## EVALUATING RELIABILITY AND VALIDITY EVIDENCE

## FOR MERRILL'S 2007 5 STAR INSTRUMENT

by

## Max Hale Cropper

A dissertation submitted in partial fulfillment of the requirements for the degree

of

## DOCTOR OF PHILOSOPHY

in

Instructional Technology and Learning Sciences

Dr. Andrew Walker
Major Professor

Dr. Byron R. Burnham
Committee Member

Dr. Anne Hunt
Committee Member

Dr. J. Nicholls Eastmond
Committee Member

Dr. Mark R. McLellan
Vice President for Research and
Dean of the School of Graduate Studies

UTAH STATE UNIVERSITY Logan, Utah

2011

Copyright © Max Hale Cropper 2011

All Rights Reserved

#### **ABSTRACT**

Evaluating Reliability and Validity Evidence for Merrill's 2007 5 Star Instrument

by

Max Hale Cropper, Doctor of Philosophy

Utah State University, 2011

Professor: Dr. Andrew Walker

Department: Instructional Technology and Learning Sciences

Merrill's 2007 5 Star instrument, hereafter called the M-5 Star, is based on Merrill's well recognized First Principles of Instruction. However, the instrument has not been tested for reliability and validity. In a pilot study, Cropper's version of his instrument (C-5 Star) showed some reliability and validity evidence, but the M-5 Star needed similar evidence. To address this gap in the literature, the purpose of this study is to assess the reliability and validity evidence for M-5 Star. Raters were drawn from a graduate course in online course evaluation and asked to rate a sample (N = 6) of exclusively online university classes using M-5 Star and three comparison instruments. The comparison instruments also purport to examine course quality but lack the emphasis on instructional strategies in M-5 Star. Interrater reliability evidence for the M-5 Star and the comparison instruments was moderate to substantial (M5-Star ICC = .56, p = .001); Texas IQ ICC = .43, p = .001; WebCT ICC = .75, p = .001: SREB ICC = .53, p = .001).

However, interrater reliability was tentative because rater pair scores were averaged, biasing the scores toward agreement and inflating ICC. Low correlations between M-5 Star (the criterion) and the comparison measures indicate divergent validity support that M-5 Star is measuring a different core concept of quality in online classes. M-5 Star correlation with WebCT was r = .39, p = .44 ( $r^2 = .15$ ), with WebCT was r = .44, p = .38 ( $r^2 = .20$ ), and with SREB was r = .43, p = .39 ( $r^2 = .19$ ). In addition to divergent validity analysis, a content validity index (CVI) analysis was undertaken using experts in the area of First Principles. According to First Principles experts, other than a few items on the rubric, the vast majority of the M-5 Star CVI results support close alignment with Merrill's First Principles of Instruction. Of the 63 M-5 Star individual items, 56 (89%) of them received high scores on Aiken's CVI that were significant at the .10 level. Study limitations are discussed at length, alongside calls for future research and practical and scholarly significance for the research.

(245 pages)

# CONTENTS

	Page
ABSTRACT	iii
LIST OF TABLES	viii
LIST OF FIGURES	x
I: INTRODUCTION	1
Dissertation Purpose	
II: LITERATURE REVIEW	5
Merrill's First Principles of Instruction  Merrill's 2007 5 Star Instrument  Search for Comparison Scales  Typical Procedures for Instrument Development  Traditional Student Course Ratings  Summary	8 9 13
III: METHODOLOGY	16
Pilot Study	16
Pilot Phase I Pilot Phases II-III	
2007 Class Ratings of Courses Methodology	20
Merrill's 2007 M-5 Star Instrument  Comparison Instrument Selection  Course Selection  School Instrument Rater Training and Rating of Courses  M-5 Star Rater Training and Rating of Courses	22 26 28
Data Analysis Methodology	31
Interrater Reliability Analysis	32

	Content Validity Analysis	37
	Concurrent and Divergent Validity Analyses	
	Summary	42
IV:	RESULTS	43
	Interrater Reliability Analysis	44
	Interrater Reliability for the M-5 Star	44
	Interrater Reliability for the School Online Course Instruments	
	Interrater reliability evidence for the SREB.	
	Comparison of ICC Across_Courses and Instruments	
	Comparison of School Instrument ICC_Between Pilot and Dissertation	
	Content Validity Analysis	51
	Concurrent and Divergent Validity Analysis	51
	Summary	60
V:	DISCUSSION	61
	Limitations of the Study	61
	Discussion of the Findings	65
	Interrater Reliability Analysis	
	Content Validity Analysis	
	Concurrent and Divergent Validity Analysis	67
	Suggestions for Further Research	72
	Interrater Reliability Testing	72
	Content Validity Analysis	
	Concurrent Validity Study	
	Experimental or Quasi-Experimental Study	76
	Conclusion	77
API	PENDICES	83
	Appendix A. Comparison of Instruments	84
	Appendix B. Pilot Study	
	Appendix C. Merrill's 2007 5 Star Instrument (M-5 Star)	
	Appendix D. Instrument Item ICC	
	Appendix E. Merrill's First Principles Domain Outline	
	Appendix F. M-5 Star Instrument CVI Study Cover Page	141

Appendix G. Rater Information	143
Appendix H. M-5 Star Instrument CVI Rating Form	
Appendix I. Potential Expert Raters for CVI Rating	158
Appendix J. Explanation of Low CVI Ratings	163
Appendix K. Texas IQ Instrument	167
Appendix L. WebCT Instrument	185
Appendix M. SREB Instrument	204
Appendix N. Cropper 5 Star (C-5 Star) Instrument	215
Appendix O. Merrill e <sup>3</sup> Instrument	225
**	
CURRICULUM VITAE	232

# LIST OF TABLES

Table		Page
1	Instruction Evaluation Instrument Articles	11
2	Student Ratings of Courses Used and Not Used for Study	27
3	5 Star Instrument Single-Measures ICC Reliability over Principles, for Each Course Separately	45
4	5 Star Instrument Single-Measures ICC Reliability over Courses, for Each Principle Separately	46
5	Texas IQ Single-Measures ICC Reliability for Each Course Separately	47
6	WebCT Single-Measures ICC Reliability for Each Course Separately	48
7	SREB Single-Measures ICC Reliability for Each Course Separately	49
8	Comparison of Instrument ICCs Across Courses	49
9	Comparison of School Instrument ICC Between Pilot and Dissertation Studies	50
10	M-5 Star Instrument Item CVI, z Score and p-Value	52
11	Shapiro-Wilk Test of Normality	54
12	Mean Ratings and Variance of the Various Instruments for the Various Courses Evaluated.	58
13	Pearson Correlations for the Four Course Evaluation Instruments Using Standardized Scores	59
14	Comparison of Correlation Coefficients Between School Instruments and Cropper's and Merrill's 5 Star Instruments	60

Table		Page
15	Number of School Instrument Items That Match Cropper and Merrill 5 Star Instrument Items	68
16	Number of School Instrument Items Matching Other School Instruments	69
17	Standardized Pearson Correlations for Student Ratings Compared with Instruments	71
18	Comparison of 5 Star with School Rating Instruments	86
19	Shapiro-Wilk Test of Normality	102
20	Number of School Instrument Items That Match C-5 Star Items	102
21	Pearson Correlations for the Four Course Evaluation Instruments	105
22	ICC Measures of Reliability for Each Item on the M-5 Star	126
23	Single-Measures Reliability for Each Item on The Texas IQ	129
24	Single-Measures Reliability for Each Item on The WebCT	132
25	Single-Measures Reliability for Each Item On The SREB	133
26	Potential Expert Raters for Content Validity Rating	160
27	Explanation of Low CVR Ratings	165
28	First Principles Course Evaluation Rubric	228
29	Instructional Event Summary	229
30	Demonstration e <sup>3</sup> Quality Rubric	230
31	Application e <sup>3</sup> Quality Rubric	231

# LIST OF FIGURES

Figure		Page
1	Overall design and timeline of pilot study and dissertation study data gathering	18
2	Analysis Methods Used to Answer Research Questions	33
3	Scatterplot of the association between the M-5 Star mean course scores and the Texas IQ mean course scores	55
4	Scatterplot of the association between the M-5 Star mean course scores and the WebCT mean course scores.	55
5	Scatterplot of the association between the M-5 Star mean course scores and the SREB mean course scores	56
6	Scatterplot of the association between the C-5 Star mean course scores and the Texas IQ mean course scores	103
7	Scatterplot of the association between the C-5 Star mean course scores and the WebCT mean course scores	104
8	Scatterplot of the association between the C-5 Star mean course scores and the SREB mean course scores	104

#### **CHAPTER I**

#### INTRODUCTION

Online courses are being offered increasingly in post-secondary and graduate education. However, questions arise regarding their level of effectiveness. Several measures are available for assessing the quality of online courses, including student ratings and a variety of expert rating instruments. Generally the expert rating instruments assess general instructional strategy factors such as presentation, practice, and testing, as well as surface level features such as syllabus quality, instructional objectives, media, interactivity, interaction, etc. However, few instruments measure the inclusion of detailed instructional strategies. In addition, despite the number of available instruments, few have reliability and validity research data to support them.

It would be invaluable to have a reliable and valid instrument for course rating that included detailed instructional strategy criteria. Merrill's 5 Star Instrument (M-5 Star) could serve as that instrument. It is based on Merrill's *First Principles of Instruction*, which was created from a synthesis of instructional strategy best practices. M. David Merrill's First Principles of Instruction is a leading instructional design theory due to its universal nature and its comprehensive focus on both general and detailed instructional strategies. We need to be able to assess the degree to which instruction intentionally or unintentionally utilizes the instructional strategies prescribed by First Principles. Fortunately, the M-5 Star makes it possible to do so. Because of the proliferation of online instruction, there is a need to be able to evaluate instruction in online settings. The M-5 Star could be used for this purpose.

Very little is known about the reliability and validity of the M-5 Star. In the pilot study, Cropper's 5 Star instrument (hereafter called C-5 Star), which was based upon Merrill's 5 Star Instructional Design Rating checklist (Merrill, 2001), was tested for interrater reliability and validity evidence. Evidence was provided that the C-5 Star had moderate interrater reliability evidence, and some concurrent and divergent validity evidence when compared with the same comparison rating instruments used in the dissertation study. The study, however, was preliminary because only award-winning courses were evaluated (Cropper, Bentley, & Schroder, 2009). (See the pilot study summary in the Methodology chapter, and a detailed description of the pilot study in Appendix B.)

#### **Dissertation Purpose**

The purpose of this dissertation study is to evaluate reliability and validity evidence for the M-5 Star as a measure of online course quality for a range of quality of courses. While limited reliability and concurrent/divergent validity evidence was provided for the C-5 Star and the comparison instruments in the pilot study, reliability and concurrent/divergent validity evidence need to be evaluated for the M-5 Star in the dissertation study. Reliability and validity evidence can also be reevaluated for the comparison instruments. The M-5 Star also needs to be tested for content validity to determine whether Merrill's course rating instrument matches his First Principles of Instruction.

The comparison instruments in the pilot and dissertation studies are prominent course ratings used to rate online courses. Two instruments are used for K-12 online course rating, and one instrument is used for rating online college courses. Hereafter, these comparison measures are called school online course rating instruments or school instruments because they are used in school settings.

## **Research Questions**

- 1. How much interrater reliability evidence do the M-5 Star and the school online course rating instruments have?
  - 1.1 How much interrater reliability evidence does the M-5 Star have based on intraclass correlation (ICC)?
  - 1.2 How much interrater reliability evidence do the school instruments have based on intraclass correlation (ICC)?
- 2. To what extent is there validity evidence for the M-5 Star?
  - 2.1 To what extent does the M-5 Star exhibit content validity as compared with a domain outline based upon his First Principles of Instruction?
  - 2.2 To what extent does the M-5 Star have evidence of concurrent or divergent validity as shown by correlation of his instrument's overall scores with school instrument overall scores?

The Literature Review chapter provides an overview of Merrill's First Principles of Instruction, and describes the development of Merrill's 2007 5 Star instrument. It also describes the unsuccessful search for comparison scales with reliability and validity data.

It gives a description of typical procedures for instrument development and a review of reliability and validity evidence for student course evaluations.

The Methodology chapter summarizes pilot study results and outlines the dissertation methodology. The dissertation methodology includes the methodology for the evaluation of courses as well as the data analysis methodology. The Results chapter explains that tentative interrater reliably evidence was found for Merrill's 5 Star instrument, as well as divergent validity evidence when compared with school online course evaluation instruments. A content validity analysis provided some evidence that Merrill's instrument matched a First Principles of Instruction domain outline. The Discussion chapter summarizes study limitations, then discusses the implications and significance of the research as well as possibilities for future research.

#### **CHAPTER II**

#### LITERATURE REVIEW

The literature review provides an overview of Merrill's First Principles of Instruction, and briefly describes the development of his 2007 5 Star instrument. It also describes the search for validated comparison instruments, which unfortunately were not found. Because most course quality measures are unvalidated, a description of standard procedures for development of unvalidated instruments is provided. Because student course ratings were used to select the online courses evaluated for the study, a review of validity evidence for student course ratings is given.

## **Merrill's First Principles of Instruction**

M. David Merrill's First Principles of Instruction is a leading instructional design theory because of its universal nature and its comprehensive focus on both general and detailed instructional strategies. "Instructional theory is concerned with two primary considerations: What to teach and how to teach." (Merrill, 1999, p. 400) *Instructional design theory* focuses on the second consideration, how to teach. According to Reigeluth, instructional design theories are design oriented (or goal oriented) as they describe methods of instruction and the situations in which those methods should be used; the methods can be broken into simpler component methods; and the methods are probabilistic (Reigeluth, 1999). In other words, instructional design theories are

that will be effective in given situations. *Instructional strategies* will be the term used to represent instructional methods or techniques that are used in the instructional process.

Merrill, one of the prominent instructional design theorists, analyzed many instructional design theories, identified the overarching instructional design principles, and developed a comprehensive instructional design theory called the First Principles of Instruction (Merrill, 2002). His First Principles of Instruction consist of core instructional strategies as well as detailed sub strategies that he believes will lead to efficient, effective, and engaging teaching and learning. Merrill maintains that instruction should (1) be *centered around real-world tasks*, (2) include *activation* of the learner's prior knowledge, (3) provide *demonstration* of the tasks being taught, (4) provide the opportunity for the learner to *apply* what is being taught, and (5) require or encourage the learner to *integrate* learning in the real world.

When Merrill created his First Principles of Instruction, he adopted the use of real-world whole tasks from constructivist instructional design theories without abandoning the demonstration and practice recommended by behaviorist and cognitivist instructional design theories. His levels of instructional strategy design illustrate how he has surpassed traditional presentation, demonstration, and practice of smaller tasks, to the demonstration and application of real-world tasks (Merrill, In Press). He recommends using meaningful, real-world tasks for demonstration, practice, and testing (Merrill, 2006, 2007b). While Merrill emphasizes the use of real-world whole tasks for instruction, he also provides detailed instructional strategy prescriptions for demonstration and practice for all types of content including facts, concepts, procedures, principles, and processes.

Merrill's First Principles of Instruction is accepted as a significant instructional design theory by many instructional design experts. In *Instructional-Design Theories And Models: Building A Common Knowledge Base* (Reigeluth & Carr-Chellman, 2009) the authors acknowledge that Merrill's First Principles of Instruction are universal principles of instruction that apply to all types of learning, while claiming that there are also situational principles that provide more specific strategies for different instructional situations.

Some research has been done to provide validity evidence for Merrill's First Principles of Instruction. Cropper has done a synthesis of research that supports Merrill's individual principles (Cropper, 2009). Various authors have cited studies that support Merrill's individual principles (Merrill, 2007a). A few studies provide evidence of the effectiveness of Merrill's principles being used in concert with each other. A Thomson NETg job impact study comparing traditional computer-based instruction with instruction based on First Principles showed an improvement in learning significant at the .001 level (Thomson NETg, 2003). Multiple studies by Frick and his colleagues demonstrated a positive correlation between the use of First Principles of Instruction and student satisfaction and success (Frick, Chadha, Watson, Wang, & Green, 2007, 2009; Frick, Chadha, Watson, & Zlatkovska, 2010). In addition, the studies conducted by Frick and his colleagues found a strong relationship between global ratings (overall instructor and course quality 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" (their successful engagement in tasks, activities, and problems that were related to course objectives) (Frick, et al., 2009, p. 714; Frick, et al., 2010).

#### Merrill's 2007 5 Star Instrument

Online course quality experts have maintained that core instructional strategy standards are still required to achieve online course effectiveness (Graham, Cagiltay, Lim, Craner, & Duffy, 2001; Hirumi, 2002; Naidu, 2007; Piña, 2005; Ritchie & Hoffman, 1997; Sherry, 2003). With the exception of Hirumi, they propose very general instructional strategies. Hirumi and Merrill are the only ones who propose specific instructional strategies for online instruction. Their specific methods make it easier to prepare effective online instruction for all types of learning outcomes. Hirumi's guidelines have not been operationalized into a standard or instrument, but Merrill's instructional strategies have been made measurable with an instrument that he developed.

Merrill's 5 Star instrument has undergone several revisions. Merrill's original 5 Star Instructional Design Rating form in 2001 consisted of yes/no questions for each of his five principles and their sub-principles (Merrill, 2001). In 2005, for the pilot study, Cropper added a five point Likert scale to Merrill's questions and also added a brief description for each possible rating value for each question (Cropper, et al., 2009). (See the C-5 Star instrument in Appendix N.) The pilot study provided some reliability and concurrent and divergent validity evidence for the C-5 Star. (See the pilot study summary in the Methodology chapter and a detailed description of the pilot study in Appendix B.) Merrill, after using The C-5 Star as a rater during Phase III of the pilot

study, created a new, more detailed M-5 Star instrument in 2007 for use in the dissertation study. (See Merrill's 2007 5 Star [M-5 Star] instrument in Appendix C.) The C-5 Star and M-5 Star have had limited use in studies and with classes taught by Merrill, and have not been distributed or published. After the data was gathered for the dissertation study, Merrill developed and published a simplified e<sup>3</sup> instrument, which is also based on First Principles (Merrill, 2009). (See e<sup>3</sup> instrument in Appendix O.)

#### **Search for Comparison Scales**

Initially instruments were sought that could be used for concurrent validity evidence. Ideally comparisons instruments should have reliability and validity data. An EBSCO search of ERIC and a search of Wilson's Web were conducted using various combinations of the terms *reliability*, *validity*, *reliable*, *valid*, *instruction*, *instrument*, *course*, *quality*, *evaluation* and *scale*.

Approximately 150 articles were found. Of those articles, 41 dealt with instruments with reliability and validity data that were used to measure the effectiveness of instruction or instructional programs. Of the 41 articles, two articles dealt with 22 scales used for *student* ratings of clinical classroom instruction which had varying amounts and types of reliability and validity data (Beckman, Cook, & Mandrekar, 2005; Beckman, Ghosh, Cook, Erwin, & Mandrekar, 2004). Because instruments for expert ratings were being sought, these scales were unacceptable for use in the study.

The remaining 39 articles dealt with individual instruments that had been developed for ratings of instruction or instructional programs. (See Table 1 for a

breakdown of what types of instruction the 39 instruments were used to evaluate and who performed the ratings.) Of the 39 articles and their corresponding instruments, only 2 dealt with expert evaluation of computer-based instruction. The other 37 were excluded from the study because they were not designed for expert rating of online instruction.

Of the 37 articles, 4 dealt with program evaluation and development of classroom instruction, 17 dealt with rating of special characteristics or types of instruction, 2 dealt with *instructor* ratings of *classroom* instruction, 13 dealt with *student* and *instructor* ratings of *classroom* instruction, 2 dealt with *expert* ratings of *classroom* instruction, and 1 dealt with *student* ratings of online instruction.

The remaining 2 articles dealt with *expert* ratings of *computer-based instruction*. One of those dealt with rating of self-contained computer and video-based courses (Barrett, 1990). Because it did not deal with interactive online courses, which are the type of online courses evaluated in this study, it was not appropriate for comparison.

The only instrument that was used for rating of interactive online instruction was the e-Learning Courseware Certification Program (eLCC) instrument. Although it had some evidence for reliability and content validity as a scale for expert rating of e-learning (Chen, 2009), it had no evidence for concurrent or divergent validity. Therefore it could not serve as a validated comparison instrument. It was also inappropriate for comparison because its items are very general, and do not relate directly to Merrill's instructional strategies. Only eLCC's "practice and feedback" item relates directly to Merrill's "application" principle and sub-principles.

Table 1

			Program	Rates	Eval. Class-	Student	Instructor	Expert	Evaluates
1	Author	Year	eval.	aspect	room Instr.	Ratings	Ratings	Ratings	e-learning
1	Rose	1977	X						
2	Bassoppo-Moyo	2006	X						
3	Eddy	2007	X						
4	Ginns	2009	X	***					
5	Foster	1975		X					
6	Brandenburg	1979		X					
7	Dolmans	1991		X					
8	Frymier	1995		X					
9	Miejer	2001		X					
10	Bogo	2002		X					
11	Doherty	2002		X					
12	Terrill	2002		X					
13	Flowers	2003		X					
14	Hafner	2003		X					
15	Fehnel	2004		X					
16	Li	2004		X					
17	Gersten	2005		X					
18	Papaioannou	2007		X					
19	Pillay	2007		X					
20	Kay	2009		X					
21	Tsai	2009		X					
22	Masters	1977			X	X			
23	Inglis	1978			X	X			
24	Witte	1981			X	X			
25	Ramsden	1991			X	X			
26	Evans	1993			X	X			
27	Tuan	2000			X	X			
28	Waldrip	2009			X	X			
29	Mitchelmore	1973			X	X	X		
30	Marsh	1979			X	X	X		
31	Marsh	1987			X	X	X		
32	Bastick	2001			X	X	X		
						Λ			
33	MacCuish	1986			X		X		
34	Shim	2001			X		X	**	
35	Doherty	2002			X			X	
36	van de Grift	2007			X			X	
37	Chaney	2007				X			X
38	Barrett	1990						X	X
39	Chen	2009						X	X
	Total		4	17	15	12	6	4	3

*Note.* Columns with X's are listed in progressive order based on increasing potential of instrument inclusion in the study.

Some of the remaining eLCC items relate only indirectly to Merrill's principles and sub-principles.

- The eLCC's "Media design" relates somewhat to Merrill's appropriate use of media found under demonstration and application for kinds-of, how-to, and what-happens instruction.
- The eLCC's "Assessment" relates indirectly under a sub-principle of Merrill's integration principle, "Do learners have an opportunity to publicly demonstrate their new knowledge or skill?"
- The eLCC's "Instructional presentation" doesn't quite match Merrill's "demonstration."

The rest of the eLCC's items do not relate to the M-5 Star. Because few items are closely related to the M-5 Star items it is not acceptable as a concurrent validity comparison instrument. However, because of the differences between the eLCC and the M-5 Star, it would seem that it could be used as a divergent validity comparison measure. However, the eLCC items are too general and vague for a useful comparison instrument. The criteria (i.e., "media design," "assessment," instructional presentation,") are so vague and ill-defined that it would make it extremely difficult for raters to give reliable and valid ratings. No further criteria or training information is available that would make the items clearer or easier to rate. Therefore, the eLCC is unacceptable as a comparison instrument, and was also eliminated from the current study.

## **Typical Procedures for Instrument Development**

During the search for comparison scales, few validated instruments for expert rating of online courses were found. A majority of instruments developed for expert rating of online courses are not validated. Typically, course quality instruments are developed by individual experts or teams of experts who formally or informally develop questionnaires that represent what they believe are the criteria for successful courses. They may develop original questionnaires or imitate other questionnaires.

Individuals or group members create, review, and refine the item categories and the items within those categories, then finalize their questionnaire. Generally the questionnaires are not reviewed for validity by other experts, nor are they subjected to item analysis procedures or tested for interrater reliability. Although the questionnaires have no reliability or validity data, they are sometimes implemented under the assumption that they are reliable and valid.

For example, at Chico State University, The Committee for Online Instruction (COI) was formed to address the need for demonstrating quality in online instruction, and for setting some guidelines for developers of online teaching. This committee (originally called CEEOC), comprising faculty, four staff, two administrators, and one student, represented a cross section of the teaching and learning environment of Chico State. The committee followed a process similar to the one described above to create the *rubric for online instruction* (ROI). In order to draw from the expertise of the scholarly community, the committee reviewed existing best practices, learning styles, and standards (e.g., Graf

and Caines' WebCT Exemplary Course Rubric, Bloom's Taxonomy, Chickering & Gamson's 7 Good Teaching Practices in Undergraduate Education). Although Chico State's ROI rubric is currently used by approximately 90 colleges and universities (Chico State University, 2010), the developers are not aware of anyone having done a study to prove its effectiveness (L. Sederberg, personal communication, February 3, 2009).

## **Traditional Student Course Ratings**

Because courses for the dissertation study were selected based upon student ratings, it is important to know whether the university student course rating form is a reliable and valid measurement of course quality. University student ratings of course and instructor effectiveness were used to select the online courses for the 2007 evaluation. However, are student ratings a reliable and valid measurement of course quality? Students are probably not experts in online course construction, but overall instructor and course ratings given by students have some evidence of validity (Frick, et al., 2009).

The USU Faculty Senate (2009) found high internal reliability (consistency in responses to questions) for the university rating form, which was the basis for selecting courses for the dissertation study. However, no questions on the current form have been tested for validity. In lieu of the expense of conducting a validity analysis, the faculty evaluations committee made recommendations to pilot a validated commercial instrument for faculty evaluation (USU Faculty Senate, 2009).

Frick et al. (2009) conducted a literature review which revealed that "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" (Frick, et al., 2009, p. 713). Global items are items such as "Overall, this was an outstanding course," "Overall, this was an outstanding instructor," and "I would recommend this instructor to others" (pp. 713-714).

Their findings can be used as evidence that the USU course global ratings have some validity as ratings for course quality. The USU student rating form has a rating for instructor effectiveness in teaching and a rating for overall quality of a course. When course selection was done for the 2007 class ratings, these two ratings were averaged to create a single course quality factor. This factor provided an overall course quality rating, which was used to select a range of quality of courses.

## **Summary**

The literature review describes Merrill's First Principles of Instruction theory, including empirical support for the theory. It briefly describes how his 2007 M-5 Star instrument was developed, and the search for validated comparison instruments, which unfortunately were not found. It was discovered that most course quality measures are unvalidated. Therefore a description of standard procedures for development of unvalidated instruments is provided. Because student course ratings were used to select the online courses evaluated for the study, a review of validity evidence for student course ratings is given.

#### **CHAPTER III**

#### **METHODOLOGY**

#### **Pilot Study**

A pilot study was conducted to determine whether Merrill's 5 Star Instructional Design Rating had reliability and validity evidence when evaluating award-winning courses. Cropper added a 5 point Likert scale to each of Merrill's 26 items, as well as a description of each of the rating levels for each item. Cropper's instrument is called the C-5 Star.

Specifically, the three-phase pilot study was conducted to assess the interrater reliability evidence of the C-5 Star, a predecessor to the M-5 Star rubric, and some comparison instruments. The study was also designed to determine whether the comparison instruments provided convergent or divergent validity for each other. Awardwinning courses were selected from both academic and commercial entities because they were assumed to represent a variety of high quality courses. These high-rated courses would be expected to score high on the C-5 Star as well as on comparison instruments. Evaluators varied from a single pair of expert raters, to an additional three expert raters drawn from a pool, and finally in Phase III, Merrill himself.

To ensure an adequate statistical foundation, a significance level of .10 was selected based upon the exploratory research purpose of the study and the small sample sizes of courses. It is inappropriate to set a stringent significance level in exploratory research (a .10 level is acceptable). Likewise, it is inappropriate to set a lenient

significance level in confirmatory research (a .05 or .01 level is required) (Garson, 2000). Alpha levels of .10 significance were accepted throughout the pilot and dissertation studies because of their exploratory nature.

The three-phase pilot study is exploratory because it is the first time that a version of a 5 Star instrument has been evaluated for interrater reliability evidence and concurrent and divergent validity. It is also exploratory because the comparison instruments have no previous reliability or validity data. The dissertation study is exploratory rather than confirmatory inasmuch as the M-5 Star instrument has been revised and a range of quality of courses was evaluated rather than exclusively award-winning courses. (See Figure 1 for an overview of the design of all phases of the pilot study and the dissertation study.)

The pilot study, which is described in detail in Appendix B, provided some evidence of the interrater reliability of the C-5 Star and the school online course instruments as measures for evaluating award-winning courses. Interrater reliability evidence between two raters was calculated using intraclass correlation (ICC), which was used because it accounts for differences between judges (Howell, 2002). According to Garson (2010b), intraclass correlation (ICC) is interpreted similarly to kappa. By convention, a kappa > .70 is considered acceptable, but this depends highly on the researcher's purpose. Another rule of thumb is that K < 0 is poor agreement, K < .20 is slight agreement, K = .20 to .39 is fair agreement, K = .40 to .59 is moderate agreement, K = .60 to .79 is substantial, and K = .80 to 1.0 is outstanding (Garson, 2010b; Landis &

Koch, 1977). Because this was an exploratory study, the lower thresholds for kappa (and ICC) were used.

		Pilot Study	Dissertation Study Data Gathering		
	Fall 2005	Spring	Summer 2007		
	Phase I	Phase II			
Raters	2	3 experts added	Merrill added	10 ID experts including Merrill	
Course type	3 college 5 commercial	2 college 4 commercial		10 university courses	
Course quality		Award-winning	Full range of quality		
Baseline instrument		Cropper C-5 Star			
Other instruments	3 school 2 commercial 1 motivational	None		3 school	

Figure 1. Overall design and timeline of pilot study and dissertation study data gathering.

## Pilot Phase I

The ICC interrater reliability evidence was fair for the Texas IQ instrument (ICC = .39, p = .001), moderate for the SREB (ICC = .40, p = .001) and the C-5 Star instrument (ICC = .59, p = .001), and substantial for the WebCT (ICC = .60, p = .001).

Correlations were calculated between the instruments to determine whether they provided concurrent or divergent validity evidence for each other.

Correlations of r > .70 (p < .10) ( $r^2 > .49$ ) were accepted for concurrent validity evidence. Correlations of r < .50 (p > .10) ( $r^2 < .25$ ) were accepted for divergent validity evidence. Correlations between .50 and .70 would provide neither concurrent nor divergent validity evidence. Concurrent and divergent validity with school instruments were mixed.

The Texas IQ provided concurrent validity evidence (r = .90, p = .002) for the C-5 Star. While Pearson's r tells us the amount of correlation, its square ( $r^2$ ), which is called the coefficient of determination, is the effect size, and tells how much effect a variable has on another. The effect size,  $r^2$ , is interpreted as percent of variance explained by the independent variable. In the case of the Texas IQ and the C-5 Star instrument, r = .90 and  $r^2$  is .80, therefore the independent variable (the Texas IQ rating) is said to explain 81% of the variance in the dependent variable (the C-5 Star rating) or vice versa (Garson, 2010a).

Both the WebCT (r = .45, p = .26) ( $r^2 = .20$ ) and SREB (r = .04, p = .93) ( $r^2 = .001$ ) provided divergent validity evidence for the C-5 Star. A similar mixture was found when comparing school instruments. The WebCT was correlated at a significant level with the Texas IQ (r = .64, p = .04) ( $r^2 = .412$ ), and SREB (r = .69, p = .03) ( $r^2 = .48$ ), providing concurrent validity for the WebCT. However, the Texas IQ and SREB had a low correlation (r = .34, p = .21) ( $r^2 = .11$ ), indicating divergent validity evidence for one

another. Because the interrater reliability evidence of the instruments varied from fair to substantial, some of the validity evidence is tentative.

#### Pilot Phases II-III

In pilot Phase II, three additional expert raters were added and only the C-5 Star was used for course rating. The ICC for the C-5 Star was lower, but fair, and still statistically significant (ICC = .34, p = .001). During pilot Phase III, with Merrill added as a rater, the ICC for the C-5 Star was lower than in Phase II, but still significant (ICC = .24, p = .001). It seems that the addition of expert raters during the phases may have polarized the ratings between purists (such as Merrill) and more lenient raters and may have therefore led to the decrease of interrater reliability, although all levels of interrater reliability were significant at the .001 level.

As only award-winning courses were used in the pilot study, it was difficult to derive reliability and validity evidence. The use of award-winning courses resulted in observations clustered around the high end of each ratings scale, reducing variance, covariance and correlation, as well as making correlation scores particularly suspect to outliers. In this dissertation study, a wide range of quality of courses was evaluated, providing the necessary variance to assess interrater reliability and divergent or convergent validity in a more meaningful way.

## 2007 Class Ratings of Courses Methodology

For the dissertation, rater pairs were drawn from a summer 2007 *Evaluating*Online Courses graduate class to assess a range of quality of online courses. Baseline

raters included Cropper, Joel Gardner (a teaching assistant), and Merrill. The M-5 Star was used, as well as the school online course rating instruments that were used in the pilot study. Two baseline raters and two randomly assigned raters were assigned to each of the four instruments (M-5 Star, Texas IQ, WebCT, and SREB). The randomly assigned rater pairs were assigned to only one instrument to avoid rating bias across instruments. Raters were trained on their assigned instruments, and then evaluated the courses using their assigned instrument.

#### Merrill's 2007 M-5 Star Instrument

The C-5 Star was used through all three phases of the pilot study. A key feature of the C-5 Star includes equal emphasis and value on each of Merrill's five principles. After serving as a rater during Phase III of the pilot study, Merrill initiated several changes reflected in the 2007 M-5 Star instrument. Merrill's changes reflected an increased emphasis on task-centeredness and task-centered instructional strategy (worth 30 points with one *task centered* item worth 15 points, another worth 10 points, and other items, if applicable, worth 5 points each), a continued emphasis on demonstration and application (each principle worth 25 points, with items worth 5 points each), and a decreased emphasis on activation (worth 10 points, with items worth 2 points each) and integration (worth 10 points, with items worth 3 or 4 points each). Finally, Merrill eliminated principle summary items and provided additional detailed items for each principle. (See actual items and points possible on his 2007 M-5 Star instrument in Appendix C.)

### **Comparison Instrument Selection**

The plan for this study was to compare the M-5 Star with expert rating instruments that had detailed instructional strategy criteria. In the pilot study, some prominent school online course instruments were used for comparison with the C-5 Star. (See Appendix B.) In the dissertation study, the same school online course rating instruments were used for comparison with the M-5 Star.

When instrument selection was done during the first phase of the pilot study, it was decided to use Merrill's M-5 Star Instructional Design Rating checklist as the baseline standard. Cropper added a 5-point Likert scale and descriptions of each of the five possible ratings for each item to create the C-5 Star. Other reputable instruments with similar levels of instructional strategy detail were sought as comparison instruments.

The author was aware of online course rating instruments that had been developed by WebCT, Brigham Young University (BYU), Michigan Virtual University (Michigan Virtual University, 2002) and American Society for Training and Development (ASTD). Hirumi, an online course standards expert, was contacted regarding possible instruments. He recommended the Texas IQ online course rating instrument, which he helped develop as part of the Investigating Quality of Internet Course (IQ) Project under the direction of the Texas Education Agency (TEA). In addition, a Google search was conducted using various combinations of *online course, standard, instrument*, and *evaluation*. The Checklist for Online Interactive Learning (COIL) (Sunal, Sunal, Odell, & Sundberg, 2003) and the Southern Regional Education Board's (SREB) Criteria for Evaluating Online Courses (Southern Regional Education Board, 2005) were found. While checking

with representatives of Brandon Hall about evaluating courses that had won their award, they also offered the use of their instrument for the pilot study.

During the Google search, additional online course standards were found, but no additional instruments were found other than those already mentioned. Because the study was focusing on the M-5 Star, which included items based upon detailed instructional strategies, existing instruments which focused on detailed instructional strategies were selected for inclusion in the study.

The types of instructional strategy criteria included core instructional strategies (i.e., activation, demonstration, practice, testing), media use, interactivity, communication, collaboration, organization, objectives, etc. In other words, the study included factors that were implemented by the instructor or instruction that could directly influence the effectiveness of the learning experience. Administrative factors were excluded because the study focused on instructional strategy factors. In some cases, when comparison instruments focused on both instructional strategies and administrative factors, only the portions of instruments which focused on instructional strategies were used. This made final instruments closer matches because they only contained instructional strategy criteria.

Another selection criterion was that the instrument existed in, or could easily be converted to, a checklist style format. The checklist format generally provided simple, easy to understand items. The checklist format also made it possible to add a Likert scale to each item, which made it possible to have variance in item response. An additional selection criterion was that the instrument shouldn't have an inordinate number of items

for evaluators to rate so they could complete course ratings in a reasonable amount of time. The rating forms that best met the selection criteria, and were therefore selected for the dissertation study, were:

- Merrill's M-5 Star instrument;
- the Texas IQ instrument;
- the WebCT Exemplary Course Project 2005 Nomination Instructions and Form; and
- the Southern Regional Education Board's (SREB) Criteria for Evaluating
   Online Courses.

The Texas IQ rating form was developed as part of the Investigating Quality of Internet Course (IQ) Project, which was initiated under the direction of the Texas Education Agency (TEA) during the fall of 2001. The purpose of the project was to develop a tool that could be used to improve the quality of internet-based courses for Texas students. (See the Texas IQ instrument in Appendix K.)

The WebCT instrument (Exemplary Course Project 2005 Nomination Instructions and Form), hereafter called WebCT, was used by course authors to evaluate their course in preparation for submission for the award. The same criteria were used to judge courses for the award. (See the WebCT instrument in Appendix L.) The Southern Regional Education Board's (SREB) Criteria for Evaluating Online Courses were based on the SREB *essential principles of quality* and were designed to assist states in determining the quality and effectiveness of web-based courses. (See SREB instrument in Appendix M.)

A 5-point Likert scale was added to all of the instruments so all instrument ratings would be done on the same 5-point scale. The five point scale would also make it easier to discern variations in course quality. A description for each rating value (from 1-5) for each question was created so raters could more easily understand what each rating value meant.

A comparison of the C-5 Star and M-5 Star instruments with the school online course rating instruments reveals that the various instruments provide emphasis on different areas. (See Appendix A.) For example, the 5 Star instruments provide emphasis on Merrill's task-centered focus, activation, the specific strategies for demonstration and application for various kinds of content, and integration in the real world. The school instruments emphasize more general instructional strategies and surface features including: objectives, course requirements, appropriate use of media, content, practice consistent with objectives, practice followed by corrective feedback, collaboration, and effective use of online technology.

