California Polytechnic State University
University/GE Assessment of Quantitative Reasoning (AY 2016)

Introduction

In AY 2013-14, the Office of Academic Programs and Planning initiated the first campus-wide assessment of quantitative reasoning at Cal Poly. WASC had revised their Handbook of Accreditation in 2013 directing institutions to measure student learning and achievement of this core competency. WASC defined quantitative reasoning as "the ability to apply mathematical concepts to the interpretation and analysis of quantitative information in order to solve a wide range of problems, from those arising in pure and applied research to everyday issues and questions. It may include such dimension as the ability to apply math skills, judge reasonableness, communicate quantitative information, and recognize the limits of mathematical or statistical methods" (WASC Handbook 2013, p 55). Unlike the competencies for writing, critical thinking, oral communication and information literacy, quantitative reasoning was seen as more varied across disciplines at Cal Poly with a greater emphasis on reasoning with numerical evidence over computational skills and abilities.

Investigation (AY 2014-2015). The research phase of the University/GE assessment of quantitative reasoning began with the establishment of a learning community comprised of faculty across campus from Engineering, Science and Math, Liberal Arts, Business, Architecture and Agriculture as well as representatives from Kennedy Library and the Office of Academic Programs and Planning. The goals of the QR Learning Community (QRLC) were to develop a working definition for quantitative reasoning, design a QR rubric based on the definition, survey the presence of QR in courses across campus, and to evaluate existing measures of student learning and achievement.

After some investigation, the QRLC used the Carleton College definition of QR as the basis for an expanded Cal Poly definition: "The pursuit and critical application of quantitative information for the purpose of constructing, communicating, and critiquing arguments in public, professional, and personal constructs". Constructing, communicating, and critiquing were significant elements of the definition as hey explain QR as a form of critical thinking that uses quantitative information as evidence. The "public, professional, and personal" formulation was significant because it illustrated QR as having several dimensions. While student achievement of QR in the professional dimension would take a discipline-specific form that would be difficult to benchmark at the university level, student achievement of QR in the public and personal dimensions would take a broadly shared form that could be benchmarked and addressed in GE as an expectation for all Cal Poly graduates.

As valuable as this discussion might have been for the initial campus understanding of QR, there was some growing awareness that the proposed Cal Poly definition was a complicated one that would also be difficult to remember. After further examination, an alternative definition of QR was proposed and agreed upon:

> "Quantitative reasoning is the ability to make (or critique) a persuasive argument about a real-world or discipline-specific problem based on numerical evidence".

Based on this definition, the learning community developed a four-point quantitative reasoning rubric with four traits: Problem Identification, Quantitative Analysis, Visual Presentation and

Oral/Written Communication. The community conducted a survey of QR in existing courses, looking for places where QR *is taught* or *should be* taught based on the existing outcomes, but also where QR *might be* taught, based on other course descriptors. This effort was complicated by the fact that QR is not explicitly addressed in either the University Learning Objectives or the GE Objectives and Criteria.

The survey, which was originally meant to cover courses in both General Education and the majors, eventually focused on GE, where it established contextualized, foundation-level courses in Math and Economics, as well as foundation-level courses in Statistics and the social sciences, specifically Psychology, as good candidates for a university-wide assessment. For senior level work, building on the assessment model established through the University/GE writing assessment, the QRLC called for senior level assignments from across campus and is currently planning to focus on a cross-section assessment of upper division programs and courses in 2017-18, ideally one program from each college, to conduct major-specific assessments of QR at the mastery level.

Evaluation (AY 2015-2016). In Spring 2016, over 240 student assignments were collected from four lower-division GE courses (PSY 202, ECON 222, MATH 112, STAT 130) with 60 artifacts chosen randomly from each course. The QR rubric had been shared in advance with faculty members who teach these courses in order to test the rubric and to ensure the assignments contained the criteria called for in the rubric. Psychology was the only assignment that did not contain all four criteria as the assignment had been administered prior to the completion of the final draft of the rubric. The psychology essay called out only two traits from the rubric (Problem Identification and Oral/Written Communication).

