Education Tech Research Dev (2009) 57:705–720
DOI 10.1007/s11423-007-9079-9

DEVELOPMENT ARTICLE

College student perceptions of teaching and learning quality

Theodore W. Frick · Rajat Chadha · Carol Watson · Ying Wang · Pamela Green

Published online: 11 December 2007

Abstract Numerous instructional design models have been proposed over the past several decades. Instead of focusing on the design process (means), this study investigated how learners perceived the quality of instruction they experienced (ends). An electronic survey instrument containing nine a priori scales was developed. Students responded from 89 different undergraduate and graduate courses at multiple institutions ($n = 140$). Data analysis indicated strong correlations between student self-reports on academic learning time, how much they learned, First Principles of Instruction, their satisfaction with the course, perceptions of their mastery of course objectives, and global course ratings. Most importantly, these scales measure principles with which instructional developers and teachers can evaluate their products and courses, regardless of design processes used: provide authentic tasks for students to do; activate prior learning; demonstrate what is to be learned; provide repeated opportunities for students to successfully complete authentic tasks with coaching and feedback; and help students integrate what they have learned into their personal lives.

T. W. Frick (✉) · R. Chadha · C. Watson · Y. Wang · P. Green
Department of Instructional Systems Technology, School of Education, Indiana University Bloomington, Bloomington, IN, USA
e-mail: frick@indiana.edu
URL: http://www.indiana.edu/~tedfrick

R. Chadha
e-mail: rchadha@indiana.edu

C. Watson
e-mail: watsonc@indiana.edu

Y. Wang
e-mail: wangying@indiana.edu

P. Green
e-mail: pagreen@indiana.edu

Keywords Course evaluation · Higher education · Student ratings · Student learning achievement · Student satisfaction · Academic learning time · Student engagement · Instructional quality · Pattern analysis · First principles of instruction · Instructional theory

Problem

Numerous instructional design theories and models have been proposed in the literature (Reigeluth 1983, 1999; Visscher-Voerman and Gustafson 2004). Eight models and theories were described by their authors in Reigeluth (1983). Twenty-three were discussed in Reigeluth's second volume (1999). Visscher-Voerman and Gustafson (2004) indicated that Andrews and Goodson (1991) and Gustafson and Branch (2002) had evaluated 40 such models. Visscher-Voerman and Gustafson (2004) did a qualitative study of 24 design practitioners to see whether there is a gap between theory and practice. They analyzed and compared design processes that these practitioners used and documents they produced. These practices were grouped by Visscher-Voerman and Gustafson according to instructional paradigm or philosophy: instrumental, communicative, pragmatic and artistic.

Presumably, these various design theories and practices must be working to some extent, or else instructional designers would not be using them. Do any of these theories or practices result in higher student learning achievement—i.e. result in more effective instruction?

Maccia (1987) proposed an epistemology in which he described three kinds of cognition: (1) knowing that, (2) knowing that one, and (3) knowing how (cf. Estep 2003). He identified six kinds of know-how (cf. Frick 1997). What is noteworthy in Maccia's categories of know-how is the distinction between *means* and *ends*. Some ends can be achieved only one way, such as opening a combination lock. There is only one way to do it correctly. Other ends can be achieved in different ways, such as multiple routes to a distant city—e.g. by air, by auto, or by train. Instructional design theories and models are means to an end; they are prescriptions for how to design instruction. The end is an instructional product, a workshop, course, courseware, etc. Clearly, there is more than one means to such an end, as evident in the numerous instructional design theories and models.

Rather than focusing on which design process is more effective or preferable, Merrill (2002) identified commonalities among these various means, which he referred to as First Principles of Instruction:

> The premise ... is that there is a set of principles that can be found in most instructional design theories and models and even though the terms used to state these principles might differ between theorists, the authors of these theories would agree that these principles are necessary for effective and efficient instruction. This premise also assumes that these design principles apply regardless of the instructional program or practices prescribed by a given theory or model. If this premise is true, there will be a decrement in learning and performance when a given instructional program or practice violates or fails to implement one or more of these first principles. (p. 44).

Merrill (2002) identified five First Principles:

(1) learning is promoted when learners are engaged in solving real-world problems;
(2) learning is promoted when existing knowledge is activated as a foundation for new knowledge;

(3) learning is promoted when new knowledge is demonstrated to the learner;
(4) learning is promoted when new knowledge is applied by the learner;
(5) learning is promoted when new knowledge is integrated into the learner's world. (pp. 44–45)

Merrill (2007, in press) has subsequently suggested a "pebble in the pond" approach to instructional design in which learners are engaged in a series of simple-to-complex, whole, real-world tasks. This approach is yet another means to an end, but based on application of First Principles of Instruction and to some extent the Elaboration Theory of Instruction (Reigeluth and Stein 1983).