Because the other instruments measure different types of instructional strategies, they may provide divergent validity evidence for the M-5 Star. These school online course instruments selected had the most instructional strategy questions of checklist type instruments available. Therefore, it is believed that they were the best instruments available for comparison with the M-5 Star. Because the Texas IQ, WebCT, and SREB instruments are intended for rating of online school course, they will be called the school instruments.

The instruments used in the dissertation study did not have any reliability or validity data prior to the pilot study. In the pilot study, the C-5 Star and the selected online course rating instruments did exhibit some interrater reliability evidence and concurrent validity. (See the pilot study summary in the Methodology chapter and the detailed description of the pilot study in Appendix B.) The pilot study results were used as preliminary evidence for the dissertation study analysis. However, the M-5 Star, which was used in the dissertation study, was significantly different from the C-5 Star, so the dissertation results would likely be different.

#### Course Selection

Online courses were used in the pilot and dissertation studies for a number of reasons. Online instruction provides more of a tangible entity to evaluate than classroom instruction because more course material is accessible. Online materials may also be less subject to variation and change, which provides more consistency for rating over time.

Ten courses were selected from available online courses at Utah State University (USU) using a stratified random sample of courses based upon student ratings.

Specifically, student ratings of the overall instructor quality and overall course quality were used as a proxy for course design quality. While certainly not producers of instructional designs, students are consumers of them and their opinions may reflect the instructional strategies used within each class. In addition, student course ratings are accessible data that may be used to select courses of varying quality.

To ensure high variability in course quality, three of the lowest, four mediumrated, and three of the highest rated classes were selected for ratings. This selection methodological shift from the high-rated classes in the pilot was done in an attempt to observe the full range of the ratings scale and improve interrater reliability. The courses covered a wide range of topics, including Biology, Communication Disorders, Computer Networking, English, Health, HTML, Philosophy, Psychology, Statistics, and Theatre. When the raters evaluated the 10 courses, they discovered that four classes did not have enough online material to be evaluated effectively. For example, courses may have referred to demonstrations, practice, and hands-on tests, but did not provide access to the demonstrations, practice exercises or tests, nor did they provide sufficient descriptions of the practice or test activities. The four classes with insufficient course material (Communication Disorders, Health, Computer Networking, and Philosophy) were not included in the study results. (See Table 2.)

Table 2
Student Ratings of Courses Used and Not Used for Study

Course	Quality	Included
Communication	High	No
Theatre	High	Yes
HTML	High	Yes
Statistics	Medium	Yes
Psychology	Medium	Yes
No Health	Medium	No
English	Medium	Yes
Biology	Low	Yes
Networking	Low	No
Philosophy	Low	No

The six remaining courses were Biology, English, HTML, Psychology, Statistics, and Theatre. Of these remaining courses, one was low-rated, three were medium-rated,

and two were high-rated. The six remaining courses included a variety of types of learning content (i.e., facts, concepts, procedures, principles, and systems) as well as a variety of types of cognitive processes (i.e., understanding, remembering, creating/designing, low- and high-level problem solving, and evaluating).

Course domain or topic may influence the variation in course quality, as might the level (undergraduate to graduate), or type of course (survey, applied). The instrument criteria were applied equally to all domains, levels, and types of courses, based on their use of instructional strategy criteria found in the instruments. While all courses theoretically have an equal chance to score high on instructional strategies used, some domains, levels and types of courses are especially susceptible to low scores on the instruments. For example, survey courses tend to be lecture oriented, and survey course tests are generally recall-type tests. They would score low on the M-5 Star and the other instruments because of lack of demonstration, practice, and real-world task-centered examples.

### **School Instrument Rater Training and Rating of Courses**

Pairs of raters were trained on each of the school online course instruments (WebCT, Texas IQ, and SREB) by one of three instructors. The instructors were Joanne Bentley, Cropper, and Gardner, a teaching assistant. Each instructor provided similar training, guiding their assigned pair of raters through a collaborative rating of an online T-Tapp exercise course using their pre-assigned rating form. Next, each rater evaluated an Educational Research course individually, and then reconciled scores with their partner and the instructor within one point on a 5-point scale. The constraint to reconcile

within one point was chosen because some items were difficult to rate because of complexity or ambiguity (ConQir, 2010; Stemler, 2004). Reconciling helped raters evaluate more carefully, as they educated each other regarding the reasons for their ratings. Rater pairs were able to standardize the way they interpreted each question as well as what each of the possible ratings meant. This reconciliation approach was also used throughout the study by all rater pairs.

After receiving training, raters assigned to each school instrument were given a few weeks to evaluate the 10 online courses. After rating each course individually, partners reconciled their ratings for each item within 1 point on a 5-point rating scale. Cropper and Gardner served as the baseline raters for the school instruments. Their average rating for each question on each instrument served as the baseline rating for the three school online course rating instruments. This provided an expert standard rating with which the other rater pairs could be compared. During the data analysis, the average rating of the rater pairs for each of the school instruments was compared with the baseline rating.

#### M-5 Star Rater Training and Rating of Courses

After the raters for the school instruments had evaluated all 10 courses, Merrill himself trained all raters, including the school instrument raters. All of the raters had read some of Merrill's recent articles. The school instrument raters voluntarily commented that the M-5 Star items were much more substantive than the items on the instruments which they had been using and asked why they had not been using the M-5

Star instead of the one they were assigned. It was explained that the school instruments were used because they were being compared with the M-5 Star.

When Merrill provided training on his instrument, he walked the raters through each item on the instrument, describing what it meant and how to rate the item.

However, he did not provide a sample rated course as an example. Because of the complexity of the M-5 Star and the lack of a worked example, the raters assigned to his instrument initially struggled with the rating form, especially with the definition of whole tasks and task-centered strategy. This led to dramatic scoring variation on task-centeredness, which represents 30% of the total possible score. Another critical issue that surfaced was that some courses have a progression of significant tasks or assignments that may have lead to a whole task at the end. The M-5 Star penalized those courses because it required that a course use a progression of whole tasks for demonstration, guided practice and unguided practice. To deal with this issue, Merrill revised his instrument to provide some task-centered strategy points for a progression of partial tasks, but more points for having a progression of whole tasks.

Merrill gave additional training to the raters assigned to his instrument, then provided clarification to the raters of his emphasis on "whole-tasks" and "task-centered instructional strategy." Merrill also rated the selected courses, providing a standard with which the other raters could be trained and compared.

Merrill and Cropper served as baseline raters using the 2007 M-5 Star instrument.

Merrill was the first choice as a baseline rater because he authored the instrument.

Cropper was selected because he is also an expert on the M-5 Star. He has studied and

taught Merrill's instructional design theories for many years. He also developed the C-5 Star instrument based on Merrill's original 5 Star Instructional Design Rating and conducted all three phases of the pilot study.

After Merrill and Cropper completed ratings of the courses, they reconciled their ratings. Cropper had to lower his ratings significantly, especially on the problem-centered factor, to more closely align with Merrill's scores.

Gardner and a class member served as the other pair of raters assigned to the M-5 Star. They also rated courses significantly higher than Merrill. Cropper retrained the rater pair, and had them compare and reconcile their ratings within one point or 20% of Merrill's ratings on each item for two of the courses. The pair of raters subsequently rerated the rest of the 10 courses individually, and then reconciled their ratings within one point or 20% of each other on each item.

## **Data Analysis Methodology**

Analyses of the data from the 2007 Evaluating Online Courses class were conducted in order to evaluate reliability and validity evidence for the M-5 Star as a measure of online course quality. First, an analysis of interrater reliability of the various instruments was completed. Next, a content validity analysis of the M-5 Star was conducted to determine if it matched a First Principles of Instruction domain outline. Finally, concurrent and divergent validity analyses of the M-5 Star were conducted to determine if the M-5 Star and the school course rating instruments were measuring similar or different constructs. These analyses answer the research questions that were

stated in the Introduction chapter. A summary of the research questions and analysis methods used to answer the research questions is found in Figure 2.

### **Interrater Reliability Analysis**

Research question 1: How much interrater reliability evidence do the M-5 Star and the school online course rating instruments have? To determine interrater reliability evidence, intraclass correlation (ICC) was used.

Four types of reliability evidence can be gathered for a rating instrument.

- Internal consistency: how consistent an instrument is across its own items
- Split-half reliability: consistency between two equivalent versions of a rating form
- Test-retest reliability: Estimation based on the correlation between two (or more) administrations of an instrument given at different times, locations, or to different populations
- Interrater reliability: Interrater reliability is a measure of consistency between raters, based on the correlation of scores between two or more raters who rate the same item, scale, or instrument (Garson, 2010b).

This exploratory study will focus on interrater reliability. Reliability, which is the consistency of assessment results (Linn & Gronlund, 2000), must be established before validity can be established. In other words, reliability is required to have validity, but it is possible to have reliability (consistent results) without having validity (accurate results).

#	Research Questions	<b>Analysis Method</b>	Rationale	Data Source
1.1	How much interrater reliability evidence does the M-5 Star have?	Intraclass correlation (ICC)	ICC used for two or more raters when data can be considered interval. Also takes differences due to judges into account	Mean item ratings of rater pairs
1.2	How much interrater reliability evidence do the school instruments have?	Intraclass correlation (ICC)	ICC used for two or more raters when data can be considered interval. Also takes differences due to judges into account	Mean item ratings of rater pairs
2.1	To what extent does the M-5 Star exhibit content validity evidence?	Aiken's Content validity index (CVI)	Aiken's CVI accounts for partial matches, and can be transformed into a <i>z</i> score with its corresponding <i>p</i> value.	Item content validity ratings by experts
2.2	To what extent does the M-5 Star have evidence of concurrent or divergent validity as shown by correlation of M-5 Star overall scores of courses with school instrument overall scores of courses?	Pearson's r	Concurrent or divergent validity can be established using correlations. Pearson's <i>r</i> is the usual measure of parametric correlation	Mean course ratings per instrument

Figure 2. Analysis methods used to answer research questions.

Intraclass correlation is a scale reliability measure used for two or more raters that takes rater differences into account (Howell, 2002). Single-measures intraclass correlation is used because it assesses whether the ratings of one judge are apt to be the same as for another judge (Garson, 2010b).

ICC interrater reliability was calculated between the consensus scores of the baseline rater pair and the other rater pair for each online course rating instrument. One approach to consensus rating is adjacent rating which allows raters to rate within one point of each other. The advantage to this method is that it relaxes the strict criterion that the raters agree exactly. The distinct disadvantage is that it can also lead to an inflated interrater reliability estimate if there are only a limited number of categories from which to choose (e.g., a 1-4 scale) (ConQir, 2010; Stemler, 2004).

Dissertation study raters were required to do an adjacent rating, reconciling within one point on each item, because of the difficulty of rater pairs achieving consensus on some items. However, because the limited 1-5 Likert scale likely would likely lead to inflation of interrater reliability, rater-pair scores were averaged and each rater-pair score was used as a single consensus score.

Averaging scores that were reconciled within one point could potentially lead to a non-exact consensus rating. However, the averaged pair scores might be very close to an actual consensus rating because raters gave an exact consensus rating to a majority of instrument items.

While averaged rater-pair scores will give more accurate interrater reliability than individual scores from adjacent ratings, the interrater reliability will still be somewhat

inflated. The average baseline rater-pair score will be closer numerically to the other average rater-pair score than individual ratings would have been.

Research question 1.1: How much interrater reliability evidence does the M-5 Star have based on intraclass correlation?

For the M-5 Star, ICC interrater reliability was calculated over items to determine which items provide consistent results and which items do not. ICC was also calculated for each principle separately to determine the interrater reliability across Merrill's five principles, for each course separately to determine the interrater reliability across courses, and for the overall M-5 Star instrument.

Research question 1.2: How much interrater reliability evidence do the school instruments have based on intraclass correlation?

For each of the school instruments, ICC was calculated over items to determine the interrater reliability of the individual items, across courses to determine their interrater reliability across courses, as well as for each instrument overall to determine the overall interrater reliability of each instrument.

### **Gathering Validity Evidence**

Research question 2: To what extent is there validity evidence for the M-5 Star? An instrument is valid if it actually measures what it claims to. The types of validity include construct validity, content validity, and internal validity. Measures have construct validity when their items match the construct they are designed to represent. Construct validity includes convergent validity and discriminant or divergent validity (Garson, 2010c).

Convergent validity includes internal consistency, criterion validity, and external validity. Internal consistency seeks to assure that there is correlation among the indicators for a concept. Criterion validity is the correlation between scale or instrument measurement items and known and accepted standard measures or criteria. Criterion validity can be concurrent, when a measure is compared with a criterion concurrently, or predictive, when a measure is to be compared with a future criterion. Another type of convergent validity, external validity, asks whether findings using a measure are generalizable. Discriminant, or divergent, validity involves providing evidence that a criterion and a measure are not measuring the same construct (Garson, 2010c).

Content validity, also called face validity, is used to determine whether items seem to measure what they claim to, including whether the items measure the full domain implied by their label. Internal validity has to do with defending against internal sources of bias arising in research design, so that variables other than the independent ones being studied won't be responsible for part or all of the observed effect on the dependent variable(s) (Garson, 2010c).

The types of validity evidence sought in this exploratory study are content validity, as well as concurrent and divergent validity. Content validity evidence is needed to determine whether the M-5 Star matches Merrill's First Principles of Instruction. This evidence can be gathered by having experts on Merrill's First Principles of Instruction compare the M-5 Star with a domain outline of his First Principles of Instruction, to see how well the M-5 Star instrument matches.

Concurrent or divergent validity evidence can be provided for the M-5 Star by comparing other course rating instruments to Merrill's instrument. This is accomplished by showing a correlation between scale or instrument measurement items and known and accepted standard measures or criteria. Ideally these criteria are direct, objective measures of what is being measured. Where direct objective measures are unavailable, the criteria may be merely closely associated (Garson, 2010c). Concurrent validity evidence between the M-5 Star and the school instruments would suggest that the school instruments measure the same construct as the M-5 Star, in other words, similar instructional strategies.

Divergent validity evidence, also called discriminant evidence, could be used to show evidence that the M-5 Star measures something different than other instruments. Low correlations between the M-5 Star and other instruments would provide discriminant validity evidence. Discriminant validity evidence would suggest that the other measures are measuring something different than the M-5 Star, and that the M-5 Star does not represent a construct other than the one for which it was devised (Kaplan & Saccuzzo, 2005).

#### **Content Validity Analysis**

Research question 2.1: To what extent does the M-5 Star exhibit content validity as compared with the First Principles of Instruction domain outline?

A content validity analysis was completed to determine the degree to which the M-5 Star matches his First Principles of Instruction. Three major statistical approaches are used to assess the fit of individual items to a content domain (Crocker, Miller, &

Franks, 1989). They are: (1) Hambleton's index of item-objective congruence (Hambleton, 1980), (2) Lawshe's content validity ratio (CVR) (Lawshe, 1975), and (3) Aiken's content validity index (CVI) (Aiken, 1980). Aiken's CVI analysis was used in this study because the item CVI can be converted to a *z* score, with a corresponding *p* value (the probability that the *z* score was obtained by chance). Aiken's CVI also accounts for partial matches.

To aid in the CVI analysis, a First Principles of Instruction domain outline was created that summarized the key elements of Merrill's principles from his First Principles articles. (See Appendix E.) A content validity rating scale was added to M-5 Star items. (See Appendix H.) The M-5 Star content validity rating form and the First Principles of Instruction domain outline were given to a number of individuals who had a range of expertise on First Principles of Instruction. (See list of potential raters in Table 27, Appendix I.) Raters were asked to rate the extent to which the items in the M-5 Star instrument content validity rating form were included in the First Principles of Instruction domain outline.

**Identification of validity raters.** Potential expert raters for the content validity analysis included faculty and students who were identified as experts on First Principles of Instruction by Merrill, Frick, Bentley, and Cropper. Potential expert raters included

- faculty experts who have published on First Principles of Instruction and other instructional design theories;
- doctoral students and Ph.D. graduates at Indiana University (IU) who assisted
   Ted Frick with research using First Principles of Instruction rating scales;

- graduate students at various universities who had developed 5 Star courses, or done research and written articles about First Principles of Instruction;
- graduate students at USU who had taken online course evaluation classes and had used the M-5 Star; and
- graduate students at Florida State University (FSU) and the University of Hawaii (UH) that had taken a First Principles class from Merrill.

All of the participants in the study are potentially biased because of their association with Merrill or one of his advocates. While USU students that have participated in M-5 Star ratings may potentially be the most biased because they have used the M-5 Star, all potential raters have been taught or trained by Merrill or his advocates, so all potentially have a bias. Regardless of bias, there are a limited number of First Principles experts that were considered knowledgeable enough to do an accurate content validity rating.

Rater expertise level was calculated using factors that provided evidence of First Principles of Instruction expertise. Expertise factors included: being a faculty member, being an instructional design theory expert, having taken a course evaluation class, having taken a First Principles class, having been a 5 Star instrument evaluator, having done research on First Principles of Instruction, having published papers on First Principles, and using First Principles in their instruction. Individual factors were summed to identify an overall level of First Principles expertise for each potential rater. (For the complete list of potential expert raters, in order of estimated First Principles expertise, see Table 26 in Appendix I.)

Content validity study methodology. Four electronic documents were sent to all 64 potential raters that had been identified by Bentley, Merrill, Frick, and Cropper. (See list of potential raters in Table 26, Appendix I.) The documents included

- a cover page with instructions (Appendix F);
- a rater information form (Appendix G);
- a M-5 Star instrument content validity rating form (Appendix H); and
- a First Principles of Instruction domain outline (Appendix E).

A tutorial video on YouTube (see Appendix F for link) was also provided, with a demonstration of how to complete some sample ratings. Raters were given approximately a week to complete and return the rater information form and the M-5 Star instrument content validity rating form. The opportunity to win a \$50 gift certificate was offered as an incentive for completing the CVR ratings. Only 10 CVR rating forms were returned the first week. As a result, a second request was sent, and another 12 ratings were returned, making a total of 22 raters, which represented 34% of the 64 potential raters. A high percentage (approximately 90 %) of potential raters with high First Principles expertise participated in content validity ratings. This could possibly be because of their high level of interest in Merrill's First Principles of Instruction.

Aiken's CVI was calculated for each of the M-5 Star items. Corresponding z scores were calculated for each of the CVI ratings with their accompanying p values. These calculations provided evidence of the level of content validity for each of the M-5 Star items.

#### **Concurrent and Divergent Validity Analyses**

Research question 2.2: To what extent does the M-5 Star have evidence of concurrent or divergent validity as shown by correlation of M-5 Star overall mean scores of courses with school instrument overall mean scores of courses?

In the pilot study, the Texas IQ provided significant concurrent validity evidence for the C-5 Star. The WebCT showed some divergent validity evidence, and the SREB provided substantial divergent validity evidence for the C-5 Star. The M-5 Star was substantially different from the C-5 Star, and even more different from the school instruments; it eliminated principle summary questions and added detailed sub-principle questions. (For a full comparison between instruments see Appendix A.) However, it was still uncertain whether concurrent or divergent validity evidence would be found between the M-5 Star and the school instruments. Therefore, an analysis was done to determine whether there was convergent or divergent validity evidence between the M-5 Star and the school instruments.

Concurrent or divergent validity evidence for the M-5 Star can be established using correlations. Pearson's r is the usual measure of parametric correlation, sometimes called Pearson *product-moment correlation*. Pearson's r is a measure of linear relationship between X and Y values of the instrument scores. The correlation coefficient (r) varies from -1 to +1, with 0 indicating no relationship, 1 indicating a perfect positive relationship and -1 representing a perfect negative relationship (Garson, 2010a, p. 1). Pearson's r was calculated to determine if there was concurrent or divergent-related evidence for validity between the mean course ratings of the M-5 Star and the mean

course ratings of each of the school online course instruments (Texas IQ, WebCT, and SREB).

## **Summary**

The pilot study provided tentative reliability and validity evidence for the C-5 Star instrument and the school online course rating instruments. However, the M-5 Star needed to be evaluated for reliability and validity evidence. In the dissertation study, an interrater reliability analysis was conducted for the M-5 Star instrument and for the school instruments. Content validity and divergent validity analyses were also conducted. The upcoming Results chapter will provide the results of these analyses.

#### **CHAPTER IV**

#### RESULTS

This chapter provides the results of the interrater reliability and validity analyses that were conducted for the M-5 Star and for the school instruments. The interrater reliability analysis of the M-5 Star provided interrater reliability support. For the content validity analysis, First Principles of Instruction experts compared the M-5 Star to the First Principles domain. The content validity evidence was substantial. A majority of the M-5 Star items were shown to be valid.

For the concurrent/divergent validity analysis, the school-online course rating instruments provided some divergent validity support for the M-5 Star. In other words, the school instruments were shown to be different, or uncorrelated with the M-5 Star, demonstrating that the M-5 Star measures an alternate construct (detailed instructional strategies) as compared with the very general instructional strategies and surface features that the school rating instruments measure.

The investigation also provided some reliability and concurrent validity support for the school online course instruments (Texas IQ, WebCT, and SREB). These instruments showed moderate to substantial interrater reliability evidence. They were also correlated with one another, manifesting strong concurrent validity, thus confirming that they are measuring similar types of general instructional strategies and surface features such as objectives, syllabi, presentation, practice, structure, interaction, interactivity, and so forth.

#### **Interrater Reliability Analysis**

Intraclass correlation was used to analyze the interrater reliability of the M-5 Star and the school instruments. The ICC for each instrument was calculated comparing the average or consensus score of the baseline rater pair with the average score of the other rater pair. Consensus ratings do not provide a true measure of interrater reliability between individual raters because ratings have been reconciled and averaged. However, the ICC between averaged rater-pair scores does provide some indication of reliability between rater pairs; in other words, this is the reliability that might be expected when rater pairs reconcile scores within one point, then scores are averaged (ConQir, 2010).

### **Interrater Reliability for the M-5 Star**

Research question 1.1: How much interrater reliability evidence does the M-5 Star have based on intraclass correlation?

The model used for the ICC analysis was a two-way random effects model. One pair of raters was randomly selected from class members, so the raters were considered to be random. The courses were also randomly selected. As was mentioned earlier, intraclass correlation (ICC) is interpreted similarly to kappa (Garson, 2010b). By convention, a kappa > .70 is considered acceptable, but this depends highly on the researcher's purpose. For our exploratory purposes K < 0 is poor agreement, K < .20 is slight agreement, K = .20 to .39 is fair agreement, K = .40 to .59 is moderate agreement, K = .60 to .79 is substantial, and K = .80 to 1.0 is outstanding (Garson, 2010b; Landis & Koch, 1977).

M-5 Star instrument reliability for each course separately. Single-measures ICC reliability was calculated over principles, for each course separately. The Psychology course had slight ICC (ICC = .07, p = .28). The Statistics course had fair ICC (ICC = .21, p = .04). The HTML (ICC = .52, p = .001) and English (ICC = .56, p = .001) courses had moderate ICC. The Biology (ICC = .64, p = .001) and Theatre (ICC = .76, p = .001) courses had substantial ICC. (See column 2 of Table 3.) Five of six courses had statistically significant ICC.

Table 3

M-5 Star Instrument Single-Measures ICC Reliability over Principles, for Each Course Separately

			95% Confidence Interval		rith True ie 0
	Intraclass Correlation	Lower Bound	Upper Bound	Value	Sig
Psychology	.07	18	.31	1.16	.283
Statistics	.21	03	.44	1.54	.045
HTML	.52	.31	.68	3.14	.001
English	.56	.36	.71	3.53	.001
Biology	.64	.47	.77	4.62	.001
Theatre	.76	.64	.85	7.49	.001

*Note.* A two-way random effects model was used, where people effects are random and measures effects are random. Degrees of freedom (1, 2) for all items = 62.

M-5 Star instrument reliability for each principle separately. When single-measures ICC reliability was calculated over courses, for each principle separately, there was low to exceptional ICC among raters. (See Table 4.) The integration principle had only slight ICC (ICC = .16, p = .25). The demonstration principle had fair ICC (ICC = .38, p = .001), which is low but significant at the .001 level. The application (ICC = .43, p = .001) and activation (ICC = .51, p = .001) principles had moderate ICC significant at

the .001 level. The task-centered principle had exceptional ICC (ICC = .91, p = .001) significant at the .001 level. (See intraclass correlation coefficients in column 2 of Table 4.) Four of five principles had significant ICC. Task-centered ICC is probably highest because of the focus on the task-centered principle during retraining and reconciling. Integration ICC is probably the lowest because of confusion over the definitions of the "defend new knowledge and skill," and "publicly demonstrate their new knowledge or skill." For example, does publicly mean before classroom peers or out in public, or both? Also, sometimes evaluators assume these activities take place when no mention is made in course materials.

Table 4

M-5 Star Instrument Single-Measures ICC Reliability over Courses, for Each Principle Separately

			95% Confidence Interval		F Test with True Value 0		
	Intraclass Correlation	Lower Bound	Upper Bound	Value	df1	df2	Sig
Integration	.16	31	.58	1.39	17.0	17	.250
Demonstration	.38	.20	.54	2.24	95.0	95	.001
Application	.43	.24	.58	2.50	89.0	89	.001
Activation	.51	.18	.73	3.05	29.0	29	.002
Task-Centered	.91	.81	.95	20.58	29.0	29	.001

*Note.* A two-way random effects model was used, where people effects are random and measures effects are random.

Tentative interrater reliability evidence has been established for five of six courses, and for four of five principles. The M-5 Star also shows moderate overall reliability significant at the .001 level (ICC = .56, p = .001). See M-5 Star item ICC in Appendix D).

# Interrater Reliability for the School Online Course Instruments

Research question 1.2: How much interrater reliability evidence do the school instruments have based on intraclass correlation (ICC)?

Interrater reliability evidence for the Texas IQ instrument. When single-measures ICC reliability was calculated for each course separately for the Texas IQ, the Statistics course had slight ICC (ICC = .15, p = .12). The Biology (ICC = .39, p = .001) and HTML (ICC = .40, p = .001) courses had fair ICC, but significant at the .001 level. Theatre (ICC = .45, p = .001), English (ICC = .50, p = .001), and Psychology (ICC = .54, p = .001) had moderate ICC, significant at the .001 level. (See Table 5.) Five of six courses showed significant ICC. The Texas IQ showed moderate overall reliability (ICC = .43, p = .001), significant at the .001 level. (See Texas IQ item ICC in Appendix D.)

Table 5

Texas IQ Single-Measures ICC Reliability for Each Course Separately

			95% Confidence Interval		t with alue 0
	Intraclass Correlation	Lower Bound	Upper Bound	Value	Sig
Statistics	.15	10	.38	1.36	.117
Biology	.39	.16	.58	2.29	.001
HTML	.40	.17	.59	2.33	.001
Theatre	.45	.23	.62	2.62	.001
English	.50	.29	.67	3.02	.001
Psychology	.54	.34	.69	3.35	.001

*Note.* A two-way random effects model was used, where people effects are random and measures effects are random. Degrees of freedom (1, 2) = 62.

Interrater reliability evidence for the WebCT instrument. When single-measures ICC reliability was calculated for the WebCT for each course separately, the Theatre (ICC = .58, p = .001) course had moderate ICC significant at the .001 level. The Statistics (ICC = .73, p = .001), HTML (ICC = .79, p = .001), and Biology (ICC = .78, p = .001) courses had substantial ICC significant at the .001 level. Psychology (ICC = .80, p = .001) and English (ICC = .81, p = .001) had exceptional ICC, significant at the .001 level. (See the ICC ratings in column 2 of Table 6.) All six courses showed significant ICC. The WebCT also showed substantial overall reliability (ICC = .75, p = .001). (WebCT item ICC is shown in Appendix D.)

Table 6

WebCT Single-Measures ICC Reliability for Each Course Separately

			95% Confidence Interval		st with Tr	ue Valu	ue 0
	Intraclass Correlation	Lower Bound	Upper Bound	Value	df1	df2	Sig
Theatre	.58	.30	.77	3.78	31.0	31	.001
Statistics	.73	.50	.86	6.35	28.0	28	.001
HTML	.79	.61	.89	8.65	30.0	30	.001
Biology	.78	.59	.89	8.24	29.0	29	.001
Psychology	.80	.62	.90	9.21	27.0	27	.001
English	.81	.63	.90	9.38	29.0	29	.001

Interrater reliability evidence for the SREB instrument. When single-measures ICC reliability was calculated for each course separately, Statistics (ICC = .34, p = .08) and English (ICC = .38, p = .06) had fair ICC, significant at the .10 level. Theatre (ICC = .42, p = .04), and HTML (ICC = .42, p = .05) had moderate ICC significant at the .05 level. Psychology (ICC = .64, p = .002) had substantial ICC

significant at the .01 level. Biology (ICC = .70, p = .001), had substantial ICC, significant at the .001 level. (See column 2 of Table 7.) All six courses had significant ICC at the .10 level. The SREB also showed moderate overall reliability (ICC = .53, p = .001) significant at the .001 level. (ICC for SREB items is found in Appendix D.)

Table 7

SREB Single-Measures ICC Reliability for Each Course Separately

			95% Confidence Interval		F Test with True Value 0				
	Intraclass Correlation	Lower Bound	Upper Bound	Value	df1	df2	Sig		
Statistics	.34	15	.70	2.04	16.0	16	.082		
English	.38	12	.73	2.25	15.0	15	.064		
Theatre	.42	06	.74	2.43	16.0	16	.042		
HTML	.42	08	.75	2.43	15.0	15	.048		
Psychology	.64	.24	.85	4.54	16.0	16	.002		
Biology	.70	.31	.89	5.71	14.0	14	.001		

Table 8

Comparison of Instrument ICCs Across Courses

		M-5	Star	Texa	as IQ	Wel	b CT	SF	REB
Course	Quality	ICC	Sig	 ICC	Sig	 ICC	Sig	 ICC	Sig
Theatre	High	.76	.001	.45	.001	.58	.001	.42	.042
HTML	High	.52	.001	.40	.001	.79	.001	.42	.048
Statistics	Medium	.21	.045	.15	.117	.73	.001	.34	.082
English	Medium	.56	.001	.50	.001	.81	.001	.38	.064
Psychology	Medium	.07	.283	.54	.001	.80	.001	.64	.002
Biology	Low	.64	.001	.39	.001	.78	.001	.70	.001
Overall ICC		.56	.001	.43	.001	.75	.001	.53	.001

# Comparison of ICC Across Courses and Instruments

As shown in Table 8, there is interrater consistency for each individual instrument across a majority of the courses. However, there isn't much consistency of reliability for each course across all of the instruments. Each instrument is unique enough that it has its own level of interrater reliability. Nor does course quality (high, medium, and low, as shown in column 2) seem to affect ICC consistency. Only the instruments themselves affect the ICC consistency.

## Comparison of School Instrument ICC Between Pilot and Dissertation

In the pilot study the school instruments received moderate to substantial interrater reliability evidence with two individual raters. In the dissertation study the school instruments received tentative confirmation with two rater-pair consensus scores. This confirmation is questionable due to consensus ratings used in the dissertation study being biased toward consistency, thus inflating interrater reliability. (See Table 9.)

Table 9

Comparison of School Instrument ICC Between Pilot and Dissertation Studies

	Texas IQ	WebCT	SREB
Pilot Study	ICC = .39	ICC = .60	ICC = .40
	p = .001	p = .001	p = .001
Dissertation	ICC = .43	ICC = .75	ICC = .53
	p = .001	p = .001	p = .001

#### **Content Validity Analysis**

Research question 2.1: To what extent does the M-5 Star exhibit content validity as compared with the First Principles of Instruction domain outline?

First Principles of Instruction experts compared the M-5 Star content validity rating form (Appendix H) with a First Principles of Instruction domain outline (Appendix E). Of 64 potential raters, 22 (34%) completed the content validity analysis. Raters gave a rating of 0 (not included), 1 (partially included), or 2 (fully included) to indicate to what extent each item on the M-5 Star was included in the First Principles domain outline.

Aiken's CVI and subsequent z score were calculated for each of the M-5 Star items. Of the 63 items, 56 (89%) were significant at the alpha = .10 level. (See Table 10.) The average CVI for all items was very high (Ave. CVI = .84, Ave. z score = 3.7, p = .001).

Only five items (45, 50, 37, 29, and 60) failed to achieve significant content validity at the .10 level. (See shaded items in Table 10.) The five items with low CVI ratings are shown with matching First Principles of Instruction domain items in Table 27, in Appendix J. An explanation is also provided regarding why each of the items may have received low CVI ratings.

### **Concurrent and Divergent Validity Analysis**

Research question 2.2: To what extent does the M-5 Star have evidence of concurrent or divergent validity as shown by correlations of M-5 Star overall mean scores of courses with school instrument overall mean scores of courses?

Table 10

M-5 Star Instrument Item CVI, z Score and p-Value

M-5 Star Item	CVI	z score	p value
28	1.00	5.61	0.001
62	1.00	5.61	0.001
11	0.97	5.31	0.001
23	0.97	5.31	0.001
36	0.97	5.31	0.001
46	0.97	5.31	0.001
55	0.97	5.31	0.001
61	0.97	5.31	0.001
63	0.97	5.31	0.001
4	0.95	5.01	0.001
6	0.95	5.01	0.001
7	0.95	5.01	0.001
8	0.95	5.01	0.001
13	0.95	5.01	0.001
15	0.95	5.01	0.001
16	0.95	5.01	0.001
17	0.95	5.01	0.001
18	0.95	5.01	0.001
21	0.95	5.01	0.001
24	0.95	5.01	0.001
27	0.95	5.01	0.001
49	0.95	5.01	0.001
52	0.95	5.01	0.001
1	0.92	4.71	0.001
10	0.92	4.71	0.001
12	0.92	4.71	0.001
20	0.92	4.71	0.001
22	0.92	4.71	0.001
47	0.92	4.71	0.001
3	0.89	4.40	0.001
9	0.89	4.40	0.001
25	0.87	4.10	0.001
33	0.87	4.10	0.001
51	0.87	4.10	0.001
2	0.84	3.80	0.001
14	0.84	3.80	0.001
19	0.84	3.80	0.001
56	0.84	3.80	0.001
34	0.82	3.50	0.002

(continued)

M-5 Star Item	CVI	z score	<i>p</i> value
44	0.82	3.50	0.001
38	0.79	3.19	0.001
48	0.79	3.19	0.001
26	0.79	3.06	0.001
5	0.76	2.89	0.002
31	0.76	2.89	0.002
43	0.76	2.89	0.002
54	0.76	2.89	0.002
32	0.71	2.29	0.011
57	0.71	2.29	0.011
30	0.69	2.10	0.018
42	0.33	-2.04	0.020
40	0.68	1.99	0.024
58	0.68	1.99	0.024
59	0.68	1.99	0.024
35	0.66	1.68	0.097
53	0.66	1.68	0.097
45	0.60	1.08	0.140
50	0.60	1.08	0.140
37	0.60	1.08	0.142
29	0.55	0.47	0.319
60	0.55	0.47	0.319
Mean	0.84	3.74	

Convergent/divergent validity-related evidence was sought between the M-5 Star and the school online course instruments (Texas IQ, WebCT, and SREB). To determine whether a parametric or non-parametric measure should be used to analyze convergent/divergent validity evidence, a Shapiro-Wilk's W test of normality was calculated across all four instruments. As shown in Table 11, each instrument contained a nonsignificant W value statistic, indicating that normality assumptions were met, and justifying the use of Pearson correlations.

Table 11
Shapiro-Wilk Test of Normality

		Shapiro-Wilk				
	Statistic	df	Sig.			
Mean M-5 Star	.93	6	.623			
Mean Texas IQ	.87	6	.224			
Mean WebCT	.86	6	.179			
Mean SREB	.92	6	.535			

Scatterplots between the M-5 Star and the school instruments. Before looking at correlation coefficients, scatterplots can be used to graphically show the relationship between the M-5 Star and each of the three school instruments. (See Figures 3-5.) Note that the scatterplot points are fairly linear, each with an outlier that is the primary reason for divergent correlation. The Theatre class is the consistent outlier. It is the highest point vertically in each scatterplot. For example, in Figure 3, the Theatre class M-5 Star mean score is high (mean score = 3.47, z = 1.59) while the Texas IQ mean score is lower (mean score = 3.2, z = -.45). Because of the high score on the M-5 Star and medium low score on the Texas IQ, the theatre course scatterplot point is farthest from the line of best fit. The scatterplot comparisons between the M-5 Star and WebCT and SREB also show a high score on the M-5 Star compared to medium low scores on the Texas IQ and SREB, so again the Theatre class is an outlier.

The Theatre class probably scored higher on the M-5 Star because it was the only course that incorporated a progression of real-world tasks (analyzing theatrical events). These "real-world tasks" raised the scores significantly on the M-5 Star without doing so on the more generic school instruments.

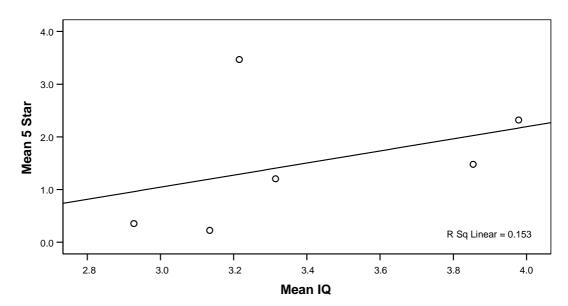


Figure 3. Scatterplot of the association between the M-5 Star mean course scores and the Texas IQ mean course scores. Note that the points are fairly linear, with a Theatre class outlier. The actual correlation is low and provides some divergent validity evidence for the M-5 Star (r = .39, p = .44)  $(r^2 = .15)$ .

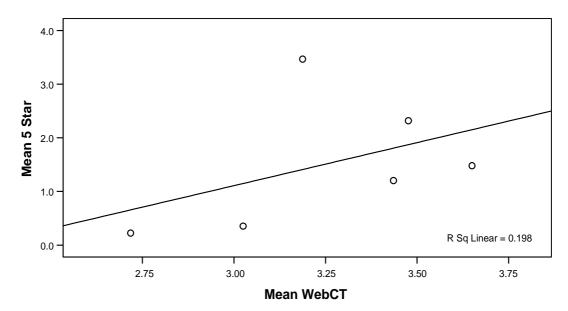


Figure 4. Scatterplot of the association between the M-5 Star mean course scores and the WebCT mean course scores. Note that the points are fairly linear, with a Theatre class outlier. The actual correlation is low (r = .44, p = .38)  $(r^2 = .20)$ , and provides some divergent validity evidence for the M-5 Star.

