ACM-Teagle Collegium on Student Learning, Final Report (September 24, 2010)
The Use of Metacognitive Prompts in a Sampling-Distribution Exercise
Joy Jordan, Lawrence University

Project Summary

In introductory statistics, the idea of a sampling distribution is an essential building block, yet a conceptually difficult idea for students. It's a slippery idea that students blithely think they understand, yet struggle to explain. Bingo! This appeared a juicy class space to add metacognitive prompts. Specifically, I created a group-work activity with problems centered on the sampling distribution of the mean. These problems incrementally asked for more metacognition from the students. I also created a short assessment to give both before and after the activity.

Knowing the difficult nature of in-class research, I still formed a challenging research question: Can a group-work activity that engages students' metacognition improve students' understanding (based on test performance) about the sampling distribution of a sample average?

Project Context

I conducted my research all three terms of last academic year (2009-10). The first term was a pilot study in a 200-level statistics course that had 26 students. Based on their feedback and performance, I modified the activity and used that modified activity (and pre- and post-assessments) during the next two terms, both of which were 100-level introductory statistics courses (with 31 and 16 students, respectively). All of these courses had an additional computer lab meeting each week. I used one of those lab sessions for the students to complete the exercise in groups.

Of special note is my winter-term statistics class. This course is filled completely with psychology majors, who need the class as a pre-requisite for research methods. In preparation for my project, I collected information from my winter-term class in 2009 (e.g., common mistakes on exams). Then I collected the same information from my winter-term class in 2010—this is the group that received the "treatment" of the sampling-distribution exercise. Because the winter-term class, regardless of year, is filled with similar students—psychology majors—I found it my closest link to a controlled-experiment environment. That is, making comparisons between winter-term classes is reasonable (albeit not comparisons that can indicate causation).

Effect on Teaching Practice

In the classroom, my work with metacognition focused solely on the sampling-distribution exercise. Since this was my only intervention, and because I wanted to compare to the previous year, I specifically didn't mention metacognition in other parts of the class. (I thought this the cleanest design.) That said, I regularly spoke with students—in class or in office hours—about methods to gauge their understanding and about "reasonable checks" that can be applied to many problems. This has always been a part of how I teach. But I didn't name it (to the students) as "metacognition."

Evidence and Conclusions

Student Feedback

I received helpful student feedback during the fall-term pilot. Specifically, my original activity was too long, and the students didn't reach the last problems—the ones that were most metacognitively challenging. Furthermore, it was clear from the pre- and post-assessment results that one of my assessment questions was worded in a confusing way. The changes I then made (shortening the activity and re-wording the assessment) were incredibly important for a seamless use of the intervention during winter term.

During winter and spring term, the sampling-distribution exercise got positive reviews from the students. In terms of helpfulness towards understanding, 30% of students said it was very helpful, 64% said it was somewhat helpful, and only 6% said it was not helpful. In terms of providing new problem-solving tools, 20% of students said they learned multiple new strategies, 66% said they learned one new strategy, and 14% said they learned no new strategies.

The written comments from students were generally positive, yet focused on the helpfulness of practice problems and working in groups (not on the metacognitive process). A few of the comments were particularly interesting; they corroborated the commonly-found issue that students with weak understanding (as measured by class performance) often have the weakest metacognitive skills—yet this is the group who would most benefit from metacognition. For example, an A-student wrote, "I think writing an explanation in words was helpful to ensure that I actually knew what I was doing and not just plugging in numbers." Yet a C-student wrote that the *least helpful* part of the exercise was "the explanations of the processes behind it. It's very easy to stop thinking and just act on auto pilot, getting the right answer in the right way, but not thinking about it." (He thought the auto-pilot route was a good thing.)

Summary of Data Analysis

At the conclusion of my project there was much information to process and data to analyze. On the assessments, I not only posed content questions to students, but also asked for confidence judgments. That is, for each question, every student provided a confidence percentage in her answer. I also had performance data on both my winter-term 2009 students (control group) and my winter-term 2010 students (treatment group). Lastly, I returned to my winter-term students midway through the very next term—while they were in research methods class—and they again completed the assessment (content questions and confidence judgments).

After sifting through the copious amounts of data, these are the most interesting findings:

- Considering all 43 students (winter and spring term together) there was no significant increase—between post-activity and pre-activity—in percentage correct on the assessment. This assessment was four multiple-choice questions created by me; each question directly related to the topic of the sampling distribution of an average. Yet the assessment was not validated in any way. In fact, the non-validated assessment instrument limited my results (yet sped up the process of my work—that is, actually bringing this exercise to the classroom).

  One-Sample T: Assessment Percentage Increase (Paired Data)

  Test of mu = 0 vs not = 0 (note: normality condition is met)

Variable	*N*	*Mean*	*StDev*	*SE Mean*	*95% CI*	*T*	*P*
Test Percentage Increase	43	2.91	26.84	4.09	(-5.35, 11.17)	0.71	**0.482**

- Interestingly, there was a significant increase (between post-activity and pre-activity) in *average confidence* in answers on the assessment. Note this is on students' average confidence over four questions (not on individual questions).