Prior to the scoring sessions, readers were sent four sample artifacts in advance along with the assignment descriptions and answer keys as needed. Two of these sample artifacts were pre-scored and the other two were sent for the readers to score individually in order that the they become familiar with the assignments and the rubric language prior to the scheduled norming and scoring sessions.

On June 27 and 28, 2016, twenty faculty members participated in four scoring sessions led by Jack Phelan, the Director of Academic Assessment and Dawn Janke, the Director of the Writing and Rhetoric Center. The readers were comprised of Associate Deans, Academic Assessment Council members, learning community members and other faculty from across campus who were interested in the assessment of Quantitative Reasoning. Readers were divided into two rooms according to subject matter. Psychology and Math were scored on one day and Economics and Statistics were scored on day two. Ten-month faculty were paid a $200 stipend for each day of service.

Norming and scoring of the artifacts was completed using the four-point Cal Poly QR Rubric. The four levels of performance on the rubric range from 1 for “Limited Proficiency” to 4 for “High Proficiency”. During norming, the sample artifacts and the QR rubric were discussed in depth, then the readers individually scored additional artifacts until consensus scoring had been achieved. When discussions were complete and the readers were declared normed, each of the artifacts was scored twice. As a measure of inter-rater reliability, any artifact with a discrepancy score greater than 1 point on any given trait was scored by a third reader.

Whenever there was a two-point discrepancy in scores on any trait between the two readers of an artifact, a third reading was required. If the third reading coincided with either of the first two reader's scores, the third reading would replace the non-matching score. If the third reading of the artifact scored between the first two readings, a coin would be flipped to decide whether the score should be rounded up or rounded down. If the coin said to round up, the third score would replace the lower of the first two and vice versa for rounding down. 21 artifacts (approximately 9%) required a third reading. Of these 21 artifacts, 12 contained a discrepancy score in one trait, 7 contained discrepancies in two traits, and 2 contained a discrepancy in three traits. The following table (Table 1) shows the breakdown of the 32 total discrepancies for each trait by subject.

Table 1
Discrepancy scores for each trait by subject:

	Problem Identification	Quantitative Analysis	Visual Presentation	Oral/Written Communication
PSY	5	N/A	N/A	3
MATH	3	9	7	3
ECON	0	0	1	1
STAT	0	0	N/A	0
Total	8	9	8	7

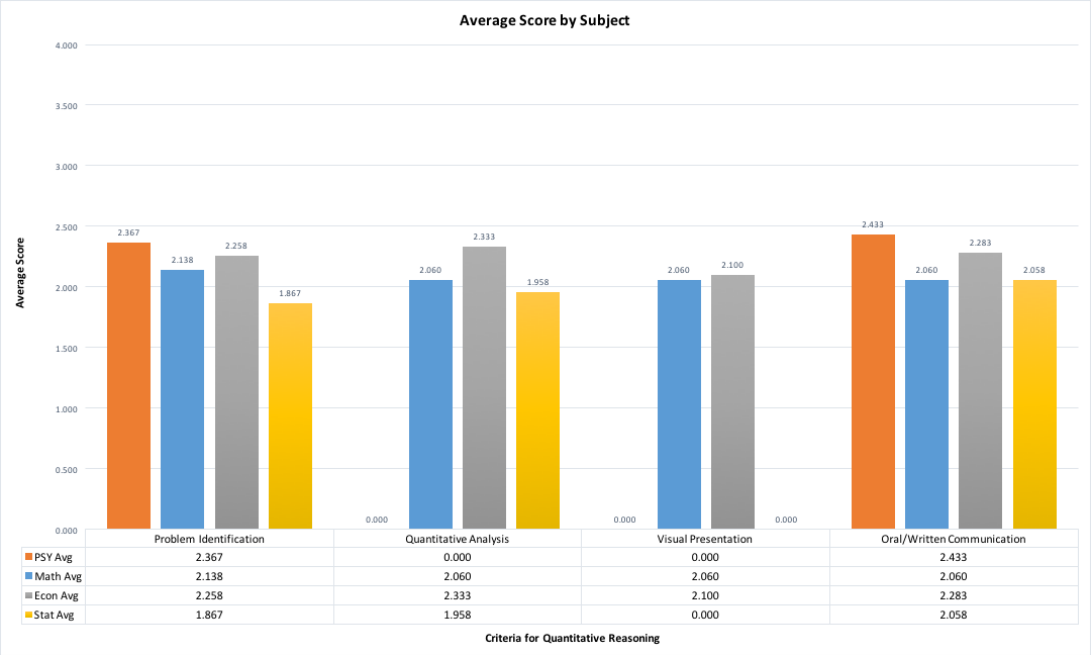
Analysis of Results. An abbreviated description of the scoring session results follows.

