Improving Course Evaluations to Improve Instruction and Complex Learning in Higher Education

Theodore W. Frick

Rajat Chadha

Carol Watson

Emilija Zlatkovska

School of Education

Indiana University Bloomington

Paper submitted for review for the

Outstanding Featured Research Paper Award

September 10, 2008

Abstract

Recent research has touted the benefits of learner-centered instruction, problem-based learning, and a focus on complex learning. Instructors often struggle to put these goals into practice as well as to measure the effectiveness of these new teaching strategies in terms of mastery of course objectives. Enter the course evaluation, often a standardized tool that yields little practical information for an instructor, but is nonetheless utilized in making high-level career decisions, such as tenure and monetary awards to faculty.

The present researchers have developed a new instrument to measure teaching and learning quality (TALQ). In a study of 464 students in 12 courses, if students agreed that they experienced academic learning time (ALT) and that their instructors used First Principles of Instruction, then students were nearly 4 times more likely achieve *high levels* of mastery of course objectives, according to independent instructor assessments. TALQ can measure improvements in use of First Principles in teaching and course design. The feedback from this instrument can assist teachers who wish to implement the recommendation made by Kuh et al. (2006) that universities and colleges should focus their assessment efforts on factors that influence student success.

Introduction

Complex learning has been defined as involving "the integration of knowledge, skills and attitudes, the coordination of qualitatively different constituent skills and the transfer of what is learned in school or training to daily life and work settings" (van Merriënboer & Kirschner, 2007, p. 4). Van Merriënboer, Clark and de Croock (2002) note that there is a need for students to be able to transfer complex cognitive skills to “an

increasingly varied set of real-world contexts and settings." They suggest that "inadequate [instructional] design may cause learning problems" (p. 39). But how is one to know if their instructional design is inadequate or how a course could be improved?

Course evaluations traditionally used in higher education have few items that are empirically related to student learning achievement. In meta-analyses of studies that have examined this relationship, global items such as "This was an outstanding course." or "The instructor of this course was outstanding." correlate moderately with student achievement (average correlations of 0.47 and 0.43, respectively—cf., Cohen, 1981; Feldman, 1989; Kulik, 2001). While these global items predict increased student achievement, items such as these do not indicate how to improve teaching.

Frick, Chadha, Watson, Wang and Green (in press, 2007) developed a new course evaluation instrument for assessing Teaching and Learning Quality (TALQ). TALQ includes rating scales on:

- use of five First Principles of Instruction in a course (Merrill, 2002; 2008; Merrill, Barclay & van Schaak, 2008),
- perceived student academic learning time (ALT) (Rangel & Berliner, 2007; Berliner, 1990; Fisher et al., 1978; Kuh et al., 2007),
- perceived learning progress (Cohen, 1981),
- self-reported mastery of course objectives (Mager, 1997),
- satisfaction with the course and instructor (Kirkpatrick, 1994); and
- global ratings of course and instructor quality (Cohen, 1981).

In a study of 140 students in 89 unique courses from a wide range of disciplines, Frick et al. (in press, 2007) found highly significant positive correlations among these

scales. Based on course ratings, *if* students agreed or strongly agreed that instructors used First Principles of Instruction *and* those students also agreed or strongly agreed that they were engaged successfully in course activities (ALT), *then* they were much more likely to: 1) report mastery of course objectives, 2) agree that they learned a lot (made learning progress), 3) agree that they were satisfied, and 4) agree that the course and instructor were outstanding. In a somewhat larger study of 190 students in 111 different courses, Frick et al. (2008) found similar patterns among TALQ scales derived from student ratings.

First Principles of Instruction are relevant to complex learning of authentic, real-world, whole tasks. Based on a synthesis of instructional design theories, Merrill (2002) claimed that student learning will be promoted when: 1) instruction is problem- or task-centered, 2) student learning is activated by connecting what they already know or can do with what is to be newly learned, 3) students are exposed to demonstrations of what they are to learn, 4) they have opportunities to try out what they have learned with instructor coaching and feedback, and 5) they integrate what they have learned into their personal lives. If one or more of these First Principles are missing during instruction, Merrill argues that learning will be negatively impacted.

