1. Changes in Pre- and Post-Training Self Assessment Surveys

Most survey instruments yield quantitative data from closed questions or rating/agreement scales. These data are generally analysed to compare pre- and post-tests for frequencies, such as percentages and averages.[1] A comparison of how students score against survey questions and tests at the end compared to before an activity gives a sense of 'distance travelled' although it doesn't prove causality.

Worked example

It is possible to calculate change across the group in a range of different ways, as illustrated in Table 1 below):

1. The share who gave a positive response in the post test minus the share who said this in the pre-test (i.e. the percentage point change) (column D)
2. The share of the participants whose individual response showed a positive change in the post test compared to the pre-test (i.e. the % that changed) (column E). To calculate this metric you need to be able to match the pre and post test scores to particular individuals.

The second is probably the most useful because if you only record changes across a whole group and are not able to match any pre/post responses you will not be able to see the extent to which individuals benefited. Individual-level data allows for more robust analysis, for example to show if there were any demographic sub-groups for whom the intervention was particularly beneficial. If you can't match at individual level you at least need to make sure the results were not skewed by any differences in the individuals included in the two groups being compared.

Table 1: Agreement with statements

A	B	C	D	E
Statement	% Pre* (N=103)	% Post* (N=95)	Percentage point Change	% of respondents with a positive change
People like me don't go to university (reverse scored)	30.1%	22.1%	8	27%
I feel well-prepared to make decisions about my next steps in education	59.6%	73.7%	14.1	24%
I have a good idea of what it's like to be a university student	92.2%	92.6%	0.4	0%
I understand how I can use what I'm learning in the future	76.7%	78.9%	2.1	3%
I think I have the skills I will need to be successful in my future studies	15.4%	20.0%	4.6	30%
Deciding what subject options are right for me feels overwhelming (reverse scored)	26.9%	9.5%	17.4	65%

* Note in a final report results in numbers and percentages in a table (never report only percents, without a reference to the number of cases).

2. T-Test of significance

Statistical analyses are needed to interpret the significance of differences between pre- and post-tests. People often tend to score things differently at different times, and significance tests can help to identify which changes from baseline to end are not just down to chance. Most significance tests aim for the probability of results due to chance being 5% or less. Its helpful to have access to statistical expertise to run tests, especially when the numbers of participants is greater than 50. A 'T-Test' will find out if there is a significant difference between the pre and post means, which helps to assess the survey.

Worked example

In the above case study, T-Tests were performed (with significance level $P<0.05$). Table 2 shows the means for a selection of the pre- and post-course survey statements, and the T-Test for Equity of Means results.

[1] It is contested whether responses to agreement scales should be averaged as this assumes the agreement statements are equally spaced, however it is fairly common practice to do so (to give a sense of the mean score).

Table 2: T-Test Group Statistics

					T-Test Results for Equity of Means (Equal variances assumed)		
Statement		Mean	Std. Deviation		t	df	Sig. (2-tailed)
People like me don't go to university (reverse scored)	Pre Post	2.75 2.65	1.03 1.16		0.61	196	0.543
I feel well-prepared to make decisions about my next steps in education	Pre Post	3.19 2.73	1.15 1.23		2.759	197	0.006
I have a good idea of what it's like to be a university student	Pre Post	1.88 1.99	0.97 1.13		-0.703	197	0.483
I understand how I can use what I'm learning in the future	Pre Post	2.52 2.41	1.24 1.19		0.659	196	0.511
I think I have the skills I will need to be successful in my future studies	Pre Post	4.51 4.26	1.15 1.23		1.461	197	0.146
Deciding what subject options are right for me feels overwhelming (reverse scored)	Pre Post	2.75 2.06	1.29 1.02		4.141	197	0.000

The results showed that Statement 6: Deciding what subject options are right for me feels overwhelming (reverse scored) T-Test results show a significance level of <0.000. This indicates that this result is robust because nearly all the time in any population this difference would occur following the same intervention. Results also showed that Statement 2: I feel well-prepared to make decisions about my next steps in education, had a significance level of <0.006, which means the same is true of this statement as well. The results show no significant difference in the rest of the questions, which indicates that taking part in this programme probably did not have any effect on the participant's initial observations of higher education.

3. Chi-Square Tests

Statistical tests are also potentially useful when you want to explore differences between the results of categorical data for sub-groups within your sample – to test a hypothesis for example. However, you'll need a sufficiently large sample size to do this.

The Pearson Chi-Square test (χ^2) is used to test 'goodness-of-fit' – i.e. whether the actual data differ significantly from what you might expect given the results predicted in a model. It evaluates how likely it is that any observed difference between two sets of data arose by chance (by testing the null hypothesis that the variables are independent). Basically, the test compares the data to a model that distributes the data according as if the variables are independent, and if the observed data differs then it is more likely to a dependent variable.

Worked example

Based on the above example, Table 3 shows the responses to the statement: "People like me don't go to university" (reverse scored) cross-tabulated with an indicator or whether respondents have a history of HE in their family.