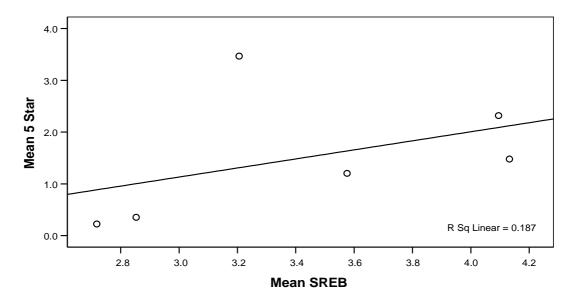


Figure 5. Scatterplot of the association between the M-5 Star mean course scores and the SREB mean course scores. Note that the points are fairly linear, with a Theatre class outlier. The actual correlation is low (r = .43, p = .39)  $(r^2 = .19)$ , and provides some divergent validity evidence for the M-5 Star.

Because of the Theatre class outliers, the correlations between the M-5 Star and the school instruments are low and provide some divergent validity evidence for the M-5 Star. If more classes had matched the M-5 Star criteria, similarly to the Theatre class, it is likely that they would have also scored high on the M-5 Star while getting medium scores on the generic school instruments. The scatterplots would have had more outliers and would have looked more scattered, the correlations would have been lower and the divergent validity evidence stronger.

Pearson correlation coefficients were calculated between each of the course rating instruments. Two-tailed tests were used for all of the correlations in view of the uncertainty whether the school instruments would have a high or low correlation with the M-5 Star or with each other. In the pilot study, the Texas IQ provided concurrent validity

evidence for the C-5 Star, while the other two instruments (WebCT and SREB) provided divergent validity evidence for the C-5 Star. Whereas Merrill made substantial changes to his M-5 Star instrument, it would be difficult to predict how the school instruments would correlate with his new instrument.

As correlation coefficients were calculated for each of the instruments, an acceptable correlation coefficient level was needed. For psychological research, a correlation coefficient of .50 or above is considered large, a coefficient of .30, is, moderate, and a coefficient of .10 is small. In many areas of psychology it is rare to find correlations that are greater than .40 (Aron, Aron, & Coups, 2009; Cohen, 1988). For this study, the correlation cutoff score was r = .70, because a larger correlation was desired to show strong concurrent validity. Correlations of r > .70 (p < .10) ( $r^2 > .49$ ) were accepted for concurrent validity evidence. Correlations of r < .50 (p > .10) ( $r^2 < .25$ ) were accepted for divergent validity evidence. Correlations between .50 and .70 were not accepted for concurrent or divergent evidence.

**Pearson correlations with standardized scores.** When there is a significant difference in the variance of instrument scores, Pearson correlations need to be calculated using standardized scores. The variance of the M-5 Star mean rating is much greater (1.52) than the variance of the Texas IQ (.18), WebCT (.12) and SREB (.37) mean ratings. (See Table 12.)

Table 12

Mean Ratings and Variance of the Various Instruments and Courses

Course	Mean M-5 Star	Mean Texas IQ	Mean WebCT	Mean SREB	Student Course Rating	Course Rating Average	Course Rating Variance
Theatre	3.47	3.21	3.19	3.21	5.60	3.74	1.10
HTML	2.32	3.98	3.48	4.09	5.70	3.91	1.49
Statistics	1.48	3.85	3.65	4.13	4.90	3.60	1.63
English	1.20	3.31	3.43	3.58	5.60	3.42	2.43
Psychology	0.35	2.93	3.02	2.85	4.80	2.79	2.52
Biology	0.22	3.13	2.72	2.72	3.50	2.46	1.67
Instrument Average	1.51	3.40	3.25	3.43	5.02	3.32	1.55
Instrument Variance	1.52	0.18	0.12	0.37	0.70		

*Note.* Instrument mean ratings for each course are based on a scale of 1-5, with some variation on Merrill's instrument, a scale of 1-5 for the Texas IQ, WebCT and SREB instruments, and a scale of 1-6 for the Student Course Ratings,.

Levene's test of equality of error variance reveals that the variances between instrument mean scores are significantly different (F = 132.71, df1 = 3, df2 = 4662, p = .001), so Pearson correlations were calculated with standardized (z) scores. The correlation between the mean M-5 Star rating and the mean Texas IQ, WebCT, and SREB ratings were all low and not significant at the .10 alpha level, as shown in Table 13. Therefore, the school instruments provide divergent validity evidence for the M-5 Star. The correlations between the school instruments (Texas IQ, WebCT, and SREB) were all high and significant at the .10 level, providing concurrent validity evidence for each other. (See Table 13.)

Table 13

Pearson Correlations for the Four Course Rating Instruments Using Standardized Scores

		Mean M-5 Star	Mean Texas IQ	Mean WebCT
Mean Texas IQ	Pearson Correlation	.39		
	Sig. (2-tailed)	.44		
	Sum of Squares and Cross-products	1.95		
	Covariance	.39		
	N	6		
Mean WebCT	Pearson Correlation	.44	.78	
	Sig. (2-tailed)	.38	.07	
	Sum of Squares and Cross-products	2.23	3.89	
	Covariance	.44	.78	
	N	6	6	
Mean SREB	Pearson Correlation	.43	.94(**)	.94(**)
	Sig. (2-tailed)	.392	.005	.004
	Sum of Squares and Cross-products	2.161	4.694	4.726
	Covariance	.432	.939	.945
	N	6	6	6

<sup>\*\*</sup> Correlation is significant at the .01 level (2-tailed).

When correlations between the M-5 Star instrument and the school instruments are compared with correlations between the C-5 Star instrument and the school instruments, it is apparent that the Texas IQ and SREB correlated very differently with the two 5 Star instruments. While all school instruments have low nonsignificant correlations with the M-5 Star, the Texas IQ had a high, significant correlation, the WebCT had a low correlation, and the SREB has a very low correlation with the C-5 Star. (See Table 14.)

Table 14

Comparison of Correlation Coefficients Between School Instruments and Cropper's (C-5 Star) and Merrill's (M-5 Star) Instruments

	Texas IQ	WebCT	SREB
Cropper C-5 Star Instrument	r = .90, p = .002	r = .45, p = .26	<i>r</i> = .04, <i>p</i> = .93
Merrill M-5 Star Instrument	r = .39, p = .44	r = .44, p = .38	r = .43, p = .39

*Note.* Significance is reported at the two-tailed level.

#### **Summary**

The data analysis provided limited interrater reliability evidence for the M-5 Star, Texas IQ, WebCT and SREB instruments. The reliability evidence was limited because rater pair scores were averaged, biasing toward consistency and inflating interrater reliability. Content validity evidence was substantial, with 56 of 63 items (89%) showing significant content validity. The school instruments provided divergent validity evidence for the M-5 Star, while showing divergent validity evidence for each other. The Discussion chapter will provide interpretation of these results, and will discuss their significance. It will also acknowledge study limitations, and discuss possibilities for further research.

#### **CHAPTER V**

#### **DISCUSSION**

In the pilot study, some limited reliability and concurrent and divergent validity evidence were provided for the C-5 Star. The Texas IQ instrument provided concurrent validity evidence for the C-5 Star, while the WebCT and SREB provided divergent validity evidence. However, the pilot study was preliminary because of its lack of range in quality of courses.

Although the dissertation research was subject to a number of limitations, it provided some tentative interrater reliability and content and divergent validity support for the M-5 Star while using a wide range in quality of courses. The school instruments provided only divergent validity evidence for the M-5 Star as compared to the concurrent and divergent validity evidence provided for the C-5 Star. Why this is the case is addressed in the discussion of the findings in this chapter. An ancillary analysis revealed that student course ratings may be more closely related to the M-5 Star and the WebCT than to the other school instruments. Recommendations for further research are also given.

# Limitations of the Study

Although the comparison school instruments received some concurrent and divergent validity support in the pilot study, the fact that they have not been previously validated makes any reliability and validity evidence they provide for the M-5 Star

tenuous, at best. The limited number of rater-pair ratings, the few courses evaluated, as well as potential rater bias, threaten any reliability and validity findings.

- 1. It was difficult to find instruments to compare with the M-5 Star for the divergent validity analysis. It is ideal for a concurrent or divergent validity analysis to have comparison measures that have pre-established reliability and validity evidence. Of the expert rating forms with reliability and validity data, which were found during the literature review, only one was designed for interactive online instruction. It was unacceptable as a comparison instrument because it lacked detailed instructional strategy criteria. The comparison instruments that were selected for the study had no reliability or validity data other than that provided by the pilot study. Although the pilot study provided some reliability and validity evidence for the C-5 Star and for the comparison instruments, it would have been preferable to use comparison instruments that had previously received reliability and validity evidence.
- 2. A limited number of courses were evaluated for the study. Ten courses were evaluated, only six which had an adequate amount of instructional material to be evaluated effectively. Therefore, only the six courses with sufficient instructional materials were used in this current analysis. The selected courses included one low-rated course, three medium-rated courses, and two high-rated courses based on student ratings. Therefore there was some bias toward medium and high-rated courses. However, the courses did represent a full range of course quality.

- 3. Online-course raters were part of an Evaluating Online Courses graduate class, and, for that reason, made up a convenience sample of expert raters, as opposed to a stratified or random sample. The raters may have been biased because Cropper, Bentley, and Gardner provided extensive instruction on online-course rating as well as specific instruction on the instruments that each pair of raters used. Another problem was that there were a limited number of raters. Only four pair of raters (one pair per instrument) plus two baseline raters (Cropper and Gardner for the school instruments, Merrill and Cropper for the M-5 Star) provided ratings using each of the instruments. The rater sample became even smaller because rater-pair scores were averaged and ultimately there was only one score per rater pair or an n of two rater-pair scores per instrument.
- 4. The dissertation study used some fixed baseline raters and some randomly selected raters. The courses were also randomly selected. When random raters are used and the courses that are evaluated are random, the ICC design is called two-way random. This design led to a less-effective ICC because it incorporated the variance of the random raters as well as the variance of the randomly selected courses. ICC is especially effective for modeling rater effects when a fixed group of raters are used because the analysis can account for variance attributable to raters. When raters are fixed and courses are randomly selected, the design is two-way mixed. A mixed-effects design includes "at least one fixed-effect way and at least one random-effect way" (Fidler & Thompson, 2001). While fixed-rater effects provide more effective ICC, the results are not generalizable beyond

the raters used in the study to the population of possible raters (Garson, 2010b; Shrout & Fleiss, 1979; Weir, 2005). However, if the desired outcome is to achieve rating accuracy through a trained group of fixed raters, it is not necessary to use randomly selected raters to make the ICC results generalizable to the population of potential raters, which would be limited to First Principles experts anyway.

- 5. The M-5 Star was difficult to use for raters other than Merrill himself. When Cropper and Merrill reconciled their scores as baseline raters, Cropper had to lower his ratings to match Merrill, particularly on the problem-centered factor. The other pair of raters also gave significantly higher ratings than Merrill. Cropper retrained them, having them reconcile their ratings for the first two courses within 20% of Merrill's ratings for each M-5 Star item. Use of the reconciled ratings for the two courses provided a bias toward consistency. Of course a bias already exists because rater pairs reconciled their own ratings within one point, and then rater-pair scores were averaged. Reconciliation biases towards rating consistency, inflating ICC, and limiting interpretation of interrater reliability.
- 6. In addition to the potential bias of course raters, all potential and actual content validity raters had been taught or trained by Merrill or his students, so all have a likely bias. USU students who had participated in M-5 Star ratings may have potentially been the most biased of content validity raters because they had first hand experience evaluating courses using the M-5 Star. One of them was

Gardner, the teaching assistant for the Evaluating Online Instruction course, who may have been the most biased. However, the content validity raters were not hesitant to rate M-5 Star items as not matching or partially matching the First Principles domain outline. In some cases they seemed more biased against the instrument than for it.

#### **Discussion of the Findings**

#### **Interrater Reliability Analysis**

In the pilot study the individual raters' ICCs for the C-5 Star decreased somewhat as expert raters were added, with Merrill added last (Phase I: ICC = .59, p = .001, Phase II: ICC = .34, p = .001, Phase III: ICC = .24, p = .001). This occurred because the addition of raters increased variation in the ratings and therefore lowered the interrater reliability in the pilot study. In the dissertation study, with increased training and the use of rater-pair consensus scores, the interrater reliability was high again (ICC = .56, p = .001). Of course the comparison between the pilot ICCs and dissertation ICC is tenuous because of the bias toward consistency caused by rater-pair consensus scores.

In the pilot study the school instruments received moderate to substantial interrater reliability evidence with two individual raters. In the dissertation study the school instruments received tentative confirmation with two rater-pair consensus scores. Again, this confirmation is questionable due to consensus ratings inflating interrater reliability.

The dissertation study showed high ICCs for the M-5 Star over five of the six courses, over four of the five principles, as well as for the overall M-5 Star instrument. The ICC was low for 21 (33%) of 63 individual items. (See Appendix D.) Although eliminating low-rated items is a possibility, doing so would threaten content validity. Each of the items represents an important aspect of Merrill's First Principles of Instruction. Modifying items with low ICC is also an option. However, providing additional training on the low- or negatively-rated items would most probably lead to a greater improvement in interrater reliability without sacrificing content validity.

#### **Content Validity Analysis**

Of 63 M-5 Star individual items, 56 (89%), received high scores on Aiken's CVI that were significant at the .10 level. The instrument in toto also had high content validity. The average CVI for all items high (Ave. CVI = .84, Ave.  $z\ score = 3.70$ , p = .0001).

For the seven items receiving low CVI ratings, probable reasons for low ratings included items being in a different order in the M-5 Star instrument and domain outline, items being worded differently or items being consolidated in the instrument or the domain outline. Table 28 in Appendix J lists each of the low-rated items with matching First Principles domain items. For each item, an explanation is provided regarding why expert raters may have given the item a low rating. For items that experts gave low CVI ratings, it would be important to revise the instrument items to match the domain items in wording and intent and put items in the instrument in the same order as the domain outline.

# **Concurrent and Divergent Validity Analysis**

The correlations of the M-5 Star with the school instruments were all low, providing some divergent validity support for his instrument. The Texas IQ instrument was correlated at r = .39, p = .44,  $(r^2 = .15)$ , the WebCT correlated at r = .44, p = .38 ( $r^2 = .20$ ), and the SREB correlated at r = .43, p = .39 ( $r^2 = .19$ ). (See Table 14, p. 60.) It would be expected that because the C-5 Star was also based on First Principles of Instruction, that similarly to the M-Star it would have low correlations with the school instruments. However, the correlations between the C-5 Star and the school instruments varied from very high to very low. The Texas IQ had high correlation with the C-5 Star, and provided substantial concurrent validity evidence (r = .90, p = 002) ( $r^2 = .81$ ). The WebCT had a low correlation, and therefore provided some divergent validity evidence (r = .45, p = .26) ( $r^2 = .20$ ), and the SREB had a very low correlation with the C-5 Star, and provided substantial divergent validity evidence (r = .04, p = .93) ( $r^2 = .001$ ). (See Table 14, p. 60.)

Why would two instruments based on Merrill's First Principles correlate so differently to the school instruments? One reason that the M-5 Star may correlate divergently with the school instruments is the M-5 Star includes very detailed instructional strategy criteria, as opposed to the school instruments' more generic instructional strategy and surface-feature items. In the M-5 Star, Merrill added detailed instructional strategy questions for each of his five principles, but did not include principle summary questions that had been in the C-5 Star. As a result, only seven

WebCT items, four Texas IQ items, and no SREB items match the M-5 Star items. (See Table 15.)

Table 15

Number of School Instrument Items That Match Cropper's C-5 Star and Merrill M-5 Star Instrument Items

	Texas IQ	WebCT	SREB
	(63 items)	(32 items)	(17 items)
Cropper C-5 Star	34	24	12
Instrument (26 items)			
Merrill M-5 Star	4	7	0
Instrument (63 items)			

*Note*. Number of items is based on a count from the comparison of instruments in Appendix A. In some cases multiple items from a school instrument matches an item on one of the 5 Star instruments.

Because the C-5 Star was also based on Merrill's First Principles, it would seem logical that it would have similarly low correlations with the school instruments. However, the C-5 Star was shorter (26 items) than the M-5 Star (63 items), with very general questions summarizing Merrill's principles and sub-principles, compared to very specific questions in the M-5 Star. Because of its more general nature, the C-5 Star matched 34 of 63 Texas IQ instrument items, 24 of 32 WebCT items, and 12 of 17 SREB items. (See Table 15.)

Because so many of the school instrument items are similar to the C-5 Star items, it might be expected that all three school instruments would provide concurrent validity evidence for the C-5 Star. However, as was explained above, only the Texas IQ exhibited concurrent validity evidence for the C-5 Star. The WebCT manifested weak

divergent validity evidence, and the SREB revealed substantial divergent validity support for the C-5 Star.

Differences in validity evidence for school instruments. The school instruments provided convergent validity evidence for each other in this dissertation study (Texas IQ - WebCT r = .78, p = .07 ( $r^2 = .60$ ), Texas IQ - SREB r = .94, p = .003 ( $r^2 = .88$ ), WebCT - SREB r = .94, p = .004) ( $r^2 = .89$ ). However, in the pilot study the Texas IQ (r = .64, p = .04) ( $r^2 = .41$ ), and SREB (r = .69, p = .03) ( $r^2 = .48$ ) instruments provided convergent validity evidence for the WebCT, but the Texas IQ and SREB provided divergent validity for each other (r = .334, p = .21) ( $r^2 = .11$ ). (See Table 16.)

Table 16

Number of School Instrument Items Matching Other School Instruments

	Texas IQ (63 items)	WebCT (32 items)	SREB (17 items)
Texas IQ		32	16
WebCT	63		17
SREB	44	23	

*Note.* The numbers in the column under each instrument indicate how many of that instrument's items are represented in the instruments listed in the rows. For example, in the Texas IQ column, all 63 Texas IQ items are represented in the WebCT instrument, while 44 Texas IQ items are represented in the SREB. Note that one item in one instrument may be matched by many items in another instrument.

A possible explanation for the low Texas IQ – SREB correlation is the difference in total number of items (63 Texas IQ items compared to 17 SREB items), or perhaps that only 44/63 Texas IQ items are matched by SREB items, although 16 SREB items have matches among the 63 Texas IQ instrument items. (See Table 16.)

Possibility of finding measures for convergent validity evidence. While one instrument (Texas IQ) provided concurrent validity evidence for the C-5 Star, no expert rating instruments provided concurrent validity support for the M-5 Star. Only divergent validity evidence was provided for the M-5 Star. It would be preferable to find some measures with concurrent validity evidence for the M-5 Star as well. It is unlikely that concurrent validity evidence for the M-5 Star will be provided using expert rating instruments; currently there are no expert rating instruments that match the M-5 Star's level of instructional strategy detail.

It might be possible to provide concurrent validity evidence by comparing the M-5 Star with measures other than expert rating forms. As was mentioned earlier, an instrument or test can be compared with various types of measures to establish concurrent validity. For example, the M-5 Star scores could be compared with student course ratings, student performance on tests or tasks, teacher and student ratings regarding student learning and performance, etc.

Ancillary concurrent validity analysis. It is appropriate at this point to include an ancillary analysis that not part of the dissertation objectives because it compared the M-5 Star with a measure other than an expert rating form. It is a correlation of student course ratings with the M-5 Star and the school instruments. This analysis is included because of the accessibility of data, the interesting nature of the analysis, and the opportunity to preview future research possibilities.

A correlation was done comparing mean global student ratings with the mean ratings of the four course rating instruments (M-5 Star, Texas IQ, WebCT, and SREB).

The mean global rating was an average of the overall course quality and overall instructor quality ratings. The test was two-tailed because it was unknown which instruments would correlate with student ratings. As discussed earlier, Levene's test of equality of error variance revealed that the variance between instrument mean scores are significantly different (F = 132.71, df1 = 3, df2 = 4662, p = .001), so Pearson correlations were calculated with standardized scores. The Pearson correlation using standardized scores between the student ratings and the M-5 Star mean ratings was substantial (r = .72, p = .106) ( $r^2 = .52$ ) but not significant at the .10 level. The correlation between the student ratings and the WebCT rating (r = .71, p = .11) ( $r^2 = .50$ ) was also substantial, but not significant at the .10 level. (See Table 17.)

Table 17
Standardized Pearson Correlations for Student Ratings Compared with Instruments

		Mean M-5 Star	Mean Texas IQ	Mean WebCT	Mean SREB
Student Course Eval	Pearson Correlation	.72	.38	.71	.59
	Sig. (2-tailed)	.11	.46	.11	.22
	Sum of Squares and Cross-products	3.60	1.90	3.54	2.93
	Covariance	.72	.38	.71	.59
	N	6	6	6	6

The Pearson correlations of student ratings with the other school online course instrument ratings were lower, with higher p-values. The next highest correlation was between student ratings and the SREB rating (r = .59, p = .22;  $r^2 = .34$ ). The lowest

correlation was between the student ratings and the Texas IQ rating (r = .38, p = .46;  $r^2 = .14$ ). (See Table 17.)

It is interesting that student ratings correlate most closely with the M-5 Star and WebCT, which diverge from each other (r = .44, p = .38). Apparently both instruments, although divergent, measure factors that are important to students. Further studies would need to be done to confirm and expand the results. In addition, it might be possible to find other measures of learning and performance that would provide better concurrent validity evidence for the M-5 Star.

## **Suggestions for Further Research**

#### **Interrater Reliability Testing**

The C-5 Star demonstrated limited evidence of interrater reliability in the pilot study and the M-5 Star demonstrated tentative evidence of overall interrater reliability in this dissertation study. However, many items on the M-5 Star had negative- or low-interrater reliability. (See Appendix D.) It would be valuable to see if interrater reliability for those items could be improved across courses and principles, and for the overall instrument. ICC may be improved through item revision and more extensive training including additional rating practice and reconciliation of practice scores.

For one nursing school study, researchers trained, and continued retraining raters until raters achieved > .80 interrater reliability for their coding of observed behavior of patients on videotapes. If interrater reliability ever dipped below .80, raters were retrained (Haidet, Tate, Divergilio-Thomas, Kolanowski, & Happ, 2009). While > .80 is

acceptable interrater reliability for observational coding from videos, interrater reliability > .70 might be a more realistic goal for the M-5 Star. Comparing courses to the M-5 Star items is difficult enough that it will be difficult for raters to achieve more than .70 ICC using single-measure ICCs. Extensive training could be done on the M-5 Star, and then raters could be given a sample course to rate. If ICC < .70, raters could be retrained on items with lower interrater reliability, after which raters could rate another sample course and ICC would be recalculated. This process could be repeated until raters had achieved > .70 interrater reliability.

## **Replication Study**

A replication study, which corrects for the limitations of the dissertation study, could be conducted using the M-5 Star and the school instruments. The replication study would be conducted with a different and larger sample of online courses with a range of quality. Analyses could be done to determine whether the interrater reliability and divergent and convergent validity ratings were similar to those achieved in this dissertation study.

The study would need to correct the flaws discussed in the limitations of the study section by making the following changes:

- Look for alternative comparison measures, including validated measures, if available.
- 2. Increase the number of courses evaluated, still including a wide range of quality.

- 3. Use individual ratings for ICC, which will increase the n of ratings. Also add additional expert raters. While bias is inevitable when using Merrill experts, encourage objectivity through training.
- 4. Use fixed effects for raters.
- 5. Provide additional training, practice, and reconciliation of scores on practice courses for Merrill's instrument.
- 6. While it will be necessary to use experts on Merrill's First Principles, eliminate as much bias as possible by encouraging raters to accurately rate partial and non-matching items.

## **Content Validity Analysis**

Merrill has developed a condensed version of his 5 Star instrument, called the e<sup>3</sup> (effective, efficient, and engaging instruction) instrument. While it is also based on First Principles of Instruction, it is a short form that allows for only binary responses which reflect whether First Principles features are present or absent in a course. One of the content validity raters, who had also done ratings using the M-5 Star, has developed e<sup>3</sup> rating instruction and has done ratings using the e<sup>3</sup> instrument. She commented that the e<sup>3</sup> instrument was easier to use, but lacked examples of how to do the ratings until Merrill came out with a newer paper that had better explanations. It is possible the e<sup>3</sup> instrument will be simple enough to use because it is part of a paper that has detailed instruction about First Principles of Instruction (Merrill, 2009). Training could be done using the e<sup>3</sup> instrument and its corresponding paper to see how much additional training and practice would be required to achieve .70 ICC. (See e<sup>3</sup> instrument in Appendix O.)

In the quest for a simpler 5 Star or First Principles instrument, it would be useful to compare the C-5 Star and the M-5 Star with Merrill's e<sup>3</sup> instrument and the instrument developed by Frick and his colleagues (Frick, et al., 2007, 2009; Frick, et al., 2010). Frick's instrument is based on First Principles of Instruction as well as some other principles of course effectiveness. It is intended for student ratings of classroom instruction. Frick and his colleagues simplified the questions somewhat so students could use the instrument more readily (Frick, et al., 2009).

A face validity comparison could be done by comparing the e<sup>3</sup> rating form, the form developed by Frick et al, the C-5 Star, the M-5 Star and the First Principles of Instruction domain outline. This would make it possible to determine which items from the 5 Star instruments and the First Principles of Instruction domain outline are included and which are excluded on the e<sup>3</sup> and Frick forms. A content validity analysis of the e<sup>3</sup> and Frick instruments could also be conducted, comparing these instruments with the First Principles of Instruction domain outline, probably using the same experts that completed the M-5 Star instrument content validity analysis.

#### **Concurrent Validity Study**

A concurrent validity study could be done comparing expert ratings using the M-5 Star with (1) expert ratings using Merrill's e<sup>3</sup> instrument, and (2) expert and student ratings using Frick's instrument. In conjunction with this study, interrater reliability testing could be done on the e<sup>3</sup> instrument and Frick's instrument to see if their interrater reliability compares with the C-5 Star and the M-5 Star. The concurrent validity study could be done with randomly selected online courses, or in conjunction with an

experimental or quasi-experimental study using courses developed to have prescribed levels of course quality.

## **Experimental or Quasi-Experimental Study**

A Thomson NETg job impact study compared various types of instruction and various levels of instructional quality, including traditional behavioral computer-based instruction and scenario-based instruction based on First Principles of Instruction. The scenario-based instruction showed a significant improvement in learning (p = .001) (Thomson NETg, 2003).

It would be worthwhile to do a study similar to the Thomson study, comparing courses developed to each of Merrill's levels 0, 1, 2 and 3 of instructional strategy (Merrill, 2006). The study would provide support for whether varying levels of inclusion of Merrill's principles bring corresponding levels of performance on real-world tasks.

Group and individual student performance could be evaluated using

- instructor and student ratings of student learning and performance;
- actual student performance on significant real-world task(s);
- instructor and student rating of student engagement in a 5 Star approach to learning; and
- student grades.

As was mentioned above, this (quasi) experimental study could be done in conjunction with the concurrent validity study comparing expert ratings using Merrill's e<sup>3</sup> instrument, and expert and student ratings using Frick's instrument. This combination of

studies could provide evidence that ratings and student performance match the level of quality of instruction.

#### Conclusion

This study contributes to the literature in a meaningful way as it is the first study conducted using the M-5 Star to evaluate course quality as defined by instructional design strategies. It provides only limited reliability evidence for the M-5 Star and school instruments because reconciliation of rater-pair scores biased consistency of ratings and therefore inflated interrater reliability. It provides content validity evidence for a majority of the M-5 Star items. It provides some divergent validity support between the M-5 Star and the school instruments, meaning that the M-5 Star and the school instruments are measuring different things. It also shows that student course ratings are more closely correlated to the M-5 Star and WebCT than the SREB and Texas IQ, although the correlation is not significant at the .10 level. This is interesting because the M-5 Star and WebCT are related divergently. For some reason, student ratings of course and instructor quality encompass characteristics of the M-5 Star detailed instructional strategies and the WebCT more generic criteria.

Further research needs to be done to confirm and expand these findings to determine to what extent the M-5 Star criteria, and other criteria, are the best measures of online course effectiveness. As an online course quality standard is established, it will make it possible to refine the process of evaluating courses, which, in turn, will provide a standard for improving the instructional design process and the courses themselves.

#### REFERENCES

- Aiken, L. R. (1980). Content validity and reliability of single items or questionnaires. *Educational and Psychological Measurement, 40*(December 1980), 955-959.
- Aron, A., Aron, E. N., & Coups, E. J. (2009). *Statistics for psychologists* (5th ed.). Upper Saddle River, NJ: Prentice Hall Higher Education.
- Barrett, A. J. (1990). Objectively determining the educational potential of computer and video-based courseware; or, producing reliable evaluations despite the dog and pony show. Paper presented at the International Conference on Technology and Education, Brussels, Belgium.
- Beckman, T. J., Cook, D. A., & Mandrekar, J. N. (2005). What is the validity evidence for assessments of clinical teaching? *Journal of General Internal Medicine*, 20(12), 1159-1164. doi: 10.1111/j.1525-1497.2005.0258.x
- Beckman, T. J., Ghosh, A. K., Cook, D. A., Erwin, P. J., & Mandrekar, J. N. (2004). How reliable are assessments of clinical teaching? A review of the published instruments. *Journal of General Internal Medicine*, 19, 971-977.
- Chen, M. P. (2009). An evaluation of the ELNP e-Learning quality assurance program: Perspectives of gap analysis and innovation diffusion. *Educational Technology & Society*, *12*(1), 18-33.
- Chico State University. (2010). Rubric for online instruction (ROI) history. Retrieved from http://www.csuchico.edu/celt/roi/history.shtml
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Earlbaum.
- ConQir. (2010). Literature review of interrater reliability. Retrieved from http://www.conqir-idr.org/literature/LR\_InterraterReliability\_JT.pdf
- Crocker, L. M., Miller, M. D., & Franks, E. A. (1989). Quantitative methods for assessing the fit between test and curriculum. *Applied Measurement in Education*, 2(2), 179-194.
- Cropper, M. (2009). Providing empirical support for Merrill's first principles of instruction: A meta-analysis/synthesis of instructional strategy research. Unpublished manuscript, Utah State University, Logan, UT.
- Cropper, M., Bentley, J. P. H., & Schroder, K. (2009). How well do high-quality online courses employ Merrill's first principles of instruction? In M. Orey, V. J.

- McClendon & R. Branch (Eds.), *Educational media and technology yearbook* (Vol. 34, pp. 121-140). Boston, MA: Springer.
- Fidler, F., & Thompson, B. (2001). Computing correct confidence intervals for ANOVA fixed- and random-effects effect sizes intervals for ANOVA fixed- and random-effects effect sizes. *Educational and Psychological Measurement 61*, 575. doi: 10.1177/0013164401614003
- Frick, T., Chadha, R., Watson, C., Wang, Y., & Green, P. (2007). *Theory-based course evaluation: Nine scales for measuring teaching and learning quality* Unpublished manuscript, Indiana University, Bloomington, IN.
- Frick, T., Chadha, R., Watson, C., Wang, Y., & Green, P. (2009). College student perceptions of teaching and learning quality. *Education Technology Research and Development*, 57(3), 705-720.
- Frick, T., Chadha, R., Watson, C., & Zlatkovska, E. (2010). Improving course evaluations to improve instruction and complex learning in higher education. *Educational Technology Research and Development*, *58*(2), 115-136.
- Garson, G. D. (2000). Significance. *Syllabus for PA 765-766: Quantitative research in public administration* Retrieved from http://www2.uta.edu/sswmindel/S6347/Significance/Intro%20to%20statistical%2 0significance.pdf
- Garson, G. D. (2010a). Correlation. Stat Notes. Retrieved from http://faculty.chass.ncsu.edu/garson/PA765/correl.htm
- Garson, G. D. (2010b). Reliability analysis. Retrieved from http://faculty.chass.ncsu.edu/garson/PA765/reliab.htm
- Garson, G. D. (2010c). Validity. Stat Notes. Retrieved from http://faculty.chass.ncsu.edu/garson/PA765/correl.htm
- Graham, C., Cagiltay, K., Lim, B.-R., Craner, J., & Duffy, T. M. (2001). Seven principles for effective teaching: A practical lens for evaluating online courses. *The Technology Source Archives*, (March/April). Retrieved from http://technologysource.org/article/seven\_principles\_of\_effective\_teaching/
- Haidet, K. K., Tate, J., Divergilio-Thomas, D., Kolanowski, A., & Happ, M. B. (2009). Methods to improve reliability of video-recorded behavioral data. *Research in Nursing & Health*, 32, 465-474.
- Hambleton, R. K. (1980). Test score validity and standard setting methods. In R. Berk (Ed.), *Criterion-referenced measurement: State of the art* (pp. 80-123). Baltimore, MD: John Hopkins University Press.

- Hirumi, A. (2002). A framework for analyzing, designing, and sequencing planned elearning interactions. *The Quarterly Review of Distance Education*, 3(2), 141-160.
- Howell, D. C. (2002). *Statistical methods for psychology* (5th ed.). Pacific Grove, CA: Duxbury.
- Kaplan, R. M., & Saccuzzo, D. P. (2005). *Psychological testing: Principles, applications, and issues* (6th ed.). Belmont, CA: Thomson Wadsworth.
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*, 159-174.
- Lawshe, C. H. (1975). A quantitative approach to content validity. *Personnel Psychology*, 28(4), 563-575.
- Linkov, F., Lovalekar, M., & Laporte, R. (2007). Quality control of epidemiological lectures online: Scientific evaluation of peer review. *Croatian Medical Journal*, 48(2), 249-255.
- Linn, R. L., & Gronlund, N. E. (2000). *Measurement and assessment in teaching* (8th ed.). Upper Saddle River, NJ: Prentice Hall, Pearson Education.
- Magnusson, D. (1967). Test theory. Reading, MA: Addison-Wesley.
- Merrill, M. D. (1999). Instructional transaction theory (ITT): Instructional design based on knowledge objects. In C. M. Reigeluth (Ed.), *Instructional design theories and models: A new paradigm of instructional technology* (Vol. 2, pp. 397-423). Mahwah, NJ: Erlbaum.
- Merrill, M. D. (2001). 5 star instructional design rating. Retrieved from http://cito.byuh.edu/merrill/text/papers.htm
- Merrill, M. D. (2002). First principles of instruction. *Educational Technology Research* and Development, 50(3), 43-59. doi: 10.1007/BF02505024
- Merrill, M. D. (2006). Levels of instructional strategy. *Educational Technology*, 46(4), 5-10.
- Merrill, M. D. (2007a). First principles of instruction: a synthesis. In R. A. Reiser & J. V. Dempsey (Eds.), *Trends and issues in instructional design and technology* (2nd ed., pp. 62-71). Upper Saddle River, NJ: Merrill/Prentice Hall.
- Merrill, M. D. (2007b). A task-centered instructional strategy. *Journal of Research on Technology in Education*, 40(1), 33-50.

- Merrill, M. D. (2008). Converting e<sub>3</sub> learning to e<sup>3</sup> learning: An alternative instructional design method. In S. Carliner & P. Shank (Eds.), *The e-learning handbook: Past promises, present challenges* (pp. 359-400). San Francisco, CA: Pfeiffer.
- Merrill, M. D. (2009). Finding e3 (effective, efficient, and engaging) instruction. *Educational Technology*, 49(3), 15-26.
- Piña, A. (2005). *Guiding principles for distance/online learning course development*. Center for Teaching and Learning, Northeastern Illinois University, Chicago, IL.
- Reigeluth, C. M. (1999). What is instructional-design theory and how is it changing? In C. M. Reigeluth (Ed.), *Instructional-design theories and models: A new paradigm of instructional theory* (Vol. II, pp. 5-29). Mahwah, NJ: Erlbaum.
- Reigeluth, C. M., & Carr-Chellman, A. A. (2009). Situational principles of instruction. In C. M. Reigeluth & A. A. Carr-Chellman (Eds.), *Instructional-design theories and models: Building a common knowledge base* (Vol. 3, pp. 57-71). New York, NY: Routledge.
- Ritchie, D. C., & Hoffman, R. (1997). Incorporating instructional design principles with the world wide web. In B. Khan (Ed.), *Web based instruction* (pp. 135-138). Englewood Cliffs, NJ: Educational Technology Publications.
- Scholfield, P. (2010). Some stats and SPSS pointers. Retrieved from http://privatewww.essex.ac.uk/~scholp/statsquibs.htm
- Sherry, A. C. (2003). Quality and its measurement in distance education. In M. G. Moore & W. G. Anderson (Eds.), *Handbook of distance education* (pp. 435-459). Mahway, NJ: Erlbaum.
- Shrout, P. E., & Fleiss, J. L. (1979). Intraclass corelations: Uses in assessing rater reliability. *Psychological Bulletin*, 86(2), 420-428.
- Southern Regional Education Board. (2005). Criteria for evaluating online courses. Retrieved from http://www.sret.sreb.org/criteria/online.asp
- Stemler, S. E. (2004). A comparison of consensus, consistency, and measurement approaches to estimating interrater reliability. *Practical Assessment, Research & Evaluation*, 9(4). Retrieved from http://PAREonline.net/getvn.asp?v=9&n=4
- Sunal, D. W., Sunal, C. S., Odell, M. R., & Sundberg, C. A. (2003). Research supported best practices for developing online learning. *The Journal for Interactive Online Learning*, *2*(1), 1-40. Retrieved from http://www.ncolr.org/jiol/issues/pdf/2.1.1.pdf

- Taylor, P. J. (2009). An introduction to intraclass correlation that resolves some common confusions. Programs in Science, Technology & Values, Critical & Creative Thinking, and Public Policy. University of Massachusetts, Boston, MA.
- Thomson NETg. (2003). *Thomson job impact study (phases I and II): The next generation of corporate learning*. Unpublished report, Naperville, IL.
- USU Faculty Senate. (2009). Faculty evaluations committee annual report to the faculty senate executive committee. 4-59. Retrieved from http://www.usu.edu/fsenate/FS/2009-2010/Agenda/FS%20Agenda%20Nov.%2030,%202009.pdf
- Varnell, S. P., Murray, D. M., Janega, J. B., & Blitstein, J. L. (2004). Design and analysis of group-randomized trials: A review of recent practices. *American Journal of Public Health*, *94*(3), 393-399.
- Weir, J. P. (2005). Quantifying test-retest reliability using the instraclass correlation coefficient and the *SEM*. *Journal of Strength and Conditioning Research*, 19(1), 231-240.

