

Knowledge surveys: Students ability to self-assess

Jon Clauss¹ and Kevin Geedey²

Abstract: Active learning theory indicates the importance of helping students take control of their learning, monitor their understanding, and assess learning strategies. Knowledge surveys are intended to improve a student's ability to self-assess and cover the content and the full range of cognitive levels of a course. We explore an interdisciplinary data set of knowledge surveys and exams, asking whether student's ability to self-assess differs at different cognitive (Bloom) levels. Although self assessment accuracy was strongly related to the Bloom level of exam questions, this dependence was not simple. Students had the most difficulty self-assessing at intermediate Bloom levels.

Keywords: metacognition, self-efficacy, Bloom's taxonomy, student learning.

I. Knowledge surveys: Students ability to self-assess.

Recognizing the importance of metacognition, and the teaching techniques that support a metacognitive approach, how can a teacher interested in active learning pedagogies make students more aware of their own strengths and weaknesses as learners? How can we help students assess both their strengths and weaknesses and monitor their progress towards meeting course objectives? And, do students' self-assessment skills vary with the kind of task being assessed? In this article we will present our experience with a tool known as a knowledge survey that we used to address these questions.

We use the term *metacognition* to “[refer] to people’s abilities to predict their performances on various tasks... and to monitor their current levels of mastery and understanding” (National Research Council, 1999, p. 12). Of most interest to us is that, from studies dating back to the 1980’s involving written composition (Scardamalia, Bereiter, and Steinbach, 1984), reading comprehension (Palinscar and Brown, 1984) to mathematical problem solving (Schoenfeld, 1985), metacognitive approaches to teaching “have been shown to increase the degree to which students transfer their learning to new settings and events” (National Research Council, 1999, p. 12). In these and other subject areas, metacognitive strategies have also been shown to improve understanding of the material being taught (White and Frederiksen, 1998).

We present the results of a year-long exploration of the relationship between student self-assessment skills and different cognitive levels. In order to categorize the cognitive levels of our course objectives we have used Bloom’s taxonomy of educational objectives: *knowledge* (recall of facts), *comprehension* (e.g. describing in one’s own words), *application* (applying information, problem solving, etc.), *analysis* (analyzing underlying structure, identifying component parts), *synthesis* (combination of ideas, creation of a unique or original product) and *evaluation* (making decisions, resolving controversies in a broader context Bloom and Krathwohl, 1956, Krathwohl, 2002). Among the questions that we have explored, in this paper

¹ Department of Mathematics and Computer Science, Augustana College, 639 38th Street, Rock Island, IL 61210, jonclauss@augustana.edu.

² Department of Biology, Augustana College, 639 38th Street, Rock Island, IL 61210, kevingeedy@augustana.edu.

we address the following: “How accurately are students able to assess their abilities to perform at different Bloom levels?” We predicted that students would have more difficulty accurately self-assessing at higher Bloom levels. That is, as students confront more complex, less clearly-structured tasks like synthesis or evaluation, we conjectured that they would be less likely to “know what they know” and less likely to accurately self-assess.

Bloom’s taxonomy has been criticized both for its perceived emphasis on higher cognitive levels at the cost of factual knowledge (Booker, 2007), and the difficulty in unambiguously categorizing tasks to specific Bloom levels. However, Bloom’s taxonomy has the advantage of being both widely used and widely recognized both in higher education as a teaching and learning tool in disciplines as diverse as feminist philosophy (Cimitile, 2008) and biology (Crowe et al., 2008). In both biology and math, specific applications of Bloom’s taxonomy have been developed for use in discipline specific contexts (Crowe et al., 2008, Smith et al., 1986).

We used knowledge surveys to assess student’s level of confidence with course material. Early versions of knowledge surveys were described by Edward Nuhfer (Nuhfer, 1993) and subsequently developed by Nuhfer (1996), and Knipp (2001). Wirth and Perkins (2005) have used knowledge surveys as both an assessment and course design tool while Bowers, N., Brandon, M., and Hill, C. D. (2005) have explored their use as an indicator of student learning in an introductory biology course. The structure and possible uses of knowledge surveys are extensively described in *The Knowledge Survey: A Tool for All Reasons* (Nuhfer and Knipp, 2003).

