

Development of an Instrument to Assess Student Opinions of the Quality of Distance Education Courses

By: Beth Hensleigh Chaney, [James M. Eddy](#), Steve M. Dorman, Linda Glessner, B. Lee Green, and Rafael Lara-Alecio

Chaney, E.H., Eddy, J.M., et al, (2007). Development of an instrument to assess student opinions of the quality of distance education courses. Accepted for publication in *American Journal of Distance Education*. Vol. 21, No .3. (September 2007).

Made available courtesy of Taylor and Francis: <http://www.taylorandfrancis.com/>

***** Note: Figures may be missing from this format of the document**

Abstract:

The purpose of this study was to develop a culturally sensitive instrument to assess the quality of distance education courses offered at a university in the southern United States through evaluation of student attitudes, opinions, and perceptions of distance education. Quality indicators, identified in a systematic literature review, coupled with an ecological framework served as the theoretical foundation for the instrument development process. The process of test development, outlined in the *Standards for Educational and Psychological Testing* (1999), was used and combined with Dillman's (2000) four stages of pretesting to construct the instrument. Results indicated that the model constructed from the quality indicators and ecological framework provided valid and reliable measures of student attitudes, opinions, and perceptions of quality of the distance education courses.

Article:

The worldwide expansion of distance education courses begets the importance of quality assurance (Sherry 2003), emphasizing the need for rigorous evaluation of programs and courses. However, according to Stewart, Hong, and Strudler (2004), there is only a "modest amount of research pertaining to evaluation" of distance education courseware, particularly Web-based courses (13 1). To evaluate student perspectives of the quality of distance education programs and courses, an instrument that produces valid and reliable measurements of student opinions is needed.

To this end, the purpose of this study was to develop a culturally sensitive instrument to assess the quality of distance education programs and courses through evaluation of student attitudes, opinions, and perceptions of distance education. The instrument, called SASODE (Survey to Assess Student Opinions of Distance Education) was developed, pilot-tested, and used to evaluate the quality of health education courses offered via distance education at a major university in the southern United States.

Foundation of Instrument Development

The foundation of the instrument development process used was theory-driven and based on results of a systematic literature review of 160 articles and twelve books (Chaney 2006), which culminated the fourteen quality indicators listed in Table 1. Overall, these findings provided a basis for the quality indicators used to frame the development of SASODE.

The opening sentence in the 2003 *Handbook of Distance Education* states, "America's approach to distance education has been pragmatic and a theoretical" (Saba 2003, 3). The application of theory surrounding research and practice of designing, implementing, and evaluating distance education programs and courses is important. Therefore, in addition to the quality indicators identified in the literature review, the SASODE construction process was based on systems theory and models that capture changes that often occur in distance education. For the development of the SASODE, Social Ecological Model (SEM) (McLeroy et al. 1988), a commonly used systems approach to health education, was used. SEM purports that student opinions are affected by the following multiple levels of influence: intrapersonal, interpersonal, institutional, community, and public policy

factors. In development of SASODE, the intrapersonal and institutional levels of SEM framework provided the theoretical under-pinnings. Intrapersonal level measurements included items to evaluate students' knowledge, attitudes, perceptions, and beliefs about each quality indicator. In addition, quality institutional level measurements examined any activity conducted by the university or designee of the university (i.e., faculty), such as university policies, technological support, student services, and faculty involvement.

Table 1. Common Quality Indicators of Distance Education Identified in the Literature

- Student–teacher interaction
 - Prompt feedback from instructor
 - Program evaluation and assessment
 - Clear analysis of audience
 - Documented technology plan to ensure quality
 - Institutional support and institutional resources
 - Course structure guidelines
 - Active learning techniques
 - Respect diverse ways of learning
 - Faculty support services
 - Strong rationale for distance education that correlates to the mission of the institution
 - Appropriate tools and media
 - Reliability of technology
 - Implementation of guidelines for course development and review of instructional materials
-

Note: The quality indicators listed are results of a systematic literature review.

Instrument Design Framework

The process of test development, outlined in the *Standards for Educational and Psychological Testing* (1999), was used and combined with Dillman's (2000) four stages of pretesting to construct the instrument. Figure 1 outlines the adapted framework used to develop and test the SASODE.