# **APPENDICES**

Appendix A. Comparison of Instruments

This comparison of instruments lists all of the items from the C-5 Star rating, the M-5 Star as well as categories of items from the school online course instruments (WebCT, Texas IQ, and SREB) that do not match First Principles. The numbers in each instrument frequency column represent the number of items that match the specified items from the 5 Star instruments items in columns 1 and 2 or the non 5 Star items in column 8.

Comparing the school instruments to the C-5 Star rating, 24 of 32 WebCT items, 34 of 63 Texas IQ items and 12 of 17 SREB item matched Merrill's principles as found on the C-5 Star. (See columns 1, 5, 6 and 7 in Table 18.) The C-5 Star was used throughout all phases of the pilot study.

It would seem logical that the school instruments would also compare well with the M-5 Star, since the M-5 Star is also based upon his First Principles of Instruction.

However, because Merrill eliminated heading items for each principle, and provided much more specific guidelines for each of his sub-principles, many of the direct comparisons between the school instrument items, which were very general, and M-5 Star items, which were very specific, were eliminated. When the school instruments were compared in Table 18 with the M-5 Star, only seven WebCT items, four Texas IQ items, and no SREB items matched Merrill's principles.

Table 18
Comparison of 5 Star with School Rating Instruments

Cropper C-5 Star Rating Items	Merrill 2007 M-5 Star Rating Items	C-5 Star Rating	M-5 Star Rating	Web CT	Texas IQ	SREB	Non- 5 Star items from other instruments
Problem Centered		1	1				
Learners are shown the task, rather than just given objectives		1	1	3	1	1	
	1. Is there an authentic, realworld whole task?		1				
Learners are engaged at problem or task level	2. Is there a task-centered strategy?	1	1		1		
Learners solve progression of problems	3. Is there a progression of whole tasks?	1	1				
	4. Is the whole task demonstrated? Is the demonstration consistent with the type of learning involved?		1				
	5. Are learners required to do the whole task? Is the application consistent with the type of learning involved?		1				
Demonstration							
Demonstrates (shows examples of) what is to be learned		1		2		1	
Demonstration consistent with learning goal		1			2		

*Note.* The numbers in a given row and column are the number of items in a given instrument that match the 5 Star instrument item.

Cropper C-5 Star Rating	Merrill 2007 M-5 Star Rating	C-5 Star Rating	M-5 Star Rating	Web CT	Texas IQ	SREB	Non- 5 Star items from other instruments
Relevant media used		1		4	8	1	
				1	3	1	Adequate content and resources to meet learning goals and objectives
Examples and non-examples for kinds -of (Concepts)		1					
•	6. Does the demonstration tell learners the name and definition of each category or alternative procedure?		1				
	7. Does the demonstration show learners relevant examples of each category?		1				
	8. Does the demonstration provide guidance by highlighting discriminating properties? By showing matched examples among categories?		1				
	9. Does the demonstration show a series of divergent examples of increasing difficulty?		1				
	10. Does the demonstration use media appropriately?		1				

Cropper	Merrill 2007		M-5				Non- 5 Star items from
C-5 Star	M-5 Star	C-5 Star	Star	Web	Texas		other
Rating	Rating	Rating	Rating	CT	IQ	SREB	instruments
Practice followed by corrective feedback and an indication of progress, not just right- wrong		1		1	6	5	
feedback.							
Diminishing coaching		1					
Varied problems		1					
				1	1	1	The content and requirements are as demanding as a face-to-face course.
				1			Assignments and projects that require students to make appropriate and effective use of external resources.
				1		1	Clearly communicated assignments.
				1			Explicitly communicated expectations, including deliverables

Cropper C-5 Star Rating	Merrill 2007 M-5 Star Rating	C-5 Star Rating	M-5 Star Rating	Web CT	Texas IQ	SREB	Non- 5 Star items from other instruments
Demonstrations 1	16. Does the	1	1	CI	IQ	SKED	mstruments
for how-to	demonstration	1	1				
(procedures)	show a specific						
(procedures)	instance of the task						
	and demonstrate						
	each of the steps						
	required to						
	complete the task						
	or provide the						
	opportunity for						
	learners to play						
	with the procedure						
	in an actual or						
	simulated						
	situation?						
	17. Does the		1				
	demonstration		1				
	show the						
	consequence of						
	each step and the						
	consequence of the						
	complete						
	procedure?						
	18. Does the		1				
	demonstration		1				
	provide guidance						
	by calling attention						
	to the execution of						
	each step?						
	19. Is the		1				
	procedure		-				
	demonstrated in a						
	series of						
	increasingly						
	difficult situations?						
	20. Does the		1				
	demonstration		-				
	make appropriate						
	use of media?						
	ase of inequa.						

Cropper C-5 Star Rating	Merrill 2007 M-5 Star Rating	C-5 Star Rating	M-5 Star Rating	Web CT	Texas IQ	SREB	Non- 5 Star items from other instruments
How-to	21. Does the	1	1	1			
application	application require						
requires	learners to do the						
learner to do	task by executing						
the procedure.	each step in a real or simulated						
	situation?						
	22. Does the		1				
	application provide		-				
	intrinsic feedback						
	(observing the						
	consequence of						
	one's actions) and						
	extrinsic feedback						
	(informing learners about the						
	appropriateness of						
	a given operation						
	or action)?						
	23. Does the		1				
	application require						
	a simple to						
	complex						
	progression of						
	tasks rather than a single task?						
	24. Are tasks early		1				
	in the progression		1				
	coached and is this						
	coaching gradually						
	withdrawn as for						
	successive tasks in						
	the progression?						
	25. Does the						
	application make						
	appropriate use of media?						
Visualizations	26. Does the	1					
for what-	demonstration						
happens	show the process in						
(processes)	a specific real or						
	simulated						
	situation?						

Cropper C-5 Star Rating	Merrill 2007 M-5 Star Rating	C-5 Star Rating	M-5 Star Rating	Web CT	Texas IQ	SREB	Non- 5 Star items from other instruments
	27. Does the		1				
	demonstration						
	show a specific						
	portrayal of the						
	conditions						
	necessary for each						
	event in the						
	process?						
	28. Is the		1				
	demonstration						
	repeated for a						
	progression of						
	increasingly						
	complex scenarios?						
	29. Does the		1				
	demonstration						
	provide guidance						
	by helping learners						
	relate the events in						
	the process to the						
	conditions or to						
	some conceptual						
	structure?						
	30. Does the		1				
	demonstration						
	make appropriate						
	use of media?						
Appropriate		1					
learner guidance							
Learners		1					
directed to							
relevant							
information							
Multiple		1					
representations							
used and							
compared							
Learners		1					
assisted to relate							
the new							
information to							
the structure that							
was recalled or							
provided							

Cropper C-5 Star Rating	Merrill 2007 M-5 Star Rating	C-5 Star Rating	M-5 Star Rating	Web CT	Texas IQ	SREB	Non- 5 Star items from other instruments
What-happens	31. Are learners	1	1	2	1		
application	required to predict						
requires	the consequence						
learner to	from a given set of						
predict a	conditions in a						
consequence	specific situation?						
of a process	OR are learners						
given	required to						
cognition, or	troubleshoot an						
to find faulted	unexpected						
conditions	consequence by						
given an	finding the faulted						
unexpected	conditions in a						
consequence.	specific situation?						
	32. Are learners		1				
	able to receive						
	intrinsic feedback						
	by being able to						
	test their						
	predictions or test						
	their trouble						
	shooting?  33. Does the		1				
			1				
	application make						
	appropriate use of media?						
	34. Are learners		1				
			1				
	required to make predictions or						
	troubleshoot a						
	series of						
	increasingly						
	complex problems?						
	35. Is coaching		1				
	provided for		1				
	problems early in						
	the progression and						
	gradually						
	withdrawn as the						
	progression						
	continues?						

Cropper C-5 Star   Star   Star   Star   Star   Web   Texas   Star   Star   Texas   Star   S								N
Star Rating     M-5 Star Rating     C-5 Star Rating     Star Rating     Rating     Rating     Rating     Rating     CT IQ SREB Instruments are in the content and requirements are as demanding as a face-to-face course.       Image: Company of the content and requirements are as demanding as a face-to-face course.     Image: Company of the course are as demanding as a face-to-face course.       Image: Company of the course are as demanding as a face-to-face course.     Image: Company of the course are as a face-to-face course.       Image: Company of the course are as demanding as a face-to-face course.     Image: Company of the course are as a face-to-face course.       Image: Company of the course are as demanding as a face-to-face course.     Image: Company of the course are as a face-to-face course.       Image: Company of the course are as demanding as a face-to-face course.     Image: Company of the course are as a face-to-face course.       Image: Company of the course are as demanding as a face-to-face	C	M:11 2007		M 5				Non- 5 Star
Rating Rating Rating Rating CT IQ SREB instruments and requirements are as demanding as a face-to-face course.  I			C 5 Ston		Wah	Towas		
1   1   The content and requirements are as demanding as a face-to-face course.   1   Assignments and projects that require students to make appropriate and effective use of external resources.   1   Clearly communicated assignments.   1   Explicitly communicated expectations, including deliverables   1   Explicitly communicated expectations, including deliverables   1   Explicitly communicated expectations including   1   Explic							CDED	
and requirements are as demanding as a face-to-face course.  I Assignments and projects that require students to make appropriate appropri	Kating	Raung	Rating	Raung				
requirements are as a face-to-face course.  1					1	1	1	
are as demanding as a face-to-face course.  1								
demanding as a face-to-face course.  1								
a face-to-face course.  1								
course.  Assignments and projects that require students to make appropriate and effective use of external resources.  I Clearly communicated assignments.  I Explicitly communicated expectations, including deliverables  Activation I  Helps learner I see relevance and have confidence in their ability to gain knowledge and skill  Uses a I 3 3 procedure to select the right content for each learner  Learners are given new experience for								
Assignments and projects that require students to make appropriate and effective use of external resources.  I Clearly communicated assignments.  I Explicitly communicated expectations, including deliverables  Activation 1  Helps learner 1 see relevance and have confidence in their ability to gain knowledge and skill  Uses a   1   3   3   1   3   1   3   1   3   1   3   1   3   3								
and projects that require students to make appropriate and effective use of external resources.  1					1			
that require students to make appropriate and effective use of external resources.  1					•			
students to make appropriate and effective use of external resources.  1 1 Clearly communicated assignments.  1 Explicitly communicated expectations, including deliverables  Activation 1  Helps learner 1  See relevance and have confidence in their ability to gain knowledge and skill  Uses a 1 3 procedure to select the right content for each learner  Learners are given new experience for								
appropriate and effective use of external resources.  1								
and effective use of external resources.  1								make
use of external resources.  1 Clearly communicated assignments.  1 Explicitly communicated expectations, including deliverables  Activation  1 Helps learner see relevance and have confidence in their ability to gain knowledge and skill  Uses a 1 3 procedure to select the right content for each learner Learners are given new experience for								
resources.  1 Clearly communicated assignments.  1 Explicitly communicated expectations, including deliverables  Activation 1  Helps learner see relevance and have confidence in their ability to gain knowledge and skill  Uses a 1 3								
1 Clearly communicated assignments.  1 Explicitly communicated expectations, including deliverables  Activation 1  Helps learner see relevance and have confidence in their ability to gain knowledge and skill  Uses a 1 3 3 procedure to select the right content for each learner  Learners are given new experience for								
communicated assignments.  1 Explicitly communicated expectations, including deliverables  Activation  1  Helps learner see relevance and have confidence in their ability to gain knowledge and skill  Uses a 1 3 3 procedure to select the right content for each learner  Learners are given new experience for								
assignments.  1 Explicitly communicated expectations, including deliverables  Activation  1					1		1	
1 Explicitly communicated expectations, including deliverables  Activation 1  Helps learner 1 see relevance and have confidence in their ability to gain knowledge and skill  Uses a 1 3 procedure to select the right content for each learner Learners are given new experience for								
Activation  I Helps learner see relevance and have confidence in their ability to gain knowledge and skill Uses a 1 3 procedure to select the right content for each learner Learners are given new experience for					1			
Activation  Helps learner see relevance and have confidence in their ability to gain knowledge and skill  Uses a 1 3 procedure to select the right content for each learner  Learners are given new experience for					1			
Activation  Helps learner see relevance and have confidence in their ability to gain knowledge and skill  Uses a 1 3 procedure to select the right content for each learner  Learners are given new experience for								
Activation  Helps learner see relevance and have confidence in their ability to gain knowledge and skill  Uses a 1 3 procedure to select the right content for each learner  Learners are given new experience for								
Helps learner see relevance and have confidence in their ability to gain knowledge and skill  Uses a 1 3 procedure to select the right content for each learner  Learners are given new experience for								
see relevance and have confidence in their ability to gain knowledge and skill  Uses a 1 3 procedure to select the right content for each learner  Learners are given new experience for	Activation			1				
and have confidence in their ability to gain knowledge and skill  Uses a 1 3 procedure to select the right content for each learner  Learners are given new experience for	Helps learner			1				
confidence in their ability to gain knowledge and skill  Uses a 1 3   procedure to select the right content for each learner  Learners are given new experience for	see relevance							
their ability to gain knowledge and skill  Uses a 1 3								
gain knowledge and skill  Uses a 1 3 procedure to select the right content for each learner  Learners are 1 given new experience for								
knowledge and skill  Uses a 1 3  procedure to select the right content for each learner  Learners are given new experience for								
skill  Uses a 1 3 procedure to select the right content for each learner  Learners are given new experience for								
Uses a 1 3 procedure to select the right content for each learner  Learners are given new experience for								
procedure to select the right content for each learner  Learners are given new experience for				1		2		
select the right content for each learner  Learners are 1 given new experience for				1		3		
content for each learner  Learners are 1 given new experience for								
each learner  Learners are 1 given new experience for								
Learners are 1 given new experience for								
given new experience for				1				
	knowledge							
foundation	foundation							

Cropper C-5 Star Rating	Merrill 2007 M-5 Star Rating	C-5 Star Rating	M-5 Star Rating	Web CT	Texas IQ	SREB	Non- 5 Star items from other instruments
Learners recall previous knowledge, experience	36. Are learners prompted to recall, describe or demonstrate relevant prior knowledge or skill?		1	1	1		
Learners recall structure for organizing knowledge	37. Are learners provided or prompted to develop an organizing structure for the skills be obtained?		1	1	1		
	38. Is guidance provided during demonstration that relates the knowledge or skill being demonstrated to this structure?		1				
	39. Is coaching provided during application that reminds learners of this structure to facilitate their application of their new knowledge or skill?		1				
	40. Do learners have an opportunity to reflect on how the newly acquired knowledge and skill relate to the structure?		1				
	sa ucture:				1		Support to prepare for entry level
					2		Test for entry- level skills

Cropper C-5 Star Rating	Merrill 2007 M-5 Star Rating	C-5 Star Rating	M-5 Star Rating	Web CT	Texas IQ	SREB	Non- 5 Star items from other instruments
				1	9	1	Syllabus/ Course Requirements
Reflection, discussion, defending knowledge	41. Do learners have an opportunity to reflect on, discuss or defend their new knowledge and skill?	1	1	1			
Creation, invention, exploration of ways to apply knowledge	42. Do learners have an opportunity to create, invent, or explore personal ways to use their new knowledge or skill?	1	1				
Public demonstration of knowledge	43. Do learners have an opportunity to publicly demonstrate their new knowledge or skill?	1	1				
Implementation The instruction		1		1	2		
facilitates navigation					_		
The degree of learner-control is appropriate for the learning goals and the learners		1			1		
Collaboration is used effectively		1		6	6	4	
Instruction is personalized		1					

Cropper C-5 Star Rating	Merrill 2007 M-5 Star Rating	C-5 Star Rating	M-5 Star Rating	Web CT	Texas IQ	SREB	Non- 5 Star items from other instruments
Kating	Ratting	Kattiig	Kating	4	9	SKED	Effective use of online
							technology
Integration		1				1	Provision is made for students with special needs
					1		The course provider's credentials are available for review.
					1		Has the course been developed by a qualified team consisting of content experts and instructional designers?
					1		Support to prepare for entry level
					2		Test for entry- level skills
				1	9	1	Syllabus/ Course Requirements
	Integration						
	41. Do learners have an opportunity to reflect on, discuss or defend their new knowledge and skill?		1	1			

Cropper C-5 Star Rating	Merrill 2007 M-5 Star Rating	C-5 Star Rating	M-5 Star Rating	Web CT	Texas IQ	SREB	Non- 5 Star items from other instruments
	42. Do learners have an opportunity to create, invent, or explore personal ways to use their new knowledge or skill?		1				
	43. Do learners have an opportunity to publicly demonstrate their new knowledge or skill?		1				
				4	9		Effective use of online technology
						1	Provision is made for students with special needs
					1		The course provider's credentials are available for review.
					1		Has the course been developed by a qualified team consisting of content experts and instructional designers?

Appendix B. Pilot Study

The purpose of the pilot study was to evaluate the reliability and validity of Cropper's 5 Star instrument (the C-5 Star) as a measure of the quality of award-winning online courses. Award-winning online courses were evaluated with the C-5 Star and six other recognized instruments. This was done to determine:

- the amount of interrater reliability evidence for the C-5 Star and the other rating instruments;
- the amount of concurrent or discriminant validity evidence between the C-5
   Star and other online course instruments; and
- the degree to which award-winning courses use Merrill's First Principles.

Award-winning online courses were found by searching Google for courses that had won awards during the previous two years. Permission was sought to use the identified courses for the study. Of 20 possible courses, permission was received to use seven online courses that had sufficient online material for adequate evaluation.

A description of the selection of online course rating instruments used for the pilot and dissertation studies is found in the Literature Review chapter. The three school online course rating instruments that were used in the first phase of the pilot study were later also used in the dissertation study. Three phases of the pilot study were conducted.

#### **Phase I: Rating with Multiple Instruments**

For Phase I, Cropper and a graduate student from Twente University were the course raters. The instruments used were the C-5 Star and six other instruments. These are described on the Literature Review chapter. The raters trained each other on the use of the instruments in iterative cycles of clarification. As part of the training, they both evaluated a *Research for the Classroom Teacher* course using the seven rating forms. After the initial ratings, because of ambiguity of many of the items on the instruments, descriptions were written for each of the rating values (1-5) for all of the forms. The raters then re-rated the course. They reconciled their ratings within one point on each item. Interrater reliability for their reconciled ratings of this course across the seven forms was calculated at ICC = .54, p = .001. Note that because ratings for the Research for the Classroom Teacher class were reconciled, interrater reliability was inflated for this course. The raters then rated the rest of the courses, all of which were award-winning, using all of the instruments. No reconciliation was done on the other course ratings.

#### **Interrater Reliability Testing**

Intraclass correlation was used to analyze interrater reliability because it accounts for rater differences, and can be used for two or more raters. A two-way mixed design was used because the raters were fixed. Interrater reliability evidence was fair to substantial for the instruments. The ICC interrater reliability was fair for the Brandon Hall (ICC = .28, p = .001) and Texas IQ (ICC = .39, p = .001), instruments, moderate for the SREB (ICC = .40, p = .001) and the C-5 Star (ICC = .59, p = .001) instruments, and

substantial for the WebCT (ICC = .60, p = .001), ASTD (ICC = .62, p = .001), and Motivation (ICC = .62, p = .001) instruments.

Intraclass correlation was slight to exceptional across the courses. ICC was slight for the Excel Scenario-Based (ICC = .14, p = .05) course. ICC was moderate for the Psychology of Communication (ICC = .48, p = .001), Evaluating Training Programs (ICC = .52, p = .001), Digital Craft (ICC = .53, p = .001), Landscape Design (ICC = .54, p = .001), and SAT (ICC = .59, p = .001) courses. ICC was substantial for the Cashier Training course (ICC = .68, p = .001), and exceptional for Research for the Classroom Teacher (ICC = .82, p = .001).

#### **Validity Testing**

The pilot study correlations between the C-5 Star and the school instruments are related to the quantity of school instrument items that matched the C-5 Star. The Texas IQ, which had 34 matching items, has a high correlation with the C-5 Star. The WebCT, which had 24 matching items, has a low correlation. The SREB, which had 12 matching items, has a very low correlation. (See Table 14, p. 60) Very possibly, the actual quantity of matching instrument items heavily affects the correlation coefficients in the pilot study.

A Shapiro-Wilk's W test of normality was calculated across all four instruments. As shown in Table 19, each contained a nonsignificant W value indicating that normality assumptions were met justifying the use of parametric statistics. Levene's test of equality of error variance shows that the variances between instrument mean scores are significantly different (F = 31.59, df1 = 3, df2 = 2200, p = .001), so Pearson correlations

need to be calculated with standardized scores. The Pearson correlation is the parametric test for measuring validity. Therefore, Pearson correlations were calculated to determine if concurrent or divergent validity evidence existed between the C-5 Star and the school instruments. A two-tailed test was used because the study was viewed as exploratory. A similar study had not been done previously and none of the instruments involved had reliability or validity data. Therefore, it was unknown whether the instruments would provide concurrent or divergent validity evidence for the C-5 Star.

Table 19
Shapiro-Wilk Test of Normality

	S	hapiro-Wilk	
	Statistic	df	Sig.
C-5 Star Mean Score	.962	8	.831
Texas IQ Mean Score	.949	8	.704
WebCT Mean Score	.907	8	.330
SREB Mean Score	.928	8	.501

Table 20

Number of School Instrument Items That Match C-5 Star Items

	Texas IQ	WebCT	SREB
	(63 items)	(32 items)	(17 items)
C-5 Star (26 items)	34	24	12

*Note.* Number of items is based on a count from the comparison of instruments in Appendix A.

Scatterplots between the C-5 Star and School Instruments. The number of matching items between the C-5 Star and the school instruments, as shown in Table 20, would suggest that they would be highly correlated. However, that is not necessarily the

case. Scatterplots and actual correlations reveal whether there is a linear association between them.

Scatterplots provide a preview of the relationships between the C-5 Star and the school instruments. Only the scatterplot between the C-5 Star and the Texas IQ instruments had a strong linear relationship, and provided convergent validity evidence. (See Figure 6.) The scatterplot for the C-5 Star and the WebCT has more scattered points that follow a less linear pattern, providing weak divergent validity support. (See Figure 7.) Scatterplot points for the C-5 Star and the SREB are totally scattered, have no linear pattern, and therefore exhibit substantial divergent validity evidence. (See Figure 8.)

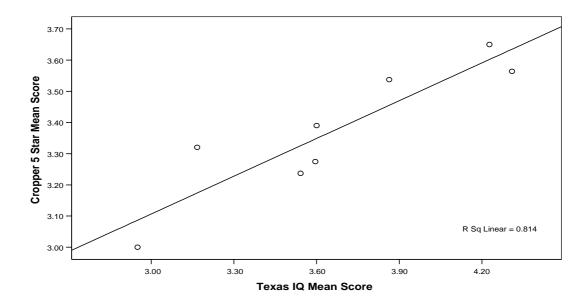


Figure 6. Scatterplot of the association between the C-5 Star mean course scores and the Texas IQ mean course scores. Note that the points follow a linear pattern. The actual correlation is high (r = .90, p = .002)  $(r^2 = .81)$  and provides concurrent validity evidence for the C-5 Star.

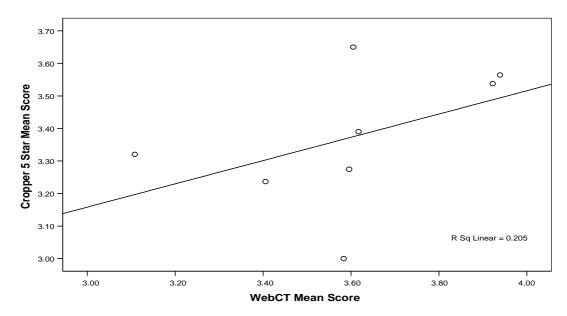


Figure 7. Scatterplot of the association between the C-5 Star mean course scores and the WebCT mean course scores. Note that the points are more scattered and follow a less linear pattern. The actual correlation is low (r = .45, p = .26)  $(r^2 = .20)$  and provides some divergent validity evidence for the C-5 Star.

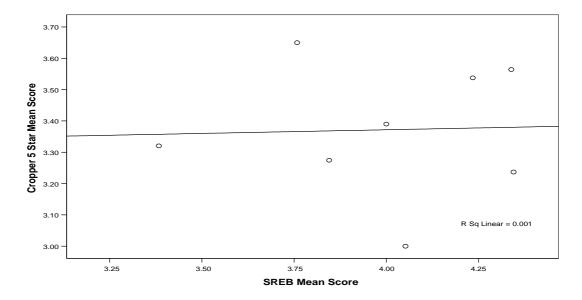


Figure 8. Scatterplot of the association between the C-5 Star mean course scores and the SREB mean course scores. Note that the points are totally scattered and have no linear pattern. The actual correlation is very low (r = .04, p = .93)  $(r^2 = .001)$  and provides substantial divergent validity evidence for the C-5 Star.

The actual correlation between the C-5 Star mean rating and the Texas IQ mean rating was very high (r = .90, p = .002) ( $r^2 = .81$ ) and significant at the .01 level. The correlation between the C-5 Star mean rating and the WebCT mean rating was low and not significant (r = .45, p = .26) ( $r^2 = .20$ ), and the correlation between the C-5 Star mean rating and the SREB mean rating was very low and not significant (r = .04, p = .93) ( $r^2 = .001$ ). (See Table 21.) The WebCT was correlated with the Texas IQ (r = .64, p = .09) ( $r^2 = .41$ ), and SREB (r = .69, p = .06) ( $r^2 = .48$ ), significant at the .10 level.

Table 21

Pearson Correlations using Standardized Scores for the Four Course Rating Instruments

		5 Star Mean Score	Texas IQ Mean Score	WebCT Mean Score
Texas IQ Mean Score	Pearson Correlation	.90(**)		
	Sig. (2-tailed)	.002		
	Sum of Squares and Cross-products	6.31		
	Covariance	.90		
	N	8		
WebCT Mean Score	Pearson Correlation	.45	.64	
	Sig. (2-tailed)	.259	.086	
	Sum of Squares and Cross-products	3.17	4.49	
	Covariance	.45	.64	
	N	8	8	
SREB Mean Score	Pearson Correlation	.04	.34	.69
	Sig. (2-tailed)	.932	.411	.057
	Sum of Squares and Cross-products	.25	2.37	4.85
	Covariance	.04	.34	.69
	N	8	8	8

<sup>\*\*</sup> Correlation is significant at the 0.01 level (2-tailed).

However the Texas IQ instrument was not correlated with the SREB (r = .34, p = .41) ( $r^2 = .11$ ). (See Table 21.)

Phase I of the pilot study provided moderate reliability evidence for the C-5 Star. It provided substantial reliability evidence for the Motivation, ASTD and WebCT, moderate reliability evidences for the SREB, and fair reliability evidence for the Texas I and Brandon Hall instruments.

The C-5 Star received exceptional concurrent validity evidence from the Texas IQ instrument and marginal to substantial concurrent validity evidence from the WebCT.

The SREB provided divergent validity evidence for the C-5 Star. The WebCT, Texas IQ and SREB provided concurrent validity evidence for each other, while the Texas IQ instrument and SREB provided divergent validity for each other.

#### Phase II: Analysis of Courses Using the C-5 Star

Phase I of the pilot study established the need to re-evaluate the reliability of the C-5 Star instrument with additional experienced raters. Phase II was also an attempt to go more in-depth in estimating the reliability for each of Merrill's principles separately. Three expert evaluators were added, each with 10 or more years of experience each as instructional designers. After being trained on the C-5 Star /First Principles of Instruction, the new evaluators evaluated sampled modules from six of the courses from the initial phase of the pilot (we eliminated the Psychology of Communication and Landscaping courses from the pilot because they were no longer available for review). In

the data analysis, we included the ratings of the three new evaluators, plus the original ratings from Phase I.

#### **Interrater Reliability Testing**

Intraclass correlation (ICC) was used to calculate interrater reliability. Intraclass correlation is a scale reliability measure that is often used for interrater reliability between two or more raters because it accounts for differences between raters. ICC over courses, for each principle separately, was negative for problem centered (ICC = -.12, p = .79), slight for application (ICC = .16, p = .13), and fair for activation (ICC = .26, p = .05), integration (ICC = .31, p = .03) and demonstration (ICC = -.36, p = .01). One reason interrater reliability may be especially low for the problem-centered principle is the definition of "real-world problems" may be confusing to raters. They must decide to what degree contrived problems, which are solved within the online class environment, can be classified as real world problems.

ICC over principles, for each course separately, was slight for the Digital Craft (ICC = .05, p = .32) and SAT course (ICC = .11, p = .21), fair for Research for the Classroom Teacher (ICC = .21, p = .10), Evaluating Training Programs (ICC = .30, p = .04) and the NETg Scenario-Based Excel course (ICC = .33, p = .03), and substantial for Cashier Training (ICC = .78, p = .001). Raters had a variety of opinions of the task-centered nature of the Digital Craft course, which truly was task-centered. They also had a problem rating the task-centeredness of the SAT course. It provides questions that are problems, but do not qualify as meaningful, real-world whole tasks; nor does the course provide a progression of problems.

Interrater reliability for the C-5 Star was negative to slight for the courses, and low to substantial for the principles. The interrater reliability for the overall instrument was fair (ICC = .34, p = .001), but significant at the .001 level.

#### **Use of First Principles**

Award-winning courses tended to use each of Merrill's First Principles with a pronounced variation in degree of application. All courses rated high and similarly on the problem-centered principle, but had wide variation in ratings on the other principles. We used ANOVAs to determine how the courses compared, how the raters compared, and how the use of Merrill's five principles compared. An ANOVA of principles, courses and principles\*courses shows that there is a significant difference between ratings of principles (F = 12.83, p = .001), between course ratings (F = 7.66, p = .001), and within the interaction of principles and courses (F = 2.72, p = .001).

A Bonferroni post hoc comparison of principles showed that the problem centered principle is used significantly more than the other principles. Integration is used significantly less than demonstration or application. There is a wide variation in the use of activation, application, and especially integration. However, most of the variation in integration can be explained by the fact that college courses can require students to apply their learning outside the classroom, and self-contained commercial courses cannot do so, because they have no way to follow up.

When college courses and commercial courses were analyzed separately, college courses scored high on integration, presumably because they can require students to integrate learning in real-life situations. Commercial courses scored lower on integration,

presumably because they are more self-contained, and cannot require learners to integrate learning into the real world. Phase II provided evidence that the C-5 Star is a reliable and valid measure of quality of award-winning courses.

A Bonferroni post hoc comparison of courses showed that the baseline NETg Excel course rated significantly higher than the Digital Craft (mean difference = .63, p = .002) and Evaluating Training Programs course ratings (mean difference = .84, p = .001). There is no significant difference between the NETg Excel course and the Research for the Classroom Teacher, SAT, and Cashier Training courses.

#### Phase III: Merrill Added as a Rater

During Phase III of the pilot study, Merrill was added as a rater, and the data was reanalyzed. Merrill evaluated the six courses using the C-5 Star. Merrill was strict in his definition of meaningful, real-world problems, and rated the courses lower than the other raters, except for Digital Craft, which he rated higher than most raters. Although the C-5 Star was based on Merrill's M-5 Star Instructional Design Rating (Merrill, 2001), Merrill felt that the equal weight given to all of the principles and sub-principles did not match the emphasis that he would give them, and began revising the instrument.

### **Interrater Reliability Testing**

The interrater reliability over courses for each principle separately, was negative for problem-centered (ICC = -.001, p = .48), slight for demonstration (ICC = .14, p = .48) and activation (ICC = .17, p = .02) and fair for integration (ICC = .22, p = .005) and application (ICC = .23, p = .001). It was especially clear that the problem-centered

principle needed significant clarification, and the other principles could also use some clarification.

The interrater reliability over principles, for each course separately, was slight for the Research for the Classroom Teacher (ICC = .02, p = .36), SAT (ICC = .14, p = .36) and Digital Craft (ICC = .15, p = .02) courses, fair for the Evaluating Training Programs (ICC = .31, p = .001), and the NETg Excel course (ICC = .34, p = .001) and moderate for Cashier Training (ICC = .52, p = .001). The overall interrater reliability for the C-5 Star was fair (ICC = .24, p = .001), but significant at the .001 level.

## **Validity Testing**

A Bonferroni post hoc comparison of principles showed that the problem-centered principle is still used significantly more than the other principles (mean difference between problem-centered and activation = .90, demonstration = .76, application = .99, and integration = 1.21, all significant at the p = .001 level. Integration is used significantly less than demonstration (mean difference = .45, p = .001), significant at the .001 level. (See a pilot study summary in the Methodology chapter of the dissertation.)

Appendix C. Merrill's 2007 5 Star Instrument (M-5 Star)

## **1st Principles Course Critique**

# © 2007 M. David Merrill mdavid.merrill@gmail.com Course parameters:

Reviewer:
Date reviewed:
Course or module title:
Course developer:
Course distributor or company:
Content area:
Audience:
Delivery mode:

#### **Critique:**

This critique evaluates an existing course to determine the extent to which the course implements First Principles of Instruction. There are five areas for evaluation: task-centered, demonstration, application, activation, and integration. However, each of these areas does not contribute equally to effective instruction (Merrill, 2006). Therefore the relative contribution of each of these areas is weighted in this scoring sheet.

The scoring yields a total score of 100. This scoring can be converted to a 50 point scale by dividing the score by 2. It can be converted to a 10 point scale by dividing by 10. It can be converted to a 5 point scale by dividing by 20.

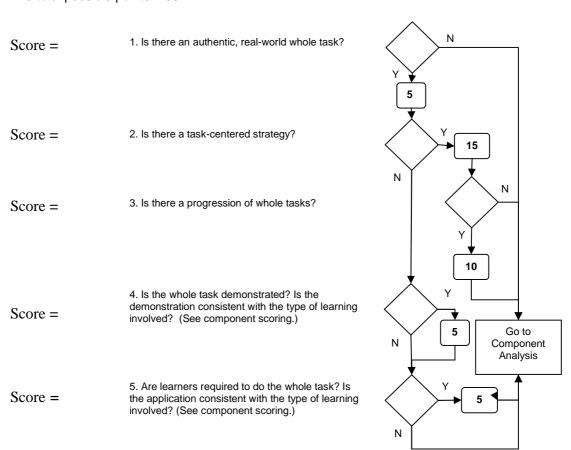
The scoring takes place in two stages: scoring for task centered and scoring for components or topics. The procedure is indicated by a flow chart with each decision represented by a question. These questions are elaborated in the attached papers.

The following scores in each category rate a star: task-centered = 25, demonstration = 20, application = 20, activation = 7 and integration = 7. Does your instruction rate 5 Stars?

## **Task-centered Analysis**

there an actual whole task? $\square$ If not what whole task is implied? $\square$	
xplain:	
escribe the actual or implied whole task.	
escription:	

Follow the flow chart to score for the task-centered principle. Add up to the number of points shown in the box for each question. The total possible points = 30.



**Total Score for Task-centered (0-30) =** 

## Component Analysis Demonstration and Application Principle List each component skill or topic.

• Classify each component skill or topic.

Component Skill	Classification	Demonstration Score	Application Score
Total Component Score =			

Score each component skill or topic using the appropriate scoring procedure below. Determine the average **demonstration** score for all the kind-of, how-to, and what-happens component skills by summing the scores and dividing by the number of these component skills (score should be 0-25). Determine the average **application** score for all the kinds-of, how-to, what-happens component skills by summing the scores for each component and dividing by the number of these component skills (score should be 0-25).

For large courses it may be sufficient to analyze a sample of the component skill lessons.

#### **Kinds-of Demonstration**

6. Does the demonstration tell Score = learners the name and definition of each category or alternative procedure? 5 Ν 7. Does the demonstration show learners relevant examples of each Score = category? Score application 8. Does the demonstration provide guidance by Score = highlighting discriminating properties? By showing matched examples among categories? Ν 9. Does the demonstration Score = show a series of divergent examples of increasing difficulty? Score = 10. Does the demonstration use media appropriately?

**Total kinds-of demonstration score (0-25) =** 

## **Kinds-of Application**

11. Does the application require learners to classify a series of new Score = examples or select among a series of new alternative procedures. 12. Does the application require learners to explain their Score = classification or selection by pointing out the presence or absence of discriminating properties? Ν Score next component 13. Does the application provide corrective feedback that focuses Score = learners' attention on discriminating properties? Or does the feedback show intrinsic consequences for selection of alternative procedures? 14. Does the application provide coaching early in the sequence by Score = helping learners see the portrayal of the discriminating attributes and gradually withdraw this coaching as 5 the application continues? Ν Score = 15. Does the application make appropriate use of media?

Total kinds-of application score (0-25) =

#### **How-to Demonstration**

16. Does the demonstration show a Ν specific instance of the task and Score = demonstrate each of the steps required to complete the task or provide the opportunity for learners to play with the procedure in an actual 5 or simulated situation? Ν 17. Does the demonstration show Score = the consequence of each step and the consequence of the complete procedure? Score application 5 18. Does the demonstration provide Score = guidance by calling attention to the execution of each step? 5 19. Is the procedure demonstrated in Score = a series of increasingly difficult situations? 20. Does the demonstration make Score = appropriate use of media?

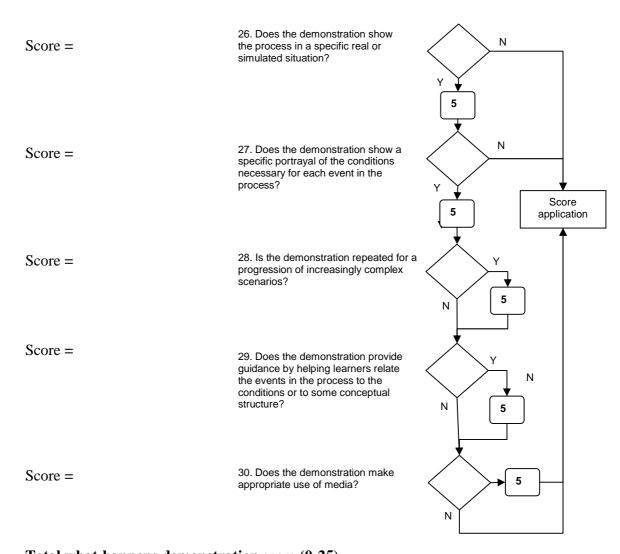
**Total how-to demonstration** score (0-25) =

## **How-to Application**

21. Does the application require Ν learners to do the task by Score = executing each step in a real or simulated situation? 5 22. Does the application provide intrinsic feedback (observing the Ν Score = consequence of one's actions) and extrinsic feedback (informing learners about the appropriateness of a given Score next operation or action)? 5 component 23. Does the application require a simple to complex progression of Score = tasks rather than a single task? 5 Ν 24. Are tasks early in the progression coached and is this Score = coaching gradually withdrawn as for successive tasks in the progression? Ν 5 Score = 25. Does the application make appropriate use of media?