Table 2 and Graph 1 show the average scores (with standard deviations) by subject for each of the four traits being measured. Although averages cannot be compared between subjects due to difference in the courses and assignments, the common average across traits and subjects hovered very closely to two on the rubric for each subject. A two is identified as "Emerging Proficiency" on the QR rubric and these results show evidence that students are "meeting expectations" at the foundational level.

Table 2
Averages and standard deviations of scores for each trait by subject:

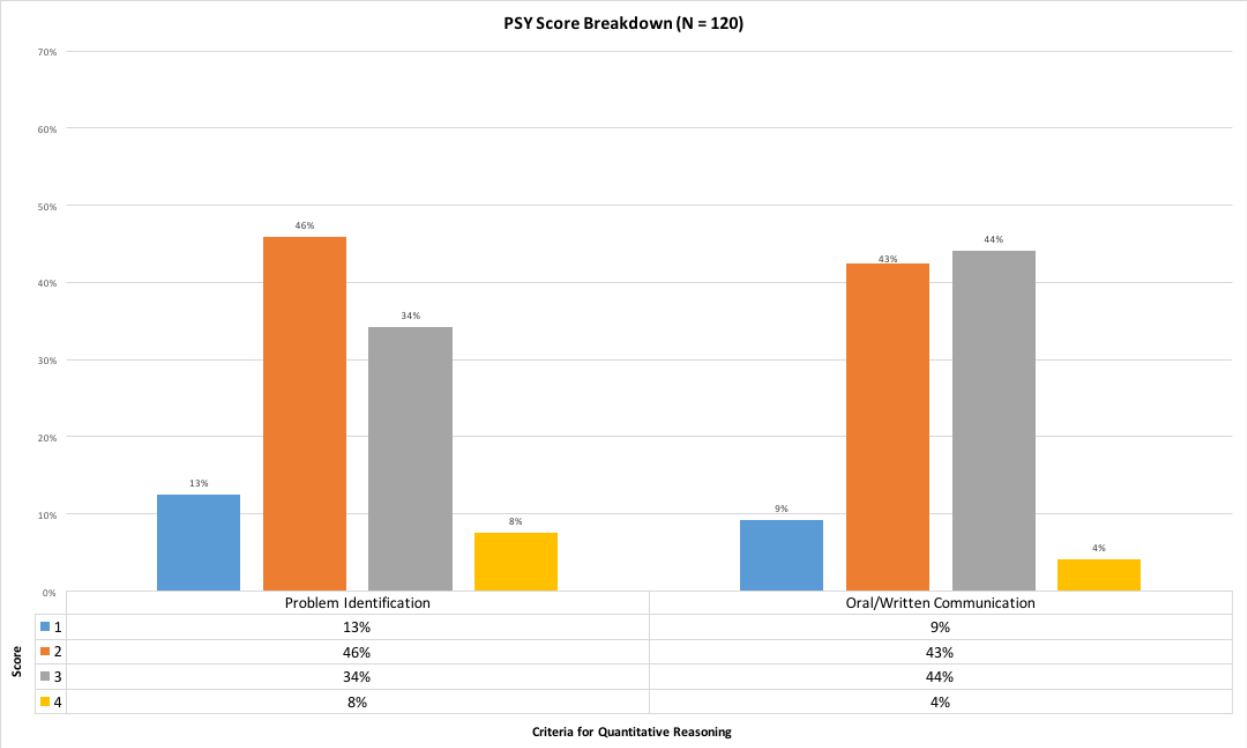
	Category	Problem Identification	Quantitative Analysis	Visual Presentation	Oral/Written Communication
PSY	Avg	2.367	N/A	N/A	2.433
	Sd	0.798	N/A	N/A	0.719
Math	Avg	2.138	2.060	2.060	2.060
	Sd	0.697	0.714	0.663	0.772
Econ	Avg	2.258	2.333	2.100	2.283
	Sd	0.615	0.760	0.824	0.676
Stat	Avg	1.867	1.958	N/A	2.058
	Sd	0.733	0.793	N/A	0.598

