Creating a Better Mousetrap: On-line Student Assessment of their Learning Gains

Elaine Seymour, Douglas J. Wiese, and Anne-Barrie Hunter
Bureau of Sociological Research,
University of Colorado, Boulder.
and
Susan M. Daffinrud,
National Institute of Science Education,
University of Wisconsin-Madison.

Paper originally presented to the National Meetings of the American Chemical Society
Symposium,
"Using Real-World Questions to Promote Active Learning,"
San Francisco, March 27, 2000

.

Abstract

This paper discusses the development of an instrument that is designed to summarize the learning gains that students perceive they have made, both as a consequence of particular aspects of class pedagogy, and of the teacher's pedagogical approach. The instrument avoids critiques of the teacher, the teacher's performance, and of teaching methods that are unrelated to student estimates of what they have gained from them. The instrument was originally designed for chemistry faculty concerned to discover the efficacy of undergraduate teaching modules. However, it may be adapted for any pedagogical approach or discipline. It can be edited on-line to reflect any set of learning objectives, is completed on-line by students, and provides faculty with a summary of results in both statistical and graphic form. Discussed are: the instrument's research origins, its on-line features, findings from two rounds of testing with faculty in different disciplines, and analysis of a sample of write-in questions and answers.

Creating a Better Mousetrap: On-line Student Assessment of their Learning Gains

Introduction

Faculty often complain that student classroom evaluations intended to judge the efficacy of their teaching offer them poor feedback. As a result, they pay less attention to the numeric scores generated by these instruments than to their students' write-in observations and advice. These comments are not normally submitted to systematic analysis, but are responded to impressionistically. Faculty dissatisfaction is heightened when student evaluation instruments "that ask the wrong questions" are the main (or sole) means by which departments evaluate faculty teaching effectiveness for tenure and promotion purposes. Standard instrument questions about how well faculty performed their teaching role, and about "the class overall" that fail to give students criteria for their judgments, yield inconclusive results that are difficult for faculty to interpret (Weise, Seymour, & Hunter, 1999).

The learning gains instrument presented here is grounded in its authors' findings:

(1) that students can make realistic appraisals of their gains from aspects of class pedagogy and of the pedagogical approach employed
(2) that this feedback allows faculty to identify course elements that support student learning and those that need improvement if specific learning needs are to be met. This is particularly useful for faculty who have revised their learning objectives and pedagogy and are seeking forms of assessment that reflect them.

Their assessment needs of classroom innovators include both appropriate and accurate tests of student learning, and more precise feedback from students on the utility of particular aspects of the class to their learning. The Student Assessment of their Learning Gains (SALG) instrument was originally developed to match the learning objectives and teaching methods of chemistry faculty who (since 1995) have introduced "modular" teaching methods into lower-division classes in over 100 two- and four-year institutions[1]. Since the SALG instrument is easily modified to meet the needs of individual faculty in different disciplines and provides an instant statistical analysis of the results, it is argued to be a powerful and useful tool. It may be found at http://www.wcer.wisc.edu/salgains/instructor.

---

[1] The chemistry module developers and adapters were participants in two, linked consortia, "ChemLinks" and "ModularChem," that were funded for five years (1995-2000) by the National Science Foundation (Division of Undergraduate Education) to develop, field-test, adapt, and disseminate "modular" materials and teaching methods that focus on learning fundamental ideas in chemistry through exploration of issues with real-life significance.

Basis for the Instrument in Findings from a Student Interview Study

The rationale for the instrument emerged from the findings of a student interview study that was part of the formative evaluation of the two modular chemistry consortia. Two hundred and ninety-two students were interviewed in a matched sample of modular and more traditionally-taught introductory chemistry classes at eight participating institutions[2]. The sample was chosen to represent the range of institutional types participating in the consortia. It comprised: two research universities, three liberal arts colleges, one community college, one state comprehensive university, and one historically black college. (Later, two more community colleges and a third research university were added to the sample, thereby increasing the student interview total to 345.)

The student interviews (both with individuals and in focus groups) were tape-recorded and transcribed *verbatim*. The resulting text files were entered into a set of computer programs (The Ethnograph, V.5, 1998) that was used to assist with the analysis. Student observations were of three types: (1) answers to interviewers' questions (2) spontaneous observations and (3) agreement with observations made by other focus group members. Across both types of class in the eight institutions, students offered 7,406 discrete observations (of all three types). These contained two broad categories of assessment: assessments of class pedagogy and of teacher performance (expressed in terms of what students "liked"), and assessment of their own learning gains from aspects of the class. Examples of what students liked or disliked were: "I thought the teacher was very organized and presented the material well;" "The tests were fair," "The teacher was very approachable," and "some of the demonstrations didn't work well." Although statements of this type may be taken to imply that organization, fairness, approachability, and technical competence have some level of impact on student learning, the connection is unstated and offer the teacher limited feedback about what may be more or less important in enabling learning. Examples of positive and negative "gains" statements include the following:

> It interests you more if there's a practical relevance to the real world. It's not just plus this equals that…I think that helps you learn, because it sticks better if you know what it relates to.
>
> I think I'm more likely to retain what I learned because you could associate some experience with it rather than just memorizing it.
>
> You'd be better off if you knew exactly what the lab stood for while you are doing it. Otherwise it's just, "Gotta get it done, so let's pour this in here, that in there. I don't know what this is, but it's an indicator, so why don't we just pour it in, and if it works, it works; if not, let's redo it." And so you don't understand the whole concept at all. So, when the lab report's due, I end up trying to figure out what it was that happened.

[2] It should be noted that the degree to which the matched comparative classes were "traditional" in their pedagogy varied considerably by institutional character. The comparative classes reflected whatever was considered the "normal" way to teach introductory chemistry classes at each institution in the sample.

> I mean, if I could just see the links a little more clearly—if someone helped me to look at them—I think I would understand the material a lot better. I'd be able to use the applications to understand what was going on. But I don't feel the links are made strongly enough so I can do that myself.

Students' observations on "what they liked" were also less useful in the aggregate than their estimates of "what they gained." Analysis of student judgments of faculty performance revealed that, when all students' observations (for *both* the modular and the comparative classes) on how well faculty performed their teaching role were compiled, 45 percent were positive (i.e., expressed what the students liked about the teacher or pedagogy), 50 percent were negative, and 5% involved mixed reviews or neutral positions (see Table1). As there was also virtually no difference between the modular and matched comparative classes in this regard, neither group of faculty got a clear picture of the overall utility of their teaching when the criterion used was the perceived quality of faculty's professional performance. This is, arguably, because students lack the knowledge or experience to make such judgments. This finding also reflects common faculty experience that asking students what they "liked" or "valued" about their classes (especially where no criteria for these judgements is offered), tells neither faculty nor their departments much about the impact of their classroom work on students.

Table 1. Comparison of what students reported that they learned with what they liked in a sample of 292 students in matched 'modular' and non-'modular' introductory chemistry classes in eight two- and four-year institutions.

Type of Student Observation on Aspects of the Class	**Frequency and Percent for Each Type of Observation**							
	Yes		**No**		**Don't know/ Mixed**		**Total #**	**Total %**
	#	**%**	**#**	**%**	**#**	**%**		
Whether students learned from the pedagogy	2,600	57	1,437	31	537	12	4,574	100
What students liked about the pedagogy and/or the instructor	1,286	45	1,416	50	130	5	2,832	100
Grand Totals	**3,886**	**52**	**2,853**	**39**	**667**	**9**	**7,406**	**100**

By contrast, in both the modular and comparative classes, students offered clear indications about what they themselves had gained from specific aspects of their classes. As is also indicated in Table 1, when all specifically gain-related student observations were totaled and divided into three types—positive (things gained), negative (things not gained), and mixed reviews (qualified assessments of gain), 57 percent of the observations for both types of class were positive, 31 percent were negative, and 12 percent were "mixed." [3]

The strong similarity between student evaluation totals for the modular and comparative classes probably reflects the early stage of development of the modules and the teachers' limited experience in using them at the time of these interviews. The issue presented here, however, is not the relative merits of modular and more traditional chemistry teaching, but the hypothesis suggested by this finding: that it is more productive to ask students how much they have gained from specific aspects of the class than what they liked or disliked about the teacher and his or her pedagogy.

Structure of the SALG Instrument

In light of these findings, the first version of the current SALG instrument was developed in 1997. Changes continue to be made in response to user suggestions, in the interest of greater clarity, consistency of language, ease of student comprehension, and in order to make the instrument flexible enough for users in different disciplines and with different learning objectives. However, the instrument structure remains exclusively focused on questions about how much students judge those aspects of the course that their teachers identify as important to their learning (by their question selection) actually enabled their learning. As already noted, the questions initially selected were those that reflected the learning objectives shared by, and of special interest to, chemistry module developers and adapters. Some of these have been retained because non-modular and non-chemistry users have chosen to keep them in their own versions of the instrument.

The "template" instrument offered on the web-site (shown in Appendix A) contains 46 sample questions, grouped in sections, such as, "class and lab activities," "resources," and "skills." SALG users are invited to use or adapt these questions in order to reflect their own course-specific learning objectives, to drop those that do not apply, and to add others. However, in paper- and-pencil versions, users are asked not to change the instrument's exclusive focus on learning gains; in the web-site version, they cannot change this focus. They are also advised to avoid questions that are vague or general "global questions," and questions that address more than one issue.