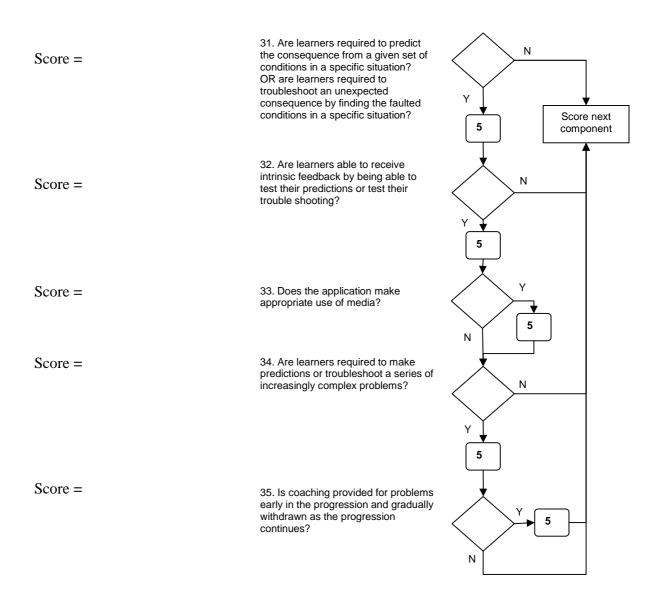
**Total how-to application** score (0-25) =

## **What-happens Demonstration**



**Total what-happens demonstration** score (0-25) =

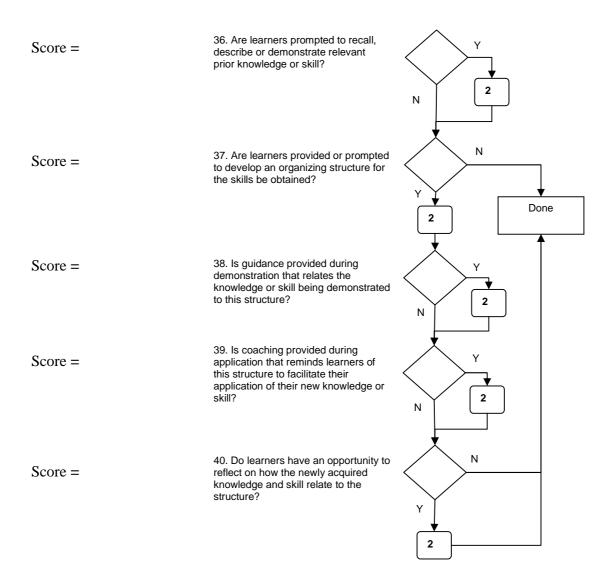
## What-happens application



Total what-happens application score = (0-25)

## **Activation Principle**

The activation principle applies to the whole course or module.



**Total activation** score (0-10) =

## **Integration Principle**

This principle applies to the course or module as a whole

Score = 

41. Do learners have an opportunity to reflect on, discuss or defend their new knowledge and skill?

42. Do learners have an opportunity to create, invent, or explore personal ways to use their new knowledge or skill?

N

Y

3

N

Done

43. Do learners have an opportunity to publicly demonstrate their new knowledge or skill?

**Total integration score (0-10) =** 

Appendix D. Instrument Item ICC

ICC was calculated for each of the items on the M-5 Star, as well as the Texas IQ, WebCT, and SREB. As was mentioned earlier, intraclass correlation (ICC) is interpreted similarly to kappa (Garson, 2010b). By convention, a kappa > .70 is considered acceptable, but this depends highly on the researcher's purpose. For our exploratory purposes K < 0 is poor agreement, K < .20 is slight agreement, K = .20 to .39 is fair agreement, K = .40 to .59 is moderate agreement, K = .60 to .79 is substantial, and K = .80 to 1.0 is outstanding (Garson, 2010b; Landis & Koch, 1977).

M-5 Star Item Interrater Reliability. The ICC is shown for each item of the M-5 Star in the second column in Table 22. The model used for the ICC analysis was two-way random because one pair of raters was randomly selected from class members, although the baseline raters were fixed. Nineteen items that are not shown (4, 5, 11, 13, 14, 21, 22, 23, 24, 25, 29, 33, 34, 35, 48, 49, 53, 54, and 55) have no ICC scores. Scholfield explains that sometimes SPSS fails to perform a procedure because of zero variance between ratings (Scholfield, 2010). Many of the items for which ICC could not be calculated had been scored very low on all of the courses, therefore had zero variance. These items likely received extremely low ratings across all of the courses because none of the courses met the item criteria. For example, the M-5 Star item 4, "Is the whole task demonstrated, is the demonstration consistent with the type of learning involved?" had little to no implementation across the courses, received a majority of zero ratings, thus had zero variance. Although raters consistently gave the courses low ratings for the item, and the ICC would be assumed high, zero variance occurred, so ICC could not be calculated.

Thirteen items (10, 44, 20, 51, 52, 31, 12, 50, 26, 27, 32, 28, and 61) had negative (poor) ICC (ICC < 0). Reliability experts claim that negative ICC estimates indicate that the true interrater reliability is low (Taylor, 2009; Varnell, Murray, Janega, & Blitstein, 2004). A negative ICC occurs when between-group variation is less than within-group variation, indicating some third (control) variable has introduced nonrandom effects on the different groups (Linkov, Lovalekar, & Laporte, 2007; Taylor, 2009). According to Garson, who cites Magnusson (Magnusson, 1967), ICC is especially susceptible to random disturbance in the data, leading to negative ICC, when true reliability approaches zero and sample size is small (Garson, 2010b).

Negative reliability coefficients can also occur because of some type of error.

Possible errors include data entry error, a data measurement problem, or a problem based on small sample or indicative of multidimensionality (Garson, 2010a). As negative reliability is rare, the data were thoroughly proofread for coding or data entry errors; none were found. A data measurement problem was unlikely due to raters being well prepared. In addition, raters made strong efforts to give objective and honest individual ratings, and were not easily swayed when reconciling ratings with their partners.

Table 22

ICC Measures of Reliability for Each Item on the M-5 Star.

			95% Confidence Interval		vith True ue 0
	Intraclass Correlation	Lower Bound	Upper Bound	Value	Sig
10	29	76	.38	.56	.803
44	22	72	.44	.64	.740
20	16	70	.49	.72	.683
51	15	69	.50	.74	.669
52	12	67	.52	.79	.637
31	12	67	.52	.79	.635
12	11	67	.53	.80	.627
50	11	67	.53	.80	.627
26	07	65	.55	.86	.585
27	06	64	.56	.89	.570
32	06	64	.56	.89	.568
28	05	63	.57	.91	.554
61	01	61	.59	.97	.517
30	.02	59	.61	1.04	.475
43	.16	49	.69	1.37	.321
56	.16	49	.69	1.38	.319
46	.17	48	.70	1.41	.308
19	.21	45	.72	1.52	.272
9	.31	36	.77	1.89	.178
45	.36	31	.79	2.12	.140
38	.36	30	.79	2.15	.135
62	.40	26	.81	2.36	.108
57	.44	22	.82	2.55	.090
39	.44	22	.82	2.57	.088
18	.44	22	.82	2.59	.086
42	.45	20	.83	2.65	.081
8	.53	11	.86	3.22	.048
41	.53	10	.86	3.29	.046
63	.53	10	.86	3.29	.045
15	.61	.02	.89	4.19	.022
6	.63	.04	.89	4.34	.02

*Note.* A two-way random effects model was used, where people effects are random and measures effects are random. Nineteen items with no ICC coefficient because of zero variance were excluded from the table. Degrees of freedom (1, 2) for all items = 9.

(continued)

			95% Confidence Interval		vith True ue 0
	Intraclass Correlation	Lower Bound	Upper Bound	Value	Sig
7	.63	.05	.89	4.41	.019
1	.65	.09	.90	4.79	.014
47	.66	.10	.90	4.97	.013
58	.68	.13	.91	5.25	.011
16	.76	.29	.93	7.36	.003
60	.80	.38	.95	9.00	.002
2	.87	.56	.97	14.39	0
36	.88	.58	.97	15.23	.001
59	.88	.60	.97	16.11	.001
3	.90	.66	.97	19.97	0
40	.90	.67	.98	20.07	.001
37	.91	.67	.98	20.75	.001
17	.92	.70	.98	23.05	.001

*Note.* A two-way random effects model was used, where people effects are random and measures effects are random. Nineteen items with no ICC coefficient because of zero variance were excluded from the table. Degrees of freedom (1, 2) for all items = 9.

The instruments are not multidimensional, so it is unlikely that multidimensionality is a cause of the negative reliability coefficients. The negative reliability coefficients were more likely a result of small sample size. With only six courses evaluated, the small sample size may have contributed to the negative reliability coefficients.

Of the items with positive ICC, four (30, 43, 56, and 46) had slight ICC (ICC = .0 - .2), four (19, 9, 45, and 38) had fair ICC (ICC = .2 - .4), eight (62, 57, 39, 18, 42, 8, 41, and 63) had moderate ICC (ICC = .4 - .6), seven (15, 6, 7, 1, 47, 58, and 16) had substantial ICC (ICC = .6 - .8), and eight (60, 2, 36, 59, 3, 40, 37, and 17) had exceptional ICC (ICC = .8 - 1.0). The top 22 items had ICCs that were significant at the .10 level or

better. The item ICC results are questionable because the reconciling of rater-pair scores causes bias towards rating consistency, and inflates interrater reliability. This will be a recurrent theme throughout the interrater reliability results for all instruments.

Items with negative or low ICC do not contribute to the discrimination between good and bad courses. Because items cannot be trusted to get consistent ratings from different raters, they cannot be trusted for accuracy in judging courses. Items may have negative or low ICC because raters were inaccurate, the various courses were difficult to rate for the specific items, items were of poor quality, items were difficult to judge because of ambiguity or complexity, or training was insufficient for difficult items.

Further research would need to be done to determine the cause of the negative (poor) ICCs, then to identify the appropriate solution. Possible solutions include revising items, providing additional training for items, or eliminating items.

Note that individual items by themselves do not provide information about course quality. Only as items are considered as a group can we get information about overall course quality. Similarly, singular item ICCs do not provide information about ICC across courses or for an instrument. Only as item ICCs are considered as a whole do we obtain evidence of interrater reliability across courses or for any instrument.

Texas IQ Item Interrater Reliability. The ICC is shown for each item of the Texas IQ instrument in the second column in Table 23. Two items could not have ICC calculated because of zero variance, but likely had similar ratings. Fourteen items (36, 59, 6, 41, 9, 62, 49, 23, 56, 31, 1, 57, 10, and 4) had negative ICC (ICC < 0) likely a combination of the low ICC and the small sample size.

Table 23
Single-Measures Reliability for Each Item on the Texas IQ

-	_		95% Confidence Interval		vith True ue 0
	Intraclass	Lower	Upper		
Item	Correlation	Bound	Bound	Value	Sig
12					
22					
36	62	94	.26	.24	.930
59	36	88	.54	.47	.789
6	34	87	.56	.50	.770
41	29	86	.59	.55	.738
9	29	86	.60	.55	.734
62	25	84	.62	.60	.702
49	20	83	.65	.66	.669
23	19	83	.66	.67	.662
56	19	83	.66	.68	.656
31	16	81	.68	.73	.630
1	14	81	.68	.74	.623
57	08	79	.72	.85	.567
10	08	79	.72	.85	.566
4	06	78	.73	.89	.550
44	.00	75	.75	1.00	.500
60	.02	74	.76	1.05	.479
42	.03	74	.77	1.06	.477
5	.06	73	.78	1.12	.450
58	.06	73	.78	1.13	.448
32	.12	70	.80	1.28	.397
34	.15	68	.81	1.35	.374
46	.17	67	.82	1.42	.356
63	.23	63	.84	1.60	.310
13	.24	63	.84	1.64	.300
25	.25	62	.84	1.67	.293
37	.26	62	.85	1.70	.287
28	.29	60	.86	1.80	.266
27	.32	57	.86	1.93	.244
3	.33	56	.87	1.99	.234
26	.34	56	.87	2.05	.225
24	.35	55	.87	2.09	.219

*Note.* A two-way random effects model was used, where people effects are random and measures effects are random. Degrees of Freedom (1, 2) = 5.

(continued)

Item         Intraclass Correlation         Lower Bound         Upper Bound         Value         Sig           43         .37        54         .88         2.16         .209           19         .39        51         .88         2.30         .191           47         .41        50         .89         2.41         .178           7         .42        49         .89         2.43         .176           30         .42        49         .89         2.46         .173           2         .47        44         .90         2.79         .143           14         .47        44         .90         2.79         .142           11         .49        42         .91         2.92         .133           33         .49        42         .91         2.93         .132           55         .49        42         .91         2.93         .132           39         .51        39         .91         3.10         .120           50         .52        38         .92         3.17         .115           40         .52        38         .92			95% Con			vith True
Item         Correlation         Bound         Bound         Value         Sig           43         .37        54         .88         2.16         .209           19         .39        51         .88         2.30         .191           47         .41        50         .89         2.41         .178           7         .42        49         .89         2.43         .176           30         .42        49         .89         2.46         .173           2         .47        44         .90         2.79         .143           14         .47        44         .90         2.79         .142           11         .49        42         .91         2.92         .133           33         .49        42         .91         2.93         .132           55         .49        42         .91         2.93         .132           39         .51        39         .91         3.10         .120           50         .52        38         .92         3.17         .115           40         .52        38         .92         3.19		Introdece			vali	ue u
43         .37        54         .88         2.16         .209           19         .39        51         .88         2.30         .191           47         .41        50         .89         2.41         .178           7         .42        49         .89         2.43         .176           30         .42        49         .89         2.46         .173           2         .47        44         .90         2.79         .143           14         .47        44         .90         2.79         .142           11         .49        42         .91         2.92         .133           33         .49        42         .91         2.93         .132           55         .49        42         .91         2.93         .132           39         .51        39         .91         3.10         .120           50         .52        38         .92         3.17         .115           40         .52        38         .92         3.19         .115           40         .52        38         .92         3.20         .113 <td>Item</td> <td></td> <td></td> <td></td> <td>Value</td> <td>Sia</td>	Item				Value	Sia
19       .39      51       .88       2.30       .191         47       .41      50       .89       2.41       .178         7       .42      49       .89       2.43       .176         30       .42      49       .89       2.46       .173         2       .47      44       .90       2.79       .143         14       .47      44       .90       2.79       .142         11       .49      42       .91       2.92       .133         33       .49      42       .91       2.93       .132         55       .49      42       .91       2.93       .132         39       .51      39       .91       3.10       .120         50       .52      38       .92       3.17       .115         51       .52      38       .92       3.19       .115         40       .52      38       .92       3.40       .103         29       .57      32       .93       3.67       .090         38       .58      31       .93       3.76       .086						
47         .41        50         .89         2.41         .178           7         .42        49         .89         2.43         .176           30         .42        49         .89         2.46         .173           2         .47        44         .90         2.79         .143           14         .47        44         .90         2.79         .142           11         .49        42         .91         2.92         .133           33         .49        42         .91         2.93         .132           55         .49        42         .91         2.93         .132           39         .51        39         .91         3.10         .120           50         .52        38         .92         3.17         .115           51         .52        38         .92         3.19         .115           40         .52        38         .92         3.40         .103           29         .57        32         .93         3.67         .090           38         .58        31         .93         3.76         .086 <td></td> <td></td> <td></td> <td></td> <td></td> <td></td>						
7       .42      49       .89       2.43       .176         30       .42      49       .89       2.46       .173         2       .47      44       .90       2.79       .143         14       .47      44       .90       2.79       .142         11       .49      42       .91       2.92       .133         33       .49      42       .91       2.93       .132         55       .49      42       .91       2.93       .132         39       .51      39       .91       3.10       .120         50       .52      38       .92       3.17       .115         51       .52      38       .92       3.19       .115         40       .52      38       .92       3.20       .113         21       .54      35       .92       3.40       .103         29       .57      32       .93       3.67       .090         38       .58      31       .93       3.76       .086         52       .59      29       .93       3.89       .081						
2       .47      44       .90       2.79       .143         14       .47      44       .90       2.79       .142         11       .49      42       .91       2.92       .133         33       .49      42       .91       2.93       .132         55       .49      42       .91       2.93       .132         39       .51      39       .91       3.10       .120         50       .52      38       .92       3.17       .115         51       .52      38       .92       3.19       .115         40       .52      38       .92       3.20       .113         21       .54      35       .92       3.40       .103         29       .57      32       .93       3.67       .090         38       .58      31       .93       3.76       .086         52       .59      29       .93       3.89       .081         48       .59      29       .93       3.94       .079         54       .62      24       .94       4.34       .067						
14       .47      44       .90       2.79       .142         11       .49      42       .91       2.92       .133         33       .49      42       .91       2.93       .132         55       .49      42       .91       2.93       .132         39       .51      39       .91       3.10       .120         50       .52      38       .92       3.17       .115         51       .52      38       .92       3.19       .115         40       .52      38       .92       3.20       .113         21       .54      35       .92       3.40       .103         29       .57      32       .93       3.67       .090         38       .58      31       .93       3.76       .086         52       .59      29       .93       3.89       .081         48       .59      29       .93       3.94       .079         54       .62      24       .94       4.34       .067         20       .63      24       .94       4.36       .066	30	.42	49	.89	2.46	.173
14       .47      44       .90       2.79       .142         11       .49      42       .91       2.92       .133         33       .49      42       .91       2.93       .132         55       .49      42       .91       2.93       .132         39       .51      39       .91       3.10       .120         50       .52      38       .92       3.17       .115         51       .52      38       .92       3.19       .115         40       .52      38       .92       3.20       .113         21       .54      35       .92       3.40       .103         29       .57      32       .93       3.67       .090         38       .58      31       .93       3.76       .086         52       .59      29       .93       3.89       .081         48       .59      29       .93       3.94       .079         54       .62      24       .94       4.34       .067         20       .63      24       .94       4.36       .066	2	.47	44	.90	2.79	.143
33         .49        42         .91         2.93         .132           55         .49        42         .91         2.93         .132           39         .51        39         .91         3.10         .120           50         .52        38         .92         3.17         .115           51         .52        38         .92         3.19         .115           40         .52        38         .92         3.20         .113           21         .54        35         .92         3.40         .103           29         .57        32         .93         3.67         .090           38         .58        31         .93         3.76         .086           52         .59        29         .93         3.89         .081           48         .59        29         .93         3.94         .079           54         .62        24         .94         4.34         .067           20         .63        24         .94         4.36         .066           61         .67        18         .94         4.91         .053 </td <td>14</td> <td>.47</td> <td>44</td> <td>.90</td> <td></td> <td>.142</td>	14	.47	44	.90		.142
55         .49        42         .91         2.93         .132           39         .51        39         .91         3.10         .120           50         .52        38         .92         3.17         .115           51         .52        38         .92         3.19         .115           40         .52        38         .92         3.20         .113           21         .54        35         .92         3.40         .103           29         .57        32         .93         3.67         .090           38         .58        31         .93         3.76         .086           52         .59        29         .93         3.89         .081           48         .59        29         .93         3.94         .079           54         .62        24         .94         4.34         .067           20         .63        24         .94         4.36         .066           61         .67        18         .94         4.91         .053           53         .68        15         .95         5.28         .046 </td <td>11</td> <td>.49</td> <td>42</td> <td>.91</td> <td>2.92</td> <td>.133</td>	11	.49	42	.91	2.92	.133
39       .51      39       .91       3.10       .120         50       .52      38       .92       3.17       .115         51       .52      38       .92       3.19       .115         40       .52      38       .92       3.20       .113         21       .54      35       .92       3.40       .103         29       .57      32       .93       3.67       .090         38       .58      31       .93       3.76       .086         52       .59      29       .93       3.89       .081         48       .59      29       .93       3.94       .079         54       .62      24       .94       4.34       .067         20       .63      24       .94       4.36       .066         61       .67      18       .94       4.91       .053         53       .68      15       .95       5.28       .046         15       .74      03       .95       6.74       .028         17       .75      015       .96       6.93       .027	33	.49	42	.91	2.93	.132
50         .52        38         .92         3.17         .115           51         .52        38         .92         3.19         .115           40         .52        38         .92         3.20         .113           21         .54        35         .92         3.40         .103           29         .57        32         .93         3.67         .090           38         .58        31         .93         3.76         .086           52         .59        29         .93         3.89         .081           48         .59        29         .93         3.94         .079           54         .62        24         .94         4.34         .067           20         .63        24         .94         4.36         .066           61         .67        18         .94         4.91         .053           53         .68        15         .95         5.28         .046           15         .74        03         .95         6.74         .028           17         .75        015         .96         6.93         .027<	55	.49	42	.91	2.93	.132
51       .52      38       .92       3.19       .115         40       .52      38       .92       3.20       .113         21       .54      35       .92       3.40       .103         29       .57      32       .93       3.67       .090         38       .58      31       .93       3.76       .086         52       .59      29       .93       3.89       .081         48       .59      29       .93       3.94       .079         54       .62      24       .94       4.34       .067         20       .63      24       .94       4.36       .066         61       .67      18       .94       4.91       .053         53       .68      15       .95       5.28       .046         15       .74      03       .95       6.74       .028         17       .75      015       .96       6.93       .027         45       .76       .01       .96       7.25       .024         18       .80       .12       .97       9.12       .015	39	.51	39	.91	3.10	.120
40       .52      38       .92       3.20       .113         21       .54      35       .92       3.40       .103         29       .57      32       .93       3.67       .090         38       .58      31       .93       3.76       .086         52       .59      29       .93       3.89       .081         48       .59      29       .93       3.94       .079         54       .62      24       .94       4.34       .067         20       .63      24       .94       4.36       .066         61       .67      18       .94       4.91       .053         53       .68      15       .95       5.28       .046         15       .74      03       .95       6.74       .028         17       .75      015       .96       6.93       .027         45       .76       .01       .96       7.25       .024         18       .80       .12       .97       9.12       .015         35       .80       .13       .97       9.26       .014	50	.52	38	.92	3.17	.115
21       .54      35       .92       3.40       .103         29       .57      32       .93       3.67       .090         38       .58      31       .93       3.76       .086         52       .59      29       .93       3.89       .081         48       .59      29       .93       3.94       .079         54       .62      24       .94       4.34       .067         20       .63      24       .94       4.36       .066         61       .67      18       .94       4.91       .053         53       .68      15       .95       5.28       .046         15       .74      03       .95       6.74       .028         17       .75      015       .96       6.93       .027         45       .76       .01       .96       7.25       .024         18       .80       .12       .97       9.12       .015         35       .80       .13       .97       9.26       .014         8       .85       .26       .98       12.29       .008   <	51	.52	38	.92	3.19	.115
29       .57      32       .93       3.67       .090         38       .58      31       .93       3.76       .086         52       .59      29       .93       3.89       .081         48       .59      29       .93       3.94       .079         54       .62      24       .94       4.34       .067         20       .63      24       .94       4.36       .066         61       .67      18       .94       4.91       .053         53       .68      15       .95       5.28       .046         15       .74      03       .95       6.74       .028         17       .75      015       .96       6.93       .027         45       .76       .01       .96       7.25       .024         18       .80       .12       .97       9.12       .015         35       .80       .13       .97       9.26       .014         8       .85       .26       .98       12.29       .008	40	.52	38	.92	3.20	.113
38         .58        31         .93         3.76         .086           52         .59        29         .93         3.89         .081           48         .59        29         .93         3.94         .079           54         .62        24         .94         4.34         .067           20         .63        24         .94         4.36         .066           61         .67        18         .94         4.91         .053           53         .68        15         .95         5.28         .046           15         .74        03         .95         6.74         .028           17         .75        015         .96         6.93         .027           45         .76         .01         .96         7.25         .024           18         .80         .12         .97         9.12         .015           35         .80         .13         .97         9.26         .014           8         .85         .26         .98         12.29         .008	21	.54	35	.92	3.40	.103
52       .59      29       .93       3.89       .081         48       .59      29       .93       3.94       .079         54       .62      24       .94       4.34       .067         20       .63      24       .94       4.36       .066         61       .67      18       .94       4.91       .053         53       .68      15       .95       5.28       .046         15       .74      03       .95       6.74       .028         17       .75      015       .96       6.93       .027         45       .76       .01       .96       7.25       .024         18       .80       .12       .97       9.12       .015         35       .80       .13       .97       9.26       .014         8       .85       .26       .98       12.29       .008	29	.57	32	.93	3.67	.090
48       .59      29       .93       3.94       .079         54       .62      24       .94       4.34       .067         20       .63      24       .94       4.36       .066         61       .67      18       .94       4.91       .053         53       .68      15       .95       5.28       .046         15       .74      03       .95       6.74       .028         17       .75      015       .96       6.93       .027         45       .76       .01       .96       7.25       .024         18       .80       .12       .97       9.12       .015         35       .80       .13       .97       9.26       .014         8       .85       .26       .98       12.29       .008	38	.58	31	.93	3.76	.086
54       .62      24       .94       4.34       .067         20       .63      24       .94       4.36       .066         61       .67      18       .94       4.91       .053         53       .68      15       .95       5.28       .046         15       .74      03       .95       6.74       .028         17       .75      015       .96       6.93       .027         45       .76       .01       .96       7.25       .024         18       .80       .12       .97       9.12       .015         35       .80       .13       .97       9.26       .014         8       .85       .26       .98       12.29       .008	52	.59	29	.93	3.89	.081
20       .63      24       .94       4.36       .066         61       .67      18       .94       4.91       .053         53       .68      15       .95       5.28       .046         15       .74      03       .95       6.74       .028         17       .75      015       .96       6.93       .027         45       .76       .01       .96       7.25       .024         18       .80       .12       .97       9.12       .015         35       .80       .13       .97       9.26       .014         8       .85       .26       .98       12.29       .008	48	.59	29	.93	3.94	.079
61       .67      18       .94       4.91       .053         53       .68      15       .95       5.28       .046         15       .74      03       .95       6.74       .028         17       .75      015       .96       6.93       .027         45       .76       .01       .96       7.25       .024         18       .80       .12       .97       9.12       .015         35       .80       .13       .97       9.26       .014         8       .85       .26       .98       12.29       .008	54	.62	24	.94	4.34	.067
53     .68    15     .95     5.28     .046       15     .74    03     .95     6.74     .028       17     .75    015     .96     6.93     .027       45     .76     .01     .96     7.25     .024       18     .80     .12     .97     9.12     .015       35     .80     .13     .97     9.26     .014       8     .85     .26     .98     12.29     .008	20	.63	24	.94	4.36	.066
15     .74    03     .95     6.74     .028       17     .75    015     .96     6.93     .027       45     .76     .01     .96     7.25     .024       18     .80     .12     .97     9.12     .015       35     .80     .13     .97     9.26     .014       8     .85     .26     .98     12.29     .008	61	.67	18	.94	4.91	.053
17     .75    015     .96     6.93     .027       45     .76     .01     .96     7.25     .024       18     .80     .12     .97     9.12     .015       35     .80     .13     .97     9.26     .014       8     .85     .26     .98     12.29     .008	53	.68	15	.95	5.28	.046
45       .76       .01       .96       7.25       .024         18       .80       .12       .97       9.12       .015         35       .80       .13       .97       9.26       .014         8       .85       .26       .98       12.29       .008	15	.74	03	.95	6.74	.028
18     .80     .12     .97     9.12     .015       35     .80     .13     .97     9.26     .014       8     .85     .26     .98     12.29     .008	17	.75	015	.96	6.93	.027
35       .80       .13       .97       9.26       .014         8       .85       .26       .98       12.29       .008	45	.76	.01	.96	7.25	.024
8 .85 .26 .98 12.29 .008	18	.80	.12	.97	9.12	.015
	35	.80	.13	.97	9.26	.014
16 .90 .44 .98 18.33 .003	8	.85	.26	.98	12.29	.008
	16	.90	.44	.98	18.33	.003

Note. A two-way random effects model was used, where people effects are random and measures effects are random. Degrees of Freedom (1, 2) = 5.

Of the items with positive ICC, eight ( 44, 60, 42, 5, 58, 32, 34, 46) had slight ICC (ICC = .0 - .2), 11 (63, 13, 25, 37, 28, 27, 3, 26, 24, 43, and 19) had fair ICC (ICC = .2 - .4), 17 ( 47, 7, 30, 2, 14, 11, 33, 55, 39, 50, 51, 40, 21, 29, 38, 52, and 48) had moderate ICC (ICC = .4 -.6), seven (54, 20, 61, 53, 15, 18, and 45) had substantial ICC (ICC = .6 - .8), and four (18, 35, 8, and 16) had exceptional ICC (ICC = .8 - 1.0). Only the fifteen items with ICCs above .57 were significant at the .10 level.

WebCT Item Interrater Reliability. The ICC is shown for each item of the WebCT in the second column in Table 24. Note that items 9 and 21 had no ICC scores. Three items had negative (poor) ICC (1, 25, and 16). Five items (29, 19, 15, 6, and 30) had slight ICC (ICC = .0 - .2). Four items (20, 23, 7, and 11) had fair ICC (ICC = .2 - .4). Seven (28, 4, 12, 32, 10, 14, and 18) had moderate ICC (ICC = .4 - .6). Four (5, 8, 27 and 26) had substantial ICC (ICC = .6 - .8), and seven (2, 22, 3, 24, 13, 31, 17) had exceptional ICC (ICC = .8 - 1.0). The items with the top 12 ICCs (≥ .57) were significant at the .10 level.

**SREB Item Interrater Reliability.** Table 25 shows the ICC coefficients for each of the items on the SREB. Of the 17 items, item 4 could not have ICC calculated. Seven items (16, 5, 1, 15, 6, 9, and 11) had negative (poor) ICC (ICC < 0)., Two (7 and 3) had slight ICC (ICC = .0 - .2), two (14 and 13) had fair ICC (ICC = .2 - .4), two (2 and 8) had moderate ICC (ICC = .4 -.6), one (12) had substantial ICC (ICC = .6 - .8), and two (17 and 10) had exceptional ICC (ICC = .8 - 1.0). Only the top three items had ICCs that were significant at the .10 level.

Table 24
Single-Measures Reliability for Each Item on the WebCT

			95% Confidence Interval		ith True e 0
	Intraclass	Lower	Upper		
Item	Correlation	Bound	Bound	Value	Sig
9					
21					
1	26	85	.62	.59	.712
25	12	80	.70	.78	.604
16	09	79	.71	.83	.579
29	.01	87	.88	1.03	.491
19	.06	73	.78	1.13	.448
15	.08	72	.79	1.17	.434
6	.09	71	.79	1.21	.421
30	.12	77	.85	1.27	.410
20	.20	65	.83	1.50	.334
23	.24	63	.84	1.64	.300
7	.31	58	.86	1.91	.247
11	.35	64	.90	2.09	.247
28	.40	89	.98	2.33	.300
4	.41	49	.89	2.41	.178
12	.45	46	.90	2.62	.157
32	.48	42	.91	2.89	.135
10	.49	42	.91	2.89	.135
14	.54	48	.94	3.39	.132
18	.57	32	.93	3.70	.089
5	.64	22	.94	4.56	.061
8	.69	14	.95	5.43	.043
27	.71	09	.95	5.95	.036
26	.78	09	.97	8.07	.034
2	.84	.25	.98	11.91	.008
22	.85	.27	.98	12.37	.008
3	.87	.33	.98	14.19	.006
24	.87	.34	.98	14.67	.005
13	.88	.39	.98	16.30	.004
31	.94	11	.10	31.00	.031
17	.95	.69	.99	38.79	.001

*Note.* A two-way random effects model was used, where people effects are random and measures effects are random.

Table 25
Single-Measures Reliability for Each Item on the SREB

	_	95% Confidence Interval		F Te	est with T	rue Value	e 0
Item	Intraclass Correlation	Lower Bound	Upper Bound	Value	df1	df2	Sig
4							
16	79	97	09	.12	5.0	5	.983
5	40	91	.61	.43	4.0	4	.784
1	31	86	.58	.53	5.0	5	.749
15	11	80	.70	.81	5.0	5	.589
6	09	79	.71	.83	5.0	5	.578
9	08	79	.72	.86	5.0	5	.565
11	06	78	.73	.88	5.0	5	.552
7	.03	74	.77	1.06	5.0	5	.475
3	.09	78	.84	1.19	4.0	4	.435
14	.23	63	.84	1.60	5.0	5	.310
13	.33	56	.87	2.00	5.0	5	.233
2	.43	58	.92	2.53	4.0	4	.195
8	.47	44	.90	2.80	5.0	5	.141
12	.75	.00	.96	7.17	5.0	5	.025
17	.87	.21	.99	14.62	4.0	4	.012
10	.89	.41	.98	17.25	5.0	5	.004

*Note.* A two-way random effects model was used, where people effects are random and measures effects are random.

Appendix E. Merrill's First Principles Domain Outline

Merrill's First Principles of Instruction domain outline was developed based upon a number of Merrill's First Principles papers which were written or published prior to completion of his 2007 5 Star instrument.

(Learning is enhanced when...)

### First Principles of Instruction Domain Topic

### Task-Centered

- 1 The courseware is presented in the context of meaningful real world tasks/problems.
  - 1.1 The courseware shows learners the whole task they will be able to do or the problem they will be able to solve as a result of completing a module or course.
  - 1.2 Students are engaged at the whole problem/task level not just the operation or action levels.
  - 1.3 The courseware involves a progression of real-world whole tasks/problems rather than a single whole task/problem.
    - 1.3.1 There is a complete worked demonstration/example of a whole-task. The demonstration is consistent with each of the types of learning involved.
      - 1.3.1.1 The component skills of the whole task are demonstrated to learners in the context of the whole task using a task-centered instructional strategy.
      - 1.3.1.2 There is a progression of at least 3 increasingly difficult whole tasks or problems demonstrated to the learners.
      - 1.3.1.3 Guidance during demonstrations shows learners how the steps in the whole task relate to an organizing structure?
    - 1.3.2 There is a progression of successively complex whole-task worked examples for guided practice (with successively decreasing guidance). The application is consistent with each of the types of learning involved.
      - 1.3.2.1 Coaching is provided for problems early in the progression and gradually withdrawn as the progression continues.
      - 1.3.2.2 Learners are able to receive intrinsic feedback on their performance by seeing the consequences of their activities
      - 1.3.2.3 Learners are required to apply the organizing structure to at least the early instances in the progression.
      - 1.3.2.4 Learners have to apply the component skills to the completion of a new whole task or problem.
      - 1.3.2.5 The application allows for peer-collaboration and peer-critique.

"Information-about" Presentation

- 2 The courseware provides names, descriptions and portrayals of features and context for a set of entities.
  - 2.1 The presentation provides the name, description and portrayal (examples) of the information elements, item by item.
  - 2.2 Learners can control pacing.
  - 2.3 Learners can repeat the presentation in any order and as many times as desired.
  - 2.4 The presentation uses appropriate media for the information being taught.

#### "Information-about" Practice

- 3 Learners are asked to provide names given descriptions or portrayals, or descriptions or portrayals given the names for the set of entities.
  - 3.1 Learners are asked to provide the name given the description or portrayal, and give the description or portrayal given the name.
  - 3.2 There is immediate, corrective feedback.
  - 3.3 A no-delay response is required.
  - 3.4 Practice provides for repetition, provides for performance scores, and encourages improvement in the score.
  - 3.5 The practice uses appropriate media for the information and portrayals.

#### "Parts-of" Presentation

- 4 The courseware provides the name and description of the parts.
  - 4.1 The system provides the name or description when the learners click on each part.
  - 4.2 The parts are clustered in chunks of 7 or fewer parts.
  - 4.3 The presentation repeats the exploration until learners are able to locate each part.
  - 4.4 The presentation avoids locations cues.
  - 4.5 The presentation uses appropriate media for the parts being taught.

#### "Parts-of" Practice

- 5 Practice requires learners to identify the part given the name or provide the name and description given the part.
  - 5.1 Practice requires learners to point to the location of the part given the name or information, or recall the name or description given the location of the part.
  - 5.2 The practice provides immediate corrective feedback.
  - 5.3 The practice requires an immediate response.
  - 5.4 The parts are presented for practice in random order or the practice avoids location cues.
  - 5.5 The practice requires a 100% criterion.
  - 5.6 The practice uses appropriate media for the parts being learned.

### Demonstration

6 The courseware provides a demonstration of the skills to be learned.

- 6.1 The demonstration provides information which is general, inclusive and applicable to many situations.
- 6.2 The demonstration also provides portrayals (examples) of what is to be learned. Portrayals are specific, limited, and applicable to one case or a single situation.
- 6.3 The demonstration should be organized around the organizing structure of the whole task in a manner that it will help build the learner's schema of the task.
- 6.4 The demonstrations are consistent with the content being taught.
  - 6.4.1 Examples and non-examples are provided for concept classification (kinds-of).
    - 6.4.1.1 Kinds-of demonstration tells the name and definition of each category.
    - 6.4.1.2 Kinds-of demonstration shows learners examples of each category.
    - 6.4.1.3 Kinds-of demonstration provides guidance by highlighting discriminating properties or by showing matched examples among categories.
    - 6.4.1.4 Kinds-of demonstration includes at least 3 examples from each category.
    - 6.4.1.5 Guidance is provided during demonstrations showing learners how the defining properties and portrayals relate to an organizing structure.
    - 6.4.1.6 Kinds-of demonstration allows peer-demonstration of new examples.
  - 6.4.2 Procedural demonstrations are provided for procedures (how-to).
    - 6.4.2.1 How-to demonstration tells learners the steps and sequence in the procedure.
    - 6.4.2.2 How-to demonstration shows a specific instance of the task and demonstrates each of the steps required to complete the task.
    - 6.4.2.3 How-to demonstration provides guidance by calling attention to the execution of each step.
    - 6.4.2.4 How-to demonstration shows the consequence of each step, focusing the learner's attention on the portrayal of the consequence, especially if the consequence is hidden from view or not obvious.
    - 6.4.2.5 The procedure is demonstrated in a progression of at least 3 increasingly difficult situations.
    - 6.4.2.6 Guidance during demonstrations shows learners how the steps in the procedure relate to an organizing structure.
    - 6.4.2.7 The steps in the procedure and their sequence are summarized.
    - 6.4.2.8 How-to demonstration allows peer-demonstration of a new instance of the procedure.
  - 6.4.3 Visualizations are provided for processes (what-happens).

