In S. Franklin, J. Marx, & K. Cummings (Eds.), **Proceedings of the 2001 Physics Education Research Conference** (pp. 41-44). Rochester, NY.

Questions First (Q1st):

The challenges, benefits, drawbacks, and results of asking students questions prior to formal instruction

William J. Leonard, William J. Gerace, and Robert J. Dufresne

Department of Physics and Scientific Reasoning Research Institute
University of Massachusetts, Amherst, MA 01003-4525 USA

The initial intent of this research study was to compare two different approaches to using a classroom communication system (CCS). Two experienced CCS users would be in charge of his own section of first-semester, calculus-based, introductory college physics for math, science and engineering majors. Electronic homework (eHW) and identical exams would be administered to both sections. But the ways in which the CCS would be used during the lecture period and the way eHW would be administered to the two sections would be notably different. In the "Questions First" (Q1st) section, the CCS would be used to stimulate discussion and motivate short lectures. In the other section (non-Q1st), it would be used after lecturing to monitor student progress and understanding. In the Q1st section, eHW would be due before each lecture period, and in the non-Q1st section, the same assignment would be due the day after the corresponding material had been covered during the lecture period.

In the end, however, this study became less about trying to pin down the effects of two different instructional styles, and much, much more about the difficulties of comparing two large sub-populations of students. Therefore, although we will report on our findings regarding the comparison of the two sections and approaches, a large fraction of this talk will focus on the development of our thinking regarding the hindrances to making definitive and reliable statements about our findings.

Philosophy behind Q1st

A popular use of a CCS is to lecture briefly (such as 10 to 15 minutes), and then to poll the class to see if the lecture was understood. There may be some discussion among students, but typically there is not a class-wide discussion. Though this approach has met with some success for some instructors, we believe that it does not take full advantage of CCS technology. Q1st is an attempt to raise the focus of instruction above what can be achieved by lecturing alone.

The Q1st approach is aimed at helping students develop higher order thinking skills. This is achieved by taking advantage of the feature that the lecture period is when all (or at least most) of the students are together under the supervision and guidance of the instructor. Students can thus use the lecture period to relate concepts to each other and to use concepts to reason about and analyze physical situations, as well as start to organize knowledge so that it is useful for communication and problem solving.

Therefore, homework is assigned prior to lecture to drive students to read the textbook and begin to gain some initial exposure to concepts and their definitions. Then, during class, questions are posed via the CCS. The focus of the period is the ensuing discussion launched by viewing a histogram of student responses. This discussion brings out the myriad interpretations and reasoning that students use to process and answer questions. If needed, a formal lecture is used to wrap up the ideas raised during the discussion.

Data

The Q1st style of instruction is expected to have consequences in many areas, such as conceptual understanding and appreciation of the role of principles for problem solving, but the only measures we have examined are performance on electronic homework and semester exams, as well as student responses to an end-of-semester (EOS) survey asking them to rate the effectiveness of various components of the course.

For homework, students were assigned 91 eHW problems, usually in assignments of 3 problems each, with 2 or 3 assignments per week during the 14-week semester. Problems were like those from any standard, calculus-based, college physics textbook.

Three tests and a final exam were administered to both sections approximately 3–4 weeks apart during the semester. Each test consisted of 24 multiple-choice questions. The questions can be divided into three basic types: (1) computa-

Table I. Some differences between sections.

Q1st	Non-Q1st
questions drive discussion and motivate lecture	questions mostly after short lecture
eHW due before lecture session	eHW due 1 day after lecture session
conceptual questions also due before lecture	repeated CCS questions due after lecture
only 29% engineering majors	mostly engineering majors (64%)
TuTh	MWF
75-minute lecture sessions	50-minute lecture sessions
only 82 students	236 students
attendance graded	attendance not graded

tional (i.e., traditional); (2) conceptual; and (3) analysis/reasoning. The computation problems can be solved either algorithmically by pattern matching and manipulating equations, or conceptually using principles. The conceptual questions usually require conceptual understanding, but students might simply know the answer from memory. The analysis/reasoning questions usually require some combination of conceptual understanding and appreciation of definitions and basic relationships, and ask students to reason about situations using concepts and equations.

The EOS survey asked students to rate each of nine components of the course on its effectiveness in helping them understand course material and prepare for the semester exams. Some of the components rated were: Lecture Sessions, CCS Questions, Problem Sets, Practice Exams, Textbook, and Review Sessions. A high percentage of students (84%) filled out this survey.