Students are asked to respond to the instrument questions on a five-point Likert-style scale. In this format, respondents choose from a set of evaluative statements (i.e., "response categories") that are ranked from low to high. Students are instructed to score each question on a scale from 1 (lowest) to 5 (highest) in terms of its perceived value in aiding their learning. There is an

[3] Again, when seen in the aggregate, there was no significant difference between the modular and comparative classes. Differences on specific items of both learning gains and student likes and dislikes did, however, appear in the breakdown by class types (i.e., modular vs. comparative).

additional option of NA (not applicable). The instrument has two slightly different versions of response categories, to reflect the nature of the questions being asked. However, all instrument questions ask students to evaluate how particular aspects of the class enabled their learning, or helped them to improve particular skills or areas of knowledge. The different response categories are summarized in Figures 1 and 2.

Figure 1: Response Categories for SALG Instrument Questions in Section 1.

Section 1 "How much did each of the following aspects of the class help your learning?"

Scale		Response Category
1	=	Was of no help
2	=	Was a little helpful
3	=	Was of moderate help
4	=	Was of much help
5	=	Was of very much help

Figure 2: Response Categories for SALG Instrument Questions in Sections 2-5.

Section 2: "As a result of your work in this class, how well do you think that you now understand the following?"
Section 3: "How much has this class added to your skills in each of the following?"
Section 4: "To what extent did you make gains in any of the following as a result of what you did in this class?"
Section 5: "How much of the following do you think you will remember and carry with you into other classes or aspects of your life?"

Scale		Response Category
1	=	I gained nothing/not at all
2	=	I gained a little
3	=	I gained somewhat
4	=	I gained a lot
5	=	I gained a great deal

Likert scales are commonly used, both for institutional classroom evaluation instruments, and in questionnaires exploring degrees of agreement or disagreement with position statements. The five-point range of values and descriptors was chosen partly because of its familiarity to students and faculty, and partly to enable students to make meaningful distinctions between the degrees of gain that lie between the extremes of "no gain" and "gained a great deal." The discriminative power and reliability of the scale have yet to be formally tested. However, high and low scores on particular items compare favorably with the percentages of positive and negative comments on the same items offered by students in the interview study. These will be discussed later in the paper.

In the first test of the (pre-web-site) instrument, the Likert scale enabled the analyst to compare questions numerically across sections. Student scores on these scales were used to calculate average scores for sections and questions. Scores of 1 and 2 were taken as low scores—that students considered aspects of the class to have contributed nothing or little to their learning. Scores of 3 and 4 were interpreted as satisfactory to high, and scores of 5 were as a strong endorsement of particular aspects of the class as having contributed to the student's learning or skills.[4]

As can be seen in the template instrument currently offered on the SALG web-site (Appendix A), questions are grouped into broad aspects of the class or lab, for example, learning gains from particular class and lab activities, from tests, graded activities, and assignments, and from resources (e.g. the text, readings, the web). Gains in skills, cognition, and attitudes toward the discipline or topics within it, and towards the students' own learning are also explored. The instrument includes some broader possible gains. For example, the question, "To what extent did you make gains in any of the following as a result of what you did in this class?" is followed by nine items that include: (1) understanding how ideas in this class relate to those in other science classes (2) feeling comfortable with complex ideas (3) ability to think through a problem or argument (4) understanding the relevance of this field to real world issues. Students are also asked to make estimates of their learning retention and adequacy of preparation for future classes offered by the current class.[5]

Development History and Testing

The pilot instrument was initially offered on the ChemLinks/ModularChem websites in December 1997 as a paper-and-pencil instrument that could also be adapted for use with Scantron forms. This followed initial testing in three chemistry departments where it was well received by both modular and non-modular faculty. The instrument was subsequently adopted as the formal end-of-semester class evaluation instrument in two of these departments.

In 1998, the Exxon Education Foundation provided the funding:

(1) to conduct tests with a sample of 17 chemistry module "adapters" at ten institutions teaching 28 classes or sections who volunteered to try out a paper-and-pencil version of the SALG instrument in their introductory modular chemistry classes in the spring and fall of 1998, and to provide the ChemLinks Evaluation team at the University of Colorado, Boulder with sets of completed instruments for analysis.

---

[4] Students scored questions using only whole numbers; calculating means, however, introduces fractions. This has to be taken into account in discussion of test results

[5] This question also derived from a comparative student interview finding that 68% of students in the modular chemistry classes, compared with 47% of those in the comparative classes, thought that the methods of teaching and learning they had experienced would help them to retain what they had learned. It was suggested to instrument users that they might collaborate with a colleague in a prior or following class to check out the validity of their students' estimate of their learning retention.

(2) to enable its development as an on-line instrument. This work was carried out by Sue M. Daffinrud, the National Institute of Science Education, University of Wisconsin-Madison who continues to monitor, maintain and improve the site in response to user suggestions.

The Exxon Education Foundation provided a second grant in 1998/1999:

(1) to gather feedback from a sample of faculty users from different disciplines on the adaptability of the on-line instrument to the needs of their disciplines, and faculty experience with student responses to on-line classroom evaluation [6]
(2) to further refine the web-site that carries the SALG instrument in light of user feed-back.

In addition, Anne-Barrie Hunter (of the Boulder team) has conducted an analysis of the nature of students' write-in comments in response to faculty's open-ended questions, including the kinds of comments that are solicited by questions of particular types.

Presentation and discussion of selected findings from these analyses are offered throughout the paper.

In addition to the SALG web-site, the instrument is available on the Field-tested Learning Assessment Guide (FLAG) web-site developed and maintained by the National Institute for Science Education at the University of Wisconsin-Madison. This site was conceived and begun in 1997-1998 by Elaine Seymour and Steve Kosciuk (of the LEAD Center) with additional funding and contribution of assessment materials by the "Establishing New Traditions Chemistry Coalition," University of Wisconsin-Madison. FLAG is designed as a resource for science and mathematics faculty seeking assessment methods appropriate to their learning objectives.[7]

The On-Line Instrument

The purpose of the SALG web-site is to streamline the use of the SALG paper instrument. A significant barrier to many instructors' use of the paper instrument is the amount of time needed for implementation within the classroom. Instructors need to modify the survey to suit their own needs within a word-processing application, produce paper copies, take class time to have students complete it, collect the data, enter it into a spreadsheet, and analyze those data. Although the feedback gained from its use is valuable, the process of getting the information from a paper-and-pencil instrument takes more time than most faculty can give. A web-site offers increased accessibility, allows easy on-line modification, enables students to complete the instrument on-line, and automatically produces a set of standard survey results. This resolves the

[6] Analysis is proceeding with data collected from recorded telephone interviews and e-mail surveys, conducted with a small sample (N=24) of registered users of the SALG on-line instrument early in 2000. The central issues were: whether users found the web-based instrument easy to use, adaptable for different learning objectives and disciplines; and whether it provided them with clear, learning-specific feedback from students. Findings will soon be available to SALG web-site users.

[7] Faculty interested in the FLAG website can locate it at: http://www.wcer.wisc.edu/nise/cl1. (Please note that the address ending is "c"-"el"-"one," *not* "c"-"one"-"one.")

problem of implementation, and allows easy dissemination of the SALG instrument to other potential users.

The SALG web-site actually contains two web-sites. It has one interface[8] for instructors (who create versions of the SALG instrument appropriate to their class learning objectives and can view the data), and a second interface[9] for students when they complete the instrument. A database[10] that underlies the SALG web-site stores all of the instructor, course, and student response data. It also stores versions of the SALG instrument developed by users who have chosen to leave them on the site for other faculty to adapt. The web-site is housed at a server at the University of Wisconsin–Madison, the Wisconsin Center for Education Research (WCER). All stored information is private. However, instructors can opt to delete all of their records from the site.

The web-site currently has 141 registered faculty users from a range of institutions and disciplines. The instructor side of the site is "hit," on average, 7 times per day.

Appendix B contains a step-by-step tour of the web-site for first-time users, including a description of the interfaces for faculty and students.

Comments from users of the SALG web-site are continuously sought, along with their suggestions for changes and additions. There is no charge for use of the site, nor for adaptation of the instrument, however, citation of the authors' copyright is requested. The authors are in correspondence with some "group adapters" (i.e., departments and institutions that are revising their standard classroom evaluations, and coalitions of classroom innovators who have adapted the SALG instrument for their own purposes. All adapters are advised that, if they introduce non-gains criteria for student response, the validity of the instrument is likely to be compromised.

Selected Findings from a Panel Test of the SALG Instrument

As mentioned above, the Exxon Education Foundation funded a panel testing of paper-and-pencil versions of the original SALG instrument in introductory chemistry classes in the spring and fall of 1998. The test panel comprised 28 modular chemistry classes (or sections) taught by 17 instructors in 10 institutions. Four types of institution were represented: 1 large research university, 1 medium-sized state university, 7 liberal arts colleges, and 1 community college. The large number of liberal arts colleges reflects the high representation of these types of institution in the two chemistry consortia. The range of class (or section) sizes was very wide—from 5 to 296 students (who completed instruments).

This (original) version of the instrument offered a choice of 49 questions grouped into the same sections as those used in the current web-site template. All of the questions were derived from modular learning objectives—though they, clearly, have wider application. The panel of testers selected and adapted questions that they considered most appropriate to their class or section

[8] http://www.wcer.wisc.edu/salgains/instructor
[9] http://www.wcer.wisc.edu/salgains/student
[10] Accessed through SQL 7.0 server.

activities. They were also offered write-in slots in some sections for extra questions specific to their classes.

Consistent with the practice of some chemistry departments represented in the sample, some faculty asked all of the questions in an open-ended format while preserving the learning gains criteria for student responses. Only those questions in numeric format that were used in common were included in the comparative numeric analysis. This reduced the number of completed instruments used for numeric comparison to 18 instruments contributed by 13 instructors in eight institutions. Student responses to non-numeric instruments and to open-ended questions in otherwise numeric instruments were subjected to a separate, later analysis. Selected results are included in this article: a full description of these findings will be reported elsewhere.