- 6.4.3.1 What-happens demonstration tells learners the conditions and consequence of the process.
- 6.4.3.2 What-happens demonstration shows the process in a specific real or simulated situation.
- 6.4.3.3 What-happens demonstration provides guidance by helping learners relate the events in the process to the conditions and consequence.
- 6.4.3.4 Demonstration of the process is repeated for a progression of at least 3 increasingly complex scenarios.
- 6.4.3.5 Guidance during demonstrations shows learners how the conditions and consequence relate to an organizing structure.
- 6.4.3.6 What-happens demonstration allows peer-demonstration of a new instance of the process.
- 6.4.4 Modeling is provided for behavior.
- 6.5 Media that are relevant to the content are used to enhance learning.

#### Application

- 7 Learners have an opportunity to practice and apply their newly acquired knowledge or skill.
  - 7.1 Opportunities for application are provided within the context of the organizing structure of the real-world tasks.
  - 7.2 The application (practice) and the posttest are consistent with the stated or implied objectives.
    - 7.2.1 Kinds-of (concept classification) practice requires learners to identify or classify new examples of each kind.
      - 7.2.1.1 Learners are asked to remember the definition.
      - 7.2.1.2 Learners are required to classify new examples.
      - 7.2.1.3 The application provides corrective feedback that focuses learners' attention on discriminating properties.
      - 7.2.1.4 Learners are required to classify a series of 3 or more divergent examples.
      - 7.2.1.5 Coaching Is provided for problems early in the progression and gradually withdrawn as the progression continues.
      - 7.2.1.6 Learners are required to apply the organizing structure to at least the early instances in the progression.
      - 7.2.1.7 The application allows for peer-collaboration and peer-critique.
    - 7.2.2 How-to (procedural) practice requires learners to do the procedure.
      - 7.2.2.1 Learners are required to remember the steps in the sequence.
      - 7.2.2.2 Learners are required to do the task by executing each step in a real or simulated situation.
      - 7.2.2.3 The application provides intrinsic feedback and extrinsic feedback
      - 7.2.2.4 Learners are required to do a progression of at least 3 simple to complex tasks.

- 7.2.2.5 Coaching is provided for problems early in the progression and gradually withdrawn as the progression continues.
- 7.2.2.6 Learners are required to apply the organizing structure to at least the early instances in the progression.
- 7.2.2.7 The application allows for peer-collaboration and peer-critique.
- 7.2.3 What-happens (troubleshooting principles, processes) practice requires learners to predict a consequence of a process given conditions, or to find faulted conditions given an unexpected consequence.
  - 7.2.3.1 Learners are required to remember the conditions and consequence of the process.
  - 7.2.3.2 Learners are required to predict the consequence OR learners are required to troubleshoot an unexpected consequence in a specific situation.
  - 7.2.3.3 Learners are able to receive intrinsic feedback by being able to test their predictions or test their trouble shooting.
  - 7.2.3.4 Learners are required to make predictions or trouble shoot a series of at least 3 increasingly complex problems.
  - 7.2.3.5 Coaching is provided for problems early in the progression and gradually withdrawn as the progression continues.
  - 7.2.3.6 Learners are required to apply the organizing structure to at least the early instances in the progression.
  - 7.2.3.7 The application allows for peer-collaboration and peer-critique.
- 7.3 The courseware requires learners to use new knowledge or skill to solve a varied sequence of problems.
- 7.4 Learners observe media that is appropriately used.
- 7.5 The application allows for peer-collaboration and peer-critique.

#### Activation

- 8 The courseware attempts to activate relevant prior knowledge or experience.
  - 8.1 The courseware directs learners to recall, relate, describe, or apply knowledge from relevant past experience that can be used as a foundation for new knowledge.
  - 8.2 If learners already know some of the content they are given an opportunity to demonstrate their previously acquired knowledge or skill.
  - 8.3 Learners are provided an effective content structure or framework that is the basis for guidance during demonstration, for coaching during application, and for reflection during integration.
  - 8.4 The courseware provides relevant experience that can be used as a foundation for the new knowledge.

### Integration

9 The courseware provides techniques that encourage learners to integrate (transfer) the new knowledge or skill into their everyday life.

- 9.1 The courseware provides an opportunity for learners to reflect-on, discuss, and defend their new knowledge or skill.
- 9.2 The courseware provides an opportunity for learners to create, invent, or explore new and personal ways to use their new knowledge or skill.
- 9.3 The courseware provides an opportunity for learners to publicly demonstrate their new knowledge or skill.

Appendix F. M-5 Star Instrument CVI Study Cover Page

#### M-5 Star Instrument CVI Study Cover Page

You have been identified by Dave Merrill, Ted Frick, Joanne Bentley or myself as someone with expertise on Merrill's First Principles of Instruction. You may also have had experience with a First Principles of Instruction rating instrument.

My dissertation study is entitled "Providing Reliability and Validity Evidence for Merrill's 2007 5 Star Instrument." As part of my dissertation study, I am conducting a content validity analysis to see how well Merrill's 2007 5 Star Instrument matches a First Principles of Instruction Domain Outline that I have created. The domain outline was created through the review of Merrill's first principles papers written between 2002 and 2007. Note that some of the papers were written in 2007, but were published later.

For this evaluation would you please complete the following forms?

- 2. Rater information (2 RaterInformation.doc)
- 3. Merrill's 2007 5 Star Instrument Content Validity Rating form (3 \_5 StarContentValidity\_Rating.doc). This evaluation has you determine if Merrill's 5 Star instrument items are included in the First Principles of Instruction Domain Outline (4 FirstPrinciplesDomainOutline.doc).

Please return it by midnight Wednesday June 2 so I can include your data in my dissertation study. By completing the rating you will be entered in a raffle to possibly win \$50.00.

A video showing how to do this rating is found at <a href="http://www.youtube.com/watch?v=tGU2sPGjM3w">http://www.youtube.com/watch?v=tGU2sPGjM3w</a>. A written explanation of how to do the rating in found at the beginning of the rating form (3\_5StarContentValidity\_Rating.doc).

Please be as thorough and honest as possible in your evaluations. It is important that you rate items accurately. As you rate items, be aware that some items are not in the same order on the different forms. You may need to search for matching items.

You can enter the information in the doc or docx and return it to me electronically, or print the pdf, complete it and scan and email it o me, or mail it to me.

If you have any questions, feel free to contact me. Also, if you are interested in receiving information on the results of this study, or a copy of my dissertation when it is completed, please let me know.

Thank you very much for your help!

Max H. Cropper
Doctoral Candidate
Instructional Technology and Learning Sciences
Utah State University
156 East 100 North #2, Logan, UT 84321
Mobile: 435-232-9130 Email:maxcropper@gmail.com

Appendix G. Rater Information

As part of the cover page instructions, raters were directed to complete the following rater information form so we would know some characteristics of the raters behind the ratings.

RATER INFORMATION
Name
Email
Phone #
Enter an X in the on the left to indicate the Institution or Organization where you are currently studying/working:
FSU
IU
OU
UGA
UH
USU
Other

Enter an X in the appropriate box on the left for the designation that currently applies:

Full Professor
Associate professor
Assistant professor
PhD Completed
PhD Candidate
PhD Student
Master's Degree Completed

Master's student
College Staff Member
Professional
Other

Enter an X by the appropriate items to indicate your experience with Merrill's First Principles of Instruction.

Evaluated courses using first principles based rating
Expert on Merrill's first principles of instruction
Taken course evaluation class based on first principles of instruction
Taken course on first principles of instruction
Conducted research on first principles of instruction
Dissertation is on first principles of instruction
Master's thesis or project on first principles of instruction
Developed course(s) using 5 Star standards
Taught/teach courses using 5 Star approach
Taught course on first principles of instruction
Expert on instructional design theories, including first principles of instruction

Number of paper(s) authored or co-authored that deal with	
first principles of instruction (total papers written for	
publication)	

# Your beliefs about Merrill's First Principles of Instruction and his 5 Star Instructional Design Rating

What are your beliefs about Merrill's First Principles of Instruction?
What experiences with First Principles of Instruction have influenced your beliefs?
What experience do you have with 5 Star evaluation forms? How has that affected your attitude toward 5 Star evaluation forms and first principles?
Do you believe that Merrill's 5 Star evaluation forms accurately represent his first principles of instruction?
Have you been involved with USU 5 Star course evaluations? To what extent? How do you think this has biased you regarding Merrill's 5 Star instrument?

Appendix H. M-5 Star Instrument CVI Rating Form

This rating form is based on Merrill's 2007 M-5 Star instrument (Appendix C). It lists each of the questions from the M-5 Star instruments and adds a content validity rating scale. It will be used by the content validity raters to determine the extent to which each item on Merrill's 2007 M-5 Star instrument is included in the first principles of instruction domain outline (Appendix E).

### MERRILL'S 2007 M-5 STAR INSTRUMENT CONTENT VALIDITY RATING FORM

Compare the items in Merrill's 2007 5 Star instrument with the First Principles of Instruction Domain Outline (4\_FirstPrinciplesDomainOutline.doc) to determine the degree to which the items found in Merrill's 5 Star instrument are included in the first principles of instruction domain outline.

In the "Matching 1<sup>st</sup> Principles Domain Item(s)" column (see example below) enter the 1<sup>st</sup> Principles domain outline number(s) that match the 5 Star Instrument item.

Enter the appropriate number in the "Enter Rating (0-2)" column depending on the degree to which it is included in the in the 1<sup>st</sup> Principles domain outline.

### The 5 Star Rating Item:

- 0 = is not included in the 1st Principles of Instruction Domain Outline
- 1 = is partially included in the 1st Principles of Instruction Domain Outline
- 2 = is fully included in the 1st Principles of Instruction Domain Outline.

For example, on the following page are some fictitious 5 Star Instrument items in a sample 5 Star Instrument Content Validity Rating form. Beneath the sample 5 Star Instrument content validity rating form is a sample 1<sup>st</sup> Principles of Instruction Domain Outline with some fictitious 1<sup>st</sup> Principles of Instruction domain items.

In the "Matching 1<sup>st</sup> Principles Domain Items(s)" column of the 5 Star Instrument Content Validity Rating Form, note that the numbers of the matching items from the Fictitious 1<sup>st</sup> Principles Domain outline have been entered. In the "Enter Rating" column you will notice that:

- Zero (0) has been entered when the 5 Star instrument item is not included at all in the 1<sup>st</sup> principles domain outline.
- One (1) has been entered when the 5 Star item is partially included in the domain outline.
- Two (2) has been entered when the item is fully included in the domain outline.

**Example**: 5 Star Instrument Content Validity Rating Form

#	Fictitious 5 Star Instrument Items:	Matching 1st Principles Domain Item(s)	Enter Rating (0-2)
70	Instruction is tailored to more than one learning style.	10	1
71	Demonstration is provided for more than one learning style.	10, 10.1	1
72	Visual demonstration is provided.	10, 10.1,10.1.1	2
73	Audio is included if necessary.	10, 10.1.2	1
74	Textual explanation and graphic examples are provided.	10, 10.1.3	2
75	Demonstration includes emotional stories.		0

**Example**: 1<sup>st</sup> Principles of Instruction Domain Outline

	Fictitious First Principles Domain Items				
10	10 Instruction is tailored to multiple learning styles (3 or more)				
	10.1 Demonstration is provided for multiple learning styles (3 or more)				
		10.1.1	Visual demonstration is provided for visual learners		
	10.1.2 Audio explanation is provided for auditory learners		Audio explanation is provided for auditory learners		
		10.1.3	Textual explanation and graphics are provided for textual learners.		

As you complete the 5 Star Instrument Content Validity Rating Form, please be as careful and unbiased as possible in your evaluations. It is important that you score items as accurately as possible.

In some cases multiple domain items will somewhat match the 5 Star item. It is preferred that you list all of them. However, be certain that you include the domain item(s) that match the 5 Star item most closely.

If the 5 Star item description is fully represented in one or more or a combination of domain items, give it full credit (2). If the 5 Star item is only partially represented, even with a combination of domain items, give it partial credit (1).

Note that some items in the documents are not in the same order. So you may need to search the domain outline for matching items.

# The 5 Star Rating Item:

- 0 = is not included in the 1st Principles of Instruction Domain Outline
- 1 = is partially included in the 1st Principles of Instruction Domain Outline
- 2 = is fully included in the 1st Principles of Instruction Domain Outline

#	5 Star Rating Item	Matching 1st Principles Domain Item(s)	Enter Rating (0-2)
	Task-Centered Analysis		
1	Is there an authentic, real-world whole task?		
2	Is there a task-centered strategy?		
3	Is there a progression of whole tasks?		
4	Is the whole task demonstrated? Is the demonstration consistent with the type of learning involved?		
5	Are learners required to do the whole task? Is the application consistent with the type of learning involved?		
	Information-about Presentation – Tell		
	(Not demonstration)		
6	Does presentation tell name, information and portrayal item by item?		
7	Can learners control pacing?		
8	Can learners repeat the presentation in any order as many times as desired?		
9	Does the presentation make appropriate use of media?		
	Information-about Practice – Ask		
	(not application)		

- The 5 Star Rating Item:
  0 = is not included in the 1st Principles of Instruction Domain Outline
- 1 =
- is partially included in the 1st Principles of Instruction Domain Outline is fully included in the 1st Principles of Instruction Domain Outline 2 =

#	5 Star Rating Item	Matching 1st Principles Domain Item(s)	Enter Rating (0-2)
10	Does practice require learners to provide the name given the description or portrayal and does the practice ask for the description or portrayal given the name?		
11	Is there immediate corrective feedback?		
12	Is a no-delay response required?		
13	Does the practice provide for repetition and does it provide performance scores and encourage improvement in the score?		
14	Does the practice make appropriate use of media?		
	Parts-of Presentation – Tell (not demonstration)		
	Given an illustration of a whole object or system:		
15	Does the system tell the name or description when learners click on each part?		
16	Are parts clustered into chunks of 7 or fewer (parts of parts)?		
17	Does the presentation avoid location cues?		
18	Does the presentation repeat the exploration until learners are able to locate each part?		
19	Does the presentation make appropriate use of media?		
	Parts-of Practice – Ask (not application)		
	Given an illustration of a whole object or system:		

- The 5 Star Rating Item:
  0 = is not included in the 1st Principles of Instruction Domain Outline
- 1 =
- is partially included in the 1st Principles of Instruction Domain Outline is fully included in the 1st Principles of Instruction Domain Outline 2 =

#	5 Star Rating Item	Matching 1st Principles Domain Item(s)	Enter Rating (0-2)
20	Does practice require learners to point to the location of the part given the name or information or recall the name or description given the location of the part?		
21	Does the practice provide immediate corrective feedback?		
22	Are the parts presented for practice in random order? Or does the practice avoid location cues?		
23	Does the practice require an immediate response?		
24	Does the practice require a 100% criterion?		
25	Does the practice make appropriate use of media?		
	Kinds-of Demonstration		
26	Does the demonstration tell learners the name and definition of each category or alternative procedure?		
27	Does the demonstration show learners relevant examples of each category?		
28	Does the demonstration provide guidance by highlighting discriminating properties? By showing matched examples among categories?		
29	Does the demonstration show a series of divergent examples of increasing difficulty?		
30	Does the demonstration use media appropriately?		
	How-to Demonstration		

- The 5 Star Rating Item:
  0 = is not included in the 1st Principles of Instruction Domain Outline
- 1 =
- is partially included in the 1st Principles of Instruction Domain Outline is fully included in the 1st Principles of Instruction Domain Outline 2 =

#	5 Star Rating Item	Matching 1st Principles Domain Item(s)	Enter Rating (0-2)
31	Does the demonstration show a specific instance of the task and demonstrate each of the steps required to complete the task or provide the opportunity for learners to play with the procedure in an actual or simulated situation?		
32	Does the demonstration show the consequence of each step and the consequence of the complete procedure?		
33	Does the demonstration provide guidance by calling attention to the execution of each step?		
34	Is the procedure demonstrated in a series of increasingly difficult situations?		
35	Does the demonstration make appropriate use of media?		
	What-happens Demonstration		
36	Does the demonstration show the process in a specific real or simulated situation?		
37	Does the demonstration show a specific portrayal of the conditions necessary for each event in the process?		
38	Is the demonstration repeated for a progression of increasingly complex scenarios?		
39	Does the demonstration provide guidance by helping learners relate the events in the process to the conditions or to some conceptual structure?		
40	Does the demonstration make appropriate use of media?		

- The 5 Star Rating Item:
  0 = is not included in the 1st Principles of Instruction Domain Outline
- 1 =
- is partially included in the 1st Principles of Instruction Domain Outline is fully included in the 1st Principles of Instruction Domain Outline 2 =

#	5 Star Rating Item	Matching 1st Principles Domain Item(s)	Enter Rating (0-2)
	Kinds-of Application		
41	Does the application require learners to classify a series of new examples or select among a series of new alternative procedures.		
42	Does the application require learners to explain their classification or selection by pointing out the presence or absence of discriminating properties?		
43	Does the application provide corrective feedback that focuses learners' attention on discriminating properties? Or does the feedback show intrinsic consequences for selection of alternative procedures?		
44	Does the application provide coaching early in the sequence by helping learners see the portrayal of the discriminating attributes and gradually withdraw this coaching as the application continues?		
45	Does the application make appropriate use of media?		
	How-to Application		
46	Does the application require learners to do the task by executing each step in a real or simulated situation?		
47	Does the application provide intrinsic feedback (observing the consequence of one's actions) and extrinsic feedback (informing learners about the appropriateness of a given operation or action)?		
48	Does the application require a simple to complex progression of tasks rather than a single task?		

- The 5 Star Rating Item:
  0 = is not included in the 1st Principles of Instruction Domain Outline
- 1 =
- is partially included in the 1st Principles of Instruction Domain Outline is fully included in the 1st Principles of Instruction Domain Outline 2 =

#	5 Star Rating Item	Matching 1st Principles Domain Item(s)	Enter Rating (0-2)
49	Are tasks early in the progression coached and is this coaching gradually withdrawn as for successive tasks in the progression?		
50	Does the application make appropriate use of media?		
	What-happens Application		
51	Are learners required to predict the consequence from a given set of conditions in a specific situation? OR are learners required to troubleshoot an unexpected consequence by finding the faulted conditions in a specific situation?		
52	Are learners able to receive intrinsic feedback by being able to test their predictions or test their trouble shooting?		
53	Does the application make appropriate use of media?		
54	Are learners required to make predictions or troubleshoot a series of increasingly complex problems?		
55	Is coaching provided for problems early in the progression and gradually withdrawn as the progression continues?		
	Activation Principle		
56	Are learners prompted to recall, describe or demonstrate relevant prior knowledge or skill?		
57	Are learners provided or prompted to develop an organizing structure for the skills be obtained?		

- The 5 Star Rating Item:
  0 = is not included in the 1st Principles of Instruction Domain Outline
- 1 =
- is partially included in the 1st Principles of Instruction Domain Outline is fully included in the 1st Principles of Instruction Domain Outline 2 =

#	5 Star Rating Item	Matching 1st Principles Domain Item(s)	Enter Rating (0-2)
58	Is guidance provided during demonstration that relates the knowledge or skill being demonstrated to this structure?		
59	Is coaching provided during application that reminds learners of this structure to facilitate their application of their new knowledge or skill?		
60	Do learners have an opportunity to reflect on how the newly acquired knowledge and skill relate to the structure?		
	Integration Principle		
61	Do learners have an opportunity to reflect on, discuss or defend their new knowledge and skill?		
62	Do learners have an opportunity to create, invent, or explore personal ways to use their new knowledge or skill?		
63	Do learners have an opportunity to publicly demonstrate their new knowledge or skill?		

**Appendix I. Potential Expert Raters for CVI Rating** 

Potential expert raters include graduate students at USU who have taken online course evaluation classes and have used the M-5 Star, graduate students and PhDs at Indiana University (IU) who have worked with Ted Frick in conducting research using First Principles of Instruction rating scales, students at Florida State University (FSU) and University of Hawaii (UH) who have taken a First Principles class from Merrill, and graduate students at various universities who have developed 5 Star courses, or done research and written articles about First Principles of Instruction. Faculty who are experts on First Principles of Instruction and other instructional design theories are also included. (See Table 26.)

Table 26

Potential Expert Raters for Content Validity Rating

	•	is joing out the	Faculty	ID Theory	Course Eval	First Princi- ples	5 Star Evalu	5 Star re-	5 Star	Uses	Total 5 Star Exper
#	University	Degree	Member	Expert	Class	Class	ator	search	Pubs	5 Star	tise
FAC1	IU	PhD	1	1			1	1	1	1	6
FAC2	USU	PhD Candidate			1		1	1	1	1	5
GS1	USU	MS			1		1			1	3
GS2	USU	MS			1		1			1	3
FAC3	Vanderbilt	PhD	1	1					1		3
GS3	USU	PhD Candidate						1	1	1	3
GS4	FSU	PhD Student				1			1	1	3
STA1	USU	MS			1		1			1	3
GS5	UGA	PhD Student			1				1	1	3
FAC4	IU	PhD	1					1	1		3
FAC5	IU	PhD	1					1	1		3
FAC6	IU	PhD	1	1							2
FAC7	Maastricht	PhD	1	1							2
FAC8	UGA	PhD	1	1							2
FAC9	UGA	PhD	1			1					2
STA2	BYUI	PhD		1			1				2
GS6	USU	MS			1		1				2
GS7	USU	MS Student			1		1				2
GS8	USU				1		1				2
STA3	IU	PhD						1	1		2
STA4	IU	PhD						1	1		2
GS9	IU							1	1		2
GS10	IU							1	1		2

*Note.* FAC = Faculty Member, GS = Graduate Student, STA = Staff Member

			Faculty	ID Theory	Course Eval	First Princi- ples	5 Star Evalu	5 Star re-	5 Star	Uses	Total 5 Star Exper
#	University	Degree	Member	Expert	Class	Class	ator	search	Pubs	5 Star	tise
FAC10	BYU	PhD						1	1		2
FAC11	FSU	PhD	1			1					2
GS11	FSU	PhD	1			1					2
GS12	USU	PhD Student								1	1
GS13	IU							1			1
GS14	IU							1			1
GS15	IU	PhD						1			1
GS16	IU							1			1
GS17	IU							1			1
STA5	USU	PhD					1				1
GS18	USU	PhD Student		1							1
GS19	USU	PhD				1					1
GS20	USU	PhD				1					1
FAC12	USU	PhD							1		1
GS21	FSU	PhD Student				1					1
GS22	FSU	PhD Student				1					1
GS23	FSU	PhD Student				1					1
GS24	FSU	PhD Student				1					1
GS25	FSU	PhD Student				1					1
GS26	FSU	PhD Student				1					1
GS27	FSU	PhD Student				1					1
GS28	FSU	PhD Student				1					1

*Note.* FAC = Faculty Member, GS = Graduate Student, STA = Staff Member

(continued)

-	Heisserite	Davis	Faculty	ID Theory	Course Eval	First Princi- ples	5 Star Evalu	5 Star re-	5 Star	Uses	Total 5 Star Exper
# GS29	University FSU	Degree PhD Student	Member	Expert	Class	Class	ator	search	Pubs	5 Star	tise
						1					1
GS30	FSU	PhD Student				1					1
GS31	FSU	PhD Student				1					1
GS32	FSU	PhD Student				1					1
GS33	FSU	PhD Student				1					1
GS34	FSU	PhD Student				1					1
GS35	FSU	PhD Student				1					1
GS36	FSU	PhD Student				1					1
GS37	FSU	PhD Student				1					1
GS38	UH	PhD Student				1					1
GS39	UH	PhD Student				1					1
GS40	UH	PhD Student				1					1
GS41	UH	PhD Student				1					1
GS42	UH	PhD Student				1					1
GS43	UH	PhD Student				1					1
GS44	UH	PhD Student				1					1
GS45	UH	PhD Student				1					1
GS46	UH	PhD Student				1					1
GS47	UH	PhD Student				1					1

Note. FAC = Faculty Member, GS = Graduate Student, STA = Staff Member

Appendix J. Explanation of Low CVI Ratings

While Table 27 illustrates that the M-5 Star items generally had matching First Principles domain outline items some raters missed those items because items were worded differently, matching items were not in the same order as the instrument, and matching items were consolidated in the domain outline or in the instrument. Therefore, for five items (45, 50, 37, 29, 60) the CVI and z scores, were lower, and the p value was not significant at the alpha = .10 level. Items are listed in descending CVI rating order.

Table 27
Explanation of Low CVI Rating.

Explan	ation of Low CVI Ratir	ngs			
Item #	M-5 Star Instrument Item	1 <sup>st</sup> Principles Domain Items	CVI	z score	p value
45	Does the application make appropriate use	7.4 Learners observe media that is appropriately used.	0.605	1.079	0.1401
50	of media?  Does the application	A couple of raters totally missed domain item 7.4.	0.605	1.079	0.1401
	make appropriate use of media?	Another couple of raters gave it a partial match because of a difference in wording.			
37	Does the demonstration show a specific portrayal of the conditions	6.4.3.1 What-happens demonstration tells learners the conditions and consequence of the process.	0.605	1.079	0.1423
	necessary for each event in the process?	A number of raters missed 6.4.3.1. Some found 6.4.3.1 but gave it a partial rating perhaps because item 37 talks about showing conditions, while domain item 6.4.3.1 talks about telling the learner the conditions			
29	Does the demonstration show a series of divergent	6.4.1.2 Kinds-of demonstration shows learners examples of each category.	0.553	0.474	0.3192
	examples of increasing difficulty?	6.4.1.3 Kinds-of demonstration provides guidance by highlighting discriminating properties or by showing matched examples among categories.			
		6.4.1.4 Kinds-of demonstration includes at least 3 examples from each category.			
		The domain items mention examples which are probably divergent examples of increasing difficulty, but it is not spelled out.			

(continued)

	M-5 Star Instrument			Z	p value
Item #	Item	1 <sup>st</sup> Principles Domain Items	CVI	score	
60	Do learners have an opportunity to reflect on how the newly acquired knowledge and skill relate to the structure?	8.3 Learners are provided an effective content structure or framework that is the basis for guidance during demonstration, for coaching during application, and for reflection during integration.  A number of raters identified incorrect matching domain items, and some did not find any matching domain items, and therefore rated the item with zeroes, as not being included in the domain. Because domain item 8.3 is written opposite of item 60, and includes the use of the content structure during demonstration, application, and integration, raters must have missed the fact that the domain item does deal with reflection (of newly acquired knowledge and skill) during integration.	0.553	0.474	0.3192

Appendix K. Texas IQ Instrument

The Texas IQ rating form was developed as part of the Investigating Quality of Internet Course (IQ) Project, which was initiated under the direction of the Texas Education Agency (TEA) during the fall of 2001. The purpose of the project was to develop a tool that could be used to improve the quality of internet-based courses for Texas students.

# Cropper Adaptation of Portions of Texas Education Agency Instructional Quality (IQ) Project Evaluation Matrix

SD = Strongly Disagree, SA = Strongly Agree, NA = Not Applicable

Guideline	Item			SD	SA	NA
I. Course Components. A. Course Design						
Guideline 2: Goals and objectives are observable and measurable.	1	Are the outcomes specified in each learning objective concrete, observable, and measurable?	5= Observable and measurable goals and objectives are specified for all parts of the course 4= 75% of the goals and objectives are observable and measurable 3= 50% of the goals and objectives are observable and measurable 2= 25% of the goals and objectives are observable and measurable 1 = None of the goals and objectives are observable and measurable	12	3 4 5	NA
Guideline 3: The course includes a comprehensive online syllabus published prior	2	(a) course requirements and structure	5= The course includes detailed course requirements and structure 4= The course includes fairly detailed course requirements and structure 3= The course includes medium detailed course requirements and structure 2= The course includes undetailed course requirements and structure 1= The course includes not course requirements or structures	12	3 4 5	NA
to course delivery. Does The syllabus include: (Even though	3	(b) time requirements	5= The course includes detailed time requirements 4= The course includes fairly detailed time requirements 3= The course includes medium detailed time requirements 2= The course includes undetailed time requirements 1= The course does not include time requirements	12	3 4 5	NA
commercial courses won't have syllabuses, they may have a course	4	(c) pre-requisite requirements	5= The course includes detailed pre-requisite requirements 4= The course includes fairly detailed pre-requisite requirements 3= The course includes medium detailed pre-requisite requirements 2= The course includes undetailed pre-requisite requirements 1= The course does not include pre-requisite requirements NA = pre-requisite requirements do not exist and are not needed.	12	3 4 5	NA
overview or introduction with some of the information	5	(d) resource requirements	5= The course includes detailed resource requirements 4= The course includes fairly detailed resource requirements 3= The course includes medium detailed resource requirements 2= The course includes undetailed resource requirements	12	3 4 5	NA

Guideline	Item			SD	) SA	<b>1</b>	NA
			1= The course does not include resource requirements				
	6	(e) contact information	5= The course includes detailed contact information 4= The course includes fairly detailed contact information	1 2	2 3 4	5	NA
		(teacher, technical, and academic support)	3= The course includes medium detailed contact information 2= The course includes undetailed contact information 1= The course does not include contact information				
	7	(f) course policies and procedures	5= The course includes detailed course policies and procedures 4= The course includes fairly detailed course policies and procedures 3= The course includes medium detailed course policies and procedures 2= The course includes undetailed course policies and procedures 1= The course does not include course policies and procedures	1 2	234	5	NA
	8	(g) learning goals and objectives	5= The course includes detailed learning goals and objectives 4= The course includes fairly detailed learning goals and objectives 3= The course includes medium detailed learning goals and objectives 2= The course includes undetailed learning goals and objectives 1= The course does not include learning goals and objectives	1 2	234	5	NA
	9	(h) testing and grading policies?	5= The course includes detailed testing and grading policies 4= The course includes fairly detailed testing and grading policies 3= The course includes medium detailed testing and grading policies 2= The course includes undetailed testing and grading policies 1= The course does not include testing and grading policies	1 2	234	5	NA
	10	Is a course scope and sequence provided?	5= The course provides a detailed course scope and sequence 4= The course provides a fairly detailed course scope and sequence 3= The course provides course scope and sequence with a medium amount of detail 2= The course provides undetailed course scope and sequence 1= The course does not provide a course scope and sequence	1 2	234	5	NA
Guideline 4: Clear and explicit alignment exists between objectives, assessments,	11	Is there alignment between the behaviors and conditions specified in objectives,	5= Total alignment between the behaviors and conditions specified in objectives, assessment items, strategies, content, and technology, 4= Approximately 75% alignment 3= Approximately 50% alignment 2= Approximately 25% alignment 1= No alignment	1 2	234	5	NA

Guideline	Item			SD	SA	NA
instructional strategies, content, and technology.		assessment items, strategies, content, and technology?				
Guideline 5: Instructional requirements for online courses are as rigorous as requirements for traditional classroom courses.	12	Are the assignments, activities, and depth of content presented online equitable to or exceed related courses provided by the same institution in a traditional classroom environment?	5= the assignments, activities, and depth of content presented online are equal to or exceed Above average related courses provided by the same institution in a traditional classroom environment  4= are equal to average related courses provided by the same institution in a traditional classroom environment  3= are a little less challenging than average related courses provided by the same institution in a traditional classroom environment  2= are less challenging than average related courses provided by the same institution in a traditional classroom environment  1= are much less challenging than average related courses provided by the same institution in a traditional classroom environment	12	3 4 5	NA
Guideline 8: The course provides appropriate types and	13	(a) provide timely and frequent feedback	5= The course provides timely feedback (immediate to 24 hr) 4= The course provides fairly timely feedback (25 - 48 hr) 3= The course provides moderately timely feedback (3 days) 2= The course provides untimely feedback (4 or more days) 1= The course provides no feedback	12	3 4 5	NA
degrees of interactions. Does the course allow for multiple forms of interactivity among participants to	14		5= The course provides (for) very frequent feedback (for all practice and test items, and all communication).  4= The course provides (for) frequent feedback (for 75% of practice and test items, and other contributions).  3= The course provides (for) somewhat feedback (for 50% of practice and test items, and other contributions).  2= The course provides (for) infrequent feedback (for 25% of practice and test items, and other contributions).  1= The course provides (for) no feedback	12	3 4 5	NA
	15	(b) establish a learning community	The same and blinks are smallered by a second it.	12	3 4 5	NA

<sup>5=</sup> The course establishes an excellent learning community. (Includes extensive team-building activities, significant group projects, extensive course-related, meaningful, discussion.)
4= The course establishes a good learning community. (Includes many

Guideline	Item			SD S	Α	NA
			team-building activities, group projects, much course-related, meaningful, discussion.)			
			3= The course establishes a good learning community. (Includes some			
			team-building activities, some group projects, and some course-related,			
			meaningful, discussion.)			
			2= The course establishes a limited learning community. (Includes limited			
			team-building activities, limited group projects, limited course-related,			
			meaningful discussion.)			
	16	(c) engage	1= The course establishes no learning community 5= The course engages students through meaningful interaction.	123	1 5	NΙΛ
	10	students	4= The course engages students through meaningful interaction.	123	+ 3	INA
		Students	3= 50% of the time			
			2= 25% of the time.			
			1= Not at all			
	17	(d) enhance	5= The course enhances student motivation through interactive methods.	123	4 5	NA
		student	4= 75% of the time			
		motivation	3= 50% of the time			
			2= 25% of the time 1= Not at all			
	18	(e) give students	5= The course gives students an active voice in their learning process by	123	1 5	NΔ
	10	an active voice in	allowing them to choose content, sequence and learning activities.	120		14/ 1
		their learning	4= 75% of the time			
		process	3= 50% of the time			
			2= 25% of the time			
			1= Not at all			
Guideline 9:	19		5= The course provides detailed information on the course developers and	123	4 5	NA
The course provider's			their credentials.  4= The course provides quite a lot of information on the course developers			
credentials are			and their credentials.			
available for			3= The course provides a medium amount of information on the course			
review.			developers and their credentials.			
			2= The course provides a minimal amount of information on the course			
			developers and their credentials.			
			1= The course provides no information on the course developers and their			
			credentials.			