The present authors empirically investigated Merrill's claim by asking students about their learning experiences in college courses. Instead of focusing on the means (the design process), we have focused on the ends (student perceptions of the quality of teaching and learning). As an initial study, we asked students about their experiences of these first principles of instruction in college courses they had recently taken or had nearly finished. We also considered Kirkpatrick's (1994) levels of evaluation, which have been used for more than five decades in non-formal educational settings such as business and industry. Those four levels of evaluation are: (1) learner *satisfaction* with the training, often referred to as a "smiles test" or reaction, (2) *learning*, (3) *transfer* of learning to the learner's job or workplace,[1] and (4) *impact* on the overall organization to which the learners belong. With respect to Level 2 (student learning), the current study utilized students' *self-reported* grades, their perceptions of their learning progress (how much they learned) and mastery of course objectives. Cohen (1981) and Kulik (2001) have indicated that many studies have identified positive correlations of self-reports with objective assessments in college courses. In Cohen's (1981) meta-analysis of 41 independent studies of 68 multi-section courses, the average correlation between self-reports of student learning progress and their learning achievement on objective common exams was 0.47. Finally, we asked students to report on their academic learning time—their *successful* engagement in tasks, activities and problems that were related to course objectives (cf. Berliner 1990; Fisher et al. 1978; Squires et al. 1983). Our research question was: What is the relationship of these factors with First Principles of Instruction?

Method

A survey instrument was constructed containing items targeting academic learning time (ALT), Merrill's First Principles, and Kirkpatrick's levels of evaluation (1 and 2), in addition to *global* ones similar to those reported in Cohen (1981), which indicated overall ratings of the course and instructor. Each set contained five[2] items intended to measure the respective construct (scale).

The paper version of the instrument was reviewed by six faculty members. Based on their feedback, wording of items considered to be confusing or ambiguous was modified. Items were randomly ordered for the nine scales and the instrument was converted to a Web survey, which can be viewed online at: http://www.indiana.edu/~edsurvey/evaluate/.

[1] It should be also noted that Kirkpatrick's Level 3 is very similar to Merrill's Principle 5 (integration). We did not attempt to measure Kirkpatrick's Levels 3 and 4.

[2] It should be noted that a minimum of two items is needed to determine internal consistency of a scale (Cronbach's α) on a single instrument administration.

Between May 2006 and January 2007, 156 respondents participated in the survey. Respondents were volunteers who learned about the survey primarily through e-mail from their instructors at several universities, and through e-mail to student organizations. Before beginning data analysis, a total of 16 cases were eliminated because they were either test cases or contained no data, leaving a total of 140 cases for analysis.

For data analysis, the Type I error rate was $p < 0.05$ for this study. Since we did 100 statistical tests, each particular statistical test was only considered significant if $p < 0.0005$. This was to prevent inflation of the Type I error rate when so many statistical tests were performed ($1 - (1 - 0.0005)^{100} = 0.0488$; cf. Kirk 1995, p. 120).

Results

Data indicated that respondents evaluated a wide range of courses with relatively few respondents from any given course. A content analysis of qualitative responses (text) to the question about the course title or content was conducted. Eighty-nine different subject areas were mentioned by 130 respondents. While courses in business (33), medicine (22), education (18), and computers and technology (13) appeared more frequently than others, a wide range of subjects was represented. In addition, students indicated that they were, by and large, taught by different instructors. This is consistent with the wide range of course topics.

In Table 1, it can be seen that 93 females and 43 males responded to the survey (4 did not report gender). While it may appear that a disproportionate number of females responded, for the scales investigated in this study, there were *no* significant associations between gender and academic learning time, learning progress, student satisfaction, first principles, grades, and course ratings. Additionally, there were no significant associations between gender and other demographics.

Table 1 also displays responses of students with respect to the grade they either expected to receive in the course they were evaluating or which they did receive. Approximately two-thirds got A's and about 21% B's. It is unclear from these data alone whether this may be evidence of grade inflation, or an indication that those respondents who were higher achieving students were more likely to respond to the survey than lower achieving students.

Since grades were not anticipated to be a discriminating factor among respondents, they were also asked: "With respect to achievement of objectives of this course, I consider myself a ____." Choices were master, partial master and nonmaster. Table 1 indicates that only 25% of respondents reported that they had mastered course objectives, even though 87% received A's or B's (see above). About 62% of students considered themselves as "partial masters" of course objectives and 12% reported themselves as "nonmasters".

Relationships among variables

Gender was not significantly related to overall course rating,[3] expected or received grade,[4] mastery level,[5] or to class standing.[6] The chi square between gender and mastery level