Methods Used to Compare Sections

Doug Wiese (of the Boulder team) conducted the quantitative analysis of this data. The most basic method used to analyze student scores for all questions was to compare the *section* means (computed by averaging all student scores on all questions for each section and from every question used. The means for the 18 sections in the sample are given in Table 2.

Table 2. Section Size, Section Means, and Their Standard Deviations by Institution.

Institution/ Section	Section Size	Section Means	Standard Deviation
1A	40	3.56	0.50
1B	5	2.70	0.24
1C	22	3.39	0.47
1D	19	3.51	0.54
1E	20	3.33	0.70
2A	236	2.79	0.56
2B	296	2.87	0.63
2C	175	2.59	0.51
3	7	3.20	0.51
4	39	3.60	0.45
5A	32	3.48	0.46
5B	31	3.23	0.48
5C	32	3.30	0.56
5D	31	3.20	0.57
6	37	3.78	0.50
7	37	3.25	0.68
8A	48	3.34	0.62
8B	29	3.54	0.78

At first glance, the means for the 18 sections indicate a good level of student satisfaction with their learning gains: all but sections 1B and 2A, 2B and 2C scored above 3.0. Five sections scored above 3.5, although no section scored 4.0 or higher. These scores, however, must be interpreted with caution. Statistical means are, by definition, a very generalized form of information. While information about individual student responses is condensed into a broad estimate of how much students gained overall from the class, detail is lost about the distribution of student responses and, thus, about student gains (or lack of them) in specific aspects of the class. A comprehensive analysis of both the SALG instrument's utility, and the level of student gains from the pedagogical approach of this group of faculty, requires a more precise analysis of individual instrument questions.

Methods Used to Compare Instrument Questions

The primary method used in this study to analyze student scores for different questions was to compare the mean averages of question scores. Two types of question mean were used in the analysis of student responses to instrument questions. The first type is the *section question mean*, that is, the mean average of scores given by all students in a section for a given question. This non-weighted mean can be used to compare how students in different sections scored the same question. The second measure is the *question mean*. For any particular question, this is the average of the question means for *all* sections where faculty used that question.

Mean averages are a useful method for condensing information. When used alone, however, they offer an incomplete picture of the actual distribution of values because of the varying numbers of students in each of the sections. (Section sizes ranged from five to 296 students.) In calculating the total question mean, each section's question mean has equal weight, regardless of class/section size. Thus, the means treat the sections themselves as the unit of analysis, and the resulting statistics do not reflect differences in section size.

To get a complete picture of the actual distribution of scores requires a measure that gives equal weight to each individual student's question score, namely, a *weighted question mean.*

The weighted question mean is calculated by multiplying each section's question mean by the proportion of the class size in relation to the total student sample.[11] Weighted means take into account varying class sizes, but can also give a distorted view of the actual distribution of scores because the variable number of students in each section potentially skews the mean. This was the case for students in the three large sections at the research university (Institution 2) who gave substantially lower scores to more questions than did students at other institutions. Used in isolation, both the non-weighted and weighted means give incomplete (and possibly misleading) pictures of the distribution of students' scores. Therefore, the two measures are best used in tandem to provide a more accurate picture of the distribution of question scores than is afforded by either of the means taken alone.

A difference between the weighted and non-weighted question means indicates that students from one or more sections scored questions higher or lower vis-à-vis other sections. For example, if the weighted mean is lower than the non-weighted mean for a particular question, it may be due to students from a small section scoring questions highly and inflating the non-

[11] The weighted question mean can also be calculated directly by averaging every student's score on a question.

weighted mean. Conversely, it could also be due to low scores from students in a large section that pull down the weighted mean. The difference between higher- and lower-than-average question means can also be a result of *both* situations. As will be seen in the following analysis of questions, where students scored questions higher or lower than average, the weighted mean is, in all but one case, lower than the non-weighted mean. In almost all cases, this is a reflection of the large class sizes of the three sections at Institution 2. Students from these sections scored nearly all of the questions about their learning gains from the class substantially lower than students from other institutions. Table 3 lists both the non-weighted and weighted question means. The generally lower scores for weighted versus non-weighted means are clearly shown.

Table 3. Non-weighted and Weighted Question Means for all SALG Instrument Questions.

	Question Content	Non-weighted Question Mean	Weighted Question Mean
Overall gains			
1-A	Focus on real-world problems	3.23	2.64
1-B	How class activities fit together	3.35	2.77
1-C	Pace of class	2.92	2.61
Gains from particular class and lab activities			
1-D1	Class presentations	3.47	2.83
1-D2	Discussions in class	3.33	2.77
1-D3	Group work in class	3.07	2.63
1-D4	Hands-on class activities	3.51	2.77
1-D5	Understanding why doing each activity	3.26	2.68
1-D6	Written lab instructions	3.45	2.66
1-D7	Lab organization	3.20	2.57
1-D8	Teamwork	3.45	2.51
1-D9	Lab reports	3.24	2.45
Gains from tests, graded activities and assignments			
1-E1	Opportunities for in-class review	3.33	2.72
1-E2	# and spacing of tests/assignments	3.18	2.67
1-E3	Fairness of tests	3.10	2.49
1-E4	Mental stretch required	3.15	2.53
1-E5	The grading system used	3.18	2.61
1-E6	Feedback received	3.11	2.60
Gains from resources			
1-F1	Module student manual	3.07	2.60
1-F2	The text	3.20	2.65
1-F3	Other reading	3.08	3.18
1-F4	Use of the web	3.08	3.17
Gains from information given about:			
1-G1	Weekly class activities	3.24	2.55
1-G2	How parts of class related to each other	3.21	2.55
1-G3	Specific assignments (write-in)	3.07	3.23
1-G4	The grading system used	3.03	2.46

Gains from support to the individual learner			
1-H1	The quality of contact with the instructor	3.47	2.91
1-H2	The quality of contact with the TAs	3.38	3.49
1-H3	Working w/ peers outside of class	3.54	3.50
General gains			
1-K	Way class was taught overall	3.46	2.95
Gains in skills			
3-1	Solving problems	3.16	2.97
3-2	Writing papers	2.34	2.71
3-3	Designing lab experiments	3.04	2.85
3-4	Finding trends in data	3.20	3.02
3-5	Critically reviewing articles	2.92	2.46
3-6	Working effectively with others	3.22	3.05
3-7	Giving oral presentations	2.61	2.21
Cognitive and affective gains			
4-1	Understanding main concepts	3.49	3.33
4-2	Understanding relations between concepts	3.41	3.20
4-3	Understanding relation to other science and math	3.34	3.03
4-4	Understanding relevance to real-world issues	3.68	3.24
4-5	Understanding the nature of chemistry	3.33	3.07
4-6	Appreciating the methods of chemistry	3.23	2.90
4-7	Ability to think through a problem or argument	3.22	2.97
4-8	Confidence in your ability to do chemistry	3.14	2.77
4-9	Feeling comfortable with complex ideas	3.07	2.78
4-10	Enthusiasm for chemistry	2.97	3.31
Class contribution to learning retention and future preparation			
5A-1	Predict will retain understanding of main concepts	3.48	3.17
5B-1	Was well-prepared for other classes	3.58	3.84

Both the weighted and non-weighted means provide useful information about question scores. However, visual representations offer an additional method to assess the distribution of scores. The charts in Appendix C show the effects that small and/or large class size can have on non-weighted and weighted means. Quantitative comparison of the two types of means, along with their visual representation provides a good picture of the distribution of question scores.[12]

Analysis of Student Responses to the Instrument Questions

In order to identify aspects of the modular sections in this sample that were either more problematic or successful, the analysis then focused on questions for which scores were higher or lower than the question average. The majority of questions (63%) had total question means between 3.0 and 3.4. A non-weighted mean below 3.0 was selected as the cut-off for low-score questions, and a non-weighted mean of 3.4 or above for high-score questions. As the efficacy of modular chemistry is not the focus of this article, commentary is offered below only for two low-

[12] The charts in Appendix C compare only non-weighted means across sections. Weighted means cannot be computed for individual sections, only for particular questions across all sections.

scoring, two high-scoring, and two problematic questions that may be of interest to faculty at large. The analysis also illustrates the possibilities of combining scores for the SALG instrument for any group of faculty with shared learning objectives. Full findings from this testing may be read elsewhere (Wiese, Seymour, and Hunter, 1999).

To validate and interpret the SALG self-report data, comparative findings are offered from other parts of the modular chemistry consortia evaluation. These include:

(1) interview data from module developers and adapters
(2) comparative student interview data (described earlier)
(3) early findings from a second round of student interviews (conducted spring and fall 1999) in classes where the best-developed modules were taught by the most experienced modular teachers
(4) a study that explores the nature of students' difficulties in traditionally-taught science and mathematics classes at seven four-year institutions (Seymour & Hewitt, 1997).

As will be seen in the discussion that follows, some scores are difficult to interpret. This highlights the importance for faculty of including write-in sections below any question that has produced low scores in a previous use of the SALG instrument. As the analysis of write-in answers also revealed, providing students with gains-related criteria for their written observations is critical if faculty are to understand why some aspects of their class appear problematic to students according to the numeric scores that they award.

Questions With Lower Than Average Scores

Table 4 lists the five questions which students rated lowest (on both types of total question means) in terms of their learning gains. (They are ranked in ascending order according to non-weighted means.) These were the only questions to score a non-weighted question mean below 3.0.

Table 4. Questions for which Students Scored the Lowest Learning Gains in a Sample of 18 Modular Chemistry Sections.