Guideline	Item			SD SA	NA
	20	Has the course	5= The course has been developed by a qualified team consisting of	12345	NA
		been developed	content experts and instructional designers		
		by a qualified	5= the course has been developed by a fairly qualified team consisting of		
		team consisting	content experts and instructional designers		
		of content	3= the course has been developed by a medium qualified team consisting		
		experts and	of content experts and instructional designers		
		instructional	2= the course has been developed by a minimally qualified team		
		designers?	consisting of content experts and instructional designers		
			1= the course has been developed by an unqualified team		
B. Course	21	Is content	5= Content is presented in units, modules, and paragraphs appropriate in	12345	NA
Content		presented in	length and format to facilitate online learning.		
Guideline 1:		segments (e.g.,	4= 75% of the content is presented in units, modules, and paragraphs		
Course content		units, modules,	appropriate in length and format to facilitate online learning.		
is organized in		paragraphs)	3= 50%		
a manner to		appropriate for	2= 25%		
facilitate		online learning?	1= None of the content is presented in units, modules, and paragraphs		
learning.			appropriate in length and format to facilitate online learning.		
Guideline 2:	22	Is content	5= All of the content is factual, free of error, and up-to-date	12345	NA
Course content		factual, free of	4= 75% of the content is factual, free of error, and up-to-date		
is accurate,		error, and up-to-	3= 50% of the content is factual, free of error, and up-to-date		
current, free of		date?	2= 25% of the content is factual, free of error, and up-to-date		
bias, and			1= None of the content is factual, free of error, and up-to-date		
sufficient to	23	Is content	5= All of the content is culturally diverse (is appropriate to audiences of all	12345	NA
meet learning		culturally	cultures, includes cultural and international examples, allows for		
goals and		diverse?	international and culturally diverse projects)		
objectives.			4= 75% of the content is culturally diverse		
•			3= 50% of the content is culturally diverse		
			2= 25% of the content is culturally diverse		
			1= None of the content is culturally diverse		
			NA = Cultural diversity is not needed.		
	24	Is there	5= There is 100% of the information needed to meet learning goals and	12345	NA
		sufficient	objectives		
		information	4= There is 75% of the information needed		
		provided to meet	3= There is 50% of the information needed		
		learning goals	2= There is 25% of the information needed		

Guideline	Item			SD	SA	NA
		and objectives?	1= There is none of the information needed			
Guideline 3: The course contains carefully selected links that are reliable	25	Is there an appropriate number of links?	5= There is 100% of needed links 4= There is 75% of needed links 3= There is 50% of needed links 2= There is 25% of needed links 1= There is 0% of needed links NA = Links are not needed	12	3 4 5	5 NA
and valid.	25	Are all of the links in the course working properly?	5= 100% of links work 4= 75% of links work 3= 50% of links work 2= 25% of links work 1= 0% of links work. Or links do not exist even though they are needed. NA= No links exist because they are not needed.	12	3 4 5	5 NA
	27	Do the linked sites provide accurate and current information?	5= 100% of links provide accurate and current information 4= 75% of links provide accurate and current information 3= 50% of links provide accurate and current information 2= 25% of links provide accurate and current information 1= 0% of links provide accurate and current information. Or links do not exist even though they are needed. NA= No links exist because they are not needed	12	3 4 5	5 NA
	28	Does this information clearly correlate to learning objectives?	5= 100% of information at links clearly correlates to learning objectives 4= 75% of information at links clearly correlates to learning objectives 3= 50% of information at links clearly correlates to learning objectives 2= 25% of information at links clearly correlates to learning objectives 1= 0% of information at links clearly correlates to learning objectives. Or links do not exist even though they are needed.  NA= No links exist because they are not needed.	12	3 4 5	5 NA
C. Instructional Strategies and Activities. Guideline 1: Instructional strategies support the	29	Do instructional strategies vary according to the desired type of learning? outcomes (e.g., verbal	5= 100% of instructional strategies match the types of learning outcomes based on Merrill's prescriptions 4= 75% of instructional strategies match the types of learning outcomes based on Merrill's prescriptions 3= 50% of instructional strategies match the types of learning outcomes based on Merrill's prescriptions 2= 25% of instructional strategies match the types of learning outcomes	12	3 4 5	5 NA

Guideline	Item			SD	SA	NA
achievement of learning objectives based on a combination of experience, research, and theory.		information, concepts, procedures, rules, problem- solving, cognitive strategies, attitudes, psychomotor skills)?	based on Merrill's prescriptions 1= 0% of instructional strategies match the types of learning outcomes based on Merrill's prescriptions			
Guideline 2: Instructional strategies allow teachers to address the needs and preferences of individual students	30	Are instructional strategies and activities customizable to address student needs and preferences for pacing, grouping, feedback, control, communication, and assessment?	5= 100% of instructional strategies and activities are customizable to address student needs and preferences for pacing, grouping, feedback, control, communication, and assessment.  4= 75% of instructional strategies and activities are customizable 3= 50% of instructional strategies and activities are customizable 2= 25% of instructional strategies and activities are customizable 1= 0% of instructional strategies and activities are customizable	12	3 4 5	NA
Guideline 3: Sufficient pre- learning activities are available to students to ensure success.	31	• Does the course identify student readiness (e.g., readiness for online learning, reading level, computer/technol ogy skills, prior content knowledge)?	5= The course identifies all aspects of student readiness (e.g., readiness for online learning, reading level, computer/technology skills, prior content knowledge).  4= The course identifies 75% of the aspects of student readiness 3= The course identifies 50% of the aspects of student readiness 2= The course identifies 25% of the aspects of student readiness 1= The course identifies no aspects of student readiness although it is needed NA= The course does not identify any aspects of student readiness because there is no need.	12	3 4 5	NA
	32	Are there self- assessments or checklists that identify	5= There are 100% of needed self-assessments or checklists that identify prerequisite knowledge and skills 4= There are 75% of needed self-assessments or checklists that identify prerequisite knowledge and skills	12	3 4 5	NA

Guideline	Item			SD	SA	NA
		prerequisite knowledge and skills?	3= There are 50% of needed self-assessments or checklists that identify prerequisite knowledge and skills 2= There are 25% of needed self-assessments or checklists that identify prerequisite knowledge and skills 1= There are 0% of needed self-assessments or checklists that identify prerequisite knowledge and skills NA= The above are not provided because they are not needed.			
	33	• Does the course offer orientations, tutorials, job-aids, activities, policies, and procedures to help students acquire the necessary technical skills and content-specific knowledge?	5= The course offers orientations, tutorials, job-aids, activities, policies, and procedures to help students acquire the necessary technical skills and content-specific knowledge.  4= The course offers 75% of the above.  3= The course offers 50% of the above.  2= The course offers 25% of the above.  1= The course offers none of the above.  NA= The above are not provided because they are not needed.	12	3 4 5	i NA
D. Learning Community. Guideline 1: Course materials provide opportunities to facilitate and encourage the development of a learning community among students, teachers,	34	Do course materials allow and support the development of shared goals and expectations among participants?	5= Course materials extensively allow and support the development of shared goals and expectations among participants 4= Course materials mostly allow and support the development of shared goals and expectations among participants 3= Course materials to a medium extent allow and support the development of shared goals and expectations among participants 2= course materials to a limited extent allow and support the development of shared goals and expectations among participants 1= course materials do not allow and support the development of shared goals and expectations among participants NA= Sharing is not intended or needed.	12	3 4 5	5 NA

Guideline	Item			SD	S	Α	NA
administrators, community members, etc.							
	35	• Do course materials foster sharing of resources, experiences, and information among and between members of the learning community with options for monitoring and/or moderating (e.g., chat room, whiteboard, shared audio/visual, etc.)?	5= Course materials extensively foster sharing of resources, experiences, and information among and between members of the learning community with options for monitoring and/or moderating.  4= Course materials mostly foster sharing of resources, experiences, and information among and between members of the learning community with options for monitoring and/or moderating.  3= Course materials foster to a medium extent sharing of resources, experiences, and information among and between members of the learning community with options for monitoring and/or moderating.  2= Course materials foster to a limited extent sharing of resources, experiences, and information among and between members of the learning community with options for monitoring and/or moderating.  1= Course materials do not foster sharing of resources, experiences, and information among and between members of the learning community with options for monitoring and/or moderating.  NA= Sharing is not intended or needed.				NA
	36	Does the course allow these features to be disabled on demand?	5= All sharing and collaborating resources can be disabled on demand. 4= 75% of sharing and collaborating resources can be disabled on demand. 3= 50% of sharing and collaborating resources can be disabled on demand. 2= 25% of sharing and collaborating resources can be disabled on demand. 1= 0% of sharing and collaborating resources can be disabled on demand. NA= Collaborative recourses are not needed or provided or there is no way of knowing if the collaborative resources can be disabled.	1 2	234	15	NA

Guideline	Item			SD SA	. 1	A
	37	• Do course features (e.g., a virtual student break room) allow and support students to identify extracurricular discussion categories?	5= Course features (e.g., a virtual student break room) allow and support students to identify extensive extra-curricular discussion categories 4= Course features allow and support students to identify somewhat extensive extra-curricular discussion categories 3= Course features allow and support students to identify a medium amount of extra-curricular discussion categories 2= Course features allow and support students to identify a limited amount of extra-curricular discussion categories 1= Course features do not allow or support students to identify extra-curricular discussion categories NA= Collaboration, including extra-curricular discussion, is not intended or needed.	1234	5	NA
Guideline 2: The course encourages collaborative methods.	38	Does the course design encourage collaboration to improve student achievement (and meet school/district goals)?	5= The course design encourages extensive collaboration to improve student achievement.  4= The course design encourages fairly extensive collaboration to improve student achievement.  3= The course design encourages a medium amount of collaboration to improve student achievement.  2= The course design encourages a minimum amount of collaboration to improve student achievement.  1= The course design does not encourage collaboration to improve student achievement.  NA= The course does not need collaboration to improve student achievement.	1234	5	NA
E. Student Assessments Guideline 1: The course includes methods for on- going assessments of student achievement.	39	Does the course include methods for determining if students have required prerequisite skills and knowledge throughout coursework?	5= The course includes methods for determining if students have required pre-requisite skills and knowledge throughout coursework 4= The course includes methods for determining if students have required pre-requisite skills and knowledge through 75% of coursework 3= The course includes methods for determining if students have required pre-requisite skills and knowledge through 50% of coursework 2= The course includes methods for determining if students have required pre-requisite skills and knowledge through 25% of coursework 1= The course does not include methods for determining if students have required pre-requisite skills and knowledge throughout coursework, although they are needed.	1234	5	NA

Guideline	Item			SD	SA	NA
			NA= The course does not need or provide methods for determining if students have pre-requisite skills and knowledge throughout coursework.			
	40	Do course materials include guidelines and/or required timelines for providing feedback?	5= Course materials provide extensive feedback or include guidelines and/or required timelines for providing extensive feedback 4= Course materials provide fairly extensive feedback or include guidelines and/or required timelines for providing fairly extensive feedback 3= Course materials provide a medium amount of feedback or include guidelines and/or required timelines for providing a medium amount of feedback 2= Course materials provide limited feedback or include guidelines and/or required timelines for providing limited feedback 1= Course materials provide no feedback or include no guidelines and/or required timelines for providing feedback NA = There is no way of knowing the amount and quality of the feedback that are provided.	12	3 4 5	NA
Guideline 2: Quality course materials provide a wide range of assessment tools.	41	Are pre- and post-testing tools available?	5= Extensive pre- and post testing measures and/or tools are available. 4= Fairly extensive pre- and post testing measures and/or tools are available. 3= A medium amount pre- and post testing measures and/or tools are available. 2= A limited amount of Pre- and post testing measures and/or tools are available. 1= No pre- and post testing measures and/or tools are available.	12	3 4 5	NA
	42	Are measurement tools both objective and subjective in nature?	5= Measures or measurement tools are both objective and subjective in nature 4= Measures or measurement tools are both objective and subjective in nature but are limited in one area. 3= Measures or measurement tools both objective and subjective in nature but may be limited in both areas 2= Measures or measurement tools are either objective or subjective, but not both 1= No measures or measurement tools or available.		3 4 5	
	43	Do measurement tools provide	5= Measures or measurement tools provide results for students (as well as faculty) 4= 75% of measures or measurement tools provide results for students (as	1 2	3 4 5	NA

Guideline	Item			SD SA	NA NA
		results for students as well as faculty?	well as faculty) 3= 50% of measures or measurement tools provide results for students (as well as faculty) 2= 25% of measures or measurement tools provide results for students (as well as faculty) 1= Measures or measurement tools do not provide results for students (as well as faculty)		
	44	Is diagnostic or placement testing available?	5= Thorough diagnostic or placement testing is available 4= Fairly thorough diagnostic or placement testing is available 3= Medium thorough diagnostic or placement testing is available 2= Minimally thorough diagnostic or placement testing is available 1= No diagnostic or placement testing is available	1234	5 NA
	45	Do the teacher and students have access to tools such as chat rooms to provide feedback?	5= The teacher and students have complete access to tools such as chat rooms to provide feedback 4= The teacher and students have fairly complete access to tools such as chat rooms to provide feedback 3= The teacher and students have some access to tools such as chat rooms to provide feedback 2= The teacher and students have limited access to tools such as chat rooms to provide feedback 1= The teacher and students have no access to tools such as chat rooms to provide feedback NA= The course provides feedback but no feedback from the student is intended or provided for.	1234	5 NA
	46	• Is security in place to ensure student authentication?	5= Extensive security is in place to ensure student authentication 4= Good security is in place to ensure student authentication 3= Some security is in place to ensure student authentication 3= Limited security is in place to ensure student authentication 1= No security is in place to ensure student authentication NA= Security for student authentication is not needed	1234	5 NA
	47	Are multiple versions of the same test available?	5= multiple versions of the same test are available in all situations. 4= multiple versions of the same test are available in 75% of situations. 3= multiple versions of the same test are available in 50% of situations. 2= multiple versions of the same test are available in 25% of situations. 1= multiple versions of the same test are not available at all.	1234	5 NA

Guideline	Item			SD S	A	NA
	48	Are tests     available using     multiple formats     (e.g., multiple     choice, true/false,     essay, etc.)?	5= Tests are available using multiple formats (e.g., multiple choice, true/false, essay, etc.)? 4= Tests are available using multiple formats 75% of the time. 3= Tests are available using multiple formats 50% of the time. 2= Tests are available using multiple formats 25% of the time. 1= Tests are not available using multiple formats.	123	4 5	NA
	49	• Are alternative evaluation methods available (e.g., performance-based assessments such as portfolio assessment rubrics, performance checklists and product checklists)?	5= Alternative evaluation methods are used/available (e.g., performance-based assessments such as portfolio assessment rubrics, performance checklists and product checklists)?  4= Alternative evaluation methods are used/available 75% of the time.  3= Alternative evaluation methods are used/available 50% of the time.  2= Alternative evaluation methods are used/available 25% of the time.  1= Alternative evaluation methods are neither used nor available.	123		
F. Technology Integration. Guideline 1: The course uses the	50	Do technologies and media directly support and enhance learning?	5= technologies and media directly support and enhance learning 4= technologies and media directly support and enhance 75% of learning 3= technologies and media directly support and enhance 50% of learning 2= technologies and media directly support and enhance 25% of learning 1= technologies and media do not support and enhance of learning	123	4 5	NA
potentials of technology and media to facilitate and enhance	51	• Do visual elements support the content?	5= visual elements support the content 100% 4= visual elements support the content 75% 3= visual elements support the content 50% 2= visual elements support the content 25% 1= visual elements do not support the content	123		
learning.	52	Do videos, animations, audio, and graphics support rather than	5= Videos, animations, audio, and graphics support rather than detract from learning 4= Videos, animations, audio, and graphics support learning fairly well 3= Videos, animations, audio, and graphics support learning somewhat 2= Videos, animations, audio, and graphics provide minimal support to	123	4 5	NA

Guideline	Item			SD	SA	NA
		detract from learning?	learning 1= Videos, animations, audio, and graphics detract from learning			
	53	Does the course use the	5= the course uses the capabilities of available technologies in an appropriate manner	1 2	3 4 5	S NA
		capabilities of available	4= The course uses the capabilities of available technologies in a fairly appropriate manner			
		technologies in an appropriate manner?	3= The course uses the capabilities of available technologies in a medium level appropriate manner 2= The course uses the capabilities of available technologies in a			
		marrier:	minimally appropriate manner  1= The course uses the capabilities of available technologies in an			
			inappropriate manner			
	54	<ul> <li>Are media used to engage students?</li> </ul>	5= Media is used extensively to engage students 4= Media is used quite a bit to engage students 3= Media is used a medium amount to engage students	1 2	3 4 5	5 NA
		Students:	2= Media is used to a limited extent to engage students 1= Medias is not used to engage students.			
	55	Does the courseware have an autosave feature?	5= The courseware has an autosave feature that saves all student work. 4= The courseware autosaves 75% of student work 3= The courseware autosaves 50% of student work 2= The courseware autosaves 25% of student work 1= The courseware has no autosave feature.	12	3 4 5	5 NA
	56	Does the courseware allow students to access previous material?	5= the courseware allows students to access all previous material 4= the courseware allows students to access 75% of previous material 3= the courseware allows students to access 50% of previous material 2= the courseware allows students to access 25% of previous material 1= the courseware does not allow students to access previous material	12	3 4 5	5 NA
	57	Does the course require/allow the student to have access from home?	5= The course requires or allows the student to have access from home 4= The course requires or allows 75% access from home 3= The course requires or allows 50% access from home 2= The course requires or allows 25% access from home 1= The course requires or allows no access from home	12	3 4 5	5 NA
	58	Does the course have a tracking feature to show	5= The course has a tracking feature to show students where they have been 4= The course has a tracking feature to show students 75% of where they	1 2	3 4 5	5 NA

Guideline	Item			SD	SA	NA
		students where they have been?	have been 3= The course has a tracking feature to show students 50% of where they have been			
			2= The course has a tracking feature to show students 25% of where they have been 1= The course does not have a tracking feature to show students where they have been			
Guideline 2: Technology and media are seamless to ensure ease of use.	59	Is it easy for students to use technology and interpret media elements?	5= It is easy for students to use technology and interpret media elements 4= It is fairly easy for students to use technology and interpret media elements 3= It is somewhat easy for students to use technology and interpret media elements 2= It is not very easy for students to use technology and interpret media elements 1= It is difficult for students to use technology and interpret media elements			NA
Guideline 3: Instructional design addresses potential technology failures.	60	Are there alternative and suitable activities and assignments for students if technology fails?	5= Alternative and suitable activities and assignments for 100% of content are available for students if technology fails 4= Alternative and suitable activities and assignments for 75% of content are available for students if technology fails 3= Alternative and suitable activities and assignments for 50% of content are available for students if technology fails 2= Alternative and suitable activities and assignments for 25% of content are available for students if technology fails 1= Alternative and suitable activities and assignments are NOT available for students if technology fails	123	3 4 5	NA
Guideline 4: The course provides alternative delivery modes to enable students to access materials online	61	Is course content made available to the user through offline resources as needed (e.g. multimedia components distributed via CD-ROM and	5= course content is made available to the user through offline resources as needed (e.g. multimedia components distributed via CD-ROM and Internet)?  4= 75% of course content is made available to the user through offline resources as needed  3= 50% of course content is made available to the user through offline resources as needed  2= 25% of course content is made available to the user through offline resources as needed	123	3 4 5	NA

Guideline	Item			SD	SA	NA
and offline.		Internet)?	1= No course content is made available to the user through offline resources as needed			
Guideline 5: Technical requirements for effective use of the course are readily available to the targeted	62	Are minimum specifications clearly defined prior to acquisition and implementation?	5= Minimum specifications are clearly defined prior to acquisition and implementation 4= Minimum specifications are 75% defined prior to acquisition and implementation 3= Minimum specifications are 50% defined prior to acquisition and implementation 2= Minimum specifications are 25% defined prior to acquisition and implementation 1= Minimum specifications are not defined prior to acquisition and implementation	12	3 4 5	NA
	63	Are minimum specifications consistent with current technologies available to target users?	5= minimum specifications are 100% consistent with current technologies available to target users 4= minimum specifications are 75% consistent with current technologies available to target users 3= minimum specifications are 50% consistent with current technologies available to target users 2= minimum specifications are 25% consistent with current technologies available to target users 1= minimum specifications are not consistent with current technologies available to target users	12	3 4 5	NA

Appendix L. WebCT Instrument

The WebCT instrument (Exemplary Course Project 2005 Nomination Instructions and Form), hereafter called WebCT, was used by course authors to evaluate their course in preparation for submission for the award. The same criteria were used to judge courses for the award.

## WebCT

Rubric Descriptions. Please rate your course using the following scale:

- 5 **Exemplary:** a model of best practice as related to this criterion
- 4 **Accomplished:** excellent implementation; comparable to other examples
- 3 **Promising:** good implementation; however, somewhat lacking in depth or detail
- 2 **Incomplete:** partial implementation of this criterion; additional work needed; good start
- 1 **Not evident:** unable to locate examples specific to this criterion
- **NA Not appropriate:** this criterion has little or no relevance for this course

## Course Design

Heading	Item	Rating Description	Rating
Course Design addresses elements of instructional design in an online course. For the purpose of this project, course design includes such elements as the structure of the course, learning objectives, and instructional	Objectives/learning outcomes are written at the appropriate level	5= Objectives/learning outcomes match audience level 4= Objectives/learning outcomes are a little too difficult or a little too easy for the audience level 3= Objectives/learning outcomes are somewhat difficult or somewhat easy for the audience level 2= Objectives/learning outcomes are way too difficult or way too easy for the audience level 1= Objectives/learning outcomes do not match audience level NA= There is no way of knowing the audience level.	1 2 3 4 5 NA
	2. Objectives/learning	5= Objectives/learning outcomes are clearly	1 2 3 4 5 NA

Heading	Item	Rating Description	Rating
	outcomes are clearly revealed to students  as part of the syllabus/course overview	revealed in great detail to students as part of the syllabus/course overview 4= Objectives/learning outcomes are fairly clearly revealed in fair detail to students as part of the syllabus/course overview 3= Objectives/learning outcomes are somewhat clearly revealed in some detail to students as part of the syllabus/course overview 2= Objectives/learning outcomes not revealed at all clearly in much detail to students as part of the syllabus/course overview 1= Objectives/learning outcomes do not exist	
	3. Objectives/learning outcomes are clearly revealed to students  as part of each learning module	1= Objectives/learning outcomes do not exist 5= Objectives/learning outcomes are clearly revealed to students as part of each learning module 4= Objectives/learning outcomes are fairly clearly revealed to students as part of each learning module 3= Objectives/learning outcomes are somewhat clearly revealed to students as part of each learning module 2= Course Objectives/learning outcomes not revealed at all clearly to students as part of each learning module 1= Objectives/learning outcomes for each learning module do not exist	1 2 3 4 5 NA
	4. Content is made available to students in	5= Content is made available to students in manageable segments	1 2 3 4 5 NA

Heading	Item	Rating Description	Rating
	manageable segments	4= Content is made available to students in	
		mostly manageable segments	
		3= Content is made available to students in	
		somewhat manageable segments	
		2= Content is made available to students in	
		minimally manageable segments	
		1= Content is made available to students in	
		mostly un <u>manageable segments</u>	
	5. Content is presented	5= Content is presented in a <u>logical progression</u>	1 2 3 4 5 NA
	in a <u>logical progression</u>	4= Content is presented in a fairly <u>logical</u>	
		progression	
		3= Content is presented in a somewhat <u>logical</u>	
		progression	
		2= Content is presented in a minimally <u>logical</u>	
		progression	
		1= Content is presented in a il <u>logical progression</u>	
	6. Content is <u>easily</u>	5= Content is <u>easily navigated</u>	1 2 3 4 5 NA
	<u>navigated</u>	4= Content is fairly easily navigated	
		3= Content is somewhat <u>easily navigated</u>	
		2= Content is not <u>navigated</u> very <u>easily</u>	
		1= Content is NOT <u>easily navigated</u>	
		5= Content modules are enhanced through the	1 2 3 4 5 NA
		use of all appropriate affiliated tools	
		4= Content modules are enhanced through the	
	7. Content modules are	use of most appropriate affiliated tools	
	enhanced through the	3= Content modules are enhanced through the	
	appropriate use of	use of some appropriate affiliated tools	
	affiliated tools (check		
	all tools made available		

through the content

Heading	Item	Rating Description	Rating
	module action menu)	2= Content modules are enhanced through the	
	<ul><li>goals quiz</li></ul>	use of a few appropriate affiliated tools	
	<ul><li>discussion</li></ul>	1= Content modules are not enhanced through	
	<ul><li>self test</li></ul>	the use of appropriate affiliated tools	
	□ glossary		
	□ audio		
	□ notes		
	□ video		
	□ links		
	□ other		
	8. The instructor makes	11 1 / / / /	1 2 3 4 5 NA
	appropriate ancillary	resources are available online as part of the	
	resources available as	course content	
	part of the course	4= About 75% of appropriate ancillary resources	
	content.	are available as part of the course content	
		3= About 50% of appropriate ancillary resources are available as part of the course content	
		2= About 25% of appropriate ancillary resources	
		are available as part of the course content	
		1= None of the appropriate ancillary resources	
		are available as part of the course content	
		NA= No ancillary resources are needed. All	
		appropriate resources are already contained in	
		the course.	
	9. The content and	5= The content and requirements are as	1 2 3 4 5 NA
	requirements are as	demanding as an above average face-to-face	
	demanding as a face-	course	
	to-face course with	4= The content and requirements are nearly as	

Heading	Item	Rating Description	Rating
	identical or similar content.	demanding as an above average face-to-face course 3= The content and requirements somewhat less demanding than an above average face-to-face course 2= The content and requirements are less demanding than an above average face-to-face course 1= The content and requirements are much less demanding than an above average face-to-face course	
	10. Accessibility issues are addressed	5= All accessibility issues are addressed 4= 75% of accessibility issues are addressed 3= 50% of accessibility issues are addressed 2= 25% of accessibility issues are addressed 1= No accessibility issues are addressed	1 2 3 4 5 NA
Interaction & Collaboration Interaction and Collaboration can take many forms. The project criteria places emphasis on the type and amount of interaction and collaboration	11. Clearly stated expectations defining minimal levels of student participation.	5= There are clearly stated expectations defining minimal levels of student participation in terms of quantity and quality.  4= There are pretty clearly stated expectations defining minimal levels of student participation in terms of quantity and quality.  3= There are somewhat clearly stated expectations defining minimal levels of student participation in terms of quantity and quality.  2= There are not very clearly stated	1 2 3 4 5 NA

Heading	Item	Rating Description	Rating
within an online environment. In exemplary courses, learner-to-learner-to-content, and learner-to-instructor interaction and		expectations defining minimal levels of student participation in terms of quantity and quality.  1= There are unclearly or no stated expectations defining minimal levels of student participation in terms of quantity and quality, when there should be.  NA= Student interaction with the instructor and peers is not needed or intended for this course.	
collaboration are exemplified through	12. Clearly defined statements informing students what to expect in terms of instructor response time.	5= There are Clearly defined statements informing students what to expect in terms of instructor response time.  4= There are fairly Clearly defined statements informing students what to expect in terms of instructor response time.  3= There are somewhat Clearly defined statements informing students what to expect in terms of instructor response time.  2= There are not very Clearly defined statements informing students what to expect in terms of instructor response time.  1= There are unclearly defined statements or no statements informing students what to expect in terms of instructor response time.  NA= This course does not involve nor need live instructors	
	13. The degree to which students [are	5= Students (are required to) interact extensively with each other and the instructor	1 2 3 4 5 NA

Heading	Item	Rating Description	Rating
	required to] interact	to communicate about course content.	
	with each other and	4= Students (are required to) interact quite	
	the instructor to	extensively with each other and the instructor	
	communicate about	to communicate about course content.	
	course content.	3= Students (are required to) interact	
	☐ The use of real-	·	
	time features	the instructor to communicate about course	
	such as chat	content.	
	rooms and	2= Students (are not required to) interact	
	whiteboards.	very extensively with each other and the	
	☐ The use of	instructor to communicate about course	
	<u>asynchronous</u>	content.	
	tools such as	1= Students (are required at all to) interact	
	discussions and		
	email.	instructor to communicate about course	
	☐ Frequent	content.	
	instances where		
	the instructor	and peers is not needed or intended for this	
	takes an active	course.	
	role in		
	moderating		
	discussions,		
	providing		
	feedback and		
	participating in		
	other interactive		
	components.	F. Thomasia accommodalih anata attau ti	4 0 0 4 5 114
	14. A deliberate	5= There is a very deliberate attempt to	1 2 3 4 5 NA
	attempt to create a	create a learning community using such	

Heading	Item	Rating Description	Rating
	learning community	strategies as group	
	using such strategies	projects/assignments/activities.	
	as group	4= There is a deliberate attempt to create a	
	projects/assignments/	, , ,	
	activities.	group projects/assignments/activities.	
		3= There is a somewhat deliberate attempt to	
		create a learning community using such	
		strategies as group	
		projects/assignments/activities.	
		2= There is a minimally deliberate attempt to	
		create a learning community using such	
		strategies as group	
		projects/assignments/activities.	
		1= There is No attempt to create a learning	
		community using such strategies as group	
		projects/assignments/activities, although	
		there should be.	
		NA= There is no intent or need to create a	
		learning community,	4 0 0 4 5 114
	15. The course makes	5= Assignments and activities enable extensive	1 2 3 4 5 NA
	appropriate use of	critical reflection and <u>analysis</u> of content	
	inherent (WebCT [and	4= Assignments and activities enable critical	
	other online])	reflection and <u>analysis</u> of content	
	<u>technologies</u>	3= Assignments and activities enable some	
		critical reflection and <u>analysis</u> of content	
	Enables critical	2= Assignments and activities enable minimal	
	reflection and	critical reflection and analysis of content	
	<u>analysis</u> of	1= Assignments and activities do NOT enable	
	content	critical reflection and analysis of content	

Heading	Item	Rating Description	Rating
	16. The course makes appropriate use of inherent (WebCT [and other online]) technologies  Assignments/ activities requiring the use of technology clearly explain how technology is to be used by students	5= Assignments/activities requiring the use of technology clearly explain how technology is to be used by students 4= Assignments/activities requiring the use of technology fairly clearly explain how technology is to be used by students 3= Assignments/activities requiring the use of technology somewhat clearly explain how technology is to be used by students 2= Assignments/activities requiring the use of technology do not explain very clearly how technology is to be used by students 1= Assignments/activities requiring the use of technology do not explain how technology is to be used by students	1 2 3 4 5 NA
	17. The course makes appropriate use of inherent (WebCT [and other online]) technologies  Internal communication tools are used	5= Multiple internal communication tools are required to be used extensively by students and instructor to elaborate on course content 4= Some internal communication tools are required to be used quite extensively by students and instructor to elaborate on course content 3= internal communication tools are required to be used somewhat by students and instructor to	1 2 3 4 5 NA

Heading	Item	Rating Description	Rating
	by students and	elaborate on course content	
	instructor to	2= internal communication tools are required to	
	elaborate on	be used minimally by students and instructor to	
	course content	elaborate on course content	
		1= Multiple internal communication tools are	
		NOT required to be used by students and	
		instructor to elaborate on course content	
		NA= The course does not intend to have or	
		need student and instructor elaboration on	
		course content.	
	18. The course makes	5= Use of technology goes well beyond the use	1 2 3 4 5 NA
	exceptional use of	of technology for technology's sake (The	
	inherent WebCT	technology is used extremely well for	
	technologies	instructional purposes	
	_	4= Use of technology goes beyond the use of	
	Use of	technology for <u>technology's</u> sake	
	technology goes		
	beyond the use	<u> </u>	
	of technology	2= Technology is somewhat used for	
	for <u>technology's</u>		
	sake	1= Technology is used extensively for	
		technology's sake	
	19. Student connectivity	5= Content is available in a very wide variety of	1 2 3 4 5 NA
	issues are considered	formats	
		4= Content is available in fairly wide variety of	
	0 Content is	formats	
	available in a	3= Content is available in a variety of formats	
	variety of	2= Content is available in a limited variety of	

Heading	Item	Rating Description	Rating
	formats (i.e., "pdf" files)	formats 1= Content is NOT available in a variety of formats	
	available in a variety of	5= Media files are available in a wide variety of formats 4= Media files are available in a fairly wide variety of formats 3= Media files are available in a variety of formats	1 2 3 4 5 NA
	formats	2= Media files are available in a limited variety of formats 1= Media files are NOT available in a variety of formats	
	21. Student connectivity issues are considered  0 Content/media are available on CD-ROM	5= Content/media are available on CD-ROM 4= 75% of the Content/media are available on CD-ROM 3= 50% Content/media are available on CD-ROM 2= 25% of the Content/media are available on CD-ROM 1= None of the Content/media are available on CD-ROM	1 2 3 4 5 NA
	22. The course makes creative use of a variety of technologies	5= The course makes very creative use of a	1 2 3 4 5 NA

Heading	Item	Rating Description	Rating
		3= The course makes somewhat creative use of	
		a variety of technologies	
		2= The course doesn't make very creative use of	
		a variety of technologies	
		1= The course doesn't make creative use of a	
		variety of technologies	
Assessment	23. Assignments that		1 2 3 4 5 NA
	encourage students	encourage students to employ critical	
Assessment	to employ critical	thinking strategies.(such as Analysis,	
focuses on the	thinking strategies.	Synthesis, Eval	
evaluation of	24. The alignment of	5= All assignments are aligned to stated	1 2 3 4 5 NA
student work.	assignments and	objectives/learning outcomes.	
The project	stated	4= 75% of assignments are aligned to stated	
criteria speak	objectives/learning	objectives/ learning outcomes.	
to the quality	outcomes.	3= 50% of assignments are aligned to stated	
and type of		objectives/ learning outcomes.	
student		2= 25% of assignments are aligned to stated	
assessments		objectives/ learning outcomes.	
within the		1= None of the assignments are aligned to	
course, placing		stated objectives/ learning outcomes.	
particular	25. Assignments that	· · · · · · · · · · · · · · · · · · ·	1 2 3 4 5 NA
emphasis on	provide students with	' ''	
	ample opportunities	concepts and skills in realistic and relevant	
	to practice and apply	ways	
	concepts and skills in		
	realistic and relevant	ample opportunities to practice and apply	
	ways.	concepts and skills in realistic and relevant	
		ways.	
		3= 50% of assignments provide students with	

Heading	Item	Rating Description	Rating
ricaurig		ample opportunities to practice and apply concepts and skills in realistic and relevant ways.  2= 25% of assignments provide students with ample opportunities to practice and apply concepts and skills in realistic and relevant ways.  1= None of the assignments provide students with ample opportunities to practice and apply concepts and skills in realistic and relevant ways.	
	26. Assignments and projects that require students to make appropriate and effective use of external resources, including print, library, Web-based, and other electronic resources.	5= Assignments and projects require students to make appropriate and effective use of external resources, including print, library, Web-based, and other electronic resources.  4= Assignments and projects require students to make fairly appropriate and effective use of external resources, including print, library, Web-based, and other electronic resources.  3= Assignments and projects require students to make somewhat appropriate and effective use of external resources, including print, library, Web-based, and other electronic resources.  2= Assignments and projects require students to make limited use of external resources, including print, library, Web-	1 2 3 4 5 NA

Heading	Item	Rating Description	Rating
		based, and other electronic resources.  1= Assignments and projects don't require	
		students to make appropriate and effective	
		use of external resources, including print, library, Web-based, and other electronic	
		resources.	
		NA= All needed resources are contained	
		within the course. There is no need for	
		external resources.	
	27. Clearly	5= Assignments and activities have very	1 2 3 4 5 NA
	communicated	clear directions.	
	assignments.	4= Assignments and activities have fairly clear directions.	
		3= Assignments and activities have medium	
		clear directions.	
		2= Assignments and activities have	
		somewhat unclear directions.	
		1= Assignments and activities have very	
	00 F 1: :41	unclear or no directions.	4 0 0 4 5 114
	28. Explicitly	5= Rubrics/ performance criteria are made	1 2 3 4 5 NA
	communicated expectations,	available to students <u>clearly stating</u> how the Web may be used in completing assignments	
	including deliverables.	-	
	<u>moraamig</u> aamvaraaraa	available to students FAIRLY clearly stating	
	□ Rubrics/	how the Web may be used in completing	
	<u>performance</u>	<u>assignments</u>	
	criteria are	3= Rubrics/ performance criteria are made	
	made available	available to students SOMEWHAT clearly	
	to students	stating how the Web may be used in	

Heading	Item	Rating Description	Rating
	clearly stating	completing assignments	
	how the Web	2= LIMITED Rubrics/ performance criteria are	
	may be used in	made available to students WITH NOT VERY	
	completing	CLEAR EXPLANATIONS OF how the Web	
	<u>assignments.</u>	may be used in completing assignments	
	_	1= NO Rubrics/ performance criteria are	
		made available to students stating how the	
		Web may be used in completing assignments	
		NA= Web-based searching is not needed for	
		completing assignments. All needed	
		information is available through the course.	
	29. Explicitly	5= The course directs students to specific	1 2 3 4 5 NA
	communicated	Web sites to use or avoid and/or provides	
	expectations,	hints for <u>searching</u>	
	including deliverables.	4= The course directs students to quite a few	
		specific Web sites to use or avoid and/or	
	□ Directing	provides fairly detailed hints for searching	
	students to	3= The course directs students to some	
	specific Web	specific Web sites to use or avoid and/or	
	sites to use or	provides some hints for searching	
	avoid and/or	2= The course directs students to a limited	
	providing hints	number of specific Web sites to use or avoid	
	for <u>searching</u>	and/or provides limited hints for searching	
	the Web.	1= The course doesn't direct students to	
		specific Web sites to use or avoid, or provide	
		hints for <u>searching</u>	
		NA= Web-based searching is not needed for	
		completing assignments. All needed	
		information is available through the course.	

Heading	Item	Rating Description	Rating
	□ 30 Evaluating and validating Web-based information in completing assignments.	5= The course evaluates and validates Webbased information in completing assignments.  4= The course evaluates and validates Webbased information in completing assignments.  3= The course evaluates and validates Webbased information in completing assignments.  2= The course evaluates and validates Webbased information in completing assignments.  1= The course evaluates and validates Webbased information in completing assignments.	1 2 3 4 5 NA
	31. Quizzes and tests that are tied to course objectives/learning outcomes.		1 2 3 4 5 NA

Heading	Item	Rating Description	Rating
		tests are directly tied to course objectives.	
	32. Providing	5= The course Provides students with	1 2 3 4 5 NA
	students with ample	extensive opportunities for self-assessment	
	opportunities for self-	using activities such as self-reflection	
	assessment.	assignments	
		4= The course Provides students with fairly	
		extensive opportunities for self-assessment	
		using activities such as self-reflection	
		assignments	
		3= The course Provides students with some	
		opportunities for self-assessment using	
		activities such as self-reflection assignments	
		2= The course Provides students with	
		minimal opportunities for self-assessment	
		using activities such as self-reflection	
		assignments	
		1= The course Provides students with NO	
		opportunities for self-assessment using	
		activities such as self-reflection assignments	

Appendix M. SREB Instrument

The Southern Regional Education Board's (SREB) Criteria for Evaluating Online

Courses were based on the SREB *essential principles of quality* and were designed to

assist states in determining the quality and effectiveness of web-based courses.

This SREB was used during the first phase of the pilot study and in the dissertation study.

## SREB CRITERIA FOR EVALUATING ONLINE COURSES

http://www.sret.sreb.org/criteria/online.asp 4/7/05

Note: In addition to the following criteria, evaluation of all online courses must also include the criteria for evaluating any Web site: content (accuracy, appropriateness, scope) and technical aspects (navigation, presentation). These criteria are delineated in the NCDPI publication Criteria for Evaluating Web Sites.

SD = Strongly Disagree, SA = Strongly Agree, NA = Not Applicable

Teaching Site and student site assume that it is an instructor led course. What about the commercial courses that have no live instructor?

