



Reporting results

In this section of the *Inquiry Brief*, the program faculty reports the **results** of its investigation.

Program faculty members may present the results regarding their claims about their graduates' accomplishments in either a quantitative or qualitative format, but they must meet the reporting standards commonly required in quantitative or qualitative research paradigms. The results must also address each component of TEAC's *Quality Principle 1*. (Note that for security reasons, data files should never link student names with social security numbers or other personal information.)

One straightforward way to present the reliability and validity of the data is through an *inter-correlation matrix* of the measures used for each component of *Quality Principle 1* and/or claim. The faculty might also use the stability of mean grades, or other measures, over time to depict the reliability of the program's assessment system.

Table 3 below offers an example of how a program faculty might organize its quantitative results for the components of *Quality Principle 1*.

Note that although means and standard deviations are the likely entries in each cell of Table 3, frequency counts, ranks, percentages, percentiles, or whatever quantitative metric the faculty relies on could be also entered. These cell entries could also vary with regard to the power of their numerical properties entered (nominal, ordinal, interval, or ratio numbers).

Table 3: Hypothetical means (standard deviations) of a sample of program students (N=) in six categories of assessments in support of claims for *Quality Principle 1*

Outcome claims*	Categories of evidence and range of scores								
	Grade point index	Standardized tests		Faculty & Cooperating Teacher evaluations		Student self-reports	Survey of graduates and employers		Gains in pupil scores on work samples
		Praxis	NES	Fac	CT		Grad	Emp	
<i>The program's graduates have acquired...</i>	(Score range, e.g 0-4 & cut score)	(Score range & cut score)	(Score range & cut score)	(Score range & cut score)	(Score range & cut score)	(Score range & cut score)	(Score range & cut score)	(Score range & cut score)	(Score range & cut score)
Subject matter <i>(Professional knowledge)</i>									
Pedagogy <i>(Strategic decision-making)</i>									
Teaching skill <i>(Leadership skill)</i>									

*Includes measures of learning how to learn, multicultural perspectives and technology

Table 4 below gives an example of evidence used to support subject matter knowledge in mathematics with data from a sample of program students.

Table 4: Mean (and SD) GPA and Praxis II scores and pass rates in secondary mathematics for years 2005-2010 at Exemplar University

		GPA in upper level mathematics courses: math department mean: 3.1; math department standard: 2.75		Scores on Praxis II: mathematics national mean: 540 State cut score: 520			Correlation: GPA-Praxis II
Year of graduation	Number graduates/ number admitted	Mean (0-4)	SD	Graduates N	Mean and (pass rates) 400-800	SD	Pearson R
05-06	15/20	3.5	.5	15	610 (84%)	90	.89
06-07	12/21	3.4	.4	12	590 (77%)	95	.70
07-08	14/22	3.6	.5	14	615 (85%)	92	.69
08-09	20/22	3.0	1.1	20	510 (47%)	130	.71
09-10	15/20	3.5	.5	15	610 (84%)	95	.80
Total	76/105	3.4	.6	75	587 (75%)	100	.75

Note: The faculty assessed the reliability of the GPA by drawing a sample of 30 students from the five-year period and calculating the mean GPA each year. The means were within .05 of each other. ETS reports that the Praxis II examination in mathematics has a reliability of .83 for its norming group. The faculty reported that the graduates of its program match well the scores and demographics of the ETS norm group.

The results, whether quantitative or qualitative, should be truly representative of the program under review and not be idiosyncratic to a particular time period or circumstance.

The results must also be disaggregated by subcategory when an aggregated presentation would mask important differences within the groups and categories being reported. This means, of course, that the faculty need to inspect the disaggregated means to be sure they are of similar magnitudes.

In cases where a program is undergoing revisions and renewal, the results should be of a character that will support a sound prediction of what future results will be. Generally, this means that the most recent results will carry greater weight in the faculty's and TEAC's decision making.

EXERCISES from *TEAC Exercise Workbook, 2010*
pages 66-75

Exercise 47: Results: The results, whether quantitative or qualitative, should be truly representative of the program under review and not idiosyncratic to a particular time period or circumstance.

The results must also be disaggregated by subcategory when an aggregated presentation would mask important differences within the groups and categories being reported.