Attempting to cover the entire breadth of course material and spanning all levels of the cognitive domain, a knowledge survey “consists of course learning objectives framed as questions that test mastery of particular objectives” (Nuhfer and Knipp, 2003, p59). The survey does not solicit content-based answers; rather, students are asked to rate their level of confidence in attempting to answer the survey questions, indicating whether they:

- (3) could answer this question with full confidence
- (2) could get most of the credit for this question
- (1) could get some partial credit for this question
- (0) could not begin to answer this question right now³

(Examples are available at <http://dmc.augustana.edu/KS/>.) As such, these surveys directly assess student confidence but do not, by themselves, assess the accuracy of a student’s self assessment. Metacognitive approaches “focus on sense-making, self-assessment, and reflection on what worked and what needs improving” (National Research Council, 1999, p.12). Metacognition, in this context, refers to a student’s ability to accurately self-assess and the ability to recognize areas of strength and weakness and act accordingly. Metacognition is related to, but is not necessarily the same as self-efficacy, or self perceived confidence in accomplishing a task. Self-efficacy may be an important factor in explaining effort and investment in a particular task, as individuals are willing to invest more in tasks they believe they will accomplish (Strecher, et al., 1986). Developing effective tools to assess self efficacy is important, since self efficacy may be a significant factor in both vocational reflection and in the desire to pursue more complex tasks (Ebert-May, et al., 1997, Baldwin et al., 1999). Knowledge surveys given at the beginning and end of a unit tend to show strong increases in student self-efficacy from pre to post instruction

³ Unlike surveys described in (E. Nuhfer and Knipp, 2003) and elsewhere, which provide a choice of three responses, we have chosen to provide four.

survey (Wirth and Perkins, 2005, Bowers et al., 2005). However, as Bowers et al. (2005) rightly point out, self-efficacy does not necessarily translate into actual achievement. Students may be confident in their abilities (high self-efficacy) but unable to perform required tasks (low metacognitive ability).

For our study, surveys were administered at the beginning of a course (or unit within the course; details below) and re-administered shortly before an exam covering the surveyed material. We have used our online course management system (moodle) to administer the surveys. This has several advantages, including: allowing students to take the survey at their convenience without sacrificing class time and to revisit it for purposes of review; allowing for easy analysis of the data using spreadsheet and statistical analysis software; and reducing the considerable environmental impact of administering large, multi-page surveys several times during the term. We have found knowledge surveys to be an effective way to help students assess and monitor their current abilities and levels of understanding, as well as document changes in student's level of confidence with course material.

II. Method.

A. Self-Assessment and the Cognitive Level of Exam Questions.

Our primary goal was to determine whether or not the accuracy of student self-assessment (metacognitive ability) via knowledge surveys depends upon the Bloom level of the task. We considered student knowledge survey responses from four courses across two disciplines (Biology: Aquatic Biology, Ecology, and Evolution taught by one instructor; Math: Linear Algebra taught by the other instructor). These courses are all junior-senior level offerings which are typically taken by students majoring in either biology or math respectively. Eighty one different students enrolled in one or more of the four courses that year. In order to compare the results of the knowledge surveys to exam scores, we focused on the surveys that were administered before an exam and the results of the exam covering the same material. Although our knowledge surveys were intended to be comprehensive, they were not intended to function as complete study guides. Therefore, not every question on a unit exam was covered in the survey, and not all survey material was included on the exam. We selected for analysis only paired survey and exam questions, i.e. cases where there was a match between a survey question and a subsequent exam question.

All survey and exam questions were assessed for Bloom level. Each faculty member scored their own surveys and exams and assessed Bloom levels for each question. In order to attempt to achieve inter-rater reliability and minimize the subjectivity inherent in this process, we also engaged a student research assistant (a math and biology double major who did not take any of the classes during the study period) to independently assign Bloom levels to the questions. We then compared the faculty and student assigned Bloom levels. Remarkably, there were relatively few conflicts between the lists. Discrepancies were often reconciled by discussion. In a few cases, the research assistant's lack of familiarity with specific course content hindered her ability to assess the questions. In other cases, faculty had worded a survey or exam question in a way that hindered the assistant's interpretation. In the latter case, such vaguely worded questions were dropped from the analysis. In a number of cases, small wording differences between the exam question and the survey question caused the exam and survey questions to have different Bloom levels. These questions were also dropped from further analysis.