Comparison of the two sections

As mentioned before, both professors used a CCS during class to ask questions and stimulate activity. Also, both sections were assigned the same 91 eHW problems, both took the same 4 semester exams, and both filled out an EOS survey. A summary of the differences between the two sections may be found in Table I.

Results and conclusions

Results on the 4 exams and eHW are shown in Table II. The Q1st section performed consistently below the non-Q1st section on all 4 exams, with an average difference of 2 percentage points. This means that, on average, the non-Q1st section scored one-half of one multiple-choice question better than the Q1st section.

The difference in eHW scores is dramatic. The non-Q1st section outperformed the Q1st section by nearly 30 percentage points, and on average, correctly answered nearly twice as many problems.

In effectiveness for understanding the material and preparing for exams, Review Sessions and Practice Exams were rated #1 and #2, respectively, by both sections, and the Lecture Sessions were rated lowest by both. CCS Questions were rated near the bottom by both sections.

We can conclude that: (1) Q1st students did not (or could not) do the eHW, perhaps because it was due prior to class; (2) despite the large difference in eHW scores, there is little difference in the exam scores; and (3) students apparently learned the material and prepared for exams (or at least, they believe that they did) by attending review sessions and using practice exams. Further, students from both sections found the lecture sessions the least effective component of the course.

Table II. Results on exams and eHW.

	E1	E2	E3	F	$EXAM_{ave}$	eHW_{ave}
Q1st (%)	54.6	39.4	48.6	37.2	45.0	34.9
non-Q1st (%)	57.1	39.9	49.2	41.8	47.0	64.6
□ (%)	–2.4	–0.5	–0.6	–4.6	–2.0	–29.7
N (Q1st)	80	76	70	71		
N (non-Q1st)	229	223	216	215		

Looking in greater detail at exam performance.

Figure 1 shows a scatterplot comparing the performance on each of the 96 exam questions by the two sections. The dotted line indicates where a question would end up if both sections performed equally well. Points above the line indicate questions on which the Q1st section performed better, and points below the line indicate questions on which the non-Q1st section performed better.

The range of differences between the two sections is rather large, going from a minimum of −24.5% to a maximum of +15.5%. We might hypothesize that many of these differences are significant, and perhaps knowing which questions were answered more successfully by which section would lead to new insights into the two differing instructional approaches.

The difficulty is knowing how large a difference is significant. Therefore, we have computed an uncertainty σ_n in each difference Δ_n, and constructed the set of ratios $(\Delta/\sigma)_n$, where n ranges from 1 to 96. These 96 ratios are shown in Figure 2.

We might presume that all ratios larger than 1 or 2 indicate questions that are significant enough to warrant closer scrutiny (i.e., differences are more than 1σ or 2σ from the average). But that means we must ask ourselves, "If these 96 ratios were distributed Normally, what would the distribution look like?" To answer this question, we can compare the *cumulative Normal probability distribution* to the distribution above. The result is shown in Figure 3.

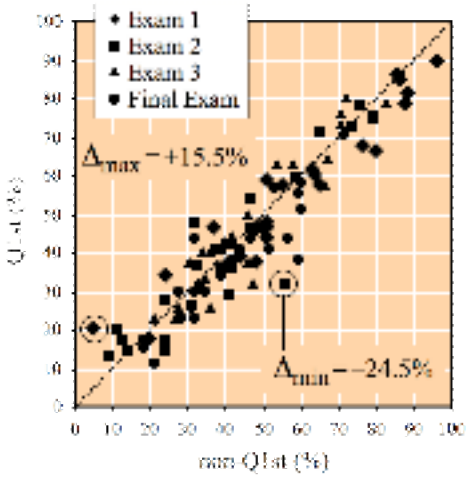

Figure 1. Comparison of exam performances by question.

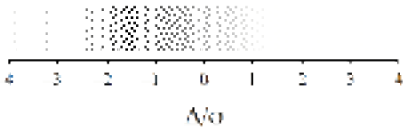

Figure 2. Distribution of Δ/σ.

The 96 values of Δ/σ appear to be distributed Normally, with a shift in the mean $(\Delta/\sigma)_{ave}$ of −0.34. The sum of squared differences between the question distribution and the Normal distribution Σ(χ^2) is about 4. These two numbers characterize the distribution.

Both numbers seem to be small, but how can we know for certain? How do we know if either the shift in the mean or the sum of squared differences is significant?

