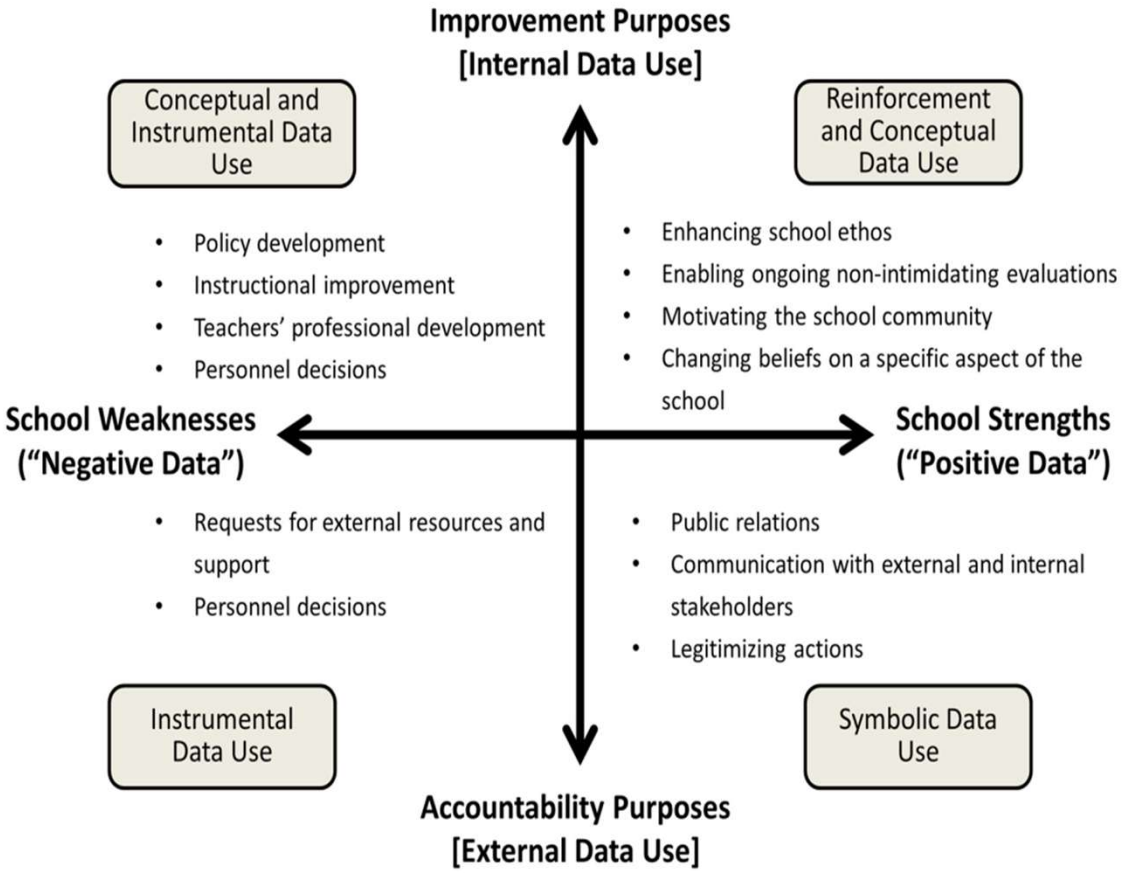
Data-informed teaching initiatives can improve teaching, learning *and* accountability when:

- 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



(Aderet-German & Ben-Peretz, 2020)

- 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.

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*.