Step 1—Purpose of Instrument

The first step of instrument design, as identified by the *Standards* (1999), is to describe “the extent of the domain, or the scope of the construct[s] to be measured” (37). For the SASODE, the quality indicators from intrapersonal and institutional levels of the SEM provided the scope of the constructs to be measured.

Step 2—Test Specifications

The second step is to design the instrument by identifying test specifications. According to the *Standards* (1999), “the test specifications delineate the format of items, tasks, or questions; the response format or conditions for responding; and the type of scoring procedures”(38). The SASODE included Likert scale questions, open-ended questions, and rank-order questions.

Issues of fairness refer to the idea “that examinees of equal standing with respect to the construct the test is intended to measure should on average earn the same test score, irrespective of group membership” (*Standards* 1999, 74). Therefore, the instrument was constructed to establish equality of measures and outcomes for respondents, regardless of gender, race, ethnicity, or any other characteristic (*Standards* 1999).

Finally, issues of bias refer to “construct-irrelevant components that result in systematically lower or higher scores for identifiable groups of examinees” (*Standards* 1999, 76). Content-related bias is a result of inappropriate test content; however, test developers can assemble a panel of diverse experts to review the instrument for content, language, and questions that might be offensive or disturbing to groups of test takers. A panel was assembled for this instrument development process and will be explained in the following steps of pretesting.

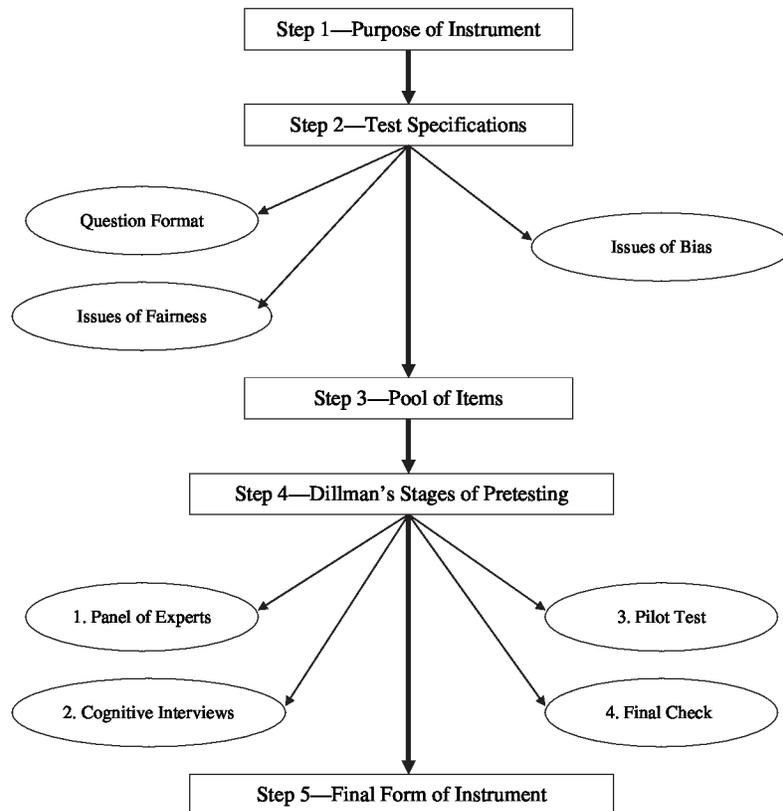


Figure 1. Instrument Design Framework.

Step 3—Development of a Pool of Items

Using the identified quality indicators, items were developed or chosen based on two identified levels (intrapersonal and institutional) of the SEM (McLeroy et al. 1988). The quality indicators assessed using this model were student–teacher interaction, prompt feedback from instructor, student support/services, student-technical assistance/instruction, evaluation and assessment, and course structure benchmarks. An initial pool of items was drawn from three sources: (1) Scanlan’s (2003) instrument to assess quality of Internet-based distance learning, (2) student evaluation forms currently used at the southern university being evaluated, and (3) questions from the *Distance Education Program in Health Studies: Student Satisfaction Survey*, developed and pilot-tested at The University of Alabama. Dillman’s (2000) Tailored Design Method was used to construct additional questions for the current study. The initial pool consisted of seventy-five items.