Survey responses were scored on a numeric scale (0, 1, 2 or 3) ranging from 0 (I could not begin to answer this question right now) to 3 (I could answer this question with full confidence). Exam answers were graded following individual course policies. After the courses ended, individual exam answers were rescaled to the same 0-3 scale used for the surveys, i.e. a full credit answer got a 3, a no credit response got a 0, with partial credit responses being scaled between 0 and 3. This created a set of paired survey and exam responses, where we could compare self assessed performance on a question (from the survey) to the subsequent teacher assessed performance (from the exam). While we did not control for differences in grading policies between instructors in the course of this study, the patterns we report below are consistent for both instructors independently.

Because exams tended to contain many questions assessed at the same Bloom level, the data set contained a large amount of replicate measurements per student (i.e. replicate responses per Bloom level per student). This replication was, perhaps not surprisingly, highly uneven across Bloom levels. Lower level knowledge and comprehension questions were quite common, while higher level synthesis and evaluation questions were quite rare. In order to avoid pseudo replication and because of non-response by some students on some questions, we randomly selected at most a single paired survey/exam response per Bloom level per student. As a result, our final sample size was 309 survey/exam result pairs. Because of this uneven replication across Bloom levels, we combined levels 4-6 for analysis.

B. Self-Assessment Ability and Overall Exam Score.

Although not central to our research question, we looked for correlations between high self-assessment ability and overall exam scores. Other authors (Wirth and Perkins, 2005, Bowers et al., 2005) have looked for correlations between knowledge survey and exam score in either an aggregate (Wirth and Perkins 2005) or question by question (Bowers et al., 2005) manner. These studies have found relatively low values of r and/or non-significant results. In order to look at these same self-assessment patterns in a different way, we ask if students who are good self-assessors (accurately predict their own exam performance good or bad) tend to score better on exams. We generated an average absolute value of actual (exam score on question) minus predicted (knowledge survey score on question) for each student, across Bloom levels and performed a least squares regression of that variable on the overall earned exam score.

C. Gains in Self-Efficacy.

Our study was designed to assess metacognitive ability as indicated by the accuracy of student self assessment. However, since previous work on knowledge surveys also shows strong pre to post instruction gains in self-efficacy (Wirth and Perkins, 2005, Bowers et al., 2005), we also report on these trends from our study to facilitate comparison. In this case, however, the author's results from biology and math classes are not directly comparable due to procedural differences in administering pre and post unit surveys, and were therefore analyzed separately. In order to see if students gained self-efficacy over the course of the term, we compared student knowledge survey responses over time using paired T-tests. This approach let us look for self-efficacy gains on a survey question by survey question basis for each student. For this test we excluded students who, for whatever reason, did not take both surveys and we excluded any survey questions that appeared on a single survey only.

In the biology classes, students were given a pre-instruction knowledge survey that had to be filled out (online) by the second class meeting. These results were compared to post-instruction, just prior to unit exam, knowledge survey results. As an example, we analyzed data from 54 questions answered by 38 students in an Evolutionary Biology class (winter 08-09).

Due to the technical nature of the material (including technical vocabulary) students were not given a pre-instruction survey in the Math class. Instead they were given a series of cumulative post-instruction, pre-unit exam surveys. We therefore compare pre-unit exam one response to the student's responses on the same questions before unit exam two and similarly compare survey responses on the second pre-unit exam survey to the pre-final exam survey. We compared responses from pre-unit exam surveys one and two (72 questions, 30 students) and also from the pre-unit exam 2 survey with the pre-final survey, (160 questions, 31 students).