[3] 2 = great, 1 = average, 0 = awful

[4] 4 = A, 3 = B, 2 = C, 1 = D, 0 = F

[5] 2 = master, 1 = partial master, 0 = nonmaster

[6] 5 = graduate, 4 = senior, 3 = junior, 2 = sophomore, 1 = freshman, 0 = other

Table 1 Descriptive statistics on respondent and course demographics

	Frequency	%
Gender		
Female	93	66.4
Male	43	30.7
Missing	4	2.9
Total	140	100.0
Class Rating: I would rate this class as:		
Great	81	57.9
Average	44	31.4
Awful	13	9.3
Missing	2	1.4
Total	140	100.0
Expected Grade: In this course, I expect to receive (or did receive) a grade of:		
A	92	65.7
B	30	21.4
C	6	4.3
D	1	0.7
N/A or Don't know	10	7.1
Missing	1	0.7
Total	140	99.9
Progress: With respect to achievement of objectives of this course, I consider myself a:		
Master	35	25.0
Partial master	87	62.1
Nonmaster	17	12.1
Unknown	1	0.7
Total	140	99.9
Class Standing: I am a:		
Freshman	23	16.4
Sophomore	19	13.6
Junior	23	16.4
Senior	19	13.6
Graduate	48	34.3
Other	7	5.0
Missing	1	0.7
Total	140	100.0
Course Setting: I took this course:		
Face to face	97	69.3
Blended	8	5.7
Online	34	24.3
Missing	1	0.7
Total	140	100.0

approached significance ($\chi^2 = 6.27$, $df = 2$, $p = 0.043$, $n = 136$). Slightly more males considered themselves to be masters than expected, and slightly fewer females considered themselves as masters than expected if there were no relationship. There was a weak relationship between gender and class standing. An ANOVA was performed, resulting in $F = 21.94$, $df = 1{,}134$, $p = 0.004$. The average male was a senior (mean = 3.81) and the average female a junior (mean = 2.96). A chi-square analysis indicated that a few more males were graduate students than expected, and a few more females were freshmen than expected. However, it was not significant ($p = 0.135$).

There was a significant association between class rating and mastery of course objectives ($\rho = 0.319$, $p < 0.0005$, $n = 138$). Students who considered themselves masters of course objectives were more likely to rate the course as "great". There was also a significant correlation between student reports of mastery level and self-reported course grades ($\rho = 0.373$, $p < 0.0005$, $n = 129$). It is also noteworthy that students' expected or received course grades were very weakly associated with their ranks of overall course quality ($\rho = 0.192$, $p = 0.030$, $n = 128$, $\rho^2 = 0.037$). None of the remaining associations among these variables was statistically significant.

Scale reliabilities

To determine the reliability of each scale, all 5 items in each scale were initially used to compute internal consistency with Cronbach's α coefficient. Items that were negatively worded had their Likert scores reversed. Since one of the goals was to reduce the number of items (to make the tool more efficient) while maximizing scale reliability, items were removed from each scale one at a time until no item could be removed without decreasing the α coefficient (Table 2). As is standard procedure for scale construction, the reliability reported for each scale is that with the remaining items. Further detail about items and reliability analyses is provided online at: http://education.indiana.edu/~frick/TALQs.pdf and Appendix A.

To examine the relationships among the scales themselves, the five First Principles were combined into a single scale score for analyses.

It can be seen in Table 3 that the combined First Principles scale correlated very highly with ALT, perceived Learning Progress, Satisfaction, and the overall Instructor/Course

Table 2 Summary of scale reliabilities

Scale	α
Academic Learning Time	0.85
Learning Progress	0.97
Global items from BEST form	0.92
Learner Satisfaction	0.94
Authentic Problems	0.81
Activation	0.91
Demonstration	0.88
Application	0.74
Integration	0.81
First Principles (combined)	0.94

Table 3 Spearman's ρ correlations among scales

		First principles	ALT	Learning progress	Satisfaction	Mastery	Class rating	BEST rating
First principles	ρ	1.000						
	N	114						
ALT	ρ	0.682**	1.000					
	N	111	137					
Learning progress	ρ	0.823**	0.602**	1.000				
	N	110	128	131				
Satisfaction	ρ	0.830**	0.515**	0.874**	1.000			
	N	112	132	128	135			
Mastery	ρ	0.341**	0.470**	0.301**	0.361**	1.000		
	N	113	136	130	134	139		
Class rating	ρ	0.735**	0.496**	0.760**	0.853**	0.319**	1.000	
	N	112	135	129	133	138	138	
BEST rating	ρ	0.867**	0.605**	0.759**	0.859**	0.386**	0.799**	1.000
	N	112	134	128	132	135	134	136

** Correlation is significant ($p < 0.0005$, 2-tailed)

Rating (BEST). The correlation with perceived student Mastery was less strong but nonetheless highly significant. Note that the Class Rating and BEST Rating (global items) are indicators of perceived overall quality by students and are also correlated very highly. As expected and consistent with past research, ALT is correlated significantly with perceptions of Learning Progress. ALT is positively correlated with Mastery Level and with Course Grade (not shown in Table 3). The ALT scale is also correlated with the BEST Rating (global items of overall course and instructor quality), with overall Class Rating, and with Learner Satisfaction.

These results are very strong as a group. ALT is correlated positively and significantly with student's self-reported Learning Progress, which is consistent with past studies of ALT. ALT is also correlated positively and significantly with students' Mastery of course learning objectives. This can be interpreted to mean that students who agreed they frequently engaged successfully in problem-solving and learning tasks also agreed that they mastered course objectives and received (or expected to receive) a high grade. Furthermore, those participants also agreed that theirs was an excellent course and they were very satisfied with it.