	Question Content	**Means**		**Difference between Non-weighted and Weighted Means**
		Non-weighted	**Weighted**	
3-2	Skills in writing papers	2.34	2.71	.37
3-7	Giving oral presentations	2.61	2.21	.40
3-5	Critically reviewing articles	2.92	2.46	.46
1-C	Pace of class	2.92	2.61	.31
4-10	Enthusiasm for chemistry	2.97	3.31	.34

1. Question 1-C: How much did the pace at which we worked help your learning?

As can be seen in Chart 1 (Appendix C), the actual distribution of scores on this question is not as low as either mean suggests. Eight out of 17 sections recorded a question mean above 3.0. The .31 difference between weighted and non-weighted means for Question 1-C is largely due to the low scores given by students of the three large sections (2A, 2B, 2C) at Institution 2 (n=707 students). The low score of the very small section 1B (n=5) also pulls down the non-weighted total question mean. Nevertheless, students in roughly half of the sections gave scores under 3.0 to this question, which merits classifying it among the lower-than-average scores.

Student difficulties with pace are also very common in classes taught by the lecture method (Seymour & Hewitt, 1997), but are brought into greater faculty and student awareness by the changes in thinking and practice required by the modules. In the (1997) study of why students leave the sciences, pace and workload issues were cited by 35 percent of "switchers" as having directly contributed to their decision to leave this group of majors; and were 41 percent of the concerns of non-switching seniors. This common student experience (described by one of our informants as "trying to drink out of a fire hose") arises because of the pressures bearing upon faculty to "cover" as much as possible of the expanding disciplinary canon. Students described "covering material" as counterproductive. They did not remember—and often had not understood—material presented to them at a pace that was too fast for absorption. Although the efficacy of presenting students with a high volume of materials at fast pace in introductory "survey" classes has not been demonstrated, some modular faculty feel under pressure from colleagues, and from their own professional socialization to cover more material than serves their learning objectives.

2. Question 3-2: How much has this class added to your skills in writing papers?

Students' rating of how the class had helped them to improve their writing skills produced the lowest non-weighted mean (Appendix B, Chart 1). Out of the 14 sections using this question, students in only two sections gave this question a score above 3.0. Students in sections that scored all or most questions higher than average, nevertheless gave this question a lower than average score. This suggests a common problem with writing skills acquisition that is independent of section-specific factors. Faculty interview data points to the inexperience of modular faculty in teaching writing and critical reading skills as an important contributory cause. However, the student interview data also reveal student attitudes (learned in high school and reinforced in college) both against writing (other than lab reports) in science classes, and against reading more widely than the text. Some students reported that they chose science classes or majors, in part, because they disliked writing and believed that less writing would be required of them in science classes. Thwarting this expectation produced complaints to their teachers and the interviewers in these early modular classes. This matched finding from both the SALG tests and the early student interviews reflects a common problem reported by faculty in early-stage evaluations of their classroom innovations. The growth of faculty competence in using "writing to learn" techniques is becoming evident in the analysis of data from the second round of student interviews in a sample of modular chemistry classes with more experienced modular teachers. In these classes, students comment on the usefulness of writing, both as a way to learn, and as an important professional skill to acquire.

Questions With Higher Than Average Scores

Table 5 lists the 10 questions with the highest scores on both types of total question means (ranked by non-weighted mean from lowest to highest). These were questions used by faculty in at least 13 sections for which non-weighted scores were 3.4 or above.[13] It should be noted that lower scores from the research university in the sample pulled down the weighted and non-weighted means from a number of otherwise good scores across the sample (e.g., the perceived contribution to learning made by hands-on activities).

Table 5. Questions for which Students Scored the Highest Learning Gains in a Sample of 18 Modular Chemistry Sections.

Question Content		Means		Difference between Non-weighted and Weighted Means
		Non-weighted	Weighted	
1-H2	Quality of contact with the TAs	3.38	3.49	.11
4-2	Understanding conceptual relationships	3.41	3.20	.21
1-D6	Written lab instructions	3.45	2.66	.79
1-K	Way class was taught overall	3.46	2.95	.51
1-D1	Class presentations	3.47	2.83	.64
1-H1	Quality of contact with the instructor	3.47	2.91	.56
4-1	Understanding main concepts	3.49	3.33	.16
1-D4	Hands-on class activities	3.51	2.77	.74
1-H3	Working with peers outside of class	3.54	3.50	.04
4-4	Relevance to real-world issues	3.68	3.24	.44

The high scores for Questions Q4-1, Q4-2 and Q4-4 are selected for commentary because of their interest to faculty in general.[14]

1. **Questions Q4-1 and Q4-2: To what extent did you make gains in understanding the main concepts of chemistry as a result of what you did in this class? To what extent did you make gains in understanding the relationship between concepts in chemistry?**

These questions are taken together because of similarity in both their content and distribution of scores. The same 14 sections used both questions, and students from the same two sections scored both questions below 3.0 (Appendix C, Charts 3 and 4). Students in the 12 other sections scored both questions above 3.0. In five sections, students scored Question 4-1 (understanding main concepts) above 3.5, with one of these sections recording scores over 4.0. For Question 4-2 (understanding relationships between concepts), students in five sections registered scores above 3.5. These high scores are consistent with the general pattern of section scores on all questions, suggesting that the results are a combination of section-specific and module-specific factors.

[13] An exception is Question 1-H2 regarding TAs, for which scores fell below 3.4 (though not by much). This question was included as a comparison to the instructor evaluation Questions 1-H1 and 1-K.

[14] Full commentary on all ten highly-scored items may be found in Wiese, Seymour, and Hunter, (1999).

Modular teachers were encouraged by these results. They also appear consistent with findings from two other sources within the evaluation. In the initial student interview round (gathered in 1996) the percentage of positive evaluative observations from students in modular classes for "seeing the relationship between concepts" was 47%, and for "understanding concepts" was 54%. Positive evaluative observations from students in non-modular courses for these items were 50% and 54% respectively. Two years later (1998) the SALG test findings indicate an increase in self-reported conceptual understanding among students in modular chemistry classes. In 1997 and 1998, Joshua Gutwill (evaluator for ModularChem) conducted controlled tests of conceptual understanding in two matched sets of modular and non-modular sections of introductory chemistry (taught by the same instructors) at two different types of institution (research university and liberal arts college). He found that modular section students out-performed those in non-modular sections on three objective measures--in-class exams, conceptual tests, and in-depth orals designed to assess scientific thinking (Gutwill, 2000). Taken together, the three studies (two based on self-reported gains, the third based on objective measures) suggest progress in students' understanding of chemistry concepts and their inter-relationships in modular classes. The validity of the SALG instrument for these items is also supported in that the data for the other two studies was independently gathered from comparable student populations also experiencing modular pedagogy.

Gutwill raises a complicating factor in reading self-report data on conceptual gains where students are experiencing new and unfamiliar forms of pedagogy. At some institutions, Gutwill found dissonance between student *perceptions* of what they had gained in understanding from the modular classes and their *actual performance*. Fewer students in the modular classes perceived that they understood the material, even though they out-performed the students in the matched non-modular classes.

Summative tests of student performance in modular chemistry classes are planned in order to clarify the extent of gains in student understanding of important chemistry concepts, and the relationships between them, in classes where well-developed and tested modules are offered by faculty who have gathered experience in using them. The degree of fit between student self-report data on their learning gains and their actual learning gains as demonstrated by formal assessments, however, looks promising. Modular chemistry users of the SALG instrument at one participating institution report high correlation between scores for SALG questions that probe student understanding of conceptual items, scores on related American Chemical Society (ACS) conceptual questions, and the instructor's examination questions exploring the same material. SALG users are invited to try this experiment for themselves.

2. Question 4-4: To what extent did you make gains in understanding the relevance of chemistry to real-world issues?

In terms of the non-weighted mean, this question produced the highest scores of any instrument question. Once again, two sections from Institution 2 (the research university) pulled down the weighted mean, and, along with section 1B, lowered the non-weighted mean (Appendix C, Chart 5). However, even students from these sections scored this question above 2.5: in one case, close to 3.0 (2.96). Students from 14 sections rated this question above 3.5; students in five of these sections scored it above 4.0. This is an encouraging finding for modular teachers who stress the relevance of the chemistry to real-world issues as a central learning objective.

This was also an early success in modular classes. Even when the first modules to be field-tested were incomplete and their teachers were inexperienced in using them, students were quick to appreciate the value of the modules' real-world applications. In the early student interview data, 81 percent of student evaluations on this item were positive, compared with 63 percent for the comparative classes.

Two Problematic Questions

1. Question 1-H1: How much did quality of contact with the instructor help your learning?

The difference between the non-weighted and weighted means for this question (.56) reflects the downward pull exerted by low scores by students from Institution 2. The distribution of scores shown in Chart 6 (Appendix C) makes this clear. Out of 18 sections, students from only four sections (section 1B and the three sections in Institution 2--all of whom scored it below 2.5) scored this question below 3.0. By contrast, students from 12 other sections rated quality of contact with their instructor as an aid to their learning at above 3.5: students from four of these sections scored this question above 4.0.

Although this is a generally positive result, it is hard to interpret. The question specifically asks about the "quality of contact" with the instructor. However, there is a danger that *any* question about "the instructor" will solicit generalized feelings about "liking" or "disliking" the instructor or the course overall. As already observed, faculty frequently comment upon this as a problem with many traditional class evaluation instruments. Some observations on this point are offered in the summary of open-ended question analysis.

Again, a degree of dissonance was noted between student self-reports of gains, and student behavior as observed and reported by their teachers in interview. Negative student comments on the class an/or their teacher that were registered in traditional departmental class evaluations, in the second round student interview data as well as in the SALG instrument responses by modular students, did not square with the observations of student behavior recorded by the two instructors who taught the three, all-modular sections in the research university (Institution 2). Although they experienced negative student feedback about the modular approach *per se*, these instructors noted that (compared with classes that they taught in more conventional fashion), the level of contact with individual students and groups of students rose dramatically. Students in the modular sections made much more frequent use of office hours, and the instructors were impressed with the high quality of the questions and discussion points that students raised in face-to-face sessions. Faculty experience does not, therefore, square with their students' scores on this question.