In cases where a program is undergoing revisions and renewal, the results should be of a character that will support a sound prediction of what future results will be. Generally, this means that the most recent results will carry greater weight in TEAC's decision making.

The exercises that follow ask you to think about ways you might present results in the *Brief* and what pitfalls you can avoid.

The four tables below describe some hypothetical and real findings related to the claim of subject matter knowledge (*Quality Principle 1.1*) for a program that prepares secondary level mathematics teachers. Consider the following tables reporting findings in support of the claim *Our graduates know their subject matter*.

Are there problems with this presentation that are serious enough for you to reject the conclusion that the program satisfies *Quality Principle I* with regard to subject matter knowledge? How could the array be made clearer? Use the space below each table for notes in preparation for the conversation.

Table 47a: Findings related to the claim of subject matter knowledge for a program to prepare secondary level mathematics teachers

Year of graduation	GPA in upper level mathematics courses: math department mean: 3.1			Scores on Praxis II: mathematics national mean: 540		
	Graduates N	Mean	SD	Graduates N	Mean	SD
06-07	15	3.5	.5	15	610	90
07-08	12	3.4	.4	12	590	95
08-09	14	3.6	.5	14	615	92
09-10	20	3.0	1.1	20	510	130
10-11	15	3.5	.5	15	610	95

Note: The average correlation across the five-year period between the two measures was .75. The faculty assessed the reliability of the GPA by drawing a sample of 30 students from the five-year period, and correlating the grades received on the odd lines on their transcript with the grades received on the even lines. The correlation was .78, providing support for a satisfactory level of reliability. ETS reports that the Praxis II examination in mathematics has a reliability of .83 for its norming group. The graduates of our program match well the characteristics of the ETS norm group.

An improved presentation of the data appears below.

Table 47b: Mean (and SD) GPA and Praxis II Scores and Pass Rates in Secondary Mathematics for Years 2006-2011 at Exemplar University

Year of graduation	GPA in upper level mathematics courses: math department mean: 3.1; math department standard: 2.75			Scores on Praxis II: mathematics national mean: 540 State cut score: 520			Correlation: GPA-Praxis II
	Number graduates/ number admitted	Mean (0-4)	SD	Graduates N	Mean and (pass rates) 400-800	SD	Pearson R
06-07	15/20	3.5	.5	15	610 (84%)	90	.89
07-08	12/21	3.4	.4	12	590 (77%)	95	.70
08-09	14/22	3.6	.5	14	615 (85%)	92	.69
09-10	20/22	3.0	1.1	20	510 (47%)	130	.71
10-11	15/20	3.5	.5	15	610 (84%)	95	.80
Total	76/105	3.4	.6	75	587 (75%)	100	.75

Note: The faculty assessed the reliability of the GPA by drawing a sample of 30 students from the five-year period and calculating the mean GPA each year. The means were within .05 of each other. ETS reports that the Praxis II examination in mathematics has a reliability of .83 for its norming group. The graduates of our program match well the scores and demographics of the ETS norm group.

The table below offers an example of how a program faculty might organize its quantitative results for the components of *Quality Principle I*.

Note that although means and standard deviations are the likely entries in each cell of the table, frequency counts, ranks, percentages, percentiles, or whatever quantitative metric the faculty relies on could be also entered.

Table 47c: Means (standard deviations) of a sample of 80 students in six categories of assessments in support of claims for *Quality Principle I*

Outcome claims*	Categories of evidence and range of scores								
	Grade point index	Standardized tests		Faculty & Cooperating Teacher evaluations		Student self-reports	Survey of graduates and employers		Gains in pupil scores on work samples
		Praxis	NES	Fac	Coop		Grad	Emp	
(Score Range, e.g 1-4 & Cut Score)	(Score Range & Cut Score)	(Score Range & Cut Score)	(Score Range & Cut Score)	(Score Range & Cut Score)	(Score Range & Cut Score)	(Score Range & Cut Score)	(Score Range & Cut Score)	(Score Range & Cut Score)	
<i>The program's graduates have acquired...</i>									
Subject matter (<i>Professional knowledge</i>)									
Pedagogy (<i>Strategic decision-making</i>)									
Teaching skill (<i>Leadership skill</i>)									

*Includes measures of learning how to learn, multicultural perspectives and technology