  One-Sample T: Average Confidence Increase (Paired Data)

  Test of mu = 0 vs not = 0 (note: normality condition is met)

Variable	*N*	*Mean*	*StDev*	*SE Mean*	*95% CI*	*T*	*P*
Average Confidence Incre	43	9.60	8.90	1.36	(6.86, 12.35)	7.07	**0.000**

- I further investigated this increase in confidence, and found it most prevalent (significantly so) with the *bottom-performing-half of the class*, in terms of course grade. As already mentioned, there wasn't a significant increase in scores on the assessment. This held up when breaking the students in half by course grade. *So the bottom-performing-half of the class was, on average, more confident (significantly so) after the activity, yet they did not perform significantly better on the post-activity assessment.*

	Confidence Judgment Increase	
Course-Performance	*Bottom Half*	*Top Half*
Bottom Half	7 (31.82%)	15 (68.18%)
Top Half	15 (71.43%)	6 (28.57%)
Pearson Chi-Square: P-Value = **0.009** Fisher's exact test: P-Value = 0.015		

	Assessment Percentage Increase	
Course-Performance	*Bottom Half*	*Top Half*
Bottom Half	13 (59.09%)	9 (40.91%)
Top Half	9 (42.86%)	12(57.14%)
Pearson Chi-Square: P-Value = **0.287** Fisher's exact test: P-Value = 0.366		

- Comparison of 2010's winter term to 2009's winter term (my "like" groups of students—of whom only 2010 students completed the activity) showed a decrease—from 34% to 17%—in the percentage of students who missed a sampling-distribution T/F exam question (common to both classes). This decrease was not statistically significant, yet still seems noteworthy (i.e., as a teacher, from a practical standpoint, I'm glad for this fairly dramatic drop). Furthermore, anecdotally, it seemed the worded answers, on a different sampling-distribution question, were better in 2010.

	Missed T/F Exam Question?	
Year	*No*	*Yes*
2009	21 (65.63%)	11 (34.38%)
2010	24 (82.76%)	5 (17.24%)
Pearson Chi-Square: P-Value = **0.129** Fisher's exact test: P-Value = 0.155		

- For the 2010 winter-term students, I was able to give the same assessment midway through the spring term. The follow-up results—comparing data from the winter-term post-assessment to the spring-term assessment—show a (non-significant) decrease in percentage correct on the assessment and a (significant) decrease in average confidence, both of which are not surprising.

One-Sample T: Post Minus Post-Post Assessment Percentage (Paired Data)

Test of mu = 0 vs not = 0 (note: normality condition is roughly met; follow-up with Wilcoxon non-parametric test agrees with t-test results)

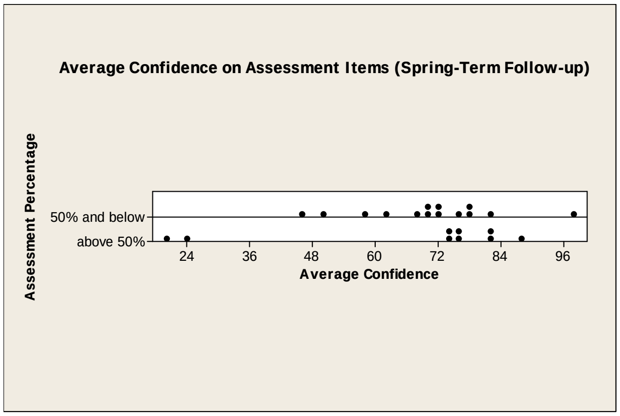
Variable	*N*	*Mean*	*StDev*	*SE Mean*	*95% CI*	*T*	*P*
Post Minus Post-Post	23	11.96	34.44	7.18	(-2.93, 26.85)	1.67	**0.110**

One-Sample T: Post Minus Post-Post Confidence (Paired Data)

Test of mu = 0 vs not = 0 (note: normality condition is met)

Variable	*N*	*Mean*	*StDev*	*SE Mean*	*95% CI*	*T*	*P*
Post Minus Post-Post	23	15.84	12.48	2.60	(10.44, 21.24)	6.09	**0.000**

- I looked more deeply at the average confidence in spring term. Even though there was a significant drop in confidence from winter term to spring term, there still appeared some overconfidence in the spring-term follow-up. Considering only the spring-term follow-up results, I separated the students who attained over 50% on the assessment and the students who received 50% or less on the assessment. There was no significant difference in average confidence for these students. In fact, the average confidence levels for these two groups were surprisingly similar. (Perhaps another indication of the lack of metacognition within lower-performing students.)



Variable	Assessment %	N	Mean	StDev	Min.	Q1	Median	Q3	Max.
Avg. Conf.	50% and below	14	69.82	13.32	45.00	61.25	70.63	77.81	97.50
	Above 50%	9	66.19	25.59	20.00	48.75	75.00	81.88	88.75

Future Implications

Overall, the sampling-distribution exercise was useful for the class. The students had interesting, deep discussions, and—for at least a brief period—seemed to have better conceptual knowledge of sampling distributions. That is, regardless of the statistical significance (or insignificance) of the results, as a teacher I found the exercise worthwhile.

What most surprised me was the repeated overconfidence in the lower-performing-half of the class. These are only initial results, but they seem meaningful. For example, when lower-performing students work in groups, do they perceive an increase in confidence even though objectively their understanding hasn't changed? This question—far afield from my initial research question—now has my interest piqued.

Annotated Bibliography

Dunlosky, J. and Metcalfe, J. (2008) *Metacognition*, Sage Publications

> This book was integral to my general understanding of metacognition (e.g., what are the different metacognitive judgments?). Furthermore, the many examples of actual research studies gave me ideas about my own project. In fact, the confidence judgments I included with my assessment (which became the most interesting piece of the results) was an idea I got from this text.

Schoenfeld, A.H. (1992), "Learning to Think Mathematically: Problem-Solving, Metacognition, and Sense-Making in Mathematics," in D. Grouws (Ed.) *Handbook for Research on Mathematics Teaching and Learning* (pp. 334—370), New York:MacMillan

> This article strongly guided my actual activity—that is, the questions I asked, the expectations I had. My exercise builds metacognitively from problem to problem, and for my last problem I used Schoenfeld's model of questioning: 1) What exactly are you doing? (Can you describe it precisely?), 2) Why are you doing it? (How does it fit into the solution?), and 3) How does it help you? (What will you do with the outcome when you obtain it?)