III. Results.

A. Self Assessment and Cognitive Level of Exam Questions.

There was a wide range of possible relationships between survey and exam response. We represented this variation by subtracting the predicted performance (from the survey) from the actual performance (from the exam) for each survey/exam result pair. The difference between survey and exam results was quite variable but were all centered around a median value of zero (actual and predicted values match) (figure 1). At lower Bloom levels, the values observed encompassed the entire possible range of results from +3 (student got a perfect score on the exam question, but had responded on the survey that they could not begin to answer that question) to -3 (student got no exam credit for the question, but had indicated full confidence on the survey). The data for each Bloom level appeared to be symmetrical around zero, which indicated that students were as likely to underestimate as they were to overestimate their exam success on the survey (see Figure 1).

Box and whisker plot of the difference between actual (exam score) and predicted (survey response) scores across Bloom levels. The midline of the box represents the median value. The upper and lower "lids" of the box represent the sills for the middle 50% of the distribution. The "whiskers" on the top and bottom of the box represent the upper and lower quartiles respectively. There is no box for Bloom level one, since the vast majority of the data fall at the median value of 0, which represents a perfect match between the survey and exam result.

We noted an apparent pattern in this data set. That is, students seemed to have a greater mismatch between survey and exam scores at some Bloom levels. To test this hypothesis, we used the absolute value of actual (exam) minus predicted (survey) data to generate a metric of similarity for each survey/exam result pair. This generated a scale from 0-3 for each survey/exam pair. A value of 0 represents a perfect match between survey and exam result, while a 3 represents cases where the student did either much better or much worse on the exam question than predicted by their survey result. Because of the low sample size at higher Bloom levels, we combined levels 4-6 into a single category for this analysis, resulting in sample sizes of 80 for each of Bloom levels 1 and 2, 78 for Bloom level 3 and 71 for Bloom levels 4-6. The value of this metric was strongly dependant on the Bloom level of the question (Kruskal-Wallis Analysis of variance, 3 df, Kruskal Wallis test statistic =35.768, $p < 0.0009$). We observed the largest

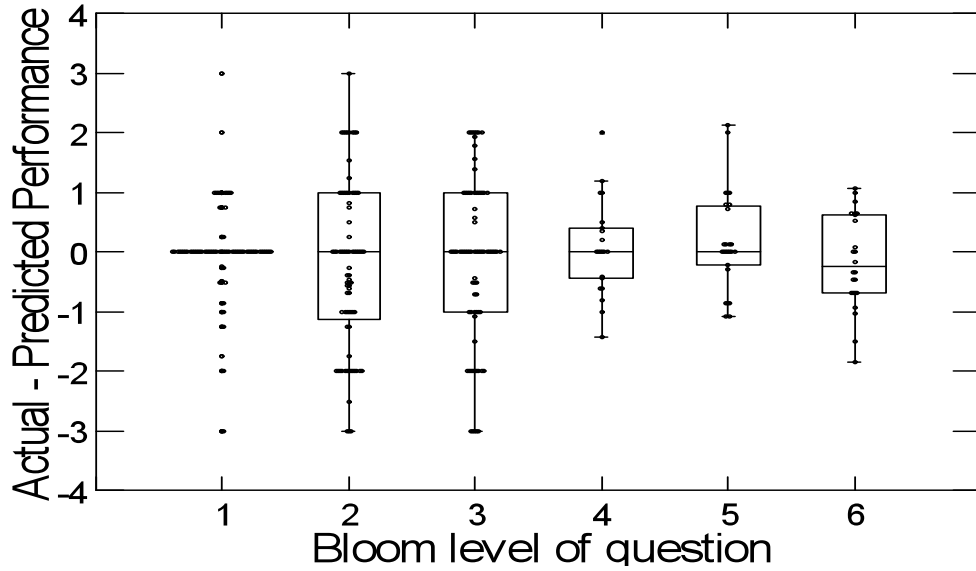


Figure 1. Actual minus predicted performance as a function of the Bloom level of exam questions.

mismatch between survey and exam results at Bloom levels 2 and 3, which are comprehension and application respectively (Figure 2).