• Prerequisites

• Teaching Site

 Access to learning resources (books, periodicals, etc.) is available equal to traditionally delivered course

5 = Access to all needed resources is available online (as compared to traditionally delivered course

4 = Access to 75% of needed resources is available online (as compared to traditionally delivered course)

3 = Access to 50% of needed resources is available online (as compared to traditionally delivered course)

2 = Access to 25% of needed resources is available online (as compared to traditionally delivered course)

1 = No access to needed resources is available online (as compared to traditionally delivered course)

SD SA NA 12345 NA

2	Monitoring plan for student chat sessions and/or student-to-student interaction is in place	5 = Monitoring plan for student chat sessions and/or student-to-student interaction is in place (Faculty, TA's, students, provide monitoring, feedback, or intent spelled out in syllabus or obvious in course.)  4 = Monitoring plan for student chat sessions and/or student-to-student interaction is 75% in place  3 = Monitoring plan for student chat sessions and/or student-to-student interaction is 50% in place  2 = Monitoring plan for student chat sessions and/or student-to-student interaction is 25% in place  1 = No monitoring plan for student chat sessions and/or student-to-student interaction is in place  NA = No chat sessions or student interactions are needed or intended.	SD SA 1 2 3 4 5	NA NA
3	Asynchronous and/or synchronous interaction between teacher and student and student-to- student guaranteed	5 = Asynchronous and/or synchronous interaction between teacher and student and student-to-student required. (Postings required daily) 4 = Postings required every 2-3 days 3 = Postings required every 4-6 days 2 = Postings required every week 1 = Postings are not required NA = No live interaction is intended or needed between teacher and student or student and student or we don't know how much interaction is required	12345	NA

•	Student Site	4	Provision is made for students with special needs	5 = Documented, planned or observable provision is made for students with special needs (i.e., Accommodation promised upon request, audio screen reader tags, descriptive title for graphics, page titles) 4 = Documented, planned or observable provision is made for 75% of students with special needs 3 = Documented, planned or observable provision is made for 50% of students with special needs 2 = Documented, planned or observable provision is made for 25% of students with special needs 1 = No documented, planned or observable provision is made for students with special needs	SD SA 12345	NA NA
•	Course Content	5	Syllabus available for review and is understandable by students	<ul> <li>5 = A detailed and clear syllabus or course overview is available for review and appears understandable for target students.</li> <li>4 = A syllabus or course overview is available for review and appears mostly understandable for target students.</li> <li>3 = A syllabus or course overview is available for review and appears somewhat understandable for target students.</li> <li>2 = A syllabus or course overview is available for review and appears minimally understandable for</li> </ul>	12345	NA
		6	Course is comparable in rigor, depth, and breadth to traditionally	target students.  1= No course syllabus or course overview exists.  5 = Course is comparable in rigor, depth, and breadth to better-than-average traditionally delivered courses (Includes challenging assignments, depth, breadth)	12345	NA

			SD SA	INA
	delivered courses	4 = Course is almost comparable in rigor, depth, and breadth to better-than- average traditionally delivered courses 3 = Course is somewhat comparable in rigor, depth, and breadth to better-than-average traditionally delivered courses 2 = Course is minimally comparable in rigor, depth, and breadth to better-than-average traditionally delivered courses 1 = Course is not at all comparable in rigor, depth, and breadth to better-than-average traditionally delivered courses		
7	Instructional and learning goals are clearly defined for student	5 = Comprehensive instructional and learning goals are clearly defined for student 4 = Nearly comprehensive instructional and learning goals are clearly defined for student 3 = Medium comprehensive instructional and learning goals are defined for student 2 = Minimal comprehensive instructional and learning goals are defined for student 1 = Very sketchy or no instructional and learning goals are defined for student	12345	NA
8	Course promotes active learning through student interaction with class peers and/or worldwide peers	5 = Course promotes active learning through extensive student interaction with class peers and/or worldwide peers 4 = Course promotes active learning through fairly extensive student interaction with class peers and/or worldwide peers 3 = Course promotes active learning through medium extensive student interaction with class	12345	NA

			SD	SА	INA
		peers and/or worldwide peers  2 = Course promotes active learning through minimally extensive student interaction with class peers and/or worldwide peers  1 = Course promotes NO active learning through student interaction with class peers and/or worldwide peers  NA = No live interaction is intended or needed			
•	Course is organized in coherent.	5 = Course is organized in coherent, sequential manner	1 2 3	8 4 5	NA
	sequential manner	4 = Course is organized in mostly coherent,			
		3 = Course is organized in medium coherent,			
		2 = Course is organized in a minimally coherent,			
		sequential manner  1 = Course is NOT organized in coherent,			
		sequential manner			
•	Course is designed to take advantage of the unique applications for online delivery	5 = Course is designed to take advantage of a majority of the unique applications for online delivery (Unique applications include: Communication tools including chat, threaded discussion, email, online submission of assignments, critical analysis of comments, documents, assignments that require online access, use of the internet, synchronous or asynchronous access to video and audio) 4 = Course is designed to take advantage of 75% of the unique applications for online delivery 3 = Course is designed to take advantage of 50%	123	3 4 5	NA
	•	<ul> <li>Course is designed to take advantage of the unique applications for</li> </ul>	minimally extensive student interaction with class peers and/or worldwide peers  1 = Course promotes NO active learning through student interaction with class peers and/or worldwide peers  NA = No live interaction is intended or needed  5 = Course is organized in coherent, sequential manner  4 = Course is organized in mostly coherent, sequential manner  3 = Course is organized in medium coherent, sequential manner  2 = Course is organized in a minimally coherent, sequential manner  1 = Course is NOT organized in coherent, sequential manner  5 = Course is designed to take advantage of a majority of the unique applications for online delivery  Communication tools including chat, threaded discussion, email, online submission of assignments, critical analysis of comments, documents, assignments that require online access, use of the internet, synchronous or asynchronous access to video and audio)  4 = Course is designed to take advantage of 75% of the unique applications for online delivery	peers and/or worldwide peers  2 = Course promotes active learning through minimally extensive student interaction with class peers and/or worldwide peers  1 = Course promotes NO active learning through student interaction with class peers and/or worldwide peers  NA = No live interaction is intended or needed  5 = Course is organized in coherent, sequential manner  4 = Course is organized in mostly coherent, sequential manner  2 = Course is organized in medium coherent, sequential manner  2 = Course is organized in a minimally coherent, sequential manner  1 = Course is NOT organized in coherent, sequential manner  5 = Course is designed to take advantage of the unique applications for online delivery  5 = Course is designed to take advantage of a majority of the unique applications include: Communication tools including chat, threaded discussion, email, online submission of assignments, critical analysis of comments, documents, assignments that require online access, use of the internet, synchronous or asynchronous access to video and audio)  4 = Course is designed to take advantage of 75% of the unique applications for online delivery	peers and/or worldwide peers  2 = Course promotes active learning through minimally extensive student interaction with class peers and/or worldwide peers  1 = Course promotes NO active learning through student interaction with class peers and/or worldwide peers  NA = No live interaction is intended or needed  5 = Course is organized in coherent, sequential manner  4 = Course is organized in mostly coherent, sequential manner  3 = Course is organized in medium coherent, sequential manner  2 = Course is organized in a minimally coherent, sequential manner  1 = Course is NOT organized in coherent, sequential manner  5 = Course is NOT organized in coherent, sequential manner  1 = Course is designed to take advantage of a majority of the unique applications for online delivery (Unique applications include:  Communication tools including chat, threaded discussion, email, online submission of assignments, critical analysis of comments, documents, assignments that require online access, use of the internet, synchronous or asynchronous access to video and audio)  4 = Course is designed to take advantage of 75% of the unique applications for online delivery

SD	SA	NA

			2 = Course is designed to take advantage of 25% of the unique applications for online delivery 1 = Course is NOT designed to take advantage of the unique applications for online delivery, although it should be.  NA= Course is NOT designed to take advantage of the unique applications for online delivery, because it does not need to.		
	11	<ul> <li>Assignments are clear and understandable to the student and site coordinators.</li> </ul>	<ul> <li>5 = Assignments, exercises, activities are very clear.</li> <li>4 = Assignments, exercises, activities are mostly clear.</li> <li>3 = Assignments, exercises, activities are medium clear.</li> <li>2 = Assignments, exercises, activities are not very clear.</li> <li>1 = Assignments, exercises, activities are NOT AT ALL clear.</li> </ul>	12345	NA
• Teacher Interaction	12	Course is taught in such a way as to promote trust and teamwork between teacher and student and among students	5 = Course is taught in such a way as to promote trust and teamwork between teacher and student and among students (tone, opportunities to share personal experience, personal information about participants, icebreaking activities that encourage unity, guidelines for tactful, positive contributions, encouraging everyone to treat everyone professionally) 4 = Course is mostly taught in such a way as to promote trust and teamwork between teacher and student and among students	12345	NA

of the unique applications for online delivery

• Various levels of teacher control are planned and teacher guides and monitors students in lower

control assignments

3 = Course is mostly taught in such a way as to promote trust and teamwork between teacher and student and among students

2 = Course is taught in such a way as to promote minimal trust and teamwork between teacher and student and among students

1 = Course is NOT taught in such a way as to promote trust and teamwork between teacher and student and among students

NA = Course does not intend to have nor does it need to have live interaction between teacher and student and among students

5 = Various levels of teacher control are planned and teacher guides and monitors students in lower control assignments (for example, teacher monitoring of threads, TA's monitoring threads, plans to do so, examples for assignments.)

4 = Various levels of teacher control are fairly well planned and teacher guides and monitors students fairly well in lower control assignments

3 = Various levels of teacher control are somewhat planned and teacher guides and monitors students somewhat in lower control assignments

2 = Various levels of teacher control are not well planned and teacher guides and monitors students very little in lower control assignments

1 = No levels of teacher control are planned and teacher does not guide and monitor students in lower control assignments, although these are needed.

12345 NA

	14	•	Teacher provides timely, specific, and authentic feedback	NA = No live teacher control or monitoring is intended or needed.  5 = Teacher guarantees specific and authentic feedback within 24 hrs.  4 = Teacher guarantees specific and authentic feedback within 24-48 hrs.  3 = Teacher guarantees feedback within 3-4 days.  2 = Teacher guarantees feedback within 5-7 days.  1 = Teacher does not guarantee feedback, although she should.	SD 1 2 3	SA 3 4 5	NA NA
	15	•	Teacher provides criteria for grading and weighting for each assignment.	NA = No live feedback is intended or needed for this course.  5 = Teacher provides complete criteria for grading and weighting for each assignment.  4 = Teacher provides mostly complete criteria for grading and weighting for each assignment.  3 = Teacher provides medium complete criteria for grading and weighting for each assignment.  2 = Teacher provides minimally complete criteria for grading and weighting for each assignment.  1 = Teacher provides NO criteria for grading and	123	3 4 5	NA
<ul><li>Evaluation</li><li>Student Evaluation</li></ul>	16	•	Evaluation is timely, fair	weighting for each assignment.  5 = Evaluation is timely and fair or there is a promise of timely and fair evaluation.  4 = Evaluation is mostly timely and fair or there is a promise of mostly timely and fair evaluation.  3 = Evaluation is medium timely and fair or there is a promise of medium timely and fair evaluation.  2 = Evaluation is not very timely and fair or there is a minimal promise of timely and fair evaluation.	123	3 4 5	NA

				SD	SA	INA
			1 = Evaluation is NOT AT ALL timely and fair or there is NO promise of timely and fair evaluation. NA = There is no way of knowing.			
7	•	Teacher and student sites have monitoring/proctori ng policies in place	5 = Teacher and student sites have effective monitoring/ proctoring policies in place (testing security controls).  4 = Teacher and student sites have fairly effective monitoring/ proctoring policies in place  3 = Teacher and student sites have medium effective monitoring/ proctoring policies in place  2 = Teacher and student sites have minimally effective monitoring/ proctoring policies in place  1 = Teacher and student sites have NO monitoring/ proctoring policies in place, although they should NA = There is no monitoring and proctoring because it is not needed.	123	3 4 5	NA

17

SD SA NA

<sup>\*</sup> NCDPI acknowledges the work of Dr. Lynne Schrum, Dr. Zane L. Berge, and the Distance Learning Task Group for the SREB's Educational Technology Cooperative in the preparation of these criteria.

Appendix N. Cropper 5 Star (C-5 Star) Instrument

When instrument selection was performed during the first phase of the pilot study, it was decided to use Merrill's 5 Star Instructional Design Rating checklist as the baseline standard. Cropper added a 5-point Likert scale and descriptions of each of the five possible ratings for each item to create the C-5 Star. The C-5 Star was used through all three phases of the pilot study. A key feature of the C-5 Star includes equal emphasis and value on each of Merrill's five principles.

# **Cropper 5 Star Rating Form (C-5 Star)**

# Based on Merrill's 5 Star Instructional Design Rating (Merrill, 2001)

SD = Strongly Disagree, SA = Strongly Agree, NA = Not Applicable

			SD SA	
Is the courseware presented in the context of real world problems?		5=The courseware is completely presented in the context of real world problems.  4=The courseware is presented in the context of real world problems to a great extent.  3=The courseware is presented in the context of real world problems to a medium extent.  2=The courseware is presented in the context of real world problems to a very limited extent.  1=The courseware is not presented in the context of real world problems.	12345	NA
	a. Does the courseware show learners the task they will be able to do or the problem they will be able to solve as a result of completing a module or course?	5=The courseware shows learners all tasks that they will be able to do.  1=The courseware does not show learners any tasks that they will be able to do,	12345	NA
	b. Are students engaged at the problem or task level not just the operation or action levels?	5=Students are engaged in performing all significant tasks and solving all significant problems representing content mastery.  4= Students are engaged mostly in complete tasks and problems, but in a few simple actions and operations or partial tasks.  3=Students are engaged in some simple actions and operations and some complete tasks and problems,  2=Students are engaged in performing mostly simple actions and operations, representing partial tasks or problems. Only a few complete tasks.  1= Students are only engaged in performing simple actions and operations, only partial tasks.	12345	NA

			SD SA	NA
	c. Does the courseware involve a progression of problems rather than a single problem?	5=The course includes a progression of problems that leads to a mastery of the overall competency.  4=The course includes a variety of problems to be solved, but they may not be in a progression, and do not lead to a complete mastery of the overall competency.  3=The course includes some problems to be solved, but they may not be in a progression, and only lead to partial mastery of the overall course competency.  2=The course includes a few problems to be solved, but they are not in a progression, and lead only to a limited mastery of the overall competency,  1=The course includes only one or two limited problems to be solved, with no progression and almost no mastery of the overall course competency.	12345	NA
2. Does the courseware attempt to activate relevant prior knowledge or experience?	a. Does the courseware direct learners to recall, relate, describe, or apply knowledge from relevant past experience that can be used as a foundation for new knowledge?	5=Extensively 4=Quite a bit 3=A fair amount 2=To a limited extent. 1=Not at all	12345	NA
	b. Does the courseware provide relevant experience that can be used as a foundation for the new knowledge?	5=The course includes a very significant relevant experience that provides a framework for the new knowledge.  4=The course includes a fairly significant relevant experience that provides a framework for the new knowledge  3=A fair amount  2=To a limited extent.  1=not at all	12345	NA

			SD SA	NA
	c. If learners already know some of the content are they given an opportunity to demonstrate their previously acquired knowledge or skill?	5=Students are given a complete opportunity to demonstrate their previously acquired knowledge or skill. 4=Students are given a fairly extensive opportunity. 3=Students are given a medium amount of opportunity. 2=Students are given minimal opportunity. 1=Students are not given an opportunity to demonstrate their previously acquired knowledge or skill.	12345	NA
3. Does the courseware demonstrate (show examples) of what is to be learned rather than merely tell information about what is to be learned? a. Are the demonstrations (examples) consistent with the content being	(1) Examples and non- examples for concepts?	5=The courseware includes extensive examples and non-examples for teaching concept classifications.  4=The courseware includes a fairly extensive amount of examples and non-examples for teaching concept classifications.  3=The courseware includes a medium amount of examples and non-examples for teaching concept classifications.  2=The courseware includes few examples and non-examples for teaching concept classifications.  1=The courseware includes no examples or non-examples for teaching concept classifications.	12345	NA
taught?	(2) Demonstrations for procedures?	5=The courseware includes good demonstrations for all procedures.  4=The courseware includes demonstrations for most procedures.  3=The courseware includes a medium amount of demonstrations for procedures.  2=The courseware includes few demonstrations for procedures.  1=The courseware includes no demonstrations for procedures.	12345	NA

				NA
	(3) Visualizations for processes?	5=The courseware includes good visualizations for all processes.  4=The courseware includes visualizations for most processes.  3=The courseware includes a medium amount of visualizations for processes.  2=The courseware includes few visualizations for processes.  1=The courseware includes no visualizations for processes.	12345	NA
	(4) Modeling for behavior?	5=The courseware includes good modeling for all behavior. 4=The courseware includes modeling for most behavior. 3=The courseware includes a medium amount of modeling for behavior. 2=The courseware includes little modeling for behavior. 1=The courseware includes no modeling for behavior.	12345	NA
b. Are at least some of the following learner guidance techniques employed?	(1) Learners are directed to relevant information?	5=For all of the tasks/modules. 4=For most of the tasks/modules. 3=For about half of the tasks/modules. 2=For a few tasks/modules. 1=Never.	12345	NA
	(2) Multiple representations are used for the demonstrations?	5=100% of needed examples are provided based on the complexity and variations of concepts, procedures, processes, principles.  4=About 75% of the needed examples are provided based on the complexity and variations of concepts, procedures, processes, principles.  3=Approximately half of the needed number of examples are provided based on the complexity and variations of concepts, procedures, processes, principles.  2=About 25% of the needed number of examples are provided based on the complexity and variations of concepts, procedures,  1=No needed examples are provided based on the complexity and variations of concepts, procedures, processes, and principles,	12345	NA

			SD SA	NA
	(3) Multiple demonstrations are explicitly compared?	5=All instances of multiple demonstrations and multiple examples are explicitly compared.  4=Approximately 75% of instances of multiple demonstrations and multiple examples are explicitly compared.  3=Approximately 50% instances of multiple demonstrations and multiple examples are explicitly compared.  2=Approximately  25% of instances of multiple demonstrations and multiple examples are explicitly compared.  1=No instances of multiple demonstrations and multiple examples are explicitly compared.	12345	NA
	c. Is media relevant to the content and used to enhance learning?	5=Video or animation is provided when motion needs to be visualized or behavior needs to be modeled, still visuals when visual depiction is important. Graphics when info needs graphic support. Audio for important explanation, etc. 4=Media is mostly relevant to content.  3=Media is somewhat relevant to content, enhances learning somewhat.  2=Media is minimally relevant to content, enhances learning a little.  1=No media other than text is used, or Media is not relevant to content, does not enhance learning.	12345	NA
4. Do learners have an opportunity to practice and apply their newly acquired knowledge or skill? a. Are the application (practice) and the posttest consistent with the stated or implied objectives?	(1) Information-about practice requires learners to recall or recognize information.	5=All information taught includes information-about practice, which has learners recall, recognize, or identify information. 4=About 75% of information taught 3=About 50% of information taught 2=About 25% of information taught 1=No information taught includes information-about practice.	12345	NA

		SD SA	NA
(2) Parts-of practice requires the learners to locate, name, and/or describe each part.	5=All parts-of information taught uses parts-of practice, which requires the learners to locate, name, and/or describe each part.  4=About 75% of parts-of information taught  3=About 50% of parts-of information taught  2=About 25% of parts-of information taught  1=No parts-of information taught includes parts-of practice.	12345	NA
(3) Kinds-of practice requires learners to identify new examples of each kind.	5=All kinds-of information (concepts) taught uses kinds-of practice, which requires learners to identify new examples of each kind.  4=About 75% of kinds-of information taught  3=About 50% of kinds-of information taught  2=About 25% of kinds-of information taught  1=No kinds-of information taught includes kinds-of practice.	12345	NA
(4) How-to practice requires learners to do the procedure.	5=All how-to (procedures) taught uses how-to practice, which requires learners to do the procedures.  4=About 75% of how-to (procedures) taught  3=About 50% of how-to (procedures) taught  2=about 25% of how-to (procedures) taught  1=No how-to (procedures) taught includes how-to practice.	12345	NA
(5) What-happens practice requires learners to predict a consequence of a process given conditions, or to find faulted conditions given an unexpected consequence.	5=All what-happens (processes, principles) taught uses what-happens practice, which requires learners to requires learners to predict a consequence of a process given conditions, or to find faulted conditions given an unexpected consequence.  4=About 75% of what-happens (processes, principles) taught  3=About 50% of what-happens (processes, principles) taught  2=about 25% of what-happens (processes, principles) taught  1=No what-happens (processes, principles) taught include what-happens practice.	12345	NA

			SD SA	NA
	B1. Does the courseware require learners to use new knowledge or skill to solve a varied sequence of problems?	5=Courseware requires learners to use new knowledge to solve a varied sequence of problems.  4=Application at 75%.  3=Application at 50%.  2=Application at 25%.  1=Courseware does not require learners to use new knowledge.	12345	NA
	B2. Do learners receive corrective feedback on their performance?	5=Learners receive corrective feedback on their performance 100% of the time. 4=75%. 3=50%. 2=25%. 1=Courseware does not give feedback.	12345	NA
	c. In most application or practice activities, are learners able to access context sensitive help or guidance when having difficulty with the instructional materials? Is this coaching gradually diminished as the instruction progresses?	5=Context-sensitive help is 100% available initially for all application and practice activities, but gradually diminishes as instruction progresses, 4=75% true. 3=50% true. 2=25% true. 1=No context-sensitive help.	12345	NA
5. Does the courseware provide techniques that encourage learners to integrate (transfer) the new knowledge or skill into their everyday life?	a. Does the courseware provide an opportunity for learners to publicly demonstrate their new knowledge or skill?	5=Assignment to do significant performance of new knowledge outside of class.  4=Strong encouragement to do significant performance of new knowledge outside of class.  3=Encouragement to do significant performance of new knowledge outside of class.  2=A little encouragement to do significant performance of new knowledge outside of class.  1=No encouragement to do significant performance of new knowledge outside of class.	12345	NA

		SD SA	NA
b. Does the courseware provide an opportunity for learners to reflect-on, discuss, and defend their new knowledge or skill?	5=Courseware provides a significant opportunity for learners to reflect-on, discuss, and defend their new knowledge or skill.  4=Courseware provides a fairly extensive opportunity for learners to reflect-on, discuss, and defend their new knowledge or skill.  3=Courseware provides some opportunity for learners to reflect-on, discuss, and defend their new knowledge or skill.  2=Courseware provides minimal opportunity for learners to reflect-on, discuss, and defend their new knowledge or skill.  1=Courseware provides no opportunity for learners to reflect-on, discuss, and defend their new knowledge or skill.	12345	NA
c. Does the courseware provide an opportunity for learners to create, invent, or explore new and personal ways to use their new knowledge or skill?	5=Courseware provides a significant opportunity for learners to create, invent, or explore new and personal ways to use their new knowledge or skill.  4=Courseware provides a fairly extensive opportunity for learners to create, invent, or explore new and personal ways to use their new knowledge or skill.  3=Courseware provides some opportunity for learners to create, invent, or explore new and personal ways to use their new knowledge or skill.  2=Courseware provides minimal opportunity for learners to create, invent, or explore new and personal ways to use their new knowledge or skill  1=Courseware provides no opportunity for learners to create, invent, or explore new and personal ways to use their new knowledge or skill	12345	NA

The consistency criterion should be applied first. If demonstrations are inconsistent then it doesn't matter if there is learner guidance or if the media is relevant. If demonstrations are consistent then additional credit should be awarded for guidance and/or relevant media.

Appendix O. Merrill e<sup>3</sup> Instrument

Merrill developed a condensed version of his 5 Star instrument, called the e<sup>3</sup> (effective, efficient, and engaging instruction) instrument. While it is also based on First Principles of Instruction, it is a short form that allows for only binary responses which reflect whether First Principles features are present or absent in a course.

### Part One: The Course e<sup>3</sup> evaluation rubric

The heart of the rubric is the First Principles Course Evaluation Rubric form (Table 28). Following is the procedure for using this form to evaluate a course.

- 1. In the header section of the form indicate the name of the course, the URL, the reviewer and the date.
- 2. In the left column of the form list each component skill taught in the course. It is often possible to determine the component skills in a course from the table of contents for the course.
- 3. Check the appropriate box in columns 2, 3, or 4 to indicate the type of skill taught by this component ... kind of, how to, or what happens (why) (See Part two: Consistent Component Skills).
- 4. For each component skill use tables 30 and 31 to evaluate the quality of the demonstration and application for this component. The cells in tables 30 and 31 correspond to the cells in the small matrix in the Course Evaluation Rubric form. The first or demonstration row in the course evaluation matrix (Table 28) corresponds to the row for that type of skill in the e<sup>3</sup> demonstration quality rubric (Table 30). The second row in the course evaluation matrix (Table 28) corresponds to the row for that type of skill in the e<sup>3</sup> application quality rubric (Table 31). If the answer to the question is *yes* check the box in the matrix. If the answer to the question is *no* leave the box empty.
  - If the instruction for a given component skill is complex or lengthy it is sometimes hard to remember all the details of the instruction. Table 29 is an instructional event summary that can help you keep track of the details of the instruction for a given instructional component.
- 5. Complete the last section of the Course Evaluation Rubric (Table 28) if the course involves the learner in a whole problem. Use the last row of the e<sup>3</sup> demonstration and application rubrics to complete the matrix for a whole problem.
- 6. Make liberal use of the comment sections of the form. After a bit of time has passed it is often difficult to remember why you made a particular rating. Your comments will help you recall your reasoning. It is impossible for someone else to know why you used a particular rating. The comments make your course evaluation more valuable to others. You may also want to use the comment section to make recommendations for course revisions to improve its e<sup>3</sup> quality and conformance to First Principles of Instruction.
- 7. For readers who may have an interest in correlating the rubric with student performance or other independent measures we suggest the following untested procedure for determining a score for individual component skills or for the course as a whole. The author would be interested in your attempts to use this scoring procedure. The proposed scoring procedure is as follows:

# Table 28 First Principles Course Evaluation Rubric<sup>1</sup>

Course Name: Reviewer: Date	URL:	Page	of						
Component Skills:	Kind?	How?	Why?			Com	ment:		
						7			
						<u> </u>			
						<u> </u>			
						<u> </u>			
						<u> </u>			
Whole Problem:						_			
						<u> </u>			
	N =	Σ	C =	P =	;	Score = (.75	x ΣC)/n	+ .25	P =
			Key:	Tell	Show	Multimedia	Guide	>=3	Structure
Comment:			itoy.	Ask	Do	Feedback	Coach	>=3	Peer

\_

a. Count 1 point for each checked box in the evaluation matrix for a given component skill or problem. The *tell* and *ask* cells score 0. This gives a

<sup>&</sup>lt;sup>1</sup> © M. David Merrill 2008

- score (C) of from 0 to 10 for each component skill and a score (P) of from 0 to 10 for the whole problem.
- b. The following formula gives you the total score for a module or course: Score =  $.75\Sigma C/N + .25P$  where:  $\Sigma C_{=}$  sum of component skill scores;  $P_{=}$  whole problem score; and  $N_{=}$  number of component skills. This formula allows component skill instruction to account for  $\frac{3}{4}$  of the total score and whole problem instruction to contribute  $\frac{1}{4}$  of the total score.

# Table 29 Instructional Event Summary<sup>2</sup>

Reviewer:

Date:

Component name:												
Instructional Events	Tell	Ask	Show	Do	Multimedia	Feedback	Guidance	Coaching	Structure	Peer	Туре	Comments
Comments:	Procedure: Use this form when a given instructional component has a large number of individual instructional events. The purpose of this form is to help you keep track of these events.  1. Identify the instructional component. 2. List the instructional events one event per line.											

An instructional event is a single tell, ask, show

- coaching, structure or peer collaboration/critique.

  5. Make liberal use of comments to help you
- Make liberal use of comments to help you remember details of events.

or do.

 Indicate the type of learning involved: what is it – kinds (K), how to do it (H), what happens (W), or whole problem (P).

Course name:

\_

Check if there is no violation of a multimedia principle. Comment on violations if they do occur.
 Check if the event includes feedback, guidance,

<sup>&</sup>lt;sup>2</sup> © M. David Merrill

Table 30 Demonstration e<sup>3</sup> Quality Rubric<sup>3</sup>

	Tell	Show	Multimedia	Guide	>3	Structure
Kinds	Does the demonstratio n tell learners the name and <b>definition</b> of each category?	Does the demonstratio n show learners examples of each category?	Does the demonstrati on use effective multimedia principles?	Does the demonstration provide guidance by highlighting discriminating properties or by showing matched examples among categories?	Does the demonstratio n include at least 3 examples from each category?	Does guidance during demonstration s show learners how the defining properties and portrayals relate to an organizing structure?
How to	Does the demonstratio n tell learners the steps and sequence in the procedure?	Does the demonstratio n show a specific instance of the task and demonstrate each of the steps required to complete the task?	Does the demonstrati on use effective multimedia principles?	Does the demonstration provide <b>guidance</b> by calling attention to the execution of each step?	Is the procedure demonstrated in a progression of at least 3 increasingly difficult situations?	Does guidance during demonstration s show learners how the steps in the procedure relate to an organizing structure?
What happens	Does the demonstratio n tell learners the conditions and consequence of the process?	Does the demonstration show the process in a specific real or simulated situation?	Does the demonstrati on use effective multimedia principles?	Does the demonstration provide <b>guidance</b> by helping learners relate the events in the process to the conditions and consequence?	Is the demonstration of the process repeated for a progression of at least 3 increasingly complex scenarios?	Does guidance during demonstration s show learners how the conditions and consequence relate to an organizing structure?
Whole Task	Does the demonstratio n describe a whole problem or task indicating some of the major steps involved?	Is the whole task or problem demonstrated to the learners?	Does the demonstrati on use effective multimedia principles?	Are the component skills of the whole task demonstrated to learners in the context of the whole task using a problem- or task-centered instructional strategy?	Is there a progression of at least 3 increasingly difficult whole tasks or problems demonstrated to the learners?	Does guidance during demonstration s show learners how the steps in the whole task relate to an organizing structure?

<sup>&</sup>lt;sup>3</sup> © M. David Merrill

Table 31 Application e<sup>3</sup> Quality Rubric<sup>4</sup>

	Ask	Do	Feedback	Coach	>3	Peer Interaction
Kinds	Are learners asked to remember the definition?	Does the application require learners to classify new examples?	Does the application provide corrective feedback that focuses learners' attention on discriminating properties?	Does the application provide coaching early in the sequence and gradually withdraw this coaching as the application continues?	Does the application require learners to classify a series of 3 or more divergent examples?	Does the application allow for peer-collaboration and peer-critique?
How to	Are learners required to remember the steps in the sequence?	Does the application require learners to do the task by executing each step in a real or simulated situation?	Does the application provide intrinsic feedback and extrinsic feedback?	Are tasks early in the progression coached and is this coaching gradually withdrawn as for successive tasks in the progression?	Does the application require learners to do a simple to complex progression of at least 3 tasks?	Does the application allow for peer-collaboration and peer-critique?
What Happens	Are learners required to remember the conditions and consequence of the process?	Are learners required to predict the consequence? OR are learners required to troubleshoot an unexpected consequence in a specific situation?	Are learners able to receive intrinsic feedback by being able to test their predictions or test their trouble shooting?	Is coaching provided for problems early in the progression and gradually withdrawn as the progression continues?	Are learners required to make predictions or trouble shoot a series of at least 3 increasingly complex problems?	Does the application allow for peer-collaboration and peer-critique?
Whole Task	Are learners asked to remember information about the whole problem or task?	Do learners have to apply the component skills to the completion of a new whole task or problem?	Are learners able to receive intrinsic feedback on their performance by seeing the consequences of their activities?	Is coaching provided for problems early in the progression and gradually withdrawn as the progression continues?	Are learners required to solve a progression of at least 3 increasingly complex whole problems or tasks?	Does the application allow for peer-collaboration and peer-critique?

<sup>&</sup>lt;sup>4</sup> © M. David Merrill

#### **CURRICULUM VITAE**

#### Max H. Cropper

#### **EDUCATION**

Ph.D. Instructional Technology and Learning Sciences	Utah State University May, 2011
M.S. Instructional Science (Spanish teaching minor)	Brigham Young University 1983
B.A. Spanish Teaching (Communications minor)	Brigham Young University 1976

**DISSERTATION** Providing Reliability and Validity Evidence for Merrill's 2007 5 Star Instrument

Advisor Andrew Walker

Merrill's 2007 5 Star instrument is based on Merrill's well recognized First Principles of Instruction. However, the instrument had not been tested for reliability and validity evidence. In a pilot study, Cropper's 5-Star instrument showed some reliability and validity evidence, but Merrill's more detailed 5 Star instrument (M-5 Star) needed similar evidence. To address this gap in the literature, the purpose of this study was to assess the reliability and validity evidence for M-5 Star. Raters were drawn from a graduate course in online course evaluation and asked to rate a convenience sample (N = 6) of exclusively online university classes using M-5 Star and three comparison instruments. The comparison instruments also purported to examine course quality but lacked the emphasis on instructional strategies in M-5 Star. Interrater reliability evidence for the M-5 Star and the comparison instruments was tentative. Correlations between M-5 Star and the comparison instruments indicate divergent validity support that M-5 Star is measuring a different core concept of quality in online classes. In addition to divergent validity analysis, a content validity index (CVI) analysis was undertaken using experts in the area of first principles. According to first principles experts, other than a few items on the rubric, the vast majority of the M-5 Star CVI results support close alignment with Merrill's first principles of instruction. The study provides limited evidence that the M-5 Star is a reliable and valid instrument of the evaluation of online. Because the study had a number of limitations, confirmatory research is needed.

#### TEACHING EXPERIENCE AND INTERESTS

USU	Helped teach Performance Systems	Fall Semester	2010
	<b>Evaluating Online Courses</b>	Summer term	2007
	Instructional Projects	Fall Semester	2005
	Instructional Theory	Summer term	
	Instructional Development	Summer term	
	Instructional Design	Summer term	
	Instructional Analysis	Summer term	
	Instructional Evaluation	Fall semester	2004

#### Courses Prepared to Teach

Performance Systems, Evaluating (Online) Instruction, Instructional Projects, Learning Theory, Instructional Design Theory, Instructional Development, Instructional Design, Instructional Analysis, Instructional Evaluation

### **PUBLICATIONS**

Cropper, M. H, Bentley, J. P. H., & Schroder, K. (2009). How well do high-quality online courses employ Merrill's first principles of instruction? In M. Orey, V. J. McClendon & R. Branch (Eds.), *Educational Media and Technology Yearbook* (Vol. 34, pp. 121-140): Springer.

- Cropper, M. H. (2009). Providing empirical support for Merrill's first principles of instruction: A metaanalysis/synthesis of instructional strategy research. Unpublished paper. Utah State University. Logan, Utah
- Cropper, M. H. (2005). *In search of a comprehensive model for instruction*. Unpublished paper. Utah State University. Logan.
- Cropper, M. H. (2000). A comprehensive instructional model for the new millennium. Paper presented at the NETg Lecture Series, University of Limerick, Limerick, Ireland.
- Cropper, M. H. (1999). *Taking best-practice instructional strategies online*. Paper presented for the Macromedia Product Tour, Palo Alto, California.

#### COURSE DEVELOPMENT

- Helped with the design and development of a prototype Learning Theory online course for an Instructional Design Certification program for the US Air Force (September 2009 June 2010).
- Helped develop participant and facilitator guides for a high performance teams program for Parker Hannifin Aerospace Group. Also produced video demos on software use. (June 2009 January 2010)
- Developed an online course for Air Force Maintenance DIFMS financial software as subcontractor for ML Technologies/SwRI. (October 2007-September 2008)
- Developed two courses for the Air Force as sub-contractor for Karta Technologies (October 2005 –
  October 2006)
- Developed and facilitated Lean process improvement workshop (Lean Thinking Value Stream Mapping) through SHIM at Hill Air Force Base. Coordinated the development of lean simulations. (May 2004 – May 2008).
- Taught six Instructional Technology courses for USU Extension Services at Weber State University (August 2004 November 2005)
- Developed two MBA-level project management courses for New Ventures of Regis University. (March-April, 2004)
- Helped develop a Change Management workshop and a Continuous Improvement workshop outline at Hill Air Force Base, (September-December 2003)
- Helped develop High School instructional design web-based training program (USU/State Office of Education) (June-December, 2003)

#### ACADEMIC WORKSHOPS AND PRESENTATIONS

- ISPI Utah, September 2010, "5 Star Course Evaluation."
- AECT, October 2010, Anaheim, California. "Evaluating Reliability and Validity Evidence for Merrill's 5 Star Instrument."
- Utah Education Network, University of Utah, June 2010, "5 Star Course Design."
- AECT, November 2009, Louisville, Kentucky. "5 Star Course Design." Presentation with Joanne Bentley.
- ISPI Utah, January 2009, "5 Star Course Design: Taking Your Instruction to the Next Level of Quality." Presentation with Joanne Bentley.
- AECT, 2008, Orlando. "5 Star Course Design: Hands on Evaluation & Re-Design for Educational Transformation." Full-day workshop with Joanne Bentley.
- AECT, 2007, Anaheim. "Where's the Oscar Award for Outstanding Online Instruction?" Presentation with Joanne Bentley and Joel Gardner.
- AECT, 2003, Anaheim. "SCORM Implemented: Show and Tell." Presentation with Mari Vawn Tinney, Andy Gibbons, and David Wiley

#### **CONSULTING WORK**

- **Prometric** (**formerly Galton Technologies**). Helped revise *Test Item Development Workbook*. Facilitated certification test development. (August 1999 August 2003).
- **Industrial Training Zone.** Created the instructional strategy framework for their web-based courses (June-August 2002).

- University of Phoenix/Apollo Press. Consulted on their Reusable Content Object Custom Course Development System (June 2002 January 2003).
- National Education Training Group (NETg). Directed eight developers in writing metadata for 422 NETg courses for the NSA/NETg "Precision Skilling" Project. Trained NETg employees in Chicago and Ireland. (February 1999-December 2000)
- Macromedia. Developed and presented the "Experts Speak" presentation for Macromedia product tour (January-February 1999).
- The Church of Jesus Christ of Latter-day Saints Church Educational Systems (CES). Assisted with in-service teacher development course for Seminary and Institute teachers (June-November 1998).
- **Oracle Corporation.** Assisted with the development of a Reusable Content Objects Strategy Guide for their Oracle Learning Architecture System (October-November 1997).
- Allen Communication. Assisted with the initial development of an Instructional Strategy Wizard (August-October 1997).
- Darryl Sink and Assoc./Hewlett Packard. Five other instructional designers and I developed and
  presented a six-week order coordination training program (one week each) for Hewlett Packard (JuneOctober 1993).
- **Brigham Young University Regional Family History Center.** Produced 3 multi-image/video programs (1985, 1990, and 1995).

## AWARDS

- Outstanding Research Assistant in Instructional Technology (2007). USU
- **ISPI Outstanding Instructional Product of the Year (1995).** Five other instructional designers and I won the award, working for Darryl Sink & Associates in conjunction with a Hewlett Packard team.

#### WORK EXPERIENCE

- *TestOut, Instructional Designer, (June 2001-January 2002)* Developed computer-based training program for Windows 2000 Professional and Server for Microsoft MCSE Certification.
- Novell, Instructional Design Manager, Advanced Technical Training (October 1996-June 1997)
   Coached high-level technical support staff on training strategies as they prepared for live and video training.
- Nu Skin International (February 1992-August 1996) Assistant Director of Human Resource
  Development, Assistant Manager and Manager of Development and Training, HR Generalist.
  Helped manage the HRD department, supervised training development, trained corporate developers
  and trainers, and coordinated training of international managers in Provo. Also developed Nu Skin's
  performance evaluation program and coordinated human resource issues for expatriates, the
  international department, and some other departments.
- Family History Department, Church of Jesus Christ of Latter-day Saints, Instructional Systems Designer (June 1980-February 1992). Did the design, development, and helped with the production of booklets and multi-image and video programs. One booklet I wrote, A Member's Guide to Temple and Family History Work, had more than 2 million copies printed.

### PROFESSIONAL AFFILIATIONS

• Association for Educational Communications and Technology (AECT)

- International Society for Performance Improvement (ISPI). Served 12 years on Utah Chapter Board.
- Former member American Society for Training and Development (ASTD).
- Formerly PHR (Professional In Human Resources) Certified (1994)

#### ADDITIONAL SKILLS

- Fluent Spanish
- Professional Photography
- Some Video and Audio Scripting and Production
- Experience managing development and training
- Extensive experience training and coaching developers and trainers
- Have developed instructional materials for all types of media: stand-up instruction, print, video, computer-based, web-based
- Over four years human resource management experience
- 30 years experience in instructional development and performance improvement
- Expert at statistical analysis and research
- Have developed comprehensive models for instruction and performance
- Expert at developing certification tests and evaluation instruments