*Journal of the Scholarship of Teaching and Learning*, Vol. 15, No. 1, February 2015, pp. 20 - 37.
doi: 10.14434/josotl.v15i1.12761

# Adding confidence to knowledge

**Ludwika Aniela Goodson[1], Don Slater[2], and Yvonne Zubovic[3]**

*Abstract: A "knowledge survey" and a formative evaluation process led to major changes in an instructor's course and teaching methods over a 5-year period. Design of the survey incorporated several innovations, including: a. using "confidence survey" rather than "knowledge survey" as the title; b. completing an instructional task analysis with an instructional designer's perspective of the Gagné framework rather than Bloom's taxonomy; and c. using a rating scale based on established measurement practices for self-efficacy surveys. Results included increased instructor-student interactions; gains in confidence scores from before to after study of course units; high value of the survey for students; changes in grades and confidence scores across teaching methods; and advancement of Chickering and Gamson's principles of good practice in undergraduate education.*

*Keywords: knowledge survey, confidence, self-efficacy, active learning, assessment, pedagogy, teaching methods, instructional task analysis, cognitive task analysis, formative assessment, formative evaluation, undergraduate education*

## Introduction

"Love the confidence surveys; more classes need them…"—Student Comment

### The Classroom Problem

Most students in the senior-level course planned to become construction site managers, a career path in which the theme for success was "If you don't know the dirt, you'll lose your shirt." Yet, they could barely care about learning how to use soils as construction material. With unwitting complicity, the instructor's own previous instructors had contributed to this problem with their use of traditional lectures and homework. The instructor had used these same methods for twenty years of classroom teaching until the day he decided to leverage his experience in teaching, over forty years in Civil Engineering, and five academic degrees toward making changes. He replaced homework with a presentation project and written report in order to require application of knowledge to actual construction and engineering realities rather than artificial routine problems. But students failed to write articulate reports. What he called a "colossal disaster" became an opportunity as he walked across campus to an instructional designer's "Active Learning" workshop. At the conclusion of this workshop, the instructor made an appointment to discuss further what was happening in his classroom. In subsequent meetings,

[1] Instructional Consultant and Designer, Center for the Enhancement of Learning and Teaching, IPFW, goodsonl@ipfw.edu
[2] Professor, Department of Construction Management and Civil Engineering, Georgia Southern University, fiverings100@gmail.com
[3] Associate Professor, Department of Mathematics, IPFW, zubovic@ipfw.edu

collaborative decisions would soon take his students to destinations never before seen in his course.

With only a couple of months before the semester began, the design and management of group work, simulations, and guided inquiry would wait for later semesters. Although a course is more than a set of learning objectives, the explicit learning outcomes expected for students would help in considering what active learning strategies would align well with those outcomes and be practical for implementing in the coming semester. The instructional designer began with document analysis. She found that the syllabus listed course goals and general statements of learning outcomes and the lecture outlines showed the main topics—suitable for a lecture guide, but indefinite about explicit outcomes. However, she found that the test items required explicit application of specific concepts and principles to realistic engineering challenges. These types of items revealed that the instructor expected deeper levels of learning than were listed in the syllabus and lecture outlines. The instructor confirmed that the test items signaled best the expectations for learning, yet students would not see those expectations until the test administration time arrived.

## Connecting Concepts

Before selecting a strategy, the instructor and instructional designer began reviewing several areas of research that seemed relevant to student engagement and motivation. These areas included the rationale for completing an instructional task analysis (Feldon & Stowe, 2009; Smith & Ragan, 2005); use of knowledge surveys in formative assessment (Nuhfer & Knipp, 2003; Wirth & Perkins, 2005); how formative evaluation can help improve instruction (Shepard, 2005, Dunn & Mulvenon, 2009; Herman, 2013); and Keller's ARCS motivation model with its elements of attention, confidence, relevance, and satisfaction (Keller, 2000, 2010).

***Instructional Task Analysis:*** To reframe expectations for students into explicit statements of outcomes, the surest method would be to proceed with an instructional task analysis. Similar to a cognitive task analysis, this process would produce more effective instruction than other ways of identifying content (Clark, Feldon, van Merriënboer, Yates, & Early, 2007; Feldon & Stowe, 2009; Gagné, 1974, 2000; Gagné, Wager, Golas, & Keller, 2005; Smith & Ragan, 2005; Jonassen, Tessmer, & Hannum, 1999; Smith & Ragan, 2005). However, conducting such analysis would require intensive instructor effort and commitment, collaboration in collecting and analyzing knowledge through interviews and documents, and cycles of review and revision until the instructor could clearly map learning outcomes to student performance and back to content.

***Knowledge Surveys:*** The value of the knowledge survey was highlighted when it was cited as one of the best practices reported in the 2001 National Survey of Student Learning (Nuhfer & Knipp, 2003). Such a survey adds value to assessment by providing greater reliability compared to faculty-made tests such as quizzes, midterm exams, and final exams (Nuhfer, n.d.; Nuhfer & Knipp, 2006). It makes a valuable addition to multiple measures, triangulation of data, and ongoing assessment with greater validity, as well as reliability (Sawchuk, 2013; Strayhorn, 2006; Yeasmin & Rahman, 2012). It helps students improve their own learning by engaging them in self-evaluation and self-monitoring (Nilson, 2013; Panadero & Alonso-Tapia, 2014).

Requirements for producing a knowledge survey appeared to be dependent on completing an instructional task analysis. Those requirements include: (1) items that cover all learning outcomes and course content in the same sequence as presented in the course; (2) major themes in the course; (3) questions like those a student might encounter on a quiz, such as "How do you perform Gaussian elimination?" (Clauss, 2006), or task statements, such as "Make a contour plot that shows the locus of points with a single value of the function" (Frary, 2009); and (4) a way for students to rate their confidence to answer a question or perform a task both before and after instruction. The instructor could look for patterns in survey results and investigate possible changes needed in the course and students could see areas on which to focus their study time.