Here is another example:

Table 47d: Mean grades (u/g & program), license tests, portfolio rating, internship ratings (mid-term and final) and course pedagogy projects for students in special education program options

	N	Minimum	Maximum	Mean	Program standard	Standard deviation
Undergrad (1-4)	43	2.60	3.88	3.29	3.00	.32
Prog GPA (1-4)	43	3.33	4.00	3.84	3.00	.14
LAST (100-300)	15	233	286	262	250*	18.45
ATS (100-300)	15	234	284	265	250*	14.74
LIT (100-300)	30	223	289	265	250*	16.11
SWD (100-300)	35	222	279	254	250*	14.18
CST (100-300)	23	202	283	259	250*	19.77
Portfolio 1 (1-3)	43	1.00	3.00	2.36	2.25	.52
Portfolio 2 (1-3)	43	1.00	3.00	2.29	2.25	.68
Portfolio 3 (1-3)	43	1.00	3.00	2.45	2.25	.60
Portfolio 4 (1-3)	43	1.00	3.00	2.00	2.25	.72
Domain 1 (1-4)	43	1.80	4.00	3.69	3.00	.42
Domain 2 (1-4)	43	2.50	4.00	3.73	3.00	.39
Domain 3 (1-4)	43	2.00	4.00	3.69	3.00	.40
Domain 4 (1-4)	43	1.80	4.00	3.81	3.00	.39
Mid Term (1-4)	42	2.30	3.70	2.92	3.00	.35
Final (1-4)	42	2.50	4.00	3.32	3.00	.35
Project 1 (1-100)	32	80.00	97.00	90.92	75.00	4.52
Project 2 (1-4)	32	3.30	4.00	3.77	3.00	.27

*State passing score is 220

🔗 Sample response:

Looking at the data in Table 47b, you might suggest to the hypothetical Brief authors that they investigate their record for the following:

- The statistics about the validity of the interpretations of the measures.
- Whether the reliability sample of 30 was representative of the population and whether the N's are representative samples of the graduates or the universe of graduates.
- The range of scores for Praxis II (what is the zero score and the maximum?).
- The completion rates in each year of the program (because in the years when fewer than 20 students graduated, the program might have had high drop-out rates, in which case the true means would be more like that of the students in year 2009-2010).
- Other mathematics measures that would contradict those in the table.
- Special features that might have been in place in 2009-2010 (e.g., changes in faculty, the curriculum, the admission standards, policy changes, etc.) to account for the lower mean and larger standard deviation.
- The range of the grades to be sure 4.0 was the maximum.
- The percentage of students who had a 3.0 or higher math GPA and the number who passed Praxis II (by the state criterion) and achieved 75% of what the top 10% of Praxis II math scorers achieved on the test.
- The comparison of the math grades of the graduates with math majors, not just the average of all the graduates at the institution.

Table 47b (repeated): Mean and sd GPA and Praxis II scores and pass rates in secondary mathematics for years 2006-2010 at Exemplar University

	GPA in upper level mathematics courses: math department mean: 3.1; math department standard: 2.75			Scores on Praxis II: mathematics national mean: 540 State cut score: 520			Correlation: GPA-Praxis II
Year of graduation	Number graduates/ number admitted	Mean (0-4)	SD	Graduates N	Mean and (pass rates) 400-800	SD	Pearson R
06-07	15/20	3.5	.5	15	610 (84%)	90	.89
07-08	12/21	3.4	.4	12	590 (77%)	95	.70
08-09	14/22	3.6	.5	14	615 (85%)	92	.69
09-10	20/22	3.0	1.1	20	510 (47%)	130	.71
10-11	15/20	3.5	.5	15	610 (84%)	95	.80
Total	76/105	3.4	.6	75	587 (75%)	100	.75

Note: The faculty assessed the reliability of the GPA by drawing a sample of 30 students from the five-year period, and correlating the grades received on the odd lines on their transcript with the grades received on the even lines. The correlation was .78, providing support for a satisfactory level of reliability. ETS reports that the Praxis II examination in mathematics has a reliability of .83 for its norming group. The graduates of our program match well the characteristics of the ETS norm group.

Are there problems serious enough for you to reject the conclusion that the program satisfies *Quality Principle I* with regard to subject matter knowledge?