Step 4—Dillman’s Four Stages of Pretesting

Following Institutional Review Board approval, the items were subjected to four stages of pretesting, identified by Dillman (2000), and assessed for culture sensitivity. Methods and results of this four-stage process are outlined as follows in sequential order.

Stage 1: Review by Knowledgeable Colleagues

Methods

The initial pool of items was sent to a panel of experts for review. The panel consisted of nine professionals from across the country, whose expertise areas included distance education, survey development and research, and cultural sensitivity in research. The main goal was “to finalize the substantive content of the questionnaire so the construction process can be undertaken” (Dillman 2000, 141). The panel was also responsible for evaluating evidence of content-related bias and cultural sensitivity issues in the instrument.

Additionally, the panel was asked to review and rank each item on a scale from 1 to 4: 1 = not important to include in survey, 2 = somewhat important to include in survey, 3 = important to include in survey, and 4 = extremely important to include in survey. To minimize the number of similar items that measured the same quality indicator, panel members were asked to label items as either “keep” or “omit.” During Stage 1, panel members evaluated the instrument for face validity (i.e., the items appear to be relevant to the constructs being investigated) (Gomm, Needham, and Bullman 2000) and content validity, which is defined “as the degree to which the scale properly [reflects] student-related dimensions of quality” in the distance education courses (Scanlan 2003, 4).

Statistical Analysis and Results

The results of the panel review for face validity and content validity revealed that twenty-three items were either redundant or did not adequately measure the intended quality indicator; therefore, the items were reduced from seventy-five to sixty. The criteria for deleting an item involved the rankings of panel members; if a majority (50% or above) indicated the item was either *important* or *extremely important* to include in the survey and suggested to *keep* the item, then it was included in the pilot study instrument. Additionally, modifications to eight questions (either with wording or separating one question into two questions) were made. Two demographic questions, regarding race and ethnicity, were added as a result of panel suggestions on cultural sensitivity of the instrument.

Modifications recommended by the panel resulted in a sixty-two-item instrument, with five parts. Part I (four items) contained general distance education items to get a sense of students’ overall experience and perception of quality of distance education (Likert scale items ranging from 1 = poor to 4 = excellent). Part II (thirty-seven items) consisted of quality indicator items, based on the identified quality indicators. The first nine items in Part II were Likert-type scale items ranging from 1 = very dissatisfied to 4 = very satisfied and 5 = not applicable. The last twenty-eight items were Likert-type scale items ranging from 1 = strongly disagree to 4 = strongly agree and 5 = not applicable. Items in Part III (ten items) were background information questions, which included items such as what channels of communication students used to reach instructors and how many distance education courses students have taken. Part IV included four open-ended questions on the strengths and weaknesses of the course. Students were also given the opportunity to provide additional comments and/or recommendations to help improve quality of the course. Finally, items contained in Part V (seven items) were demographic questions.

Stage 2: Interviews to Evaluate Cognitive and Motivational Qualities

Methods

In this stage, ten students who have either taken a distance education course in health education previously and/or are currently enrolled in one of the distance education courses were asked the sixty-two items, individually, by an interviewer. Respondents were asked to think out loud when answering questions. According to Dillman (2000), the interviewer “probes the respondents in order to get an understanding of how each question is being interpreted and whether the intent of each question is being realized” (142). Cognitive interviewing, such as this, “is designed to produce information when the respondent is confused or cannot answer a question” (Dillman 2000, 142).

Results

Cognitive interviews resulted in minor changes to the instrument. Wording on three items was modified to clarify the meaning of the question, and minor grammatical changes were made. No items were deleted during the cognitive interview process; therefore, the sixty-two-item SASODE was administered for the pilot study.

Stage 3: A Pilot Test

Methods

According to Dillman (2000), the pilot study should emulate procedures to be used in follow-up studies. To this end, 601 students enrolled in at least one of four distance education courses offered at the participant university during the spring 2006 semester were asked to complete the pilot test. The asynchronous distance education

courses selected consisted of two general health courses, Healthy Lifestyles and Women's Health, and two content-specific health courses, Human Sexuality and Consumer Health. Students were sent the SASODE and informed consent via e-mail and asked to return the SASODE with their final examination. Several follow-up e-mails were sent to encourage student input. Responses were kept confidential, and students were asked not to give any identifiable information. Five hundred sixty-eight students completed the pilot test for a response rate of 94%.