This mis-fit again raises the issue of whether it is worth including *any* question about "the instructor" in a class evaluation, even if it is carefully worded so as to reference the utility of particular aspects of the pedagogy for student learning. This may be especially critical in classes where teachers are introducing new forms of pedagogy. Whether scores are high or low, instructors have no way of knowing what criteria students are using to make judgments about their educational experiences. Omitting such questions is one option; asking students about the extent of their office hours contacts with the professor or TA, and for estimates of the value of these contacts to their learning, is another.

2. Question 1-K: How much did the way the class was taught overall help your learning?

The distribution of scores on this question is similar to that on Question 1-H1 (the quality of contact with the instructor) (Appendix C, Chart 7). This makes sense in that both questions ask students to evaluate the instructor's impact on their learning. The authors again debated whether or not to include this question in the instrument because it too might elicit responses that reflect how much students liked the teacher, rather than estimates of the impact of the teacher's methods on their learning. The question has been retained mainly because many faculty feel that it "ought" to be included.

Students from all but four of the 18 sections scored this question above 3.0. The four low-scoring sections are also those in which students gave low scores to Question 1-H1 (quality of contact with the instructor). Students from 11 sections scored this question above 3.5, and students from one of these sections rated it above 4.0. The close correlation between scores on the impact of class teaching overall and on the question exploring the quality of contact with the instructor reinforces the conclusion that questions about "the class overall" are apt to produce an evaluation of the teacher, rather than of course content or structure. Instrument users might consider dropping it for this reason.

Reliability and Validity of the Instrument

Reliability: A technique of measurement is said to be reliable if, when used repeatedly to measure the same thing, it consistently yields the same result. For the SALG instrument, tests for reliability would be either (1) with the same student group in the same class on different occasions, or (2) with comparable classes, each taught by the same teacher in the same way, either simultaneously, or serially.

(1) Many modular chemistry faculty have found it useful to administer the instrument midway through the class for formative feedback that allows them to make adjustments to their teaching methods. However, as a test for reliability, administering the instrument more than once with the same student group introduces at least two sources of bias: the act of completing the instrument may lead students to re-think their assessments; and in the time between the two evaluations, students are subject to influences that are likely to affect how they think and feel about the class (e.g. talking with other students or receiving their final grade).

Findings from several sources in the chemistry modules evaluation, indicate that, where the class pedagogy is new and unfamiliar, student perceptions of their learning gains change over time. For example, where students in the same modular class were interviewed in two groups at different times—the first at the end of the semester, the second early in the following semester (in a random division)—those students interviewed in the next semester assessed their learning gains significantly more favorably than those interviewed toward the end of the class.

(2) The second option is a “natural experiment” that introduces the normal variability of student groups and teacher behavior from semester to semester. Faculty who have decided to use the SALG instrument will test its reliability over time as they discover which form of which questions used in what combination gives them the clearest picture of student perceptions of their learning gains. This process will continue to be monitored in the process of refining the web-site instrument and service to users.

There is no way to control for the extraneous factors that may influence a student’s assessment of the class even when the instrument is administered only once. This is, however, true of all class evaluation instruments. Where criteria for students’ judgements are not provided, and questions focus on students’ evaluation of the teacher’s performance rather than on their learning gains is likely that there is even greater variability in the factors that influence the nature of their answers. On these grounds, we argue that the SALG instrument is more reliable than student evaluations based on unstated, mixed, or ill-defined criteria.

Although, for all of these reasons, it is not possible (in a strict sense) to test the instrument’s reliability, in practice, this does not pose a problem for potential users. By the end of the class, students will, for the most part, have formed their opinions about the value added to their learning. This may be, as already illustrated, a conservative estimate in a class with a new and unfamiliar pedagogy. Concerns about the adequacy of the academic preparation that it offers will be resolved soon after entry to a related subsequent class. In practice, the best that can be done to insure reliability is for faculty users to make the conditions under which the instrument is administered as isomorphic as possible, and/or to solicit student assessments of their gains from the current class in a subsequent class.

Validity: A related issue is whether the SALG instrument is a valid measure of learning gains as assessed by students themselves. A method for measuring a concept is valid only if it measures what it purports to measure. In this case, the instrument is valid if it measures students’ perceptions of their learning gains and not some other phenomenon.

To be valid, the measuring technique should have “content validity,” that is, it should cover as many aspects of the concept being measured as possible. There is no manner in which content validity can be “objectively” adjudicated. However, one can ask whether the full range of learning objectives that define a modular chemistry class has been covered by the SALG instrument. The instrument is argued to show strong content validity because it was designed to reflect all of the salient learning objectives that are promoted by modular chemistry faculty, and testers were free to adjust the questions to reflect particular learning objectives—hence the variations in instrument format across the tester panel. The main caveat to users, is, therefore, for instrument validity, they should include all of the learning objectives that define what is important about their class, and should resist the temptation to be brief rather than complete in their selection of questions. (This is enabled by a web-based classroom evaluation instrument. Students can complete more questions in a given time—the 46 questions in the current template takes approximately 10 minutes to complete. Students are also less conscious of the length of an instrument on-screen than on a printed page.)

It is also a common practice to compare one measurement technique to another in order to establish its validity. Selective discussion of findings from the comparative analysis of students' write-in comments from SALG and non-SALG instruments in the following section throw light on the content validity of a learning gains-based instrument.

The only manner in which the validity of any instrument can be definitively established is to test it against another measure. This type of testing is referred to as establishing a measurement's "criteria-based validity." It involves comparing the results derived from one measurement technique to those obtained from a different measure (i.e., one of which is also deemed a valid measure of the concept). As argued earlier, it is possible to judge the SALG instrument in this manner both by comparing students' test scores with how they rated their learning gains using the SALG instrument and by comparing these self-report data with those from the student interview study. Gutwill's (2000) finding that the students did better on their tests than they had predicted in both interviews and class evaluations is relevant here: the low level of confidence in the modular learning method expressed by many students was not reflected in their performance scores. This finding also reflects the observation of many classroom innovators—that student perceptions of the utility of a new pedagogy are apt to be unnecessarily pessimistic. This phenomenon makes the fit between student perception and student performance in itself dependent upon institutional culture and practices, and the socialized expectations of students.

For all of these reasons, the SALG instrument's overall validity must be assessed mainly in terms of its content validity.

Selected Findings from the Qualitative Analysis of Student Responses to Open-Ended Evaluation Questions

As observed earlier, the written comments of students are an important source of feedback to faculty who find the numeric data provided by their institutional classroom evaluation instrument unhelpful or insufficient in this regard. A qualitative analysis was conducted of all student comments gathered from completed sets of evaluation instruments of four different types from the same classes and sections represented in the sample in order to learn the nature of students' experiences of these classes from their written responses. The content analysis also served as an additional means to test the validity of the SALG instrument, by triangulating the quantitative analysis. The instruments used were:

- the SALG instrument
- faculty adaptations of the SALG instrument (including an all-write-in version from one institution)
- one instrument developed by an instructor
- institutional instruments

From a total of 1,530 completed student instruments, 779 (51%) yielded written responses. From these, 2,180 student observations were catalogued.

Students' responses were first sorted by topic (e.g., skills or knowledge gained, learning resources, lecture, lab, fit between elements of the class, the modular approach, the instructor, TAs, etc.) and by the nature of the opinions they expressed (i.e., a positive, mixed or negative experience). It was then discovered that almost all student responses from all types of instrument could also be categorized both according to whether or not they referenced learning gains, and into four main types of comment: the effects of the pedagogy, specific aspects of the pedagogy, the approach used in the class, and the quality of contact with instructor, TAs and peers. This finding suggests that faculty using the modular approach could be offered this classification as a matrix by which to categorize, count and make more use of the write-in comments that they get from the SALG instrument. The schema might also be adapted for use by other faculty groups who share a common pedagogical approach.

Global, open-ended questions were divided into types in order to discern what kinds of student responses were evoked by each type of question. The quality of the responses for each type of question was affected by whether or not they offered students criteria for their judgments, the "frame" or tone set by the rest of the instrument as a whole. Write-in comments for SALG and SALG-adapted instruments (where both the criteria for response and the instrument frame are consistently focused on learning gains) were also overwhelmingly focused on learning gains. By contrast, the common question pair, "What did you like best (and least) about the course" not only forces polarized answers, but offers students no criteria other than "liking" for their answers. The nature of student responses, therefore, varied widely. Such a question is also problematic for innovative teachers: when students are not given clear criteria for their judgments, they draw upon definitions of "the good teacher" and "the good class" that strongly reflect traditional teaching and learning methods. Student answers focus upon issues of class structure and management; whether and how these affected their learning is left to inference.

Answers to questions about the "strengths" and "weaknesses" of the course also require students to define what these terms might mean. Again, student answers reference conventional conceptions of "good teaching" that may or may not be helpful to instructors. The common pair to this question--concerning the strengths and weaknesses of the instructor—also elicits conventional generalities about how "good teachers" look: "enthusiastic," "open," "has a good attitude." This type of response offers poor feedback to instructors, but may be perceived as useful by departments and administrators. This is also the type of question that is prone to the kind of popularity poll response (also called the "Dr. Fox effect") that has undermined the utility of classroom evaluation for many faculty. What the reader can learn about student gains from the class from this type of question proved, again, to be limited. Both this, and the "like best and least" type of question, although they distinguish between "the course" and "the instructor" also produced some crossover in student responses, such that the distinction is not preserved in their answers.