Nuhfer and Knipp (2003) had provided a blueprint for creating and using a knowledge survey and explained how it could advance Chickering and Gamson's (1987) seven principles for good practice in undergraduate education. This explanation carried high promise for more student-faculty contact, cooperation among students, active learning, prompt feedback, time on task, and addressing diverse ways of learning. Wirth and Perkins (2005) gave an account of how a survey provides full disclosure of course content to students before instruction, the value of a survey as a learning guide, and how the survey process helps students develop self-assessment skills. Wirth and Perkins, like Nuhfer and Knipp (2003), elaborated on how the process of constructing a survey and analysis of data can lead to improved course design and teaching, and their data and analysis showed that knowledge survey scores provided reliable and meaningful measures of learning gains.

***Formative Evaluation:*** Formative evaluation involves assessments during instruction that help instructors make changes before the end of a course (Herman, 2013; Johnson, 2009; Nuhfer & Knipp, 2003; Wirth & Perkins, 2005), and they have a positive impact on student achievement, a practice that itself engages more active learning (Herman, 2013). Whether during or at the end of a course, student feedback helps to inform changes to make in the future (Gilpin, 2013; Nuhfer & Knipp, 2003; Wirth & Perkins, 2005). A knowledge survey is one example of formative assessment.

***Motivation Model:*** Important relationships unfolded during review of Keller's ARCS model (Keller, 2000; Keller, 2010): (1) a knowledge survey at the beginning of the course could gain student *attention;* (2) it would focus on student *confidence*, and furthermore, the pre- and post-survey process and use of the survey as a study guide could support confidence as students focus on what to study; (3) with survey items reflecting all of the course content and organized into thematic units, students could see the *relevance* of what they study; and (4) completing the survey with gains in confidence could allow students to experience *satisfaction* in their accomplishments.

## Survey Design

Notwithstanding the *knowledge survey* nomenclature in published studies, the survey was re-titled as a *confidence survey*. Students would see "confidence" in the title and be less likely to think of the survey as a knowledge test. Besides, the survey would be gathering ratings of confidence, not answers about knowledge levels.

### Instructional Task Analysis

The instructor listed the major topics for the course, in the sequence taught, and learning outcomes for each topic. Weekly review meetings identified ambiguities or gaps in conceptual and procedural content. Between meetings, the instructor filled in gaps or clarified expected learner performance. These cycles of review and revision are frequently needed with experts because they typically possess the desirable professional quality of "automated, unconscious knowledge" (Clark, Feldon, Yates, & Early, 2007, p. 590), yet this same quality often leads them to understate the conceptual and/or procedural knowledge in a content domain (Clauss & Geedey, 2010; Feldon & Stowe, 2009; Frary, 2009; Merrill, 2009).

Answers to the instructional designer's questions produced explicit statements of learning tasks. For example, for the topic of *Soils, Investigation, Testing and Classification*, questions included: "Would students need to define soils? What kinds of testing would they need to do? What are the possible classifications of soil?" The resulting statements of learning tasks used performance verbs to signal levels of learning as in the Gagné taxonomy of learning (Gagné, 1977; Gagné, Wager, Golas, & Keller, 2005; Smith & Ragan, 2005), such as: (1) "Describe the mechanical analysis test" (verbal information), (2) "Identify two problem soils" (concept classification), and (3) "Plan and execute a preliminary site investigation" (rule using). These learning tasks then became items in the survey.

This approach to analysis differs from Nuhfer's model, which persistently uses Bloom's taxonomy (Bloom, 1956) with varied methods of application, and varied results (Bell & Volckman, 2011; Bowers, Brandon, & Hill, 2005; Clauss & Geedey, 2010; Marshall & Nuhfer, 2013; Nuhfer & Knipp, 2003; Wirth & Perkins, 2005). The alternative Gagné taxonomy worked well with student outcomes in this course because it makes distinctions between verbal information, attitudes, and psychomotor skills, as well as the hierarchy of intellectual skills defined as concept learning, rule using, and problem solving.

Through the analysis process, the instructor had identified eleven major topic categories and listed explicit learning tasks for the first nine. The number of learning tasks within each ranged from as few as three to as many as forty-one. Because of pressing deadlines to copy the survey for the first day of class, the instructor began analysis of the remaining two topics later in the semester and completed it before the next semester began, at which time he also updated the corresponding survey. Thereafter, he gathered formative evaluation data each semester and continually reviewed and refined the analysis and parallel confidence survey items. The instructional task analysis became his course map for teaching.

**Response Scale, Terms, and Layout**

On any scale, numbers and their anchoring labels should ascend in the same pattern, from low to high (Bandura, 2006; Kasunic, 2005) such as "cannot do at all" to "highly certain can do" (Bandura, 2006, p. 312), and scales should "measure what they purport to measure" (p. 318). Bandura's scale examples tend to go from 0% to 100%. "Readability" is the critical feature, not so much the number of numbers; that is, the rater needs to see at a glance the construct and how to indicate the strength of belief depicted in the scale, whether as a checklist or a ratings continuum (Bandura, 2006; Kasunic, 2005; Tullis & Dumas, 2009).

Yet, most knowledge surveys have used complex double-barreled directions with mixed constructs in a multiple choice format (Bowers, Brandon, & Hill, 2005; Nuhfer & Knipp, 2003; Wirth & Perkins, 2005). For example, Nuhfer and Knipp's middle choice on a 3-point scale could allow a rating to show 50% confidence in answering a question or in knowing where to find the information needed within 20 or 30 minutes. But to "know" and to "find" are different constructs. Several subsequent studies followed the same complex multiple-choice approach to rating confidence (Bell & Volckman, 2011; Clauss & Geedey, 2010; Fleisher, 2008; Frary, 2009; Price & Randall, 2008).