Results from the studies by Frick et al. (in press, 2007; 2008) are consistent with Merrill's claims, according to student ratings and self-reports. Moreover, it would appear that instructors could improve their courses by implementation of First Principles of Instruction. While First Principles were drawn from apparently successful instructional theories, few empirical studies have been conducted to verify Merrill's (2002) claim that First Principles promote student learning.

Problem

The current study addresses two limitations of the Frick et al. (in press, 2007; 2008) studies:

1. The present study sought participation from whole classes to address concerns about representativeness of student ratings, since the two prior studies were based on ratings by one or a few students in each course.
2. In the first two studies, mastery of course objectives was self-reported. In the present study, an independent assessment of student mastery of course objectives was collected from their instructors.

Research questions addressed in the present study are:

1. What are the relationships among student ratings of First Principles of Instruction, student academic learning time (ALT), satisfaction with the course, student learning progress, global ratings of instructor and course quality and instructor ratings of student mastery of course objectives?
2. When students agree that First Principles of Instruction occurred, what are the odds that they agree that ALT also occurred, compared with students who did not agree that First Principles of Instruction occurred?
3. When students agree that they experienced frequent success in course activities (ALT) compared with not agreeing that ALT had occurred, what are the odds that students are rated as high masters of course objectives by their instructors?

Method

In collaboration with staff from a teaching center at Indiana University Bloomington, a recruitment email was sent to university faculty that sought volunteers to

who were willing to have the TALQ instrument used in their classes, in addition to their normal course evaluations. During the last three weeks of the fall 2007 semester, a paper version of the TALQ evaluation was administered by researchers a week or two before the standard course evaluation was given to that class. In two of the classes, only the TALQ course evaluation instrument was completed, according to that instructor's preferences, since it was administered in the last week of classes.

Items from the TALQ scales were disaggregated and randomly mixed on the course evaluation form, so that students did not know which items belonged to what scale. The TALQ instrument was administered at the beginning of a regular class period. Each evaluation form had a unique code number on the cover sheet that was repeated on the evaluation form itself. Participating students wrote their names on the top halves of the cover sheets, which were detached and given to the instructor, who then left the classroom. Students completed the TALQ course evaluation anonymously; their individual ratings were collected by the researchers and never shown to instructors.

About one month after completion of the course, instructors rated each participating student's mastery of course objectives using a 10-point scale. Ratings were based on instructor records of grades on student performance in class, completed assignments and projects, exam scores, etc. The bottom halves of the cover sheets with instructor ratings and unique code numbers were returned to the researchers. Thus, student anonymity was maintained, while researchers could pair instructor ratings of student mastery with student ratings of the course by matching the unique code numbers.

None of the courses in this study was taught by any of the researchers. Instructors were provided with summary reports of TALQ scales *after* they had submitted their

ratings of student mastery of course objectives, except for one instructor who needed the reports for her annual report and these were the sole course evaluations she used.

TALQ instrument. The first page of the TALQ contained questions about student gender, year in school, expected grade, and self-assessment of mastery of course objectives. The remaining 3 pages contained 40 Likert-scale items that were scrambled (i.e., not organized by the scales discussed below). Responses to items were later coded numerically when entered into SPSS: strongly agree=5, agree=4, undecided=3, disagree=2, and strongly disagree=1. Negatively worded items were subsequently reverse-coded using SPSS recode. Scale scores for each student were formed by computing the mean for items comprising each scale, in order to facilitate interpretation.

The directions in the TALQ instrument informed students that "*authentic problems* or *authentic tasks* are meaningful learning activities that are clearly relevant to you at this time, and which may be useful to you in the future (e.g., in your chosen profession or field of work, in your life, etc.)" (p. 2).