Questions about the "overall" value of the course did produce student comments about their own learning gains where "value" is defined in the question as "valuable to your learning (or educational) experiences."

However, the most productive types of global questions were those which solicited advice. Solicitations for student advice (about the course, or for the instructor) offered the most direct and useful feedback because students were very clear in identifying what was "missing" for them. Students framed their responses largely in terms of needing more of something: "more explanation," "more discussion," "more hand-outs," or "more review." These suggestions provide instructors with specific information about students' unmet learning needs and what actions might be taken to address them.

Among the several versions of this type of question in the sample, the analyst documented the relationship between how questions are worded and their degree of productivity. For example, the institutional instrument question, "What advice would you give your instructor?" produced one- and two-word comments (e.g., "Good job!" "Go slower"). By contrast, the most productive of the advice questions "What suggestions do you have for us in revising this module for the next class?" (used on an adapted SALG instrument) produced extra, contextual information that explained why the writer needed "more" or "less"—for example, "More lecture would help. It seems we're expected to make connections, but we're not given the information we need to answer the questions," and, "I don't like 'discovering' things in lab. I would prefer it if you covered the purpose of the lab prior to class." The invitational, personal tone of this question seems to have contributed to its effectiveness in soliciting more considered responses than any other type of advice question.

A third version of the advice question illustrates the error of asking several questions at the same time: students only responded to the last part of "Do you have any comments about your performance, or that of other students, or the instructor that would be valuable in evaluating this course?"

Including "advice" questions in instruments provided the best feedback to the instructor of all types of global questions. However, they may be seen as less helpful by departments and administrators seeking "bottom line" statements about the "quality" of any class. A central proposition underlying the development of the SALG instrument is that its focus on learning gains better serves both the faculty's need for feedback and the institution's need for assessments of classroom quality. Unfortunately, productive though they are in other ways, global advice questions reveal little about students' learning in the class.

One adapted SALG instrument deserves special mention in that it used an 'all-write-in' format. This instrument (adopted by the chemistry department in one institution and used in three sections in the sample) produced the highest number of learning gains-related comments on both specific and overall effects of the pedagogy on student learning. Students' direct references to specific learning gains: understanding concepts, the relationship between concepts, and the relation of these concepts to other areas of science accounted for 20% (n=105) of all comments.

The instrument's questions are posed in such a way that students must reflect on their learning in order to respond appropriately: "How much did this class add to your skills in solving problems, thinking critically, or working effectively with others?" "To what extent did you make gains in understanding the concepts presented in the course, and how these concepts are related to each other." The global questions "What did you like best?" and "What did you like least," were re-framed as "What contributed most to my learning?" and "What contributed least to my learning." Responses to questions framed in terms of how elements of the class affected student learning produced specific student responses. Though students still drew upon traditional observations about elements of the pedagogy (i.e., "clear lecture presentation," and "course was well organized"), student responses to questions framed in terms of learning also reflected an understanding of their own learning process: "Labs have increased my ability to think critically about problem-solving. It also helped me understand a lot of other concepts in chemistry and how they are all critical parts of my learning." "I feel I can take apart a problem, answer those parts, and then answer the entire problem much more effectively than before." "I learned to do something every day and to practice it."

This kind of feedback is argued to be equally useful to instructors, departments, and faculty. Given that open-ended questions produce a large number of comments, it is helpful to know that responses to an all-write-in instrument focusing on student learning can be quickly sorted and counted with the use of the matrix developed for this analysis.

The request for "Other Comments" at the end of the SALG instrument also elicited responses that were highly related to student learning gains. This indicates that the exclusive 'framing' of this instrument's questions in terms of how aspects of the class affected learning strongly influenced student responses. This finding also indicates that the SALG instrument, which seeks to be a more reliable measure of student learning gains than traditional institutional instruments, does, indeed, offer a high degree of reliability in its construct validity. This is also affirmed in the findings from student responses to open-ended questions, where the highest percentages of learning gains-related comments were gathered from institutions using an adapted SALG instrument.

It would seem that the first step in building a better mousetrap is to start by asking the *right* questions. This is, perhaps, self-evident. However, the findings from both the panel test and the students' comments analysis indicate that classroom evaluations must be designed with the preferred types of student response in mind if they are to be optimally useful both for faculty seeking to improve their teaching practices, and as an indicator of classroom quality for departments and institutions. Since course evaluations are generally used for *both* purposes (as well as to inform students as consumers), it is important to understand that the quality and nature of the responses solicited are absolutely linked to the provision of clear criteria, precise questions, and an overall framework that sets the tone for the kinds of answers sought.

Development of the On-line Version of the SALG Instrument

Feedback from a Panel of SALG Web-site Users in Nine Disciplines

Much of the SALG web development was undertaken in spring and summer of 1999. The first panel test (conducted entirely with modular chemistry users) was, thus, undertaken before the instrument was available on-line. Therefore, between September 1999 and April 2000, we asked 30 faculty users of the web-based SALG instrument in nine different disciplines, and three types of institution (research university, liberal arts colleges, and community colleges) to give us feedback on:

1. How well the site worked for users in different disciplines, and what modifications were needed to accommodate them;

2. How well the instrument worked when delivered by electronic medium for:

- classes of various size
- both mid-course and end-of-course evaluation purposes;.

Information was collected both by recorded telephone interviews and by e-mail surveys with open-ended responses from faculty in the following disciplines: 9 mathematics, 6 chemistry, 4 engineering, 4 biology, 2 physics, and one each in psychology, statistics, business, and communications. The panel members varied in the number of times they had used the electronic version of the instrument over the two semesters of its availability: 6 had used it once and 10 had used it twice. By using the instrument both at mid-term and end of semester, and/or in multiple sections or classes, 5 had used it three times, 6 had used it five times, 2 had used it eight times, and one professor had used it 10 times over the two semesters. Thus, the 30 faculty had collectively used the instrument 97 times. All 30 indicated that they would be using the instrument again in the following semester.

Class Size: Class sizes in which the SALG instrument had been used varied from 12 to 434 and fell into four main groups:

- 6 large lecture classes in the 150-450 range
- 11 lecture classes in the 40-55 range
- 9 lecture/discussion classes in the 20-35 range
- 4 small classes in the 12 –20 range

Panel members reported that the web-based SALG instrument resolved the problem of monitoring the progress of students in classes whose size makes this difficult to assess by personal interaction, or by paper instruments that require hand tabulation:

> "The SALG instrument is easy to use, easily accessed through the web, provides individualized assessment tools on demand, and efficiently organizes and manages the results for large-scale enrollment." (Chemistry professor)

Two professors with very large classes (one in business, the other in chemistry) had also used the instrument to evaluate the impact of their pedagogical experiments:

> "The section of 404 students received the server-based assignments as homework, while the other section of 434 students received text-based assignments. The SALG instrument was very useful in gathering data on whether the assignments were effective in forcing students to think by a questioning process…Did the homework help them prepare for exams, focus in areas of weakness, and learn at their own pace?" (Chemistry professor)

Faculty with larger classes described data management and hand calculation as a barrier to creating assessment instruments that was surmounted by the SALG instrument's on-line statistical package:

> "The web-site is an excellent way to do assessment as the results are available for reference immediately." (Mathematics professor).

> "It's an on-line statistics handling service….This is a very innovative thing. I can construct the best feedback forms that I can imagine and somebody else will do the math. That's a big deal. I'm not normally going to give an instrument with scaled questions several tines because it takes too much time for me to crunch the numbers…It's just a question of revising the questionnaire once every other week and having the students complete it, so it's a really good way to keep a thermometer on your students." (Physics professor).

Length of Edited Instrument: There are 44 questions in the template which faculty can edit to match their own requirements. No one presented the instrument unaltered: deletions were made for features that were not part of particular classes (e.g., internet work). Additions referenced very specific activities for which faculty sought feedback:

> "It was an excellent tool for my needs. I particularly liked the ability to customize and tailor it with questions about my course on which I definitely want student input. (Chemistry professor)

The number of questions used varied from 20 to 67, but most faculty (N=24) used a total of 40-45 original, edited, or additional questions, thus retaining the two-page length of the original template.

Student Response Rates: The value of results gained by the use of any instrument is directly proportional to the response rate. For almost half of the panel (N=14), the student response rate for end-of-semester use was 100% because faculty asked their students to complete the instrument during lab or computer lab time. However, for the balance of the sample (N=16), completion of the SALG instrument was voluntary, and, in 13 of these cases (including the four largest classes in the sample), was additional to the completion of the normal institutional instrument. Response rates for voluntary completion were reported as 60% to 90% (mean= 80.3%), including multiple use of the instrument in one or more classes or sections during the same semester. For mailed surveys with up to 3 reminders, a response rate of 50% is conventionally regarded as "adequate," and 70% as "good." For classroom evaluations, *per se*, Centra (1979) argues that a response rate of at least two-thirds of the class is required for validity. This sample is, of course, small, and it would be useful to monitor the voluntary response rate with a larger sample over time. The high student response rate for this on-line instrument is, however, encouraging:

> "I have two classes of over 400 students each. Our normal class evaluation is voluntary and I got a far greater response rate using the SALG than the usual form—74% and 78%, instead of less than 50%." (Chemistry professor).

We asked panel members how soon after their request to complete the SALG instrument most students responded. For those 16 faculty who did not use class time to do this, 14 replied that students responded in a time frame ranging from "almost immediately" to within 2 days. The remaining two reported that most students had responsed within one week.

> "I announced by e-mail to my 340-plus students just a couple of hours ago that I wanted them to complete it (the SALG instrument), and I have a response from over 10% of them already." (Chemistry professor)

Most faculty set a deadline for completion—the longest being 10 days—which appears to be longer than is required.