Graph 1
Averages and standard deviations of scores for each trait by subject:

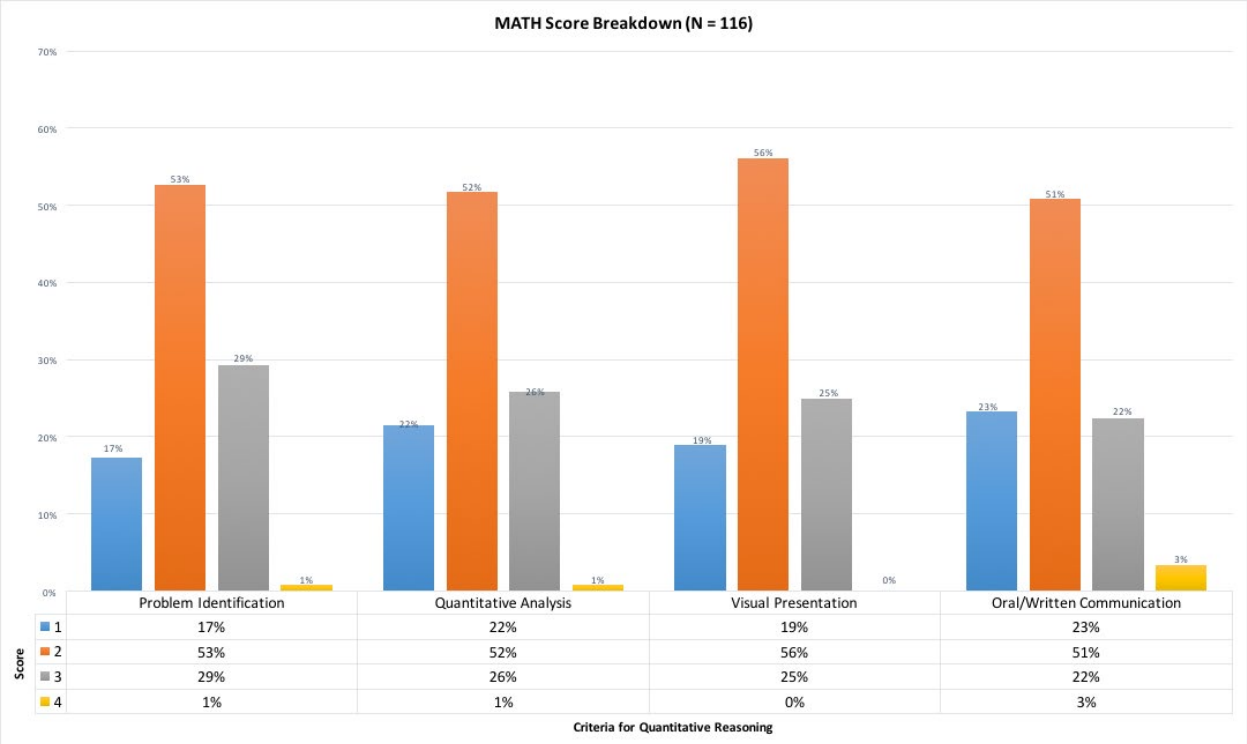


The following four graphs (Graphs 3, 4, 5, 6) show percentage distributions for each rubric trait by subject. We see the distribution of scores across subjects and traits rising at emerging proficiency with most students (80%) found to either meet or exceed the benchmark across traits and subjects. As expected for freshmen and sophomores, very few students reached a high level of proficiency indicating the validity of the rubric, which seemed to work very well.