Results

Respondents. Data were collected from 464 students in 12 different courses taught by 8 instructors in business, philosophy, history, kinesiology, social work, informatics, nursing, and health, physical education and recreation. The number of student respondents who completed the TALQ ranged from 16 to 104 in the 12 classes, and in 10 of the 12 classes the range was from 22 to 53. Response rates were very high among those students present at the beginnings of these classes, although a few students declined to participate. Approximately 56 percent of the respondents were female and 44 percent male—very similar to gender proportions on the Bloomington campus overall.

Unlike the previous two studies of the TALQ, nearly all of the student respondents were undergraduates (52 freshmen, 104 sophomores, 115 juniors and 185 seniors). In the prior studies about one-third of the respondents were graduate students. A larger percentage of juniors and seniors participated in the present study, compared with freshmen and sophomores. This was not unexpected as university faculty members are likely to teach more advanced courses, rather than introductory classes that are often taught by associate instructors. Only one of the courses was at the 100 level, while the remaining classes were at the 200-400 levels.

Mastery scales. Instructor ratings of student mastery of course objectives was correlated highly with student reports of expected grades ($\rho = 0.584$, $p < 0.0005$). Since our study was completed about two weeks prior to the end of the semester, most students did not know their exact final grades. Student mastery was rated by instructors on a 10-point scale, from low to high achievement of course objectives. Due to privacy issues, researchers were unable to obtain actual student course grades, although instructors later informed us that they based their ratings on student performance in the course. Instructor ratings were converted to three categories: Low mastery (1-5), medium mastery (6-8) and high mastery (8.5 – 10).

In Table 1, it can be seen that the association is strong ($\chi^2 = 163.3$, $p < 0.0005$), although about 21 percent of students who expected to receive *A*'s were classified at a medium level of mastery by their instructors (between 6 and 8 on a 10-point scale), and about 6.5 percent received *B*'s who were rated at a high level of mastery. About 7 percent of the students expected a grade of *B*, who were rated as low masters by their instructors.

Table 1. Crosstabulation of instructor ratings of student mastery and students' expected course grades.

			Instructor Rating of Student Mastery			
			Low (0-5)	Medium (6-8)	High (8.5-10)	Total
b. In this course, I expect to receive a grade of:	A	Count	1	93	126	220
		% of Total	.2%	20.8%	28.1%	49.1%
	B	Count	32	137	29	198
		% of Total	7.1%	30.6%	6.5%	44.2%
	C	Count	16	12	0	28
		% of Total	3.6%	2.7%	.0%	6.3%
	D	Count	1	1	0	2
		% of Total	.2%	.2%	.0%	.4%
Total		Count	50	243	155	448
		% of Total	11.2%	54.2%	34.6%	100.0%

Students also self-reported their mastery levels on a 10-point scale. The Spearman correlation between student and instructor ratings of student mastery was 0.382 and highly significant ($p < 0.0005$). After student self-ratings were recoded into low, medium and high mastery in the same manner as were instructor ratings, a crosstabulation was performed. Agreement between student and instructor ratings was highly significant when corrected for chance ($\kappa = 0.17$, $p < 0.0005$). The area of greatest discrepancy was 108 students (23.5 percent) who considered themselves to be *medium* masters, whereas their instructors classified those same students as *high* masters of course objectives. A further crosstabulation between student self-reported mastery and expected grades indicated that 153 students (34 percent) expected to receive a grade of *A*, yet considered themselves medium masters (between 6 and 8 on a 10-point scale).

In summary, 108 of the 464 students were rated by their instructors as *medium* masters, but those same students expected to receive a grade of *A* in the course. Nearly 24 percent of the students perceived themselves as *medium* masters while their instructors rate them as *high* masters of course objectives. About one-third of the students (153) expected to receive *A*'s who also considered themselves as *medium* masters.

First Principles of Instruction scale: Problem-centered. Student reports of their engagement in authentic problems (Principle #1), was indicated by three items: 3) I performed a series of increasingly complex authentic tasks in this course; 22) I solved authentic problems or completed authentic tasks in this course; and 27) In this course I solved a variety of authentic problems that were organized from simple to complex. The Cronbach α coefficient (internal consistency) of this scale was 0.690.