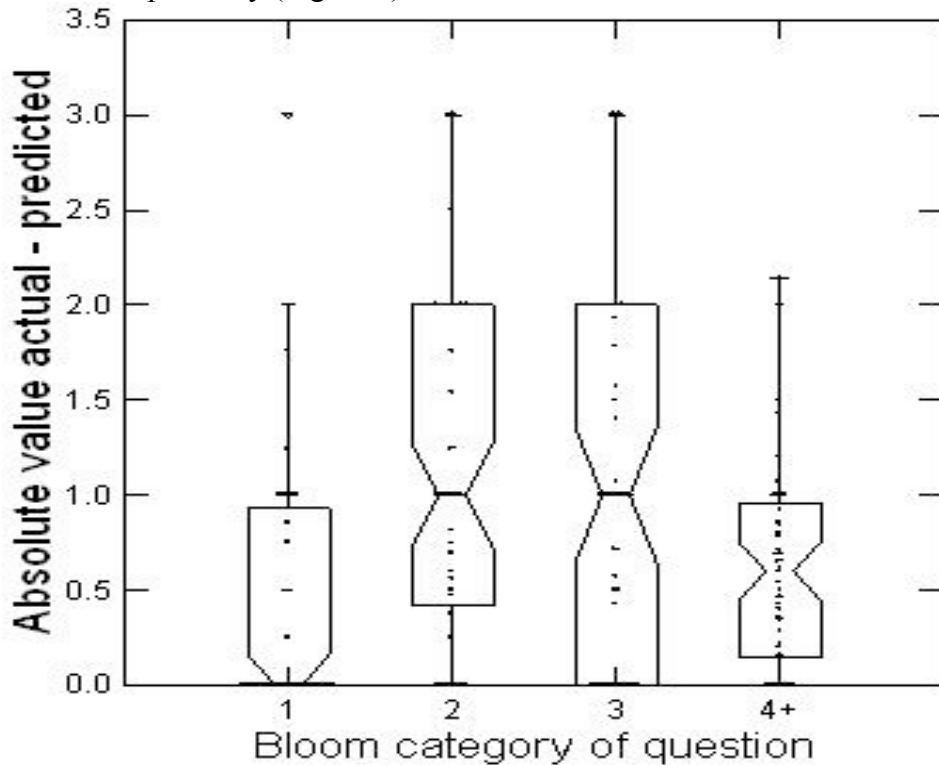


Figure 2. Absolute value of actual minus predicted performance as a function of the Bloom level of exam questions.

Box and whisker plot of the absolute value of the difference between actual (exam score) and predicted (survey response) scores across Bloom levels. Bloom levels 4-6 are combined into a single category. The midline of the box represents the median value. The upper and lower “lids” of the box represent the sills for the middle 50% of the distribution. The “whiskers” on the top and bottom of the box represent the upper and lower quartiles respectively. The “notch” on the box represents the confidence interval of the median.

B. Self-Assessment Ability and Overall Exam Score.

We assessed if students who were more accurate in their survey predictions (lower absolute value of actual minus predicted) tended to do better on the exams overall. We generated an average absolute value of actual minus predicted for each student, across Bloom levels (Figure 3), and performed a least squares regression of that variable on earned exam score (adjusted $r^2=0.143$, $f=9.98$, $p<0.003$). Given the low value of r^2 , this metric explains little of the variation in exam scores.