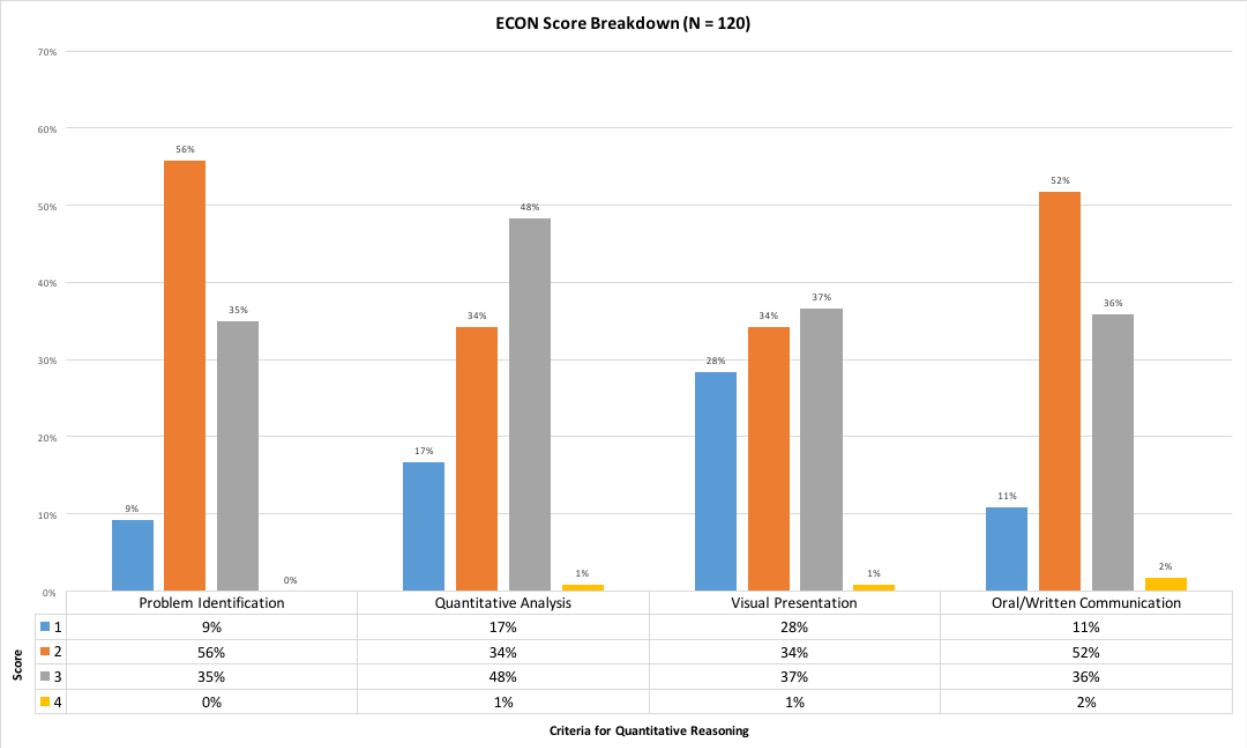
Graph 3
PSYCHOLOGY: Breakdown of scoring distribution per trait for each subject (percentiles):



Graph 4
MATH: Breakdown of scoring distribution per trait for each subject (percentiles):