Further examination of the way that these items are stated indicates student *engagement* with these authentic problems (I performed…, I solved…). Thus, it is possible that a course could have provided authentic problems for students to solve, but they did not engage in doing so. Therefore, they could disagree that they performed authentic tasks or solved authentic problems, even though they were expected to do so in the course.

Furthermore, when ratings of engagement with authentic problems were examined within each of the 12 classes, some classes were more divided than others in terms of their agreement and disagreement on this scale. This would suggest that perceptions of authentic problems may be further related to the nature of course content and types of students who are enrolled. For example, authenticity of tasks in an advanced level nursing course would be less ambiguous to nursing students, when compared with perceptions of authenticity of tasks in a history course taken as an elective by non-majors.

First Principles of Instruction scale: Activation. Items for Principle #2 were: 9) I engaged in experiences that subsequently helped me learn ideas or skills that were new and unfamiliar to me; 19) In this course I was able to recall, describe or apply my past experience so that I could connect it with what I was expected to learn; 26) My instructor

provided a learning structure that helped me to mentally organize new knowledge and skills; 35) In this course I was able to connect my past experience to new ideas and skills I was learning; and 36) In this course I was not able to draw upon my past experience nor relate it to new things I was learning (reverse-coded). The Cronbach α for this Activation scale was 0.812.

First Principles of Instruction scale: Demonstration. Items for Principle #3 were: 5) My instructor demonstrated skills I was expected to learn in this course; 14) Media used in this course (texts, illustrations, graphics, audio, video, computers) were helpful in learning; 16) My instructor gave examples and counter-examples of concepts that I was expected to learn; 17) My instructor directly compared problems or tasks that we did, so that I could see how they were similar or different; 31) My instructor did not demonstrate skills I was expected to learn (reverse-coded); and 38) My instructor provided alternative ways of understanding the same ideas or skills. The Cronbach α for this Demonstration scale was 0.830.

First Principles of Instruction scale: Application. Items for Principle #4 included: 7) My instructor detected and corrected errors I was making when solving problems, doing learning tasks, or completing assignments; 32) I had opportunities to practice or try out what I learned in this course; and 37) My course instructor gave me personal feedback or appropriate coaching on what I was trying to learn. The Cronbach α for this Application scale was 0.758.

First Principles of Instruction scale: Integration. Items for Principle #5 were: 11) I had opportunities in this course to explore how I could personally use what I learned; 24) I see how I can apply what I learned in this course to real life situations; 30)

I was able to publicly demonstrate to others what I learned in this course; and 33) In this course, I was able to reflect on, discuss with others, and defend what I learned. The Cronbach α for this Integration scale was 0.780.

First Principles of Instruction: Combined Scale. For each student and each First Principle, a scale score was computed by taking the mean value from responses to that scale by that student. These five scales were further combined into a combined First Principles scale by taking the mean of the scale means for each student. Cronbach's α for this combined First Principles scale was 0.881.

Successful student engagement: Academic Learning Time (ALT). Items comprising the ALT scale were: 1) I did not do very well on most tasks in this course, according to my instructor's judgment of the quality of my work (reverse-coded); 12) I frequently did very good work on projects, assignments, problems and/or activities for this course; 13) I spent a lot of time doing tasks, projects and/or assignments, and my instructor judged my work of high quality; and 21) 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. Cronbach's α for this ALT scale was 0.763.

Student learning progress scale. This scale was comprised of the following items: 4) Compared to what I knew before I took this course, I learned a lot; 10) I learned a lot in this course; 23) I learned very little in this course (reverse-coded); 28) I did not learn much as a result of taking this course (reverse-coded). The Cronbach α for this Learning Progress scale was 0.935.

Student satisfaction scale. The following items were used on this scale: 2) I am very satisfied with how my instructor taught this class; 6) I am dissatisfied with this

course (reverse-coded); 18) This course was a waste of time and money (reverse-coded); and 40) I am very satisfied with this course. Cronbach's α for this Satisfaction scale was 0.926.