Additionally, the evidence supports a strong relationship between ALT and First Principles of Instruction. Students who agreed that First Principles were used in the course also agreed that they were frequently engaged in successfully solving problems and doing learning tasks. It appears, from a student's perspective, First Principles being used in a course is associated with high quality instruction, students' satisfaction and their perception of learning a lot (Learning Progress). There are many other highly significant and strong correlations in Table 3. Space precludes further discussion here.

It is noteworthy that none of the demographic variables (in Table 1) was related to measures reported in Table 3. Univariate ANOVA's were performed with Gender, Class Standing, and Course Setting as independent variables and with the measures in Table 3 as dependent variables. None of the F values was significant at $p < 0.05$. Thus, perceived use of First Principles, ALT, Learning Progress, Satisfaction, Mastery, Class Rating or BEST

Table 4 MAPSAT results for the pattern: If *ALT* and *First Principles*, then *Learning Progress*

	ALT Agreement			
	No		Yes	
	First Principles Agreement		First Principles Agreement	
	No Learning Progress Agreement Count	Yes Learning Progress Agreement Count	No Learning Progress Agreement Count	Yes Learning Progress Agreement Count
No	15	1	5	1
Yes	7	9	5	64

Rating do not differ according to gender, to year in school, or to whether the course was taught face-to-face or online.

Pattern analysis (MAPSAT)

The Spearman ρ coefficients indicate correlations of ranks of ordinal measures. While there were numerous highly significant relationships, the specific patterns that show temporal relations among three or more variables are not shown. For example: *If* students agreed that ALT occurred during the course, *and if* they also agreed that First Principles occurred during the course, *then* what is the likelihood that they agreed that they learned a lot in the course (i.e., Progress)? This is a temporal pattern. MAPSAT (Mapping and Analyzing Patterns and Structures Across Time) is an alternative way of approaching data analysis (Frick 1990, 2005; Frick et al. 2006).[7] In the present study, we wanted to know: if ALT and First Principles occur, then what is the likelihood that students will learn a lot, master course objectives, or feel satisfied with their instruction?

In Table 4, results are presented for the MAPSAT pattern: If student agreement with ALT is Yes, and if student agreement with First Principles is Yes, then student agreement with Learning Progress is Yes?

It can be seen in Table 4 that there were a total of 65 occurrences of the antecedent condition. Given that the antecedent was true, the consequent (student agreement with Learning Progress is Yes), "followed" in 64 out of those 65 cases, which yields a conditional probability estimate of 64/65 or 0.985. Next we investigated: If student agreement with ALT is No, and if student agreement with First Principles is No, then student agreement with Learning Progress is Yes? It can be seen that the antecedent occurred a total of 22 times, and the consequent occurred in 7 out of those 22 cases, for a conditional probability estimate of $7/22 = 0.32$. Thus, about 1 out of 3 students agreed that they learned a lot in the course (Learning Progress), given that they did not agree that ALT and First Principles occurred.

This can be further interpreted: When both ALT and First Principles occurred, students were more than three times as likely to agree that they learned a lot in the course, compared to when ALT and First Principles were reported not to have occurred ($0.985/0.32 = 3.09$). This same pattern was repeated for the additional variables. The estimated odds of Learner

[7] Frick (1990) used Analysis of Patterns in Time (APT). MAPSAT is a more comprehensive methodology that includes mapping and analysis of temporal patterns and structure of systems relations—previously called APT&C in Frick et al. (2006).

Satisfaction when both ALT and First Principles are present compared to when both are not were about 3.6 to 1. The odds were about 4.75 to 1 that an Instructor/Course was viewed as outstanding by students when ALT and First Principles were reported as both present versus both absent. Finally, students were nine times more likely to agree that they mastered course objectives when they also agreed that both ALT and First Principles occurred during the course.

It is interesting to note that 16 out of 25 students judged themselves to be Partial Masters of course objectives although they disagreed that ALT and First Principles occurred. Thus, about 64% of students report that they achieve objectives to some degree even in the absence of both ALT and First Principles.

Discussion

Results of analysis of patterns in time (MAPSAT) are consistent with correlational results in Table 3, but make more clear the relationships between three or more variables without being restricted by linear assumptions in regression analysis. MAPSAT results indicated that students were three to five times more likely to agree or strongly agree that they learned a lot and were satisfied with courses when they also agreed that First Principles of Instruction were used *and* students were frequently engaged successfully (ALT). Students were nine times more likely to report mastery course objectives when both First Principles and ALT were reported to have occurred, compared with their absence. Results from this study provide the strongest empirical support thus far for First Principles of Instruction (M. D. Merrill, personal communication, March 2007).

How valid are course evaluations?