Statistical Analyses

Construct validity is "the results achieved from using the instrument predict those matters which the theory underlying the instrument's design says they should predict" (Gomm, Needham, and Bullman 2000, 82). In this study, construct validity was evaluated within Stage 3 by conducting a confirmatory factor analysis (CFA) to identify factor scores of the items. Additionally, predictive validity or criterion-related validity was assessed. This type of validity applies "when one wishes to infer from a test score an individual's most probable standing on some other variable called a criterion" (*Standards* 1999, 179–180). According to Scanlan (2003), for an instrument assessing quality of distance education to have predictive validity, "it should explain or predict students' perceptions of the quality" of their experience in the distance course (6). A CFA was conducted to determine if the scale "has meaningful component structure" (Scanlan 2003, 5) and to develop a measurement model of quality indicators. Then, a structural model was developed to test the relationship between the identified factors in the measurement model (based on the intrapersonal and institutional items) and the more global/institutional items (i.e., overall satisfaction with the distance education course) to assess convergent and predictive validity. Finally, Cronbach's (1984) coefficient alpha (α) was used to determine internal consistency reliability.

Results

Sample Characteristics. Demographic analyses from the pilot study, using Statistical Package for the Social Sciences (SPSS 14.0), indicated the majority of the sample was female (83.3%), white (86%), and classified as seniors (39.7%). The sample represented all nine colleges across the university, with a majority of participants in either Education and Human Development ($n=191$, 34.2%) or Liberal Arts ($n=141$, 25.2%). Refer to Table 2.

Construct Validity Measures. CFA was used to summarize the relationship among ordinal items in the Likert-type scale of Part II of the survey in a smaller number of quality indicators that the items were chosen to measure. In this measurement model, polychoric correlations, which "estimate the linear relationship between two unobserved continuous variables given only observed ordinal data," are fit in the model with Robust Weighted Least Squares (WLS), which is a method for estimating model parameters using categorical or ordinal data (Flora and Curran 2004, 467). The measurement model was estimated using the software package Mplus (Muthén and Muthén 2002).

Robust WLS requires that the distribution of ordinal data is not extremely skewed or leptokurtotic. Otherwise, the standard error of the parameter estimates will be underestimated, and the chi-square model fit test statistic will be inflated, resulting in overrejection of adequately fit models (Flora and Curran 2004). There were fifteen items excluded from the CFA analysis due to non-normality, because skewness and kurtosis were larger than 3.0. After these items were removed from the analysis, the quality indicator, *prompt feedback from instructor*, was measured only by one item in the model. A Pearson-product moment correlation indicated that *prompt feedback from instructor* was highly correlated with *student-teacher interaction* (Pearson's $r=0.852$); therefore, these two indicators were collapsed into one factor for the measurement model tests. It is important to note that the non-normal items were not deleted from the final SASODE because their inclusion in the final instrument was based on face and content validity measures.

Table 2. Demographic Analysis of Sample

	N	%
Gender		
Male	97	16.7
Female	470	83.3
Race		
White	480	86.0
Black/African American	32	5.7
Asian	6	1.1
Native Hawaiian/other Pacific Islander	3	0.5
American Indian/Alaskan Native	2	0.4
Other	35	6.2
Class rank		
Freshman	33	5.9
Sophomore	140	24.8
Junior	158	28.0
Senior	224	39.7
Other	8	1.4
College at Southern University		
Agriculture and Life Sciences	79	14.1
Architecture	4	0.7
Education and Human Development	191	33.6
Geosciences	8	1.4
Liberal Arts	141	25.2
Look College of Engineering	17	3.0
Business School	23	4.1
Science	42	7.5
Veterinary Medicine	28	5.0
Non-declared	21	3.7
Other	5	0.9