In contrast, the aim in this project would be to assess ONE construct, namely *confidence to do a task*, and to make the directions and scale easy to interpret by following the standards for self-efficacy rating scales (Bandura, 2006; Kasunic, 2005). This approach would avoid unnecessary cognitive load and split attention (Sweller & Chandler, 1994) because students would not need to focus on interpreting a complex scale while also trying to estimate confidence. The survey question for each topic focused student attention: "How much confidence do you have in your ability to accomplish each of the following objectives and tasks?" and the scale ranged from 1 (no confidence at all) to 5 (complete confidence). This scale was used in all semesters. Students were asked to enter a *score* rather than a *rating*, and instead of asking about confidence before and after *instruction*, students marked scores before and after *study*. Tasks were listed under the directions and scales.

Students could add up the total score for all tasks in a topic category before study, after study, and compute differences in their before-study and after-study scores. The bottom of the form allowed room for students to make comments. Other areas allowed room for ancillary information, such as topic name, student's name, and survey date. Figure 1 illustrates a sample survey for one of the topic categories. Regardless of number of tasks, this same format was used for all topics in the confidence surveys in all semesters.

| **Score Before Study** | **Task Number** | **How much confidence do you have in your ability to accomplish each of the following objectives and tasks?**<br>**5 = complete confidence**<br>**4 = moderate confidence**<br>**3 = some confidence**<br>**2 = a little confidence**<br>**1 = no confidence at all** | **Score After Study** | **Before /After Score Difference** |
|---|---|---|---|---|
| | 1 | List two duties of the construction surveyor. | | |
| | 2 | Sketch two ways to mark grade stakes. | | |
| | 3 | Use a hand level, Jacob staff, and folding rule to conduct a vertical survey. | | |
| | 4 | Use a tape to conduct a horizontal survey. | | |
| **Total Score** ________ | | **Please make any comments here.** | **Total Score** ______ | **Total Score Difference** _________ |

*Figure 1.* Survey for the topic of "Layout and Grade Staking"

## Formative Evaluation Measures

Over the five-year period of this project, the instructor also added a *formative evaluation* questionnaire, a *pre-evaluation* questionnaire, and a *pretest*. Altogether these measures, along with the *confidence survey*, contributed to the practice known as *formative evaluation* or *formative assessment*.

***End-of-Course Formative Evaluation:*** The end-of-course formative evaluation questionnaire collected information anonymously about the students' learning experiences. The first question asked: "How valuable were the confidence surveys for objectives and tasks?" with a response scale from 1 (no value) to 5 (very valuable). Other questions inquired about what students thought were the most and least important things they learned and what changes they would make.

***Pre-Evaluation:*** After the first semester of using the confidence survey, the instructor reframed the same questions from the end-of-course formative evaluation to ask students at the beginning of the course what they *anticipated* happening.

***Pretest:*** In the next year, the instructor copied the survey items and put them in a pretest format, thus keeping tight alignment with the survey. That is, the instructor took the same questions from the survey and simply added space to write answers under each question, added "Pretest" as the title, and provided directions for how to complete the pretest.

### Classroom Procedures

The instructor duplicated and handed out the pre-evaluation, pretest, and survey at the beginning of the course, and the formative evaluation questionnaire at the end of the course. Survey directions explained expectations for students to use the survey throughout the semester, and the survey was posted online for students to review, print, and use as a study guide. Before completing a unit of study, students entered their before-study confidence scores. After completing a unit, they entered their after-study scores and calculated differences in their scores before and after study of the unit. Students handed in the survey score sheet each time they completed it, but could retrieve the survey at any time.

Figure 2 shows the learning methods and formative assessments for five years, eleven semesters, eighteen sections, and 428 students. The confidence survey and formative evaluation were introduced in Year 3 when the instructor began his collaboration with the instructional designer. The pre-evaluation and pretest were introduced in Year 4. Years 3, 4, and 5 each had more teamwork and more quizzes. Year 5 had a process-oriented guided inquiry learning (POGIL) project. Coding for semesters shows number of students in different classes ($n_A$, $n_B$, $n_C$). Updates noted in Figure 2 occurred because of the instructor's ongoing analysis of formative evaluation measures and student performance.

## Data Analysis

Course data included survey scores, quiz points, test points, student ratings of the value of the survey, points on projects, total points earned toward final grades, and final grades.

***Students' Evaluation of the Confidence Survey:*** The instructor reviewed students' comments about the survey, observed students using the survey as a study guide, and tabulated value ratings. To determine if there were differences in students' value ratings for the survey across teaching methods, pairwise comparisons were made with the Kruskal-Wallis test, which is a nonparametric equivalent for a one-way analysis of variance (ANOVA).

***Confidence Scores:*** For confidence surveys in Years 3, 4, and 5, the instructor entered before-study and after-study confidence scores in Excel. The survey data pool for statistical analysis included items with slight variations in wording across surveys, for example, a change from *soil* to *soil mass,* and excluded topics with missing data in any teaching year and any item that had not been used in every survey. As a result, the data pool provided 83 learning tasks for analysis, about 86% of the whole set used in the surveys.

As in previous knowledge survey studies, data analysis compared the average before-study and after-study confidence scores per student per topic. Years 1 through 5 were analyzed separately. Within each year, analysis was completed for all tasks using paired t-tests to test whether the average total score after study exceeded the average total before study. A one-way ANOVA compared changes in the size of gains in confidence scores from before-study to after-study across teaching years. A Kruskal-Wallis test was used to examine the total after-study final confidence scores across the five years.