Outstanding course and instructor scale: Global Quality. Items on this scale were taken from the university's course evaluation item pool and consistent with those that Cohen (1981) had identified as being moderately correlated with student learning achievement: 8) Overall, I would rate the quality of this course as outstanding; 15) Overall, I would rate this instructor as outstanding; and 34) Overall, I would recommend this instructor to others. The Cronbach α for this Global Quality scale was 0.915.

Relationships Among TALQ Scales

Spearman ρ correlations were computed for the TALQ scales on the 464 students and the instructor ratings of student mastery, since these are all ordinal level measures. It can be seen from Table 2 that First Principles of Instruction ratings are positively and very highly correlated with Global Quality, Student Satisfaction, ALT, and Learning Progress. Similarly, remaining scales are highly correlated with each other, except for Instructor Rating of Student Mastery. The best correlation with Student Mastery is Academic Learning Time ($\rho = 0.362$, $p < 0.0005$).

Table 2. Spearman Correlations among TALQ Scales

	Combined First Principles	Global Course/ Instructor Quality	Student Satisfaction	Academic Learning Time	Learning Progress	Instructor Rating of Student Mastery[a]
First Principles	1.000	0.774	0.778	0.583	0.725	0.115[b]
Global Quality		1.000	0.848	0.528	0.664	0.180
Student Satisfaction			1.000	0.557	0.746	0.202
ALT				1.000	0.498	0.362
Learning Progress					1.000	0.136[c]
Student Mastery						1.000

a: 10-point scale used here; b: $p = 0.014$; c: $p = 0.003$; all remaining correlations are significant at $p < 0.0005$; $n = 464$.

Pattern Analyses

Analysis of Patterns in Time (Frick, 1990) was used to further investigate these relationships. With the exception of the student mastery scale (already coded as low, medium and high), remaining scales were recoded for 'agreement' = 'Yes' if the scale score was greater than 3.5, and 'agreement' = 'No' if the student's scale score was less than or equal to 3.5. The reasoning for this coding system was that on the original Likert scale, 'agree' was coded as '4' and 'strongly agree' as '5'; thus, any mean scale score that was closer to '4' or '5' was interpreted as agreement with that scale; otherwise it interpreted as *not* in agreement (strongly disagree = '1', disagree = '2', or undecided = '3').

As noted above, while significantly and positively correlated, instructor ratings of student mastery and student self-reports of their mastery were in some disagreement. About 1 out of 4 students rated himself or herself at a medium level of mastery, when the instructor independently rated him or her at a high mastery level. Thus, we believed that

a more reliable determination of student mastery was evident when both the student and instructor *independently* agreed on that student's mastery level (Low/Low, Medium/Medium, or High/High). Since no other metric of student learning achievement was available (and course grades were less discriminating, as discussed above), we selected cases in which the instructor and student ratings matched for each student. This resulted in 256 students, or about 55 percent of the original sample of 464 cases. The proportions of males and females were almost identical in the reduced sample as in the original, and proportions in other demographics also appeared to be about the same.

Table 3. Results for the APT Query: If Agreement on First Principles = ? and Agreement on Successful Engagement = ?, then Student Mastery = ?[a]

	Agreement on First Principles							
	No				Yes			
	Agreement on Successful Engagement				Agreement on Successful Engagement			
	No		Yes		No		Yes	
	Instructor Rating of Student Mastery		Instructor Rating of Student Mastery		Instructor Rating of Student Mastery		Instructor Rating of Student Mastery	
	Count	%	Count	%	Count	%	Count	%
Low (0-5)	15	31.9%	1	5.6%	1	2.3%	2	1.4%
Medium (6-8)	29	61.7%	12	66.7%	41	95.3%	112	75.7%
High (8.5-10)	3	6.4%	5	27.8%	1	2.3%	34	23.0%
Total	47	100.0%	18	100.0%	43	100.0%	148	100.0%

a: Note that the '?' signifies that categories within that classification are free to vary; thus, this is a compact way of expressing all possible queries for this 'If …. and …, then …' pattern.