An imputation method, EM algorithm, was utilized to input missing data values for items measuring the quality indicators. The statistical software NORM was used to handle missing data values (Schafer 1999). Additionally, the raw data were assessed for consistency of answers on positively worded questions and negatively worded questions. For example, one item states, “The instructor provided *prompt* feedback to my questions,” and the answer choices range from 1 = strongly disagree to 4 = strongly agree. Another item in the same section states, “The feedback to my questions was *delayed*,” with the same Likert-type scale answer choices. These negatively worded questions were included to make sure students were not simply marking the same answer all the way through the survey without reading the questions. Upon analysis of the raw data, 141 students (out of 568) indicated either they *agree* or *strongly agree* on both the positively worded and negatively worded questions that were assessing similar quality indicators or content, which created inconsistent answers. Therefore, these data were filtered and not used in the measurement model analyses.

Model Specifications

The hypothesized measurement model (Model 1), created based on the quality indicators identified in intrapersonal and institutional levels of SEM, contained five factors (latent variables representing the following quality indicators): Factor 1, student–teacher interaction (included prompt feedback from instructor because these two latent variables were collapsed); Factor 2, student support/services; Factor 3, student–technical assistance/instruction; Factor 4, evaluation and assessment; and Factor 5, course structure benchmarks. The Pearson’s product moment correlation between Factor 4 and Factor 5 was extremely high ($r = 1.00$); therefore, these factors were collapsed into one factor.

The fit indexes of Model 1 indicated the model provided poor fit to the data. The chi-square goodness-of-fit index was statistically significant ($X^2 = 529.5$, $d.f. = 65$, $p = .000$), which reveals that Model 1 is not a preferred model. However, according to Thompson (2004), chi-square statistical significance test is “not very useful in

evaluating the fit of a single model” because chi-square values are dependent on sample size. Therefore, other fit indexes were evaluated to justify fit of the model. Bentler’s (1990) comparative fit index (CFI) and the Tucker and Lewis index (TLI) (Tucker and Lewis 1973) were 0.877 and 0.941, respectively. According to Heubeck and Neill (2000), many researchers accept CFI and TLI fit indexes greater than 0.90; therefore, the TLI index is acceptable. However, Root Mean Square Error of Approximation (RMSEA = 0.113) is acceptable at 0.08 and lower, whereas Standardized Root Mean Square Residual (SRMR = 0.071) is acceptable at 0.05 or less (Heubeck and Neill 2000); both RMSEA and SRMR for Model 1 did not achieve acceptable values to ensure proper model fit. Finally, the Weighted Root Mean Square Residual (WRMR) was evaluated for acceptable rates of approximately 1.0; however, WRMR was 2.04, which indicated Model 1 does not fit the data appropriately, and therefore, modifications to the measurement model were required.

Modification indexes revealed the model would be improved by deleting two items from the survey. These items had multiple R^2 values of 0.150 and less; therefore, these did not explain much variance of the items, which means these items did not measure the quality indicators well. Additionally, modification indexes indicated that by adding an additional observed variable to Factor 2 and Factor 5, the model would better explain the data. Once the identified items were removed and additional observed variables were added to Factor 2 and Factor 5, this model (Model 2) was evaluated for model fit. Refer to Figure 2.

Fit indexes for Model 2 indicated a better fit for the data than Model 1. The chi-square goodness-of-fit ($\chi^2 = 383.311$, $d.f. = 57$, $p = .000$) was statistically significant; however, other fit indexes were analyzed for a better idea of model fit and appropriateness. CFI (0.952) and TLI (0.970) were acceptable and indicated appropriate model fit; however, RMSEA (0.116), SRMR (0.065), and WRMR (1.811) did not necessarily meet cutoff points mentioned earlier. Considering the complexity of the model and the high sensitivity of RMSEA, SRMR, and WRMR to model complexity (Potthast 1993), values of the three fit indexes were close enough that Model 2 was not rejected as a good model for the data. Therefore, after two model tests, the fit indexes were approximately satisfactory.

Table 3 provides parameter estimates and standard error for parameter estimates for Model 2. A parameter estimate to standard error ratio (Est./SE) greater than +1.96 or below -1.96 indicates factor loading is statistically significant. Two items (3j. and 3k.) did not have statistically significant factor loadings to their respective factors; however, the model became unstable and less appropriate for the data when these two items were deleted. Therefore, Model 2 remained unchanged.