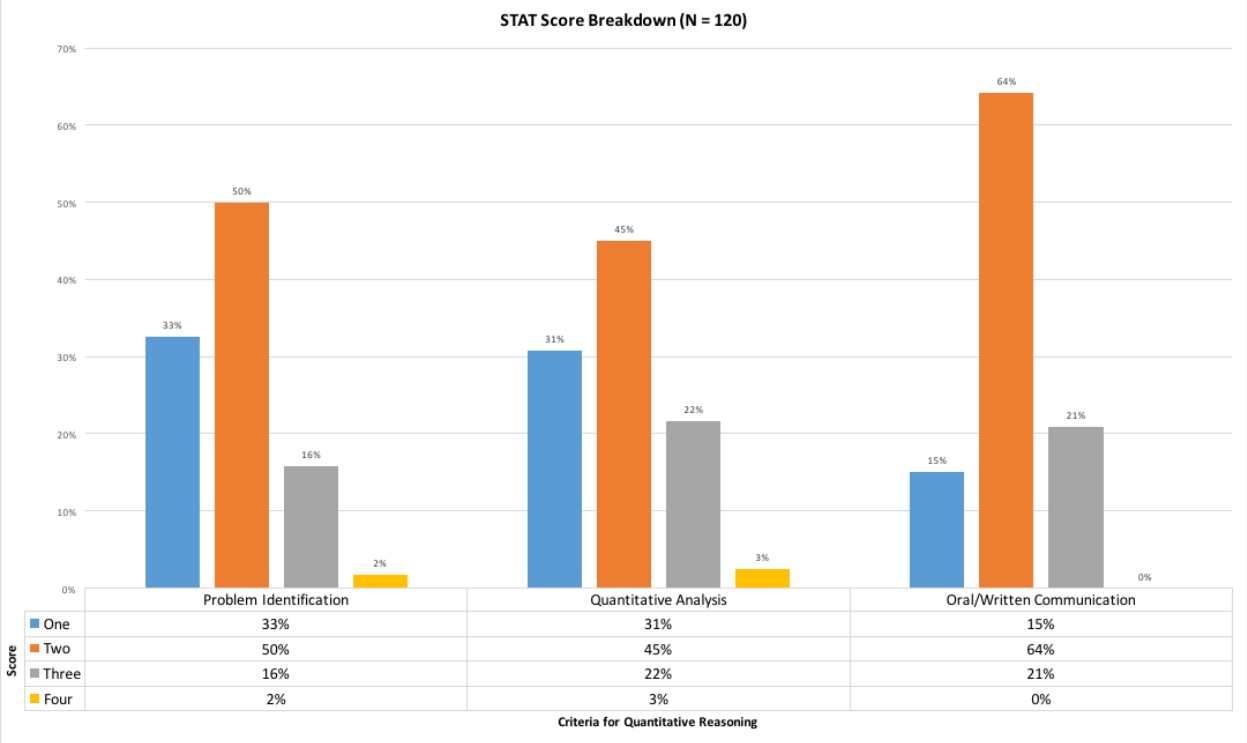
Graph 5
ECONOMICS: Breakdown of scoring distribution per trait for each subject (percentiles):



Score	Problem Identification	Quantitative Analysis	Visual Presentation	Oral/Written Communication
1	9%	17%	28%	11%
2	56%	34%	34%	52%
3	35%	48%	37%	36%
4	0%	1%	1%	2%

Criteria for Quantitative Reasoning

Graph 6
STATISTICS: Breakdown of scoring distribution per trait for each subject (percentiles):