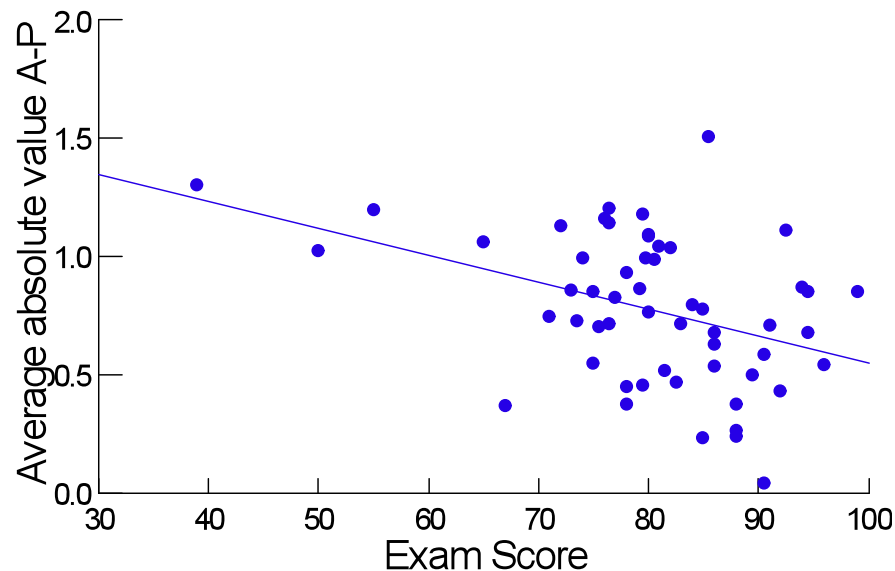


Figure 3. Relationship between self-assessment accuracy and exam score.

Mean absolute value of actual minus predicted for each student (averaged across Bloom levels) as a function of exam score. Line is estimated best linear function using Systat12 software.

C. Gains in Self-Efficacy.

In order to see if students gained self-efficacy over the course of the term, we compared student knowledge survey responses over time using paired T-tests. In the biology class, students showed dramatic increases in self-efficacy for every survey question from pre to post instruction. Median survey response values ranged from 0 to 2 for all questions in the pre-instruction survey and from 2 to 3 in the post instruction survey. Gains in self-efficacy were highly statistically significant with p values range from <0.003 to $<1 \times 10^{-20}$.

The Math class, which did not use pre-instruction surveys, necessarily showed a different pattern of self-efficacy gain. Comparing responses from pre-unit exam surveys one and two, median values ranged from 1-3 in pre-unit exam survey one from and 2-3 in pre-unit exam survey two. We found statistically significant ($p < 0.05$) increases in self-efficacy in 34 (or 47.2%) of the questions. Comparing pre-unit exam survey two with the pre-final survey, Median values in the pre-final exam survey ranged from 1.5 to 3 (with the vast majority being 3). In this contrast we found statistically significant ($p < 0.05$) increases in self-efficacy in 73 (or 45.6%) of the questions.

IV. Discussion.

We had naively predicted that a student's ability to self assess, as measured by correctly predicting their success on exams via knowledge survey questions, would decrease as the Bloom level of the question increased. That is, we predicted that students would certainly know whether or not they knew the definitions of terms and other lower Bloom level tasks, but we believed students would not "know what they knew" with more open-ended questions representing higher Bloom level tasks. Although our data do show a strong relationship between Bloom level of question and prediction accuracy, our data do not support the hypothesis that increasing Bloom level leads, in a simple way, to decreased student ability to self-assess. What we did find was that students were better self assessors at low (knowledge) and high Bloom levels (analysis, synthesis, evaluation), and poorer self assessors at intermediate levels (comprehension, application; see Figure 2). This pattern may be caused by several factors, some rooted in learning and others in teaching practice, none of which were directly addressed in this study. Comprehension and application tasks may be poorly understood or poorly diagnosed by students; that is, students may not understand how an application question differs from one that simply demands a straightforward recall of facts. Reduced self-assessment ability at these Bloom levels might also be due to poor modeling or practice of these tasks in class, or uneven evaluation of these tasks by the instructor. We may, subconsciously, grade these categories harder since we might view higher level Bloom questions as difficult and cut students some slack with partial credit, while at the same time feeling that students "really ought to know" comprehension and application type questions. However, these hypotheses run afoul of the fact that our data show that students were equally likely to underestimate success or overestimate at these Bloom levels (figure 1). If we were just strict graders on comprehension and application questions, for example, students might have been expected to overestimate their success.