Probably not, because the evidence in favor of the conclusion is persuasive. There are two measures with reliability and validity determinations within the TEAC guidelines. With regard to the sufficiency of the two measures, each represents more than 75% of the maximum score available to the graduates. All the grades in math for 80% of the years were at or above 3.0 out of 4.0 and averaged at least 3.0 every year. The preponderance of the evidence (that is, 80% of the time and for over 75% of the 76 graduates) is consistent with the claim that they know their subject matter. The graduates in most years exceeded on average the grades earned in the math department, many of who were presumably majoring in math. The table offers no evidence that would indicate that the students are not competent in their mathematics knowledge.

Exercise 48: Apple University’s claim for caring teaching skills:

The faculty members at Apple University have claimed that their graduates leave the program with the teaching skills needed to perform well in their first school placement. The evidence they share in the *Inquiry Brief* to support their claim consists of three measures:

(1) scores on the state tests showing pupil gains in performance on the state content standards; (2) ratings received from the graduates’ first year principals; and (3) comparison of hiring rates of the program with other similar programs in the state.

How credible is Apple University’s claim? What are some rival explanations for the findings reported below? Would you conclude there was sufficient evidence to support the claim the Apple faculty made?

Apple University’s evidence

Pupil gains on state tests. While the teacher education program at Apple University graduates 200 new teachers every year, only 120 of them elect to accept positions “in-state.” Others take teaching positions in other states (n = 40) while others pursue life options in graduate school, in business, or in raising families (n=40). The state examinations of pupil achievement are administered only in 3rd grade and 5th grade. Of the 120 graduates who are teaching in state, only 40 are teaching at the 3rd grade or 5th grade level. The results of the pass rates for the pupils of those teachers are reported below in Table 46a:

Table 48a: Pass rates of students in 3rd and 5th grade of Apple University first-year graduates compared to state rates overall on the state curriculum test

Grade level	N	Percent passing	Percent failing
3 rd grade: Apple teachers	25	70%	30%
3 rd grade: Overall	2,050	55%	45%
5 th grade: Apple teachers	15	65%	35%
5 th grade: Overall	2,200	58%	42%

Principal survey. The Apple University faculty surveyed the principals of the 120 recent graduates of the program who were teaching. The faculty was fortunate in that of the 87 schools in which the 120 Apple graduates were employed, 60 had principals who had either graduated from Apple University as undergraduates or had studied at Apple University for their administration credentials. This situation increased the likelihood that the principals would respond to the survey instrument. The results of the survey are disaggregated in Table 48b, below:

Table 48b: Number of Apple-trained teachers rated excellent, satisfactory, and unsatisfactory on a survey of principals disaggregated by principals' affiliation with Apple University

Principals (N)	Teachers rated as "excellent"	Teachers rated as "satisfactory"	Teachers rated as "unsatisfactory"
Principals with previous affiliation with Apple University (n = 45)	12	30	3
Principals with no previous affiliation with Apple University (n = 15)	1	9	5
Total number	13	39	8

Hiring Rates. Principals and directors of personnel in schools are very conscientious about hiring new teachers who have outstanding teaching skills. It is not enough these days to be smart or to have high grade point averages. Often candidates for teaching posts must present portfolios documenting their teaching practices, and on occasion they must demonstrate their skills by offering a lesson to pupils in the school. Hiring rates are quite significant, then, in attesting to the teaching skills of a program's graduates. Apple University faculty interviewed the directors of teacher placement at 20 sister campuses within the state with programs similar to those of Apple to determine what percentages of placement represented the hiring rates of Apple's graduates. Only 13 campuses were able to provide this information. The results are arrayed below in Table 46c:

Table 48c: Number of Apple University graduates and non-Apple graduates and percentage hired in and out of state

Institution	Number of graduates	Percentage hired in state	Percentage hired out of state	Not hired
Apple University	200	60%	20%	20%
Sister institutions	650	55%	15%	30%

Exercise 48, continued: Formulate your own response to the questions posed at the outset of the exercise:

- How credible is Apple University's claim?
- What are some rival explanations for the findings reported in the tables?
- Would you conclude there was sufficient evidence to support the claim the Apple faculty made?

Use the space below for notes.