One might be critical of student ratings of courses, as was the first author of this study—prior to conducting an extensive review of research. One may believe that course evaluations are at best “smiles tests” or “reactionnaires” (what Kirkpatrick refers to as Level 1 evaluation). Significant data from numerous empirical studies changed this erroneous belief. Findings from several decades of research indicate that *global items* are significantly related positively to student learning achievement as measured by objective tests and classroom observation instruments (Abrami 2001; Abrami et al. 1990; Arthur et al. 2003; Clayson et al. 2006; Cohen 1981; Emery et al. 2003; Feldman 1989; Koon and Murray 1995; Kulik 2001; Marsh 1984; Renaud and Murray 2004). In one of the most cited by scholarly research studies subsequently published on this issue, Cohen (1981):

> ... used meta-analytic methodology to synthesize research on the relationship between student ratings of instruction and student achievement.... The results of the meta-analysis provide strong support for the validity of student ratings [global items] as measures of teaching effectiveness. (p. 281).

Abrami (2001) cites further empirical research studies which are consistent with Cohen’s meta-analysis. Abrami also criticizes the oft-cited contrary research reports by detailing their significant methodological limitations and conclusions not warranted by evidence. The fact remains that numerous methodologically sound studies have shown positive and significant correlations between student global ratings of courses and student learning achievement. Global items are those such as used in our study (Overall, this was an outstanding course; Overall, this was an outstanding instructor; I would recommend this instructor to others). While such global items may appear to be measuring student satisfaction (Level 1), they nonetheless are predictive of student learning achievement (an indicator of Level 2).

In our present study, we also found very strong, positive relationships between *these same global ratings* and: (1) student self-reports on use of First Principles of Instruction in their courses, (2) student satisfaction with the instructors and courses, (3) self-reports on Learning Progress (learning a lot), and (4) self-reports of Academic Learning Time (see Table 3).

Self-reports on Learning Progress are not the same as objective measures of student learning achievement. However, these variables are correlated; student ratings of their Learning Progress and objective measures of learning achievement were found to correlate on average 0.47 in Cohen's (1981) meta-analysis. It should be noted that the square of this correlation is about 0.22, which means that about 78% of the variance in achievement is not predicted by self-reports of Learning Progress.

Past research has also shown that ALT is strongly and significantly correlated with student academic achievement as measured by standardized tests (cf. Berliner 1991; Brown and Saks 1986; Fisher et al. 1978; Kuh et al. 2006). In our study, student reports of ALT in their courses and their perceived Learning Progress were highly correlated ($\rho = 0.60$). ALT and self-reports of Mastery of course objectives were also significantly correlated ($\rho = 0.30$).

While students report that they believe they have learned a lot in their courses, they apparently have not learned enough. As can be seen in Table 1, 1 out of 4 students considered themselves to have achieved the course objectives. These results are consistent with a large scale study by Baer et al. (2006). Literacy skills of a random sample of 1,827 students were assessed through standardized tests administered by trained examiners. These students were nearing completion of their degrees at 80 randomly selected two- and four-year public universities and colleges. Baer et al. reported percentages of students from 2-year vs. 4-year institutions, respectively, who were *proficient* in prose literacy as 23% and 38%, in document literacy as 23% and 40%, and in quantitative literacy as 18% and 34%. This means that less than 25% of students at 2-year institutions performed *at proficiency level or higher*, and less than 40% at 4-year institutions likewise were proficient. While one may wonder how students interpreted the question of their achievement of course objectives in our study (master, partial master, nonmaster), those self reports are surprisingly consistent with objective measures of college student literacy reported by Baer et al. (2006). This type of concurrent validity lends credence to student ratings of their mastery of course objectives.

When the empirical findings from our study are taken as a whole, they do suggest relationships between student ratings of First Principles of instruction and: ALT, Learning Progress, Satisfaction, Mastery of course objectives, and overall course quality (average correlation of 0.71—see Table 3). Items from these scales were mixed randomly in the survey; students did not know what items belonged to which scales. Moreover, the MAPSAT results indicated that, according to self-reports, students were nine times more likely to consider themselves as masters of course objectives when they also agreed that First Principles and ALT occurred in their courses. They were more than three times as likely to agree that they learned a lot (Learning Progress), and about five times as likely to rate the course and instructor as outstanding when they also agreed that First Principles and ALT occurred. These results should not be surprising, since First Principles of Instruction were derived from numerous extant instructional theories and models which have been investigated with respect to their relationship to student learning outcomes (cf. Merrill 2002).

Limitations

The results of this correlational study do not permit causal inferences about student learning achievement. The results do show consistent patterns of significant positive correlations among variables that have been empirically associated with learning achievement and ALT in past studies. However, we did not measure student learning achievement in this survey. These results do suggest that further studies are warranted in which objective measures of student learning achievement are obtained, as well as objective measures of Academic Learning Time and use of First Principles.