Our results on the overall correlation between self-assessment and exam scores are in line with other published results (Bowers et al., 2005). We found no evidence that variation in self-assessment ability, as assessed by our knowledge surveys, explains a significant amount of the variation in exam scores. Others have argued that knowledge surveys, by themselves, may be a poor indicator of learning and may be a poor stand-alone assessment (Bowers et al., 2005). We do not feel that our data make the case for or against knowledge surveys as a direct learning assessment tool. However, we do not regard the current literature as definitive. If the knowledge surveys serve as a prompt to students that they have more work to do before they take an exam, then we would not expect a strong correlation between survey results and exam results. Students with excellent self-assessment skills who do poorly on the survey may end up doing better on the exam because the survey served as a wake-up call, just as intended. On the other hand, students may consider the survey as one more task demanded of them which, once gotten out of the way,

can be forgotten. If students are actually using the surveys as a learning tool, then there is every reason to expect that student mastery of course material will change between the self assessment and the faculty assessment and therefore raw survey results ought not explain much variation in exam scores. In short, a simple correlation with exam scores misses the whole goal of the surveys as a metacognitive tool for students.

Our results are also in line with other published work on the gains that students experience in self-efficacy as indicated in pre vs. post instruction knowledge surveys (Bowers et al., 2005). In the biology class, students demonstrated dramatic increases in their confidence following instruction for all survey questions. This is not surprising since the college level material covered in these classes will be unfamiliar to students at the beginning of class. The results from the math class are different, but also show significant increases in self confidence over the term. Because the math survey were all post instruction, it was much harder to demonstrate significant increases in self-efficacy. If the median survey response to question is high on the first survey, then there is nowhere to go to demonstrate increased confidence. Under these circumstances it is remarkable that a large number of these questions showed significant increases in self confidence. This suggests that students in this class continue to build on their knowledge base and continue to gain confidence as course concepts are continually applied.

Our data raise an important question that this study does not answer: How do students actually use the surveys? Anecdotally, students report that they like having access to the surveys, and that they regard them as useful study guides. Indeed, in subsequent courses we have found that students who have used them often request a knowledge survey, especially as they prepare for an exam, if one has not been provided to the class. It is clear that the surveys foster metacognition—students’ “abilities to predict their performances on various tasks... and monitor their current levels of mastery and understanding” (National Research Council, 1999, p. 12). What, however, does the act of filling out the survey give a student? Do they really help identify problem areas? Do students follow-up on the self knowledge they gain from the survey? In other words, do the surveys help students learn how to learn? Furthermore, does this particular act of metacognition “increase the degree to which students transfer their learning to new settings and events” (National Research Council, 1999, p. 12)? To date, we know of no studies that address these questions or that document the effect of knowledge surveys on student learning.

While we have no documented evidence of the effects of knowledge surveys on student learning, we have, nonetheless, found them to be effective and useful tools for a number of reasons. For the students, the simple act of asking them to assess their abilities does encourage, if not force, metacognition on their part. It also allows the students to be exposed to a much larger collection of questions, at all Bloom levels, than could be presented on any single assignment, project or exam. We also believe that it has increased interaction both between students and faculty and among students since once they have identified specific areas of weakness they are more likely to seek help in those areas either from the teacher or from their peers (combating the “I don’t even know what question to ask” syndrome). Students frequently tell us that they like using the surveys. It has been reported that they are “a good reality check” and that “they tell me what I need to work on.” Again, anecdotally, we believe that the surveys help promote responsible student behavior by providing students with a guide to the kinds of questions we expect them to be able to answer in a unit and throughout the course. It is true that initially students are often daunted by what appears to be “busy work” in filling out the surveys. Because of that, in order to help provide motivation we have attached nominal credit to survey completion throughout the course. Eventually, though, we found that when we were late in posting a

knowledge survey, students would ask for the survey to be posted as soon as possible. Overall it has been our impression that students perceive the surveys as a useful study guide.

For the instructors, one of the most obvious and significant benefits of using knowledge surveys has been that, in constructing them, we have been forced to pay more attention to the frequency at which we challenge students with tasks of various Bloom levels. We have had to design classroom experiences that integrate higher Bloom level tasks into the fabric of our classes with intentionality. We have also had to confront the challenge of designing tasks that assess students' mastery of higher level tasks in a meaningful way. For one of us, the act of constructing the survey forced us to recognize the fact that several higher level skills that we would have reported as important in our course were actually rarely assessed. As in the study by Wirth and Perkins (2005) mentioned above we have found them to be a very helpful course design tool.