One technique is to compute $(\Delta/\sigma)_{ave}$ and Σ(χ^2) for arbitrary sub-populations of the same section. Because students within a particular section experienced the same treatment, any differences between chosen sub-groups may be considered a sampling of the statistical differences we might expect when comparing other sub-groups. Table III shows values for $(\Delta/\sigma)_{ave}$ and Σ(χ^2) for various sub-groups of the larger (non-Q1st) section.

This table indicates that differences (Δ) as large as 25% should not be considered unusual. It also indicates that an average difference between two sub-groups (Δ_{ave}) of 2–4% is not unusual either. Further, a shift in the mean of Δ/σ as large as about 0.5 or a sum of squared differences of as large as about 10 should be considered within statistical bounds. (That is, for this set of choices, $(\Delta/\sigma)_{ave}$ is small when Σ(χ^2) is large and vice versa.)

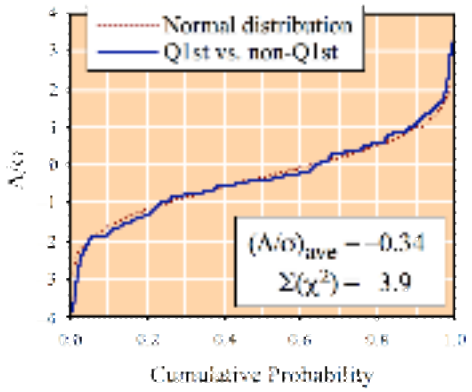

Figure 3. Comparison of question distribution to Normal distribution.

Comparing various sub-populations of the two sections

Now that we have a sense of how to interpret differences between the two sections, we can look at selected sub-groups to see how they compare to each other. Table IV summarizes the results. Based on the criteria developed in the previous section, some of these differences are probably not significant, some *might* be significant, and two are probably significant.

We can also break up each section according to performance on the electronic homework. The results (not shown) indicate that there may be significant differences between the two classes after all, depending on one's ability to do the electronic homework. The exam averages for the non-Q1st section are largely independent of which third of the class the student is in (top: 48%, middle: 48%, and bottom: 45%), while those for the Q1st section depend strongly on eHW success (53%, 46%, and 37%).

Conclusions

- It's hard to tell if the Q1st approach worked or not. Q1st students who managed to do the eHW did better on exams (as compared to the non-Q1st section), but too many students in the Q1st section did not do the homework.
- Differences in exam performance might have more to do with the distribution of students in the sections rather than the treatments they received.
- Upperclass students might benefit from the Q1st approach more than 1st year students, perhaps because the approach requires more self-motivation.
- Math and science majors did not appear to benefit from the Q1st approach.
- Trying to compare two classes or two treatments is hard work! What appears to be structure can easily be interpreted as statistical fluctuation, making it difficult to extract meaning from the "noise" of comparing two sub-populations.

Future plans

We have lots of data that we have not yet had any opportunity to analyze. Perhaps some of it will yield additional insights into the strengths and weaknesses of the Q1st approach.

Acknowledgment

Work supported in part by NSF grant ROLE-0106771. However, any opinions, findings, conclusions, etc. herein are those of the authors and do not necessarily reflect the views of NSF.

Table III. Results when arbitrary sub-groups of the non-Q1st section (N = 236) are compared.

	A–Le Li–Z	1st, last 1/4 middle 1/2	1st 3/4 last 1/4	1st 1/4 last 3/4
Δ_{min} (%)	–21.6	–17.7	–23.1	–21.6
Δ_{max} (%)	18.0	10.4	24.5	15.8
Δ_{ave} (%)	0.0	–3.7	+2.2	–3.5
$(\Delta/\sigma)_{ave}$	0.0	–0.6	+0.3	–0.5
$\Delta(\sigma^2)$	4.4	1.5	10.1	0.9

Table IV. Results when different sub-groups of the two sections are compared.

	All students	1st year students	Upperclass students	Engineering majors	Comp.Sci. majors	Math/Sci. majors	Engineering 1st years
Δ_{min} (%)	–24.5	–24.0	–26.7	–30.7	–28.6	–52.3	–29.1
Δ_{max} (%)	15.5	13.3	35.3	22.3	48.6	17.8	25.3
Δ_{ave} (%)	–2.0	–5.6	+6.1	–1.4	+9.9	–12.4	–0.3
$(\Delta/\sigma)_{ave}$	–0.3	–0.7	+0.6	–0.1	+0.6	–0.9	0.0
$\Delta(\sigma^2)$	3.9	0.9	6.9	2.1	2.9	3.0	0.9
significant?	No	Maybe	Likely	Unlikely	Maybe	Likely	Unlikely