Student Feedback on Completing the Instrument: We asked panel members about feedback that they had received from students: about the instrument itself (including its length), about answering it on-line, and their concerns (if any) about the confidentiality of their answers. Student feedback to faculty on the time they had taken to complete the questionnaire ranged from 15 to 25 minutes. Faculty reported no complaints about instrument length, but some about completing both the institutional and SALG instrument. Some faculty in the first panel test edited the paper instrument down to a single page because they feared a low response rate for a longer (voluntary) instrument. However, on screen, where students see just a few questions at a time, any negative effect created by presenting a two-page (or longer) instrument appears to be reduced. There was agreement across the panel that, although their edited versions of the SALG instrument were (in all but 3 cases) longer than their (paper or Scantron) institutional instrument, students had completed the on-line instrument in shorter time and with less complaint. This discovery was allowing faculty to be more specific in the questions they asked:

> "It was longer than the usual instrument, but they were all pleased that it took a shorter amount of time" (Chemistry professor).

Overall, students appeared to prefer the on-line mode of feedback to faculty:

> The students prefer the computer over paper." (Mathematics professor).

> "I got no complaints. Everyone was open to it, and my impression was that they liked doing it on the computer. They were all very comfortable with the instrument." (Chemsitry professor)

> "They like answering on-line better than the usual Scantron forms." (Biology professor)

Faculty were asked whether students reported any difficulties, either with gaining access to a computer, or with using this particular program. Neither were, apparently, a problem, with one exception: where students are assigned a random identification number (which is one of the two options for securing confidentiality described in Appendix B), but enter it incorrectly, one student may find that another student has already completed "their" form.

Complete confidentiality is inherently difficult to achieve with on-line instruments. However, some studies of institutional instruments have revealed that confidentiality in students' responses is sometimes breached by faculty in the administration of these instruments (CITE), and that students may be identified by their hand-written comments:

> “The SALG actually provided more anonymity than the Scantron forms because I receive the original hand-written forms. Apparently the secretary in our department doesn’t type up the comments, as is the practice in other departments. So, I can guess who wrote which Scantron comment far more easily than I can guess who wrote which SALG comments.” (Engineering professor)

As we describe in Appendix B, two methods are offered for the assignment of random student identifiers. However, we have, as yet, found no absolute means to protect student identity. We, therefore, asked if faculty had received any expressions of concern from students about the anonymity of their answers. Although none had been received, users should be aware of this issue while the search for technical means to secure anonymity continues.

Faculty did not report any student feedback about the nature of the questions asked—perhaps because they were not asked to offer such a critique, and do not expect to give one. However, faculty reported strong positive feedback on their inclusion of a number of open-ended questions about issues specific to their class to which students could type in answers. Open-ended questions are a common feature of traditional classroom evaluation instruments, and faculty have reported to us, and to other researchers, (CITE) that they habitually make more use of these observations in evaluating and revising their courses than the numeric scores they receive:

> “Students’ comments are very important to me. Through the years, I have learned a lot from statements made by students, and believe that their comments contribute to my growth as an instructor…I listen, and I do make changes on a regular basis. They usually have good advice—for example, one piece of advice was to more clearly state what material is to be handed in. There was much less confusion when I put everything in writing and they knew exactly what was expected of them.” (Chemistry professor)

However, students are not, necessarily, aware of the value that faculty place on their comments and suggestions. Indeed, as we found in earlier research (1997, Ch.3), students are commonly cynical about the utility of offering feedback to their professors, and doubt that what they say will be either read or heeded. We were, therefore, surprised to learn that students made more use of this option than in conventional instruments:

> “I felt the students appreciated the opportunity to express their opinions.” (Mathematics professor)

> “They especially liked the fact that they could type in their own comments.” (Biology professor)

> “I get more detailed information about what worked and what didn’t work in the course in a user-friendly way. And I can’t tell you how delightful it was not having to decipher bad hand-writing” (Chemistry professor with approximately 170 students)

An added value was that the instrument appears to encourage students to reflect upon their own learning:

> "The students' comments on the web were lengthier, more reflective, honest, and critical than the ones I have ever received before using paper instruments." (Chemistry professor with two sections of 434 and 404 students)

Greater student optimism about the value of offering on-line observations to their professors may be a function of the shift toward increased interaction between faculty and students enabled by e-mail communication. Why students appear to have greater faith that faculty will take more note of electronic, rather than hand-written, suggestions for the improvement of their courses is an interesting question. However, it is useful to know that they do.

What Faculty Learned: This was not a random sample of faculty. As their search for and use of this alternative evaluation instrument suggests, all of the panel members are interested in using student feedback to improve student learning in their classes, and are actively engaged in trying out different teaching methods. They described students' hand-written comments in traditional instruments as their main source of feedback about what might improve their courses. One common reason offered for trying out the SALG instrument was the hope that they would learn more than they usually do from questions to which numeric scores are given. All panel members reported that the switch in emphasis from student assessments of their performance to estimates of their own gains provided them with useful numeric information that was specific to their class goals and activities in addition to open-ended commentary:

> "I think this format us so much more positive. The students tell me what helps them and what does not, as compared with our usual assessment form. (Biology professor).

> "It was very useful in gathering data on whether they had improved their understanding of the material." (Chemistry professor.).

> "I learned what they learned, and that is very valuable—rather than how they view an instructor's effectiveness. " (Biology professor).

> "The instrument was a positive tool for my information about what the students were learning." (Chemistry professor)

> "It will help you rethink the way you structure your homework, lab, and tests." (Physics professor).

Information gathered using the SALG instrument was also regarded as more reliable, and thus more useful in negotiating changes in teaching methods with colleagues than student feedback from others sources:

> I have gained a great deal because I can see where to make changes in order to support student learning. I can also use the information as leverage to show the lab instructors how they are important for overall learning. There was a problem with one of them and it showed up in the SALG results. I must have one voice that we speak so that student

> confusion is held to a minimum. I like the positive tone of the instrument versus the adversarial one of many feedback surveys."" (Biology professor)

As indicated earlier, 23 of the panel members had used the instrument both at the end of the semester and at mid-points in their classes. Ease of web-site use and immediate statistical feedback made it possible to use the instrument for formative as well as summative purposes. Students were able to see the value of giving feedback to their teachers as aspects of their current course were reconsidered and adjusted:

> "I gave the evaluation before the exam and first grade review. The students were asked to do it again after mid-term grades and the results changed slightly. The students were very open and honest." (Chemistry professor).

> "I got good mid-term feedback that helped me restructure some of the assignments." (Chemistry professor)

> "I used it mid-term to make some adjustments in how the course was being taught." (Engineering professor)

> "Some topics will need to be adjusted, but, for the most part, I could see that the students were learning." (Chemistry professor)

There were a number of caveats, however. Some faculty (correctly) observed that student evaluation instruments, by their nature, cannot assess actual learning gains, and are limited (in the words of one chemistry professor) to "student perceptions of their understanding." However, they saw these estimates as valid and useful as a supplement to what could be learned through graded assignments and tests.

Some faculty wondered if there was merit in including some traditional questions on faculty performance, although their reasons for this view (and which performance questions they saw as valuable), were not offered. As we describe in the previous section on open-ended questions and answers, we have found that the focus of any instrument shapes the kinds of responses it elicits. We intend, therefore, to retain the exclusive focus on student "gains" unless good arguments for introducing teacher performance questions are forthcoming.

We asked panel members about the usefulness of answers to the general question" How did the way that this class was taught overall help your learning?" and whether we should retain it. It was reported to be less informative than questions on specific aspects of the class, however, the majority advised retaining the question as many faculty (and their colleagues or departments) are accustomed to expect such a question:

> "I think it is too sweeping a question compared with the rest of the questions, making it more of a popularity judgment if you are not careful. However, I did use it" (Biology professor).

.

Lastly, panel members made suggestions for other types of questions they would like to see added, and that we are currently considering. These include: a general question about what changes the teacher could make that would help them gain more; comparative assessment of the gains from two different teaching methods; and a section requesting demographic information (e.g., year in school, major, ASK SUE…).

Flexibility of the Instrument: We were concerned to know whether an instrument that was originally developed for faculty in one particular science is flexible enough for use by faculty in other sciences, mathematics and engineering, and in non-science disciplines.

For its original use in modular chemistry classes, we made it impossible for faculty to edit out questions on some issues that our research findings indicated were critical to student learning in both more and less traditional classes. Important among these were, how well the different aspects of the class fitted together into a coherent, comprehensible whole—e.g., class and lab goals and activities, curriculum and text, materials presented and materials tested. Panel members reported that coherence continues to be an important issue, particularly when aspects of class pedagogy or structure are changed, so this "required" question will be retained. However, we found other (originally required) questions that we needed to make optional, e.g., questions about labs, the use of text books, and exams—which may not be relevant to particular classes or disciplines. More recent users have appreciated these changes:

> "The required university forms always seemed inappropriate for my courses. I found the SALG a relief in that I could easily delete questions about things that I don't use in my class—like labs and textbooks." (Engineering professor).

We were interested to learn that a template of suggested questions across a wide range of possible class goals and activities is helpful to faculty in different disciplines in that it provides a starting place for thought. It gives those faculty who are not accustomed to spelling out their class learning objectives some basic ideas from which to consider what they most want students to get out of their class, and how well the content, methods, and resources serve these ends:

> It made us think through our course objectives again, which is usually a good thing for the course, no matter that the responses are." (Communications professor)

All panel members reported that the instrument was flexible enough to adapt to different learning goals, class content, and activities:

> "It's more flexible, targeted, and specific than other student evaluations instruments that I have worked with." (Mathematics professor).