There are also, of course, mechanical challenges to creating, administering and processing a knowledge survey. It is a daunting task to create one for a class the first time, particularly if it is a relatively new course for the instructor. As mentioned above, recognizing the lack of activities that address higher level tasks often requires rethinking the fundamental design of the course or at least major aspects of it. On a lighter note, we have found out the hard way the importance of having a stable, durable medium for administering and processing the surveys. While there are obvious advantages to using course management software when available and feasible, it can pose problems in an environment where the electronic tools, or the platforms for delivering them, are less stable and durable. We have found it to be a good idea to plan well ahead and work closely with our Instructional Technology department in designing and creating the surveys for use online.

Acknowledgements

The authors would like to thank the reviewers for the many helpful suggestions to improve the paper. In addition, we would like to thank Dr. Stacey Rodman and Dr. Dara Wegman-Geedey for their insightful comments as well as Courtney Cook, who was supported by an Augustana Alumni Departmental Assistantship, for her assistance during the study.

References

- Bloom, B. S., and Krathwohl, D. R. (1956). *Taxonomy of educational objectives: The classification of educational goals / (1st ed.)*. New York: Longmans, Green.
- Booker, M. (2007). A roof without walls: Benjamin Bloom's taxonomy and the misdirection of american education. *Academic Questions*, 20(4), 347-355.
- Bowers, N., Brandon, M., and Hill, C. D. (2005). The use of a knowledge survey as an indicator of student learning in an introductory biology course. *Cell Biology Education*, 4, 311-322
- Cimitile, M. (2008). The use of bloom's taxonomy in feminist philosophy. *Teaching Philosophy*, 31(4), 297-310.

Clauss, J. and Geedey, K.

Crowe, A., Dirks, C., and Wenderoth, M.P. (2008). Biology in Bloom: Implementing Bloom's Taxonomy to Enhance Student Learning in Biology. *CBE-Life Sciences Education*, 7, 368-381.

Knipp, D. (Spring, 2001). Knowledge surveys: What do students bring to and take from a class? *United States Air Force Academy Educator, Spring, 2001*. Retrieved February 23, 2010 http://www.isu.edu/ctl/facultydev/KnowS_files/KnippUSAFA/KSKNIPPUSAFA.html

Krathwohl, D. R. (2002). A revision of bloom's taxonomy: An overview. *Theory into Practice*, 41(4), 212.

National Research Council. (1999). In Bransford J., Brown A. L., Cocking R. R. and Committee on Developments in the Science of Learning (Eds.), *How people learn : Brain, mind, experience, and school*. Washington, D.C: National Academy Press.

Nuhfer, E. B. (1993). Bottom-line disclosure and assessment. *Teaching Professor*, 7(7), 8.

Nuhfer, E. B. (1996). The place of formative evaluations in assessment and ways to reap their benefits. *Journal of Geoscience Education*, 44(4), 385-385-394.

Nuhfer, E., and Knipp, D. (2003). The knowledge survey: A tool for all reasons. *To Improve the Academy*, 21, 59.

Palinscar, A. S., and Brown, A. L. (1984). Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities. *Cognition and Instruction*, 1(2), 117.

Scardamalia, M., Bereiter, C., and Steinbach, R. (1984). Teachability of reflective processes in written composition. *Cognitive Science*, 8(2), 173-190.

Schoenfeld, A. H. (1985). *Mathematical problem solving*. Orlando, Fla: Academic Press.

Smith, G., Wood, L., Coupland, M., Stephenson, B., Crawford, K., and Ball, G. (1996). Constructing mathematical examinations to assess a range of knowledge and skills. *International Journal of Mathematical Education in Science and Technology*, 27(1), 65.

White, B. Y., and Frederiksen, J. R. (1998). Inquiry, modeling, and metacognition: Making science accessible to all students. *Cognition and Instruction*, 16(1), 3.

Wirth, K. R., and Perkins, D. (2005). Knowledge surveys: An indispensable course design and assessment tool. *Innovations in the Scholarship of Teaching and Learning*. Retrieved February 23, 2010 <http://www.macalester.edu/geology/wirth/WirthPerkinsKS.pdf>