One might be concerned that only one or a few students rated each course. The implication is, apparently, that these ratings may not be representative of how each course may have been rated by all of the students. Stated another way, the single rating may not be a good estimator of the average rating for each class. It might be argued that we should have focused on a few instructors and collected more student ratings per class. Had we done so, however, we would not have received ratings of such a wide range of courses (at least 89 different college courses were evaluated by respondents in our study). Moreover, if we had ratings from just a few courses, such a design would *decrease variance* among course attributes and thus attenuate correlation coefficients. Instead of an N of 130–140 for the correlations, we might have an N of 3–5, since we would need to use an average student rating on each scale for each course. Most important, however, is that we found such high correlations across such a *wide range of course topics and instructors*. The authors view this as strength of this study and thus enhancing the generalizability of findings.

The fact that results from the present study are consistent with empirical findings in numerous research studies over the past 40 years, when measuring the same variables, supports the validity of our findings. If the rating scales used in this study were not highly reliable, then a single student rating per course would be more problematic. Such was not the case (see Table 2). Having a wide range of courses increased the variability in between-course ratings and permitted such high intercorrelations among scales to be observed.

Since this was a descriptive-correlational research study, causal inferences are not warranted. We cannot conclude that First Principles of instruction *cause* student ALT, Learning Progress, Mastery of course objectives, Satisfaction, or overall course quality. In order to answer the question about the effectiveness of First Principles, further research is needed in which variables can be controlled or manipulated experimentally and where student achievement is measured objectively. However, it should be noted that, logically, instruction does not cause student learning. Instruction is neither necessary nor sufficient for learning to occur. Clearly, students can learn on their own without instruction—e.g., by trial and error or through disciplined inquiry. The question should be: Do students learn better or learn more when instructors use First Principles, compared to when these principles are not used (or when some other methods or strategies are used)? The data from our survey of student ratings suggest that further research in this area is worth pursuing.

A further limitation of our study is that respondents were volunteers; hence whether or not respondents are representative of the population of U.S. college students is unknown. We sent e-mail to instructors and to student organizations at several institutions. We know that about 1/3 of the respondents completed the survey using computers whose IP numbers were from three different campuses at our own institution. Thus, about 2/3 of our sample used computers in domains outside of our institution, including those from .com, .net and other .edu domains. We also know that respondents reported ratings on at least 89 different courses and instructors, representing a wide range of subject matter in business, health sciences, liberal arts, education and technology courses. Thus, while sample bias may be of concern, the size of the sample and the wide range of courses evaluated tend to mitigate this concern. Furthermore, our results on student reports of low rates of mastery of course objectives (1 out of 4) are consistent with relatively low levels of objectively measured student proficiencies in literacy skills in a stratified random sample of over 1,800 students from 2- and 4-year colleges in the U.S. (Baer et al. 2006). This consistency—as well as consistency of our results with those from meta-analyses of student ratings and learning achievement—also argues against sample bias in our study. While females were over-represented in our sample, there was no statistically significant statistical association between gender and other variables measured in our survey.

Implications

The nine TALQ scales can be used by instructional designers. After formative and summative evaluation of the developed instruction with the target audience, these TALQ scales can be completed by learners in about 10 min. We also strongly recommend that learners be assessed both before and after instruction with respect to their mastery of instructional objectives. Designers can confirm whether learning a lot is learning enough. If students are not learning enough, then it is likely that parts of the instruction can be improved. Thus, the TALQ scales can be used as an evaluation tool in conjunction with objective assessment of mastery of instructional objectives and observations of student performance.

College instructors can substitute the nine TALQ scales in their course evaluations to either replace or supplement existing items. We recommend that, as a baseline, instructors use the TALQ scales in courses they teach along with objective assessments of student learning achievement. Instructors can subsequently revise their courses by attempting to implement First Principles of Instruction. They can further evaluate their revised courses with the same TALQ scales and objective assessments of student achievement. Instructors can see for themselves whether the findings from the present study are consistent with empirical data from their own courses. If increased used of First Principles of Instruction is associated with increased learning achievement, student satisfaction and greater ALT, then students will benefit.

As a final note, Glenn (2007) reported that the extant "method of using student evaluations to assess professors is flawed but fixable..." (n.p.) After an extensive review of the literature, Kuh et al. (2006) recommended that postsecondary institutions should "... *focus assessment and accountability efforts on what matters to student success*" (p. 4, italics added). Use of the TALQ scales for course evaluation would further this aim.

Appendix A

The nine TALQ scales: teaching and learning quality

1. Academic Learning Time (ALT) Scale: Cronbach $\alpha = 0.85$

- I frequently did very good work on projects, assignments, problems and/or learning activities for this course.
- I spent a lot of time doing tasks, projects and/or assignments, and my instructor judged my work as high quality.
- I put a great deal of effort and time into this course, and it has paid off—I believe that I have done very well overall.

2. Learning Progress Scale (Kirkpatrick, Level 2) : Cronbach $\alpha = 0.97$

- Compared to what I knew before I took this course, I learned a lot.
- I learned a lot in this course.
- Looking back to when this course began, I have made a big improvement in my skills and knowledge in this subject.
- I learned very little in this course. (−)[8]
- I did not learn much as a result of taking this course. (−)

3. Student Satisfaction Scale (Kirkpatrick, Level 1): Cronbach $\alpha = 0.94$

- I am dissatisfied with this course. (−)
- This course was a waste of time and money. (−)
- I am very satisfied with this course.