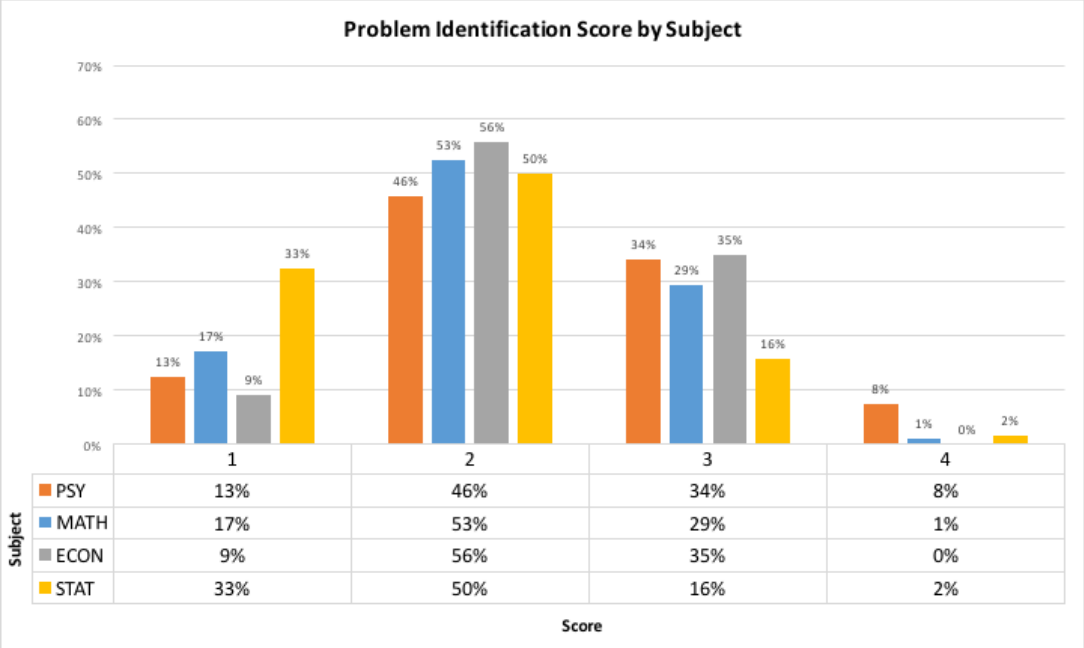


Score	Problem Identification	Quantitative Analysis	Oral/Written Communication
One	33%	31%	15%
Two	50%	45%	64%
Three	16%	22%	21%
Four	2%	3%	0%

Criteria for Quantitative Reasoning

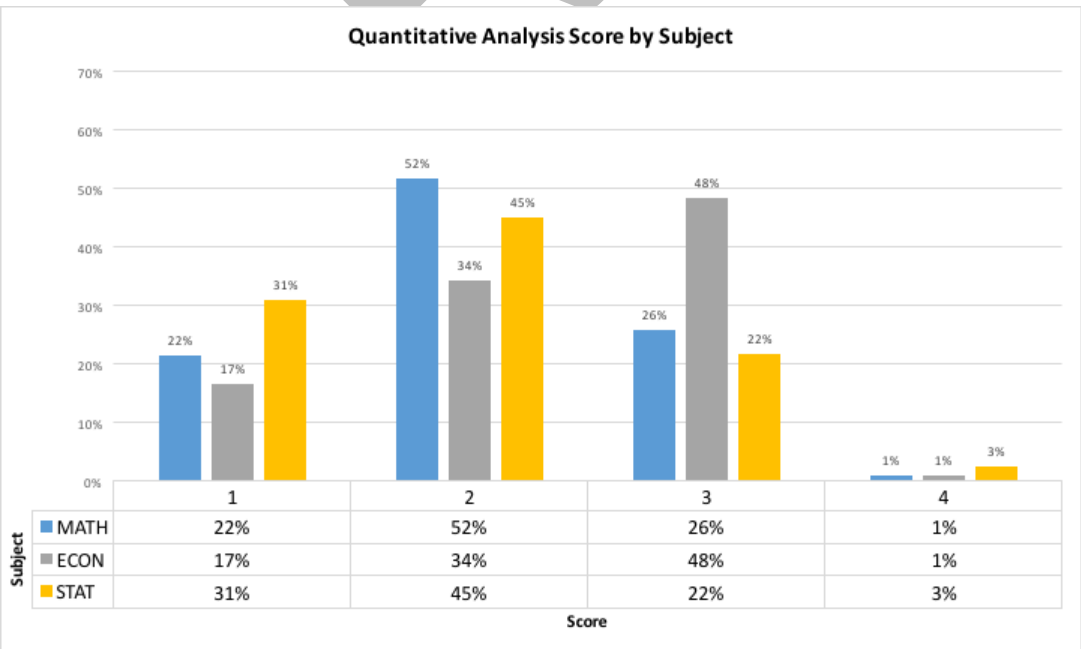
The following graphs (Graph 7, 8, 9, 10) show a side by side comparison of subjects for each trait in percentiles. The distribution of scores across traits appears to rise at emerging proficiency.