In Table 3 it can be seen that for the APT Query, 'If Agreement on First Principles is Yes and Agreement on Successful Engagement is Yes, then Instructor Rating of Student Mastery is High' is true in 34 out of 148 cases, yielding a probability estimate of 0.23. On the other hand, 'If Agreement on First Principles is No and Agreement on Successful Engagement is No, then Instructor Rating of Student Mastery is High' is true in 3 out of 47 cases, yielding a probability estimate of 0.064. Thus, students are about 3.6 times as likely to be rated by their instructors (and themselves) as high masters of course objectives when students agreed that First Principles occurred and

also agreed that they experienced ALT (successful engagement), compared with *not* agreeing that First Principles and ALT occurred. The odds of 3.6 to 1 are computed as a ratio of the two probabilities: (0.230/0.064). The odds are about 23 to 1 of being rated by their instructors (and themselves) as a *low* master of course objectives when students do not agree that First Principles and ALT occurred (0.319/0.014 = 22.8), compared with being rated as a low master when students agreed that both First Principles and ALT did occur.

One can also see in Table 3 that the likelihoods of being rated as a *medium* level of mastery are higher when either students agree that ALT occurred but not First Principles ($p = 0.667$), or they agreed that First Principles occurred but not ALT ($p = 0.953$).

Theoretically, we would expect students to be more motivated when instructors use First Principles of Instruction, because students are expected to solve authentic or real-world problems as well as to integrate what they have learned into their personal lives. In other words, what they learn is expected to be more relevant and meaningful (see Keller, 1987). If students are more highly motivated, then they would be expected to be engaged more often in learning tasks. Furthermore, if instructors demonstrate what students are expected to learn and also provide feedback and scaffolding when students themselves try, we would expect student engagement to be successful more often—i.e., more Academic Learning Time (ALT). The research on ALT indicates that the more frequently students are engaged successfully, the higher they tend to score on tests of achievement (assuming that what students engage in is similar to what they are tested on).

We conducted further analyses using APT to see if these patterns occurred in data from our study. In Table 4, it can be seen that 148 out of 191 (77.5 percent) of the students agreed that they experienced ALT when they also agreed that First Principles of Instruction occurred in their courses. On the other hand, 18 out of 65 (27.7 percent) of the students agreed that they experienced successful engagement (ALT) when they did not agree that First Principles occurred. Thus, the odds of successful engagement are 0.775/0.277 or about 2.8 to 1 that students report that they are successfully engaged when they also agree that First Principles occurred versus not having occurred.

Table 4. Results for the APT Query: If Agreement on First Principles = ?, then Agreement on Successful Engagement = ?

	Agreement on First Principles			
	No		Yes	
	Agreement on Successful Engagement		Agreement on Successful Engagement	
	Count	%	Count	%
No	47	72.3%	43	22.5%
Yes	18	27.7%	148	77.5%
Total	65	100.0%	191	100.0%

Table 5. Results for the APT Query: If Agreement on Successful Engagement = ?, then Instructor Rating of Student Mastery = ?

	Agreement on Successful Engagement			
	No		Yes	
	Instructor Rating of Student Mastery		Instructor Rating of Student Mastery	
	Count	%	Count	%
Low (0-5)	16	17.8%	3	1.8%
Medium (6-8)	70	77.8%	124	74.7%
High (8.5-10)	4	4.4%	39	23.5%
Total	90	100.0%	166	100.0%

It can be further seen in Table 5 that the odds are about 23.5/4.4, or 5.3 to 1, for students to be rated as high masters of course objectives when they agreed versus disagreed that

they were successfully engaged (ALT). On the other hand, *medium* levels of mastery are about equally likely, regardless of whether or not students agreed about their ALT.

Thus, these results from Analysis of Patterns in Time appear to be consistent with theoretical predictions from Merrill (2002) on First Principles of Instruction and with well-established empirical evidence that supports the relationship between Academic Learning Time and student achievement (e.g., Kuh et al., 2007; Rangel & Berliner, 2007).