4. BEST Scale (Global IU course evaluation items): Cronbach $\alpha = 0.92$

- Overall, I would rate the quality of this course as outstanding.
- Overall, I would rate this instructor as outstanding.
- Overall, I would recommend this instructor to others.

5. Authentic Problems Scale (Merrill, Principle 1): Cronbach $\alpha = 0.81$

- I performed a series of increasingly complex authentic tasks in this course.
- I solved authentic problems or completed authentic tasks in this course.
- In this course I solved a variety of authentic problems that were organized from simple to complex.
- Assignments, tasks, or problems I did in this course are clearly relevant to my professional goals or field of work.

6. Activation Scale (Merrill, Principle 2): Cronbach $\alpha = 0.91$

- I engaged in experiences that subsequently helped me learn ideas or skills that were new and unfamiliar to me.
- In this course I was able to recall, describe or apply my past experience so that I could connect it to what I was expected to learn.
- My instructor provided a learning structure that helped me to mentally organize new knowledge and skills.
- In this course I was able to connect my past experience to new ideas and skills I was learning.
- In this course I was not able to draw upon my past experience nor relate it to new things I was learning. (−)

7. Demonstration Scale (Merrill, Principle 3): Cronbach $\alpha = 0.88$

- My instructor demonstrated skills I was expected to learn in this course.
- My instructor gave examples and counter-examples of concepts that I was expected to learn.
- My instructor did not demonstrate skills I was expected to learn. (−)
- My instructor provided alternative ways of understanding the same ideas or skills.

[8] Items with (−) are negatively worded; thus rating scores are reversed for analysis of these items. Each item requires Likert scale ratings (strongly disagree, disagree, undecided, agree, strongly agree).

8. Application Scale (Merrill, Principle 4): Cronbach $\alpha = 0.74$

- My instructor detected and corrected errors I was making when solving problems, doing learning tasks or completing assignments.
- My instructor gradually reduced coaching or feedback as my learning or performance improved during this course.
- I had opportunities to practice or try out what I learned in this course.
- My course instructor gave me personal feedback or appropriate coaching on what I was trying to learn.

9. Integration Scale (Merrill, Principle 5): Cronbach $\alpha = 0.81$

- I had opportunities in this course to explore how I could personally use what I have learned.
- I see how I can apply what I learned in this course to real life situations.
- I was able to publicly demonstrate to others what I learned in this course.
- In this course I was able to reflect on, discuss with others, and defend what I learned.
- I do not expect to apply what I learned in this course to my chosen profession or field of work. (−)

References

Abrami, P. (2001). Improving judgments about teaching effectiveness using teacher rating forms. *New Directions for Institutional Research, 109*, 59–87.

Abrami, P., d'Apollonia, S., & Cohen, P. (1990). Validity of student ratings of instruction: What we know and what we do not. *Journal of Educational Psychology, 82*(2), 219–231.

Andrews, D. H., & Goodson, L. A. (1991). A comparative analysis of models of instructional design. In G. J. Anglin (Ed.), *Instructional technology: Past, present and future* (pp. 133–155). Englewood Cliffs: Libraries Unlimited.

Arthur, J., Tubré, T., Paul, D., & Edens, P. (2003). Teaching effectiveness: The relationship between reaction and learning evaluation criteria. *Educational Psychology, 23*(3), 275–285.

Baer, J., Cook, A., & Baldi, S. (2006). *The literacy of America's college students.* American Institutes for Research. Retrieved January 20, 2007: http://www.air.org/news/documents/The%20Literacy%20of%20Americas%20College%20Students_final%20report.pdf.

Berliner, D. (1991). What's all the fuss about instructional time? In M. Ben-Peretz & R. Bromme (Eds.), *The nature of time in schools: Theoretical concepts, practitioner perceptions*. New York: Teachers College Press.

Brown, B., & Saks, D. (1986). Measuring the effects of instructional time on student learning: Evidence from the Beginning Teacher Evaluation study. *American Journal of Education, 94*(4), 480–500.

Clayson, D., Frost, T., & Sheffet, M. (2006). Grades and the student evaluation of instruction: A test of the reciprocity effect. *Academy of Management Learning and Education, 5*(1), 52–65.

Cohen, P. (1981). Student ratings of instruction and student achievement. A meta-analysis of multisection validity studies. *Review of Educational Research, 51*(3), 281–309.

Emery, C., Kramer, T., & Tian, R. (2003). Return to academic standards: A critique of student evaluations of teaching effectiveness. *Quality Assurance in Education, 11*(1), 37–46.

Estep, M. (2003). *A theory of immediate awareness. Self-organization and adaptation in natural intelligence*. NY: Springer-Verlag.

Feldman, K. (1989). The association between student ratings of specific instructional dimensions and student achievement: Refining an extending the synthesis of data from multisection validity studies. *Research in Higher Education, 30*, 583–645.