Graph 7
PROBLEM IDENTIFICATION: Side by side comparison of subjects by trait (percentiles):



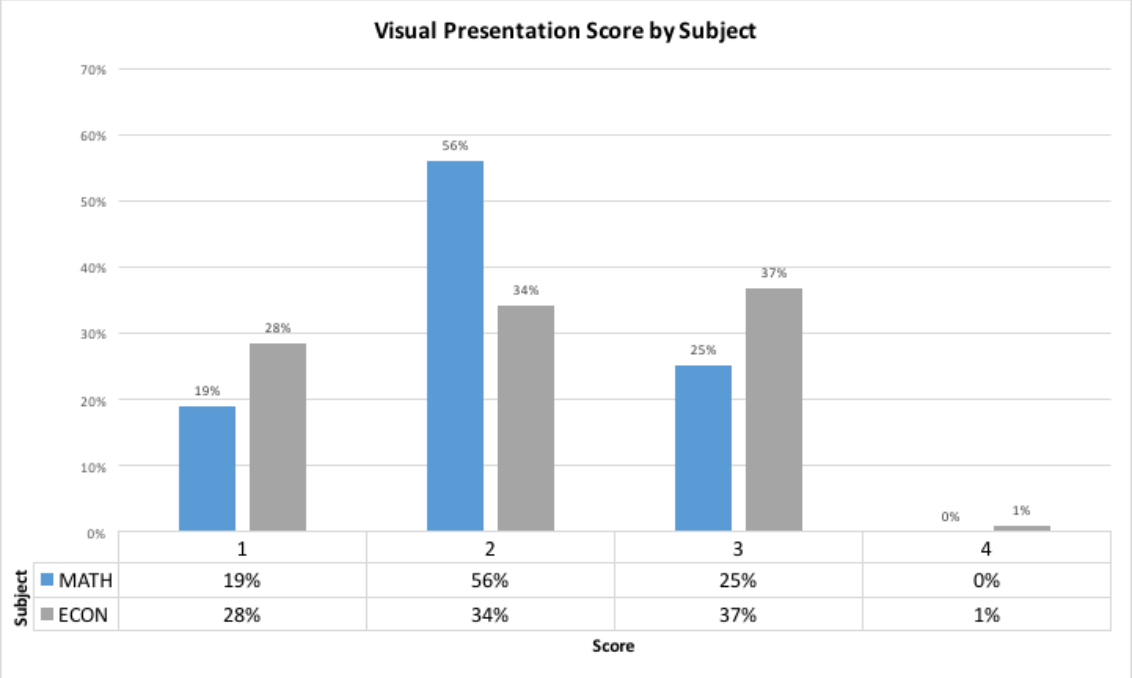
	1	2	3	4
PSY	13%	46%	34%	8%
MATH	17%	53%	29%	1%
ECON	9%	56%	35%	0%
STAT	33%	50%	16%	2%

Graph 8
QUANTITATIVE ANALYIS: Side by side comparison of subjects by trait (percentiles):



	1	2	3	4
MATH	22%	52%	26%	1%
ECON	17%	34%	48%	1%
STAT	31%	45%	22%	3%

Graph 9
VISUAL PRESENTATION: Side by side comparison of subjects by trait (percentiles):



Graph 10
ORAL/WRITTEN COMMUNICATION: Side by side comparison of subjects by trait (percentiles):