Table 3: "People like me don't go to university" cross-tabulated with HE background

	Disagree	Agree	
No HE Background in Family			
Count	17	14	31
% within HE Background	54.8%	45.2%	100.0%
% within S1	25.0%	46.7%	31.6%
% of Total	17.3%	14.3%	31.6%
He Background in Family			
Count	51	16	67
% within HE Background	76.1%	23.9%	100.0%
% within S1	75.0%	53.3%	68.4%
% of Total	52.0%	16.3%	68.4%
Total			
Count	68	30	98
% within HE Background	69.4%	30.6%	100.0%
% within S1	100.0%	100.0%	100.0%
% of Total	69.4%	30.6%	100.0%

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	4.509	1	.034

Results showed that participants who had a history of HE in their family background were more likely to disagree with the statement "People like me don't go to university", than those with no HE background. These results suggest that with HE experience in the family may also come more knowledge or understanding of higher education.

4. Group-average normalized gain

Another way of approaching analysis of change data which is quite helpful when you want to benchmark changes in different groups or activities is group-average normalised gain.[2] Group-average normalized measures the ratio of a whole group's performance to the maximum achievable improvement. It is expressed mathematically as a fraction of the maximum achievable pre-test/post-test gain. By taking account of the starting point, the formula diminishes the confounding effects of baseline characteristics including pre-course knowledge (which can be particularly important in studies which do not include a control group).

Worked example

This example is for a one-day Maths workshop, which was designed to increase HE knowledge and Maths skills. The workshop comprised educational content, learner assessments, and teaching guides and incorporated multiple modules throughout the day using different educational media and techniques (classroom-based didactic lectures, interactive audience-participation question-and-answer sessions, hands-on technical skill workshops and a problem-solving case scenario). Year 13 students participated in the workshop (randomly divided into five groups for the small group sessions that rotated through four technical modules).

A multiple-choice question test of Maths knowledge was administered comprising 20 items as a pre-test and post-test (the maximum score for each test was 20). The delivery partners hypothesised that participation would result in improved group and individual student learning gains in Maths skills. A target group average normalized gain of 30% was agreed as defining the minimum value at which the educational intervention could be regarded as effective.

Individual single-student normalized gain (gi) is the actual gain divided by the maximum gain achievable by each student, and is calculated as follows:

$$gi = [\%post\text{-}test - \%pre\text{-}test] / 100\% - \%pre\text{-}test]$$

[2] The formula was introduced by Hake 1998 and is commonly described as the amount students learned divided by the amount they could have learned.

As a measure of course effectiveness, the class average normalized gain ⟨g⟩ was calculated.

g = (group average post test – group average pre test) / 100 – (group average pre test)

The class average normalized gain does not account for where participants score less post test than pre test (which could therefore inflate the perception of effectiveness). Therefore having first calculated the normalized gain for each student, an average of gains was also calculated g(ave).

g(ave) = <(Post - Pre)/(100 - Pre)>

Results: Pre- and post-test scores and learning gain (N = 24)

Pre-test scores	48% (9.6/20 ± 2.58)
Post-test scores	66% (13.2/20 ± 2.53)
p value	0.043
Absolute gain	18%
Relative gain	37%
⟨g⟩	34%
g(ave)	29% ± 33

5. Measuring longer term effects of the knowledge gained

Ideally, the same students will be picked up in follow-up surveys or focus groups at a later date to explore medium-term outcomes (rather than just immediate-term outcomes collected in the post-programme survey).

Worked example

In this example three months after an outreach intervention, a questionnaire was emailed to each student. The follow-up survey sought to establish: whether a transfer of the learning from the workshop to the college situation had occurred; whether the workshop prompted the participants to continue their individual Maths learning; and whether the participants were encouraged to become an advocate for Maths subjects. Table 4 shows the percent of agreement/disagreement on statements about the long-term effect of the workshop on respondents knowledge, attitudes or behavior.

Table 4: Three Month Post-Training Assessment Survey

Statement:		
My attendance at the Maths workshop added to my knowledge about Mathematical techniques.	% Disagree % Agree	0.0 100.0
My attendance at the Maths workshop changed my assumptions about studying Maths.	% Disagree % Agree	20.5 79.5
My attendance at the Maths workshop affected my feelings in a positive manner about studying Maths.	% Disagree % Agree	9.9 90.1
My attendance at the Maths workshop motivated me to gather more information about options for studying Maths.	% Disagree % Agree	34.3 65.7
My attendance at the Maths workshop encouraged me to consider studying Maths in future at University.	Yes No	47.1 51.4
I have used what I learnt when I attended the Maths Workshop in my college work.	Yes No	74.6 25.4

Cross-tabulations were made of the results against demographic categories and HE background factors. Only one category demonstrated significant data (Gender). Statement 5: "My attendance at the workshop affected my feelings in a positive manner about studying Maths" had a chi-square value of $P<0.054$. This shows that females were much more likely than males to agree with this statement. The results of the cross-tabulation and chi-square data are given in Table 5.

Table 5: Cross-tabulation of Gender with Statement "My attendance at the Maths workshop affected my feelings in a positive manner about studying Maths"

		disagree	agree	
Male	Count	3	9	12
	% within Gender	25.0%	75.0%	100.0%
	% within S5	42.9%	14.1%	16.9%
	% of Total	4.2%	12.7%	16.9%
Female	Count	4	55	59
	% within Gender	6.8%	93.2%	100.0%
	% within S5	57.1%	85.9%	83.1%
	% of Total	5.6%	77.5%	83.1%
Total	Count	7	65	71
	% within Gender	9.9%	90.1%	100.0%
	% within S5	100.0%	100.0%	100.0%
	% of Total	9.9%	91.1%	100.0%

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	3.725	1	.054