Conclusion

Although Merrill (2008) stated that the real value of the First Principles was in the design of instruction, he also pointed out that "learning from a given program will be facilitated in direct proportion to its implementation of these principles" (p. 175). Indeed, this was born out in our study. While academic learning time (ALT) is under the control of the student, use of the First Principles of Instruction in a classroom is something that college instructors can control. Data from this study indicate that when both ALT and First Principles were reported to occur, the likelihood of a high level of student mastery of course objectives (according to instructor evaluation of student performance) is about 3.6 times greater than the likelihood of high mastery when neither First Principles nor ALT were reported to occur.

On a typical course evaluation, low scores on global items do not tell instructors anything about how to improve their teaching in ways that are likely to also improve student learning achievement or student mastery of objectives. On the other hand, the TALQ scales on the First Principles of Instruction can be used to identify areas in which teaching and course design can be improved—i.e., instructors can incorporate authentic,

real-world problems for students to solve, activate student learning, perform additional demonstrations about what is to be learned, provide opportunities for students to successfully solve problems with coaching and feedback, and help students integrate what they have learned into their personal lives. Incorporation of these First Principles of Instruction in courses is strongly associated with high ratings of academic learning time, student satisfaction, student perceptions of learning a lot, and student ratings of overall instructor and course quality.

References

Berliner, D. (1990). 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.

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.

Feldman, K. A. (1989). The association between student ratings of specific instructional dimensions and student achievement: Refining and 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.W., Chadha, R., Watson, C., Wang, Y., & Green, P. (2007, in press). College student perceptions of teaching and learning quality. *Educational Technology Research and Development*.

Frick, T.W., Chadha, R., Watson, C., Wang, Y., & Green, P. (2008). Theory-based course evaluation: Implications for improving student success in postsecondary education. Paper presented at the American Educational Research Association conference, New York.

Keller, J. M. (1987). The systematic process of motivational design. *Performance & Instruction, 26*(9), 1-8.

Kirkpatrick, D. (1994). *Evaluating training programs: The four levels.* San Francisco, CA: Berrett-Koehler.

Kuh, G., Kinzie, J., Buckley, J., Bridges, B., & Hayek, J. (2007). Piecing together the student success puzzle: Research, propositions, and recommendations. *ASHE Higher Education Report, 32*(5). San Francisco: Jossey-Bass.

Kulik, J. A. (2001). Student ratings: Validity, utility and controversy. *New Directions for Institutional Research, 109*, 9-25.

Mager, R. (1997). *Measuring instructional results.* Atlanta, GA: The Center for Effective Performance.

Merrill, M. D. (2002). First Principles of Instruction. *Education Technology Research & Development, 50*(3), 43-59.

Merrill, M. D. (2008). What makes e^3 (effective, efficient, engaging) instruction? Keynote address at the AECT Research Symposium, Bloomington, IN.

Merrill, M. D., Barclay, M., & van Schaak, A. (2008). Prescriptive principles for instructional design. In J. M. Spector, M. D. Merrill, J. van Merriënboer & M. F. Driscoll (Eds.), *Handbook of research on educational communications and technology* (3rd ed.). New York: Lawrence Erlbaum Associates, 173-184.

Rangel, E., & Berliner, D. (2007). Essential information for education policy: Time to learn. *Research Points: American Educational Research Association*, *5*(2), 1-4.

van Merriënboer, J. J. G., Clark, R. E., & de Croock, M. B. M. (2002). Blueprints for complex learning: The 4C/ID model. *Education Technology Research & Development*, *50*(2), 39-64.

van Merriënboer, J. J. G. & Kirschner, P. A. (2007). *Ten steps to complex learning: A systematic approach to four-component instructional design*. Hillsdale, NJ: Lawrence Erlbaum Associates.

van Merriënboer, J. J. G., Kirschner, P. A., & Kester, L. (2003). Taking the load off a learner's mind: Instructional design for complex learning. *Educational Psychologist*, *39*(1), 5-13.