| |
|---|
| **Year 1 "Old Methods"** |
| ♦Lecture ♦Homework 1, 2, 3 ♦ Quizzes 1-5 ♦ Exams 1, 2 ♦ Extra Credit Project, 1-3 points |
| Total 26: Semester 1 $n_A$=26 |
| **Year 2 "Written Report, Presentations, Peer Review"** |
| ♦Lecture Updated ♦Semester Project (replaced homework sets) ♦Written Report<br>♦Presentations 1, 2 ♦Peer Review ♦Quizzes 1-5 ♦Exams |
| Total 81: Semester 1 $n_B$=15 $n_C$=21; Semester 2 $n_A$=17 $n_B$=28 |
| **Year 3 "1st Confidence Survey, Team Building, Formative Evaluation"** |
| ♦Lecture Updated ♦Confidence Survey ♦Lab Report ♦Team Building Activity<br>♦Four More Quizzes♦Formative Evaluation |
| Total 81: Semester 1 $n_A$=22 $n_B$=21; Semester 2 $n_A$=22 $n_B$=16 |
| **Year 4 "Completed Survey, Pretest, Pre-Evaluation, Formative Evaluation"** |
| ♦Lecture Updated♦Pre-Evaluation (Aligned with Formative Evaluation)<br>♦Pretest (Aligned with Survey) ♦Confidence Survey (Updated) ♦Team Project<br>♦Two More Quizzes♦"Bonus" Quiz♦Three Presentations with Peer Review<br>♦Formative Evaluation (Updated) |
| Total 105: Semester 1 $n_A$=17 $n_B$=24; Semester 2 $n_A$=15 $n_B$=*17*; Semester 3 $n_A$=32 |
| **Year 5 "Addition of POGIL"** |
| ♦Lecture Updated♦Pre-Evaluation (Aligned with Formative Evaluation, Updated)<br>♦Pretest (Aligned with Survey, Updated)♦Confidence Survey (Updated)<br>♦Kept Teamwork♦Two More Quizzes♦Three Presentations with Peer Review<br>♦Process Oriented Guided Inquiry Learning Team Project<br>♦Student Reviews of Team Projects♦Formative Evaluation (Updated) |
| Total 135: Semester 1 $n_A$=28 $n_B$=32; Semester 2 $n_A$=44; Semester 3 $n_B$=31 |

*Figure 2*. Learning methods and formative assessments

***Grades and Points Earned in the Course:*** The student-topic-average was matched to instructor records of final grading points and letter grades. (Final grading points=numerical total of points earned in the semester. Letter grades=A, B, C, D, or F assigned based on a percent of total possible points, that is, 90-100%=A, 80-89=B, 70-79=C, 60-69=D, less than 60=F.) The number of letter grades earned in each year was counted to examine differences in the grade distribution across teaching methods. Final points earned toward the assigned grades were compared using the Kruskal-Wallis test. To examine the correlation between final points earned in the course and after-study confidence scores, correlation coefficients were computed for each task separately.

***Pre-Evaluation and Formative Evaluation:*** The instructor reviewed students' comments on the pre-evaluation and formative evaluation questions. No statistical analysis was applied to those evaluations.

## Results

### Students' Evaluation of the Confidence Survey

Students indicated they gained *relevance* and *satisfaction* with comments such as "…Showed me how much I really learned;" "Gives you a degree of accomplishment;" "…showed a difference of what I thought I knew versus what I learned." Correlations of after-study confidence scores with final grading scores would be expected to further increase student satisfaction. Students also gave high value ratings for the confidence survey as shown in Figure 3. Other charts generated for each assigned grade showed a similar high-value pattern.

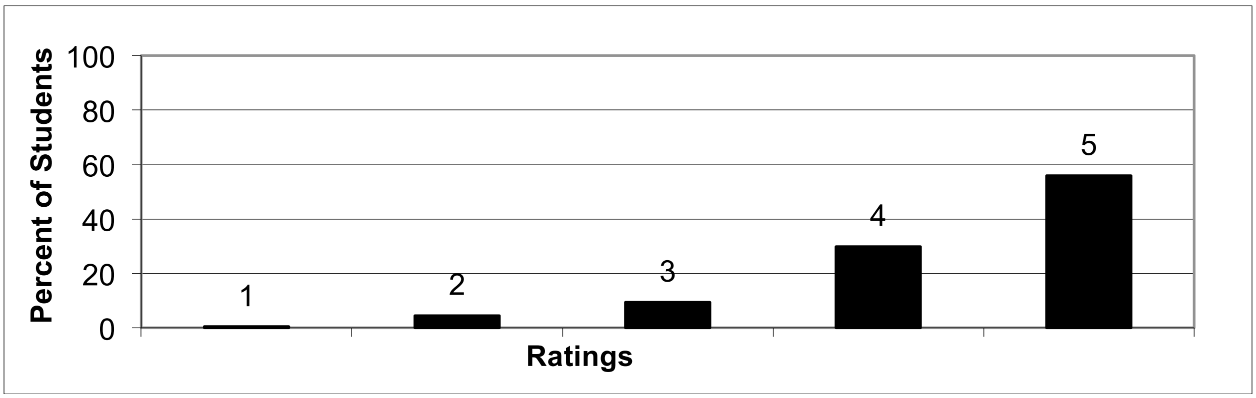


*Figure 3*. Student ratings for value of confidence surveys; 1=No Value; 5=Very Valuable

The Kruskal-Wallis test showed differences in the value ratings across semesters (H=61.43 with $p < .001$). Pairwise comparisons based on Mann-Whitney tests, showed higher value ratings in Year 5 (POGIL added) than in the two previous years—Year 3 (incomplete confidence survey without pretest; $p < .0001$) and Year 4 (complete survey, pretest added; $p < .0001$). There was no significant difference between Years 3 and 4 ($p = .1239$).

### Confidence Scores

Charts like the one shown in Figure 4 allowed easy visual comparison of before-study and after-study confidence scores. In this example, before-study was higher for Topic 4 than for the other three topics, signaling the need for some investigation and possible re-evaluation of the planned teaching on this topic. This investigatory process is similar to descriptions by Nuhfer and Knipp (2003) and Wirth and Perkins (2005). In this project, however, separate formative evaluation comments from students also helped to guide the instructor's investigations.