Fisher, C., Filby, N., Marliave, R., Cohen, L., Dishaw, M., Moore, J., & Berliner, D. (1978). Teaching behaviors: Academic Learning Time and student achievement: Final report of Phase III-B, Beginning Teacher Evaluation Study. San Francisco: Far West Laboratory for Educational Research and Development.

Frick, T. (1990). Analysis of patterns in time (APT): A method of recording and quantifying temporal relations in education. *American Educational Research Journal, 27*(1), 180–204.

Frick, T. (1997). Artificial tutoring systems: What computers can and can't know. *Journal of Educational Computing Research, 16*(2), 107–124.
Frick, T. (2005). *Bridging qualitative and quantitative methods in educational research: Analysis of patterns in time and configuration (APT&C).* Proffitt Grant Proposal. Retrieved March 4, 2007: http://education.indiana.edu/~frick/proposals/apt&c.pdf.
Frick, T., An, J., & Koh, J. (2006). Patterns in education: Linking theory to practice. In M. Simonson (Ed.), *Proceedings of the Association for Educational Communication and Technology*, Dallas, TX. Retrieved March 4, 2007: http://education.indiana.edu/~frick/aect2006/patterns.pdf.
Glenn, D. (2007). Method of using student evaluations to assess professors is flawed but fixable, 2 scholars say. *Chronicle of Higher Education Daily*. Retrieved May 29, 2007 from http://chronicle.com/daily/2007/05/2007052901n.htm.
Gustafson, K., & Branch, R. (2002). *Survey of instructional development models* (4th ed.). Syracuse: Syracuse University, ERIC Clearinghouse on Information Resources.
Kirk, R. (1995). *Experimental design: Procedures for the behavioral sciences* (3rd ed.). Pacific Grove: Brooks/Cole.
Kirkpatrick, D. (1994). *Evaluating training programs: The four levels*. San Francisco: Berrett-Koehler.
Koon, J., & Murray, H. (1995). Using multiple outcomes to validate student ratings of overall teacher effectiveness. *The Journal of Higher Education, 66*(1), 61–81.
Kuh, G., Kinzie, J., Buckley, J., & Hayek, J. (2006, July). *What matters to student success: A review of the literature (Executive summary)*. Commissioned report for the National Symposium on Postsecondary Student Success. Retrieved January 20, 2007: http://nces.ed.gov/npec/pdf/Kuh_Team_ExecSumm.pdf.
Kulik, J. (2001). Student ratings: Validity, utility and controversy. *New Directions for Institutional Research, 109*, 9–25.
Maccia, G. S. (1987). Genetic epistemology of intelligent natural systems. *Systems Research, 4*(1), 213–218.
Marsh, H. (1984). Students' evaluations of university teaching: Dimensionality, reliability, validity, potential biases, and utility. *Journal of Educational Psychology, 76*(5), 707–754.
Merrill, M. D. (2002). First principles of instruction. *Education Technology Research & Development, 50*(3), 43–59.
Merrill, M. D. (2007). A task-centered instructional strategy. *Journal of Research on Technology in Education, 40*(1), 33–50.
Reigeluth, C. M. (1983). *Instructional-design theories and models: An overview of their current status.* Mahwah: Lawrence Erlbaum.
Reigeluth, C. M. (1999). *Instructional-design theories and models: A new paradigm of instructional theory* (Vol. II). Hillsdale: Lawrence Erlbaum.
Reigeluth, C. M., & Stein, F. S. (1983). The elaboration theory of instruction. In C. M. Reigeluth (Ed.), *Instructional-design theories and models: An overview of their current status* (pp. 335–382). Mahwah: Lawrence Erlbaum.
Renaud, R., & Murray, H. (2004). Factorial validity of student ratings of instruction. *Research in Higher Education, 46*(8), 929–953.
Squires, D., Huitt, W., & Segars, J. (1983). *Effective schools and classrooms: A research-based perspective*. Alexandria, VA: Association for Supervision and Curriculum Development.
Visscher-Voerman, I., & Gustafson, K. (2004). Paradigms in the theory and practice of education and training design. *Educational Technology Research & Development, 52*(2), 69–89.

Theodore Frick is an Associate Professor in the Department of Instructional Systems Technology, School of Education, Indiana University Bloomington. His current research interests include improvement of teaching and learning, simulations and games for understanding educational systems, and predicting patterns in educational systems.

Rajat Chadha is a doctoral student in the School of Education, Indiana University Bloomington. His areas of specialization include instructional systems technology, educational measurement, and statistics in educational research.

Carol Watson is the Program Manager, Fundamentals and Online Training Development at the Eppley Institute for Parks and Public Lands at Indiana University Bloomington.

Ying Wang is a doctoral candidate in the School of Education, Indiana University Bloomington. Her areas of specialization are instructional systems technology, educational inquiry methodology, and technology preparation for teachers.

Pamela Green is a doctoral student in the School of Education, Indiana University Bloomington. Her areas of specialization include educational measurement, application of instructional design theories in corporate e-learning courses, and systemic change school reform research.