> "I really liked the dynamic nature of the questionnaire, and the ability to change the questions quickly to fit a new situation. I was glad to have the sample questions as a guide, and I rewrote virtually all of them." (Engineering professor)

"The editing was very easy and took less than an hour…I used it for the modules, but I think it would be good for any course. I have recommended it to colleagues in math and physics.." (Chemistry professor)

"Of all the assessments, I like the SALG best, and it's the easiest to administer." (Chemistry professor)

Web-site development: Practical feedback, both from panel members and from other users, about specific difficulties—either in instrument structure, or in the way the web-site worked—has allowed us to improve the faculty editing process and to address other difficulties with use of the web-site, largely through extensive e-mail discussion. For example, users requested, and have responded positively to the option to insert an invitation to students to type in their observations after any question. Although the web-site looks very similar to the original version, a number of other functions have been added:

- the capacity to view student comments, either by themselves, or embedded within the numerical data
- two options by which to protect students' anonymity—either by requesting the system to automatically assign student identifiers, or to assign identifiers outside of the system
- a "guest button" that allows potential users to visit the site without registering a course.

Web development also involved optimizing, cleaning, and rewriting code (in SQL 7.0 language) in order to optimize processing time.

In light of our finding from the analysis of students' open-ended comments, that such comments can be coded, classified, and counted by faculty users with the help of an analytic matrix, we are also considering the addition of such an option to the site. This would allow those faculty who would like to do more than simply read what students write a way to quantify the weight of additional student observations on particular issues. We canvassed the 30 panel members as to whether they would find such an option useful: 18 indicated that they would use this option if it were available. It would have the advantage of giving the same weight to observations making the same point, no matter how strongly, eloquently, or amusingly individual comments were expressed.

Current and Future Users: Currently the SALG web-site has 170 registered users, representing approximately 100 institutions and a range of disciplines. It is visited, on average, 12 times a day. It is known to be linked to 12 other web-sites, including the Field-tested Learning Assessment Guide (FLAG).[15] The on-line version of the SALG instrument is set up for use by single faculty for direct personal feedback from their students. However, from its early appearance as a paper-and-pencil instrument, it has been adapted for use by groups of users

[15] The concept for the FLAG, and its prototype, were also collaboratively developed as a response to the classroom assessment needs of innovative faculty in the chemistry consortia and other science and mathematics reform initiatives (http://www.wcer.wisc.edu/nise/cl1/flag) by Elaine Seymour, in collaboration with Susan B. Millar and Steve Koskuik--both of the New Traditions Chemistry Consortia. The FLAG, which is a web-based repository of classroom-generated and tested science and mathematics assessment methods, was further developed (and is now edited and maintained) by the National Institute for Science Education, the University of Wisconsin-Madison.

seeking comparative estimates of students learning for a variety of purposes. Group users who are in current discussion with the instrument author include: 4 higher education projects that are using the SALG instrument as part of their evaluation strategies, and a further 2 projects who have made the instrument available to their participants for individual use; 3 academic departments or divisions using the instrument for end-of-semester faculty evaluation purposes; 2 learning centers testing the instrument for cross-campus use; and 1 center of distanced learning who are using it for student feedback across different courses.

We are now discussing, therefore, the technical possibilities that would allow us to offer a separate version of the on-line SALG instrument for groups of users.

Acknowledgements: Finally, we would like to thank those who have helped up to develop, test, and disseminate the SALG instrument, both in its original form, and as an on-line instrument:

- ChemLinks and ModularChem Chemistry consortia, and their funders, the NSF Division of Undergraduate Education. The SALG instrument was originally developed as part of a set of formative evaluation strategies to meet the needs of chemistry faculty introducing modular chemistry into their undergraduate classrooms. Participating faculty and departments were its first users, and the first panel tests were conducted with a sample of "modular chemists" as part of the consortia evaluation. These two consortia also contributed funding for the development of the SALG web-site and offered it to other users through their web-sites, workshops, and conference presentations.

- The Exxon-Mobil Corporation, who have provided two grants to enable the development of the on-line version of the SALG instrument, and two panel tests. They have also encouraged its use by project groups funded by Exxon-Mobil as part of their evaluation strategies.

- A third chemistry consortium, Chemistry New Traditions, based at the University of Wisconsin-Madison, also contributed funding for the development of the SALG web-site, and its participating faculty have been active in testing its use and providing feedback to the developers

- The National Institute for Science Education, also at the University of Wisconsin-Madison, has funded presentations of the SALG web-site, included the on-line instrument within the FLAG web-site, and recently created a descriptive brochure for its wider dissemination.

- The Director, Susan B. Millar, of the Learning through Evaluation, Adaptation, and Dissemination (LEAD) Center at the University of Wisconsin-Madison has also contributed to the development of the on-line SALG instrument, and has made both local and national presentations on its use.

Ref**rences**

Gutwill, Joshua P. (2000) "The impact of active and context-based learning in introductory chemistry courses: an early evaluation of the modular approach." *Journal of Chemical Education (forthcoming).*

Seidel, John V. (1998). *The Ethnograph, V. 5: A User's Guide*, Qualis Research for Scolari Sage Publications, Thousand Oaks, CA.

Seymour, Elaine and Nancy Hewitt. 1997. *Talking About Leaving: Why Undergraduates Leave the Sciences.* Westview Press: Boulder, CO.

Wiese, D., Seymour, E., and Hunter, A.B. (1999). "Report on a panel testing of the student assessment of their learning gains instrument by faculty using modular methods to teach undergraduate chemistry." Report to the Exxon Education Foundation. Bureau of Sociological Research, University of Colorado, Boulder.

Appendix A

Appendix B

Web-Site Tour

Instructor Interface

The features of the instructor side of the SALG web-site are presented as a walk-through of the sequence for first-time users.

Step 1: Registering: Although faculty interested in the site may simply view its features, to use the SALG web-site services, an instructor needs to register one or more courses. This involves completion of an on-line form that requires an e-mail address, password, and other information that allows the web-site to remember instructors and their associated courses. When revisiting the site, instructors need only enter their e-mail addresses and passwords to gain access to their own course information.

Step 2: Adding a new course: Once registered, the instructor can add a course. This involves a second on-line form asking for information that will allow students and instructor to identify the course: course name, number, department, institution, semester, and year. Additionally, the instructor supplies a course password, and the system assigns a unique identifying course number. The course ID and password are used by students when they log into the system to complete the SALG instrument.

Step 3: Finding an appropriate version of the SALG instrument: When a course is added to the system, it automatically provides instructors with a template SALG instrument that they can modify to their own course needs. As already described, this "generic" template reflects the needs and use by previous users, beginning with chemistry faculty using modules. Instructors may also browse a set of variations of the SALG instruments that have been developed by other instructors and may adapt one of these as a base from which to work.

Step 4: Tailoring the SALG instrument to the course: Working with the chosen template, the instructor tailors the SALG instrument to their own course. All of this work is done on the web-site, using on-line forms. Instructors can add, edit, or delete any course aspects listed in the template; they may also place explanatory text-boxes at the end of any question. These changes are made immediately and are visible on-screen. To keep the focus of the on-line SALG instrument on learning gains, instructors cannot change the general structure of the instrument, the labels or size of the instrument scales, the five major questions, or the categories within them.[16]

Step 5: Assigning student identifiers: The instructor then has the option of assigning unique identifiers to students, or having the system assign these automatically. Each approach has its own strengths and weaknesses, as follows:

A. *Identifiers external to the web-site.* When instructors assign unique identifiers to

[16] As the instrument has begun to be used by faculty in disciplines that do not include laboratory work, this category will be made optional.

students, they are prevented from answering the SALG instrument more than once. This process is done *outside* the system, that is, the instructor determines a set of ID values (code names and numbers) and distributes them to students. Identifiers must be alphanumeric strings with less than 15 characters. The instructor can either *randomly assign* these identifiers to students, or can assign a *specific identifiers to specific students*. The instructor can then view a list of IDs for those students who have completed the instrument at any stage. However, the web-site does not link these IDs with the data.

B. *Identifiers as an automatic part of web-site.* The instructor can also ask the system to assign identification numbers automatically. This option saves time, both in and out of class. One weakness of this approach is that students could complete the SALG instrument multiple times, since the system will assign new identification numbers each time they log in.

Step 6: Tell the students: Once the issue of identifiers is resolved and the SALG instrument is ready, students need to know where to go and what they need to do to complete the instrument. Students complete the instrument from the student-side of the web-site, at http://www.wcer.wisc.edu/salgains/student/. From that URL, students must enter the course ID and password. If the instructor has assigned unique identifiers, they also need to enter these. To make it easier to check how many students have completed the instrument, the instructor should assign a date by which the instrument should be completed.

Step 7: View the results: Once the students have completed the SALG instrument, instructors can view the responses in multiple ways. They can see who has responded, although this information is not linked with student responses. They can also get a frequency distribution, with averages and standard deviations of the responses to all questions, and raw scores. They can also download the raw and tabulated data in tab-delimited form.

Step 8: Finishing up: When the analysis is done and feedback has been received, instructors can delete their courses, delete student responses, and (if they wish), offer their version(s) of the SALG instrument for others to use.

Student Interface

For students, completing the SALG instrument on-line is a linear process. First, students register their course, using the course ID and password. If faculty have chosen to assign unique identifiers, then student also enters that number in order to access the instrument. Where the system assigns identifiers, students may complete the instrument after successful entry of the course ID and password alone. Upon completion, each registered student is notified that the data has been saved.

Software and Hardware Used

The site is completely server-based and lives on a SQL 7.0 server. It is an extensively database-driven application and almost every page calls the database for some sort of information--either about the user, the course of the user, or the student data associated with the course. The scripting language which interfaces between the web pages and with the database is Active Server Pages (ASP) 2.0, which was developed by Microsoft for that purpose. The benefit of developing a server-based web application is that any computer, using any operating system or browser, can access the web-site.