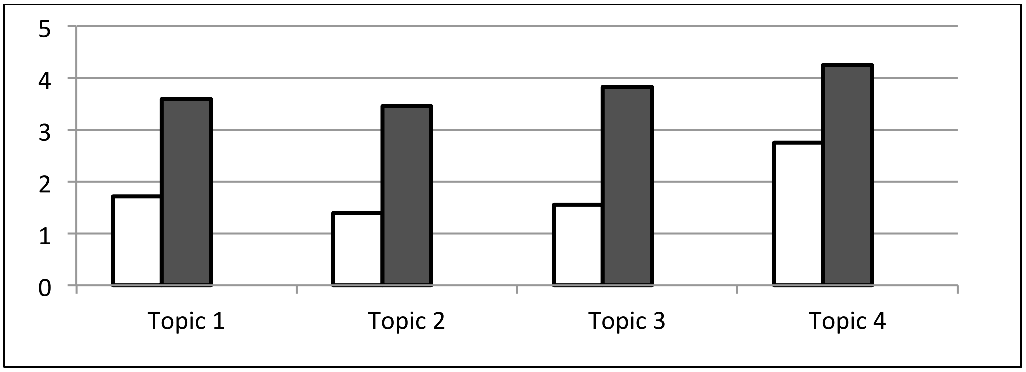


*Figure 4*. Before-study and after-study average confidence scores for first four tasks, Year 5, Semester 3; White=Before Study, Black=After Study

Besides patterns in records and charts, statistical analysis confirmed the significance of increases in confidence scores from before to after study and showed their correlation with final grading points. For total confidence scores summed across topics, results of paired t-tests for after study minus before study were T=28.4, 30.41 and 42.33 for Years 3, 4 and 5, respectively with $p < .001$ in each case. Similarly, within each topic, the confidence scores increased from before to after study with paired t-test statistics ranging from T=14.77 to T=38.40 with each $p < .001$. Spearman's correlation coefficients for after-study total confidence score and final grading points were significant for Year 3 ($r = .294$ with $p = .008$) and Year 4 ($r = .421$ with $p < .0001$), but not for Year 5 ($r = .025$ with $p = .773$).

In Years 3, 4, and 5, gains in average confidence ranged from 1.92 to 2.76 on the 5-point scale, with the Years 4 and 5 (pretest years) showing the greatest increase ($p < .001$). The size of gains for half of the learning tasks was greatest in the year when the pretest was first introduced (significant differences across years for Task 1 with $p = .040$, Task 4 with $p = .036$, Task 5 with $p = .006$, and Task 6 with $p = .011$), but otherwise showed no differences. The same high-level total after-study final confidence scores occurred across Years 3, 4, and 5.

## Grades and Points Earned in the Course

Grades and final points shifted across methods. In Years 1 and 2, before introducing the confidence survey, grades were 30-35% A's, 40% B's, and 20-25% C's. In year 3 when the confidence survey was introduced, grades shifted by 10% from A's to B's, with 25-30% A's, 50% B's. In years 4 and 5, with the introduction of the pretest, grades shifted to more A's and B's and fewer C's: approximately 60% A's, 30% B's and 5% C's. The percentage of students earning below a C was less than 4% in any year, with the highest percentage of D's occurring in Years 1 and 2. The distribution of A, B, and C grades differed significantly across years (Chi Square =72.6 for d.f. = 8 with $p < .0001$). Figure 5 shows the percentage and counts of the grades.

In Years 4 and 5 with the completed revised survey and the pretest, final points earned toward the assigned grades were significantly greater than in Years 1, 2, and 3 (using ANOVA with F=23.65, $p < .001$). No statistically significant difference was found in Years 2 and 3. No topic had a statistically significant correlation of confidence scores to final points in Year 5 when the POGIL project was added.

| Years | Letter Grades | | | | Totals |
|---|---|---|---|---|---|
| | A's | B's | C's | D's | |
| **Year 1** | 34.62% (9) | 42.31% (11) | 19.23% (5) | 3.85% (1) | 100.00% (N=26) |
| **Year 2** | 31.76% (27) | 40.00% (34) | 25.88% (22) | 2.35% (2) | 100.00% (N=85) |
| **Year 3** | 21.25% (17) | 50.00% (40) | 27.50% (22) | 1.25% (1) | 100.00% (N=80) |
| **Year 4** | 62.50% (65) | 33.65% (35) | 3.85% (4) | 0.00% (0) | 100.00% (N=104) |
| **Year 5** | 62.96% (85) | 31.85% (43) | 3.70% (5) | 1.485% (2) | 100.00% (N=135) |

*Figure 5*. Percentage and counts of A, B, C, and D grades by teaching year

### Pre-Evaluation and Formative Evaluation

Multiple channels of student input from the confidence survey and the other formative evaluation measures helped to guide changes in the course. Students' responses allowed the instructor to clarify misconceptions or affirm alignment with what was actually planned for course content and activities. Where student confidence scores seemed pervasively low, or students wrote comments such as "I have no clue about this topic," the instructor would expand instruction; where confidence scores were high, the instructor would lean it up. Added dialogue with students about their formative evaluation responses in Year 3 led the instructor to envision the pre-evaluation questionnaire launched in Year 4. Like his use of the pre-study confidence scores, he used pre-evaluation responses as the basis for discussions with students and to make adjustments early in the course.

## Discussion

This discussion includes a recap of the context of change, how this project advanced the principles of good practice in undergraduate education, the power of instructional task analysis to clearly identify learning outcomes and improve course alignment, alternative learning taxonomies for coding levels of learning, alternatives for survey implementation, value of the confidence survey, and the use of a confidence survey and other formative assessments as a catalyst for scholarship.

### The Context of Change

The initial classroom problem concerned lack of student engagement and motivation to learn. Teaching methods at that time were limited to "old methods" of traditional lecture, homework, and tests. A first attempt at change, though a "colossal disaster," prompted investigation of active learning methods in an instructional designer's workshop. From there, the instructor embraced research-based concepts including instructional task analysis to identify major themes and explicit learning outcomes in his course, the confidence survey, and an end-of-course-formative evaluation. The instructional task analysis required for survey construction produced an organized content map for teaching and learning. Over time, the instructor added more formative assessments and realized his vision for adding more active learning methods. None of the added assessment methods replaced formal grading, but all provided insights to improve instruction and learning.

## Principles of Good Practice in Undergraduate Education

Integration of the confidence survey process with the other formative assessments and new teaching methods further advanced the principles for good practice in undergraduate education (Chickering & Gamson, 1987), similar to what Nuhfer and Knipp (2005) advocated for the use of the knowledge survey.

- **Student-faculty contact:** Students' survey ratings and comments inspired the instructor to initiate dialogue on many topics. For example, the instructor discussed with students the reasons for confusion of the terms "stabilization" and "modification" and made subsequent changes on the survey and in the course. In response, students communicated more frequently with the instructor about other issues. Student input on the pre-evaluation and end-of-course formative evaluation also prompted more instructor-student interactions. Students engaged in face-to-face discussions with the instructor before class, during class, after class, and in his office. They sought clarifications and expressed how much more aware they had become of additional knowledge they needed to explore as a result of the before-study survey and pretest.
- **Reciprocity and cooperation among students:** Team-building activities and projects added by the instructor advanced this principle and supported particular learning outcomes the instructor had identified in the instructional task analysis. Teams formed in class became study and peer tutoring groups. In peer-review activities, students evaluated each other's presentations and provided positive communications.
- **Active learning:** Students not only completed the confidence survey, but used it as a study guide. With the pre-evaluation, students wrote about what they anticipated learning, and with the end-of-course formative evaluation, they wrote about their learning experiences and changes they would make in the course. The whole experience of teaching seemed to improve when the instructor went to team projects with in-class presentations instead of written reports. This began in Year 3, which was the first year of the confidence survey and continued in Years 4 and 5.
- **Prompt feedback:** Completing the pretests and survey items gave prompt feedback to students about their knowledge and skill levels. By using the survey as a study guide they could track and record their progress. Their added interactions with the instructor also gave them immediate responses to their questions.
- **Time on task:** The instructor observed students using the survey throughout the course, indicating task-focused time. Student comments disclosed they used the survey to prepare for class, as well as for quizzes and tests. In addition, as the instructor began to teach the content in a more organized and logical manner, students indicated they learned and retained more.
- **Communicating high expectations:** The pretest, with items identical to the survey, contained the same explicit expectations for learning as the survey items. Students reported they could see in the confidence survey what they were expected to study and the different types of tasks for the whole course.
- **Diverse ways of learning:** With the instructional task analysis, the instructor reflected upon the teaching and learning activities that work well for the types of students in his course. For example, he found that the presentations and dialogue played to students' strengths in ways that matched up with the roles they would encounter in their future jobs. His reflections led to deployment of an increasing variety of activities, giving

students the opportunity to learn in more diverse ways.

## Instructional Task Analysis

Before developing a survey, standards of practice require an analysis to be completed in such a way that different levels of task demands within a domain become clear (Bandura, 2006; Nuhfer, 2003; Nuhfer & Knipp, 2003). Instructional task analysis, the process used by instructional design practitioners (Gagné, Wager, & Golas, 2005; Smith & Ragan, 2005), fulfills this survey analysis requirement. The instructional task analysis produced the content for the confidence survey and pretest, and prompted changes in the instructor's organization of the course. The analysis process included: document analysis, unstructured interviews, learning hierarchy analysis, and "a multi-stage interview technique that captures the automated and unconscious knowledge" of the content expert (instructor) (Clark, Feldon, van Merriënboer, Yates, & Early, 2007, p. 106).

As a result, the instructor reflected upon and articulated the significant learning in the course (Fink, 2007) along with implications for what teaching and learning activities should be developed. Other published studies, with variable methods of analysis, have reported similar impact of the survey development process on instructor reflections about course content and learning outcomes (Bell & Volckman, 2011; Bowers, Brandon, & Hill, 2005; Clauss & Geedey, 2010; Frary, 2009; Wirth & Perkins, 2005). Furthermore, survey design was not a one-shot process. For example, after the instructional task analysis in Year 3, the instructor prepared for Year 4 by using student input and further reflections to refine his analysis and reorganize the survey. Years 4 and 5 both had updates, but Year 4 had the greatest changes in survey items with the re-organization of tasks in the first topic and the completed analysis for the last two topics.

In this project, the pretest and survey had tight alignment with each other, and the final exam and projects required students to integrate the knowledge and skills articulated therein. This alignment could account for the positive shift in grades, a persistent phenomenon reported in a much earlier review of studies by Cohen (1989). In several knowledge survey studies, the alignment of analysis with surveys and exams has varied. On the one hand, Bell and Volckman (2011) reported that authors of the survey also were authors of the tests, and the resulting survey aligned well with the same distribution of topics and levels of learning as taught in the course. On the other hand, Bowers, Brandon, and Hill (2005) reported that different instructors developed different exams for their different sections of the course. And, Clauss and Geedey (2010) reported that faculty wording of survey items produced some confusion for the research assistants who did the actual coding of levels of learning. While dividing the analysis workload may be practical, it may not produce tight alignment of the analysis with the survey and assessments.

To support instructional task analysis, this project used the Gagné framework. The learning task statements simply signaled explicit learning outcomes. Some were information level, such as "Name the four cycles of particle angularity." Others were higher order, such as "Draw and interpret a PI/LL chart." The focus was on explicit clarity without manipulation to add more high-level learning outcomes beyond those identified through the analysis process. This has not been the case in other studies which reported the practice of adding more high-level learning outcomes after the analysis was completed (Bell & Volckman, 2011; Bowers, Brandon, & Hill, 2005; Clauss & Geedey, 2010; Frary, 2009; Nuhfer & Knipp, 2003; Wirth & Perkins, 2005). In those instances, the researchers may have used the process to prompt consideration of how high-level outcomes might fit with the course, but sometimes it seemed the push was to add more simply for the sake of having more.

## Learning Taxonomies

Coding of levels of learning might be distinctively different with different taxonomies. The Gagné framework (Gagné, 1977; Gagné, Wager, & Golas, 2005) could be compared to Bloom's taxonomy or to the "new Bloom" (Anderson & Krathwohl, 2001; Krathwohl, 2002). Future research could advance the coding of learning levels by making a logical choice among taxonomies to fit the type of content in a course. For example, learning that involves safety, health, and life as in the case of nursing (Harper, 2007), often depends on psychomotor skills and affective dispositions not addressed in Bloom's cognitive taxonomy. However, if choosing Bloom's taxonomy, using the full set of levels and sub-levels from the original publication (1956) could produce greater reliability of coding than by using only its major categories and general descriptions. Other choices include content-specific taxonomies like an *engineering taxonomy* (Girgis, 2010) or a pocketful of other taxonomies identified by Anderson and Krathwohl (2001) and Moseley, Baumfield, Elliott, Higgins, Miller, Newton, and Gregson (2005).

## Survey Design

Confidence survey items tightly aligned with the instructional task analysis. The confidence survey used a Likert-type scale for the single construct of *confidence* to do each task within each topic category. Future studies could compare results for a single construct on a simpler 3-point continuum, a checklist, or the longer 0 to 100% continuum often used for efficacy scales (Bandura, 2006). The 0 to 100% scale could yield visual displays in line with final student scores based on a 100-point scale. If an instructor prefers the Nuhfer model and thinks getting the information is an important level of "confidence," he or she could split out the mixed construct item format into separate items—one for each construct.

## Survey Implementation

As predicted by Nuhfer (2003), the survey offered time-efficient comprehensive assessment, allowing students to score confidence for many items in a "very short time span" (p. 59). Frary (2009) suggested: "The instructor's comfort level with each medium and the length of the survey will determine the best method" (p. 8) for implementation. But methods could be compared, such as the pen and pencil ratings on printed surveys, as in Bell and Volckman's (2011) study and in this project, versus surveys completed on an Apple-based mobile device, like an iPad, on a personal computer, a smart-phone, or with other technology tools.

## Value of Confidence Survey

Developing and using the confidence survey was of high value to the instructor and using it was of high value to students. Students used it as a study guide, an observation similar to other anecdotal reports (Bowers, Brandon, & Hill, 2005; Clauss & Geedey, 2005) and student confidence increased from before to after study. However, changes in points earned in the course and redistribution of grades may be due to the combination of formative assessments and more active learning methods. Although Year 5 showed higher value ratings, it had lower confidence correlation with grades. But final points and grades were better, and the lower correlation might be due to the added effort that often accompanies POGIL (Vanags, Pammer, & Brinker, 2013).

Another consideration is that taking the pretest before the survey in Years 4 and 5 may have allowed students to make a more honest appraisal of their before-study confidence levels than in Year 3. In other words, students may have had improved metacognitive confidence (better evaluative judgment) which was the knowledge survey focus for Bell and Volckman (2011). The pretest, survey, and study process together may have contributed to improved judgment of after-study confidence. Changing the time and place for taking the survey also might improve metacognitive confidence as suggested by Nuhfer and Knipp (2003), namely: "The best results occur when survey items clearly frame specific content, and students take the survey home to complete it with plenty of time for self-reflection" (p.5). A useful future study would be to find verification for this proposition compared to other methods.

### A Catalyst for Scholarship

Nuhfer and Knipp (2003) and Wirth and Perkins (2005) reported that the process of making a confidence/knowledge survey improves course organization and preparation. This study shows the same impact. Whether with an instructional designer or a colleague, the process can engage reflection about what content should be taught, in what sequence, how students learn, and what teaching methods to use. Integrating design and use of a confidence survey with other formative assessments allows the instructor to gain more knowledge about students and to accommodate teaching methods to address their needs. For instructors who care about the scholarship of teaching and learning, this endeavor can be a worthy investment, as it was for this instructor and for all of his students.

## References

Anderson, L.W., & Krathwohl, D. R. (Eds.). (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's Taxonomy of Educational Objectives*. New York: Addison, Wesley, Longman, Inc.

Bandura, A. (2006). Guide for constructing self-efficacy scales. In T. Urdan & F. Pajares (Eds.), *Self-Efficacy Beliefs of Adolescents*, (pp. 307-337). Charlotte, NC: Information Age Publishing.

Bell, P., & Volckmann, D. (2011). Knowledge surveys in general chemistry: Confidence, overconfidence, and performance. *Journal of Chemical Education, 88*(11), 1469-1476. doi: 10.1021/ed100328c

Bloom, B. S. (Ed.). (1956). *Taxonomy of educational objectives, the classification of educational goals—Handbook I: Cognitive domain*. New York: David McKay Company, Inc.

Bowers, N., Brandon, M., & Hill, C.D. (2005). The use of a knowledge survey as an indicator of student learning in an introductory biology course. *Cell Biology Education: a Journal of Life Science Education, 4*(4), 311-322. doi: 10.1187/cbe.04-11-0056

Chickering, A. W. & Gamson, Z. F. (1987). Seven principles for good practice in undergraduate education. *American Association of Higher Education Bulletin, 39*, 3-7.

Clark, R. E., Feldon, D. van Merriënboer, J. J. G., Yates, K. & Early, S. (2007). Cognitive task analysis. In J.M. Spector, M.D. Merrill, J.J.G. van Merriënboer, & M.P. Driscoll (Eds.), *Handbook of research on educational communications and technology* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Clauss, J. (2006). *Moodle questionnaire*. Retrieved from http://dmc.augustana.edu/KS/KSFinMA329.pdf

Clauss, J., & Geedey, K. (2010). Knowledge surveys: Students ability to self-assess. *Journal of the Scholarship of Teaching and Learning, 10*(2), 14-24.

Dunn, K. E. & Mulvenon, S. W. (2009). A critical review of research on formative assessment: The limited scientific evidence of the impact of formative assessment in education. *Practical Assessment, Research & Evaluation, 14*(7).

Fink, D. L. (2007). The power of course design to increase student engagement and learning. *Peer Review*, *9*(1).

Fleisher, S.C. (2008). *Knowledge survey: Psychology 211 Survey of Psychology (Cognition and Learning)*. Retrieved from http://familymedicine.medschool.ucsf.edu/paetc/resources/asilomar_2009/psychology.pdf

Feldon, D. F. & Stowe, K. (2009). A case study of instruction from experts: Why does cognitive task analysis make a difference? *Technology, Instruction, Cognition and Learning, 7,* 103-120.

Frary, M. (2009). *Knowledge surveys*. Retrieved from http://scholarworks.boisestate.edu/cgi/viewcontent.cgi?article=1000&context=ctl_teaching

Gagné, R. M. (1974). Task analysis—its relation to content analysis. *Educational Psychologist*, *11*(1). doi:10.1080/00461527409529118

Gagné, R. M. (1977). Analysis of objectives. In L.J. Briggs, (Ed.), *Instructional Design: Principles and Applications*. Englewood Cliffs, NJ: Educational Technology Publications.

Gagné, R. M., Wager, W. W., Golas, K., Keller, J. M. (2005). *Principles of instructional design,* (5th Ed). Belmont, CA: Wadsworth/Thomson Learning.

Gilpin, L. (2013). Enhancing teaching and learning. *Mountain Rise: The International Journal for the Scholarship of Teaching and Learning, 8(1).*

Girgis, M. (2010). A new engineering taxonomy for assessing conceptual and problem-solving skills. Proceedings of the American Society for Engineering Education Conference, USA, AC 2010-1793.

Harper, S. P. (2007). *Instructional design for affective learning in online nursing education*. (Doctoral dissertation). Capella University. Retrieved from http://udini.proquest.com/view/instructional-design-for-affective-goid:304722378/

Herman, J. L. (2013). Formative assessment for next generation science standards: A proposed model. Resource Paper No. 16. National Center for Research on Evaluation, Standards, and Student Testing. Retrieved at https://www.cse.ucla.edu/products/resource/cresst_resource16.pdf

Johnson, E. (2009). Formative and summative assessment. Retrieved at http://www.education.com/reference/article/formative-and-summative-assessment/

Jonassen, D. H., Tessmer, M., & Hannum, W. H. (1999). *Task analysis methods for instructional design*. NJ: Lawrence Erlbaum Associates.

Kasunic, M. (2005). Designing an effective survey. Carnegie Mellon University. Retrieved at http://www.sei.cmu.edu/reports/05hb004.pdf

Keller, J. M. (2000, February). *Applying the ARCS model of motivational design*. Paper presented at VII Semanario, Santiago, Cuba.

Keller, J. M. (2010). *Motivational design for learning and performance: The ARCS model approach*. New York, NY: Springer.

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

Marshall, S. & Nuhfer, E. B. (2013, November). *Knowledge surveys and course-based learning outcomes for soil science courses*. Presentation at Soil Science Society of America Symposium, Tampa, Florida.

Merrill, M.D. (2009). *What makes e3 (effective, efficient and engaging) instruction*? Presented at World Conference on Educational Multimedia, Hypermedia and Telecommunications 2009. Retrieved from http://www.editlib.org/p/31461

Moseley, D., Baumfield, V., Elliott, J., Higgins, S., Miller, J., Newton, D. P., & Gregson, M. (2005). *Frameworks for thinking: A handbook for teaching and learning*. Cambridge, UK: Cambridge University Press.

Nilson, L. B. (2013). *Creating Self-Regulated Learners: Strategies to Strengthen Student Self-Awareness and Learning Skills*. Sterling, VA: Stylus Publishing, LLC.

Nuhfer, E. (n.d.). Knowledge surveys (KS). *MERLOT ELIXR*. California State University Channel Islands. Retrieved from http://elixr.merlot.org/assessment-evaluation/knowledge-surveys/knowledge-surveys2

Nuhfer, E. & Knipp, D. (2003). The knowledge survey: A tool for all reasons. *To Improve the Academy*, *21*, 59-78.

Nuhfer, E. B. & Knipp, D. (2006). Re: The use of a knowledge survey as an indicator of student learning in an introductory biology course. *CBE Life Science Education, 5*(4): 313–314. doi: 10.1187/cbe.06-05-0166

Panadero, E. & Alonso-Tapia, J. (2014). How do students self-regulate? Review of Zimmerman's cyclical model of self-regulated learning. Anales de Psicología, 30(2), 450-462.

Price, B. A. & Randall, C. H. (2008, May/June). Assessing learning outcomes in quantitative courses: Using embedded questions for direct assessment, *Journal of Education for Business*, 288-294. doi: 10.3200/JOEB.83.5.288-294#preview

Sawchuk, S. (2013). Combined measures better at gauging teacher effectiveness, study finds. *Education Week*, *32*(17), 1-16.

Shepard, L. A. (2005, October). *Formative assessment: Caveat emptor*. Presented at ETS Invitational Conference: The Future of Assessment: Shaping Teaching and Learning, New York.

Smith, P. L. & Ragan, T. J. (2005). *Instructional design* (3rd ed.). Hoboken, NJ: John Wiley & Sons, Inc.

Strayhorn, T. L. (2006). *Frameworks for assessing learning and development outcomes*. Retrieved from Council for the Advancement of Standards in Higher Education website: http://nau.edu/Student-Affairs/_Forms/FALDO/

Sweller, J., & Chandler, P. (1994). Why some material is difficult to learn. *Cognition & Instruction, 12*(3), 185-233.

Tullis, T. & Dumas, J. (2009, May). *Rating scales: What the research says*. Presented at the Boston Usability Professionals' Association Mini-UPA Conference. Retrieved from http://www.measuringuserexperience.com/Mini_UPA-Dumas-Tullis.ppt

Vanags, T., Pammer, K., Brinker, J. (2013). Process-oriented guided-inquiry learning improves long-term retention of information. *Advances in Physiology Education, 37*, 233-241. doi: 10.1152/advan.00104.2012

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

Yeasmin, S. & Rahman, K. F. (2012). 'Triangulation' research method as the tool of social science research. *Bangladesh University of Professionals Journal, 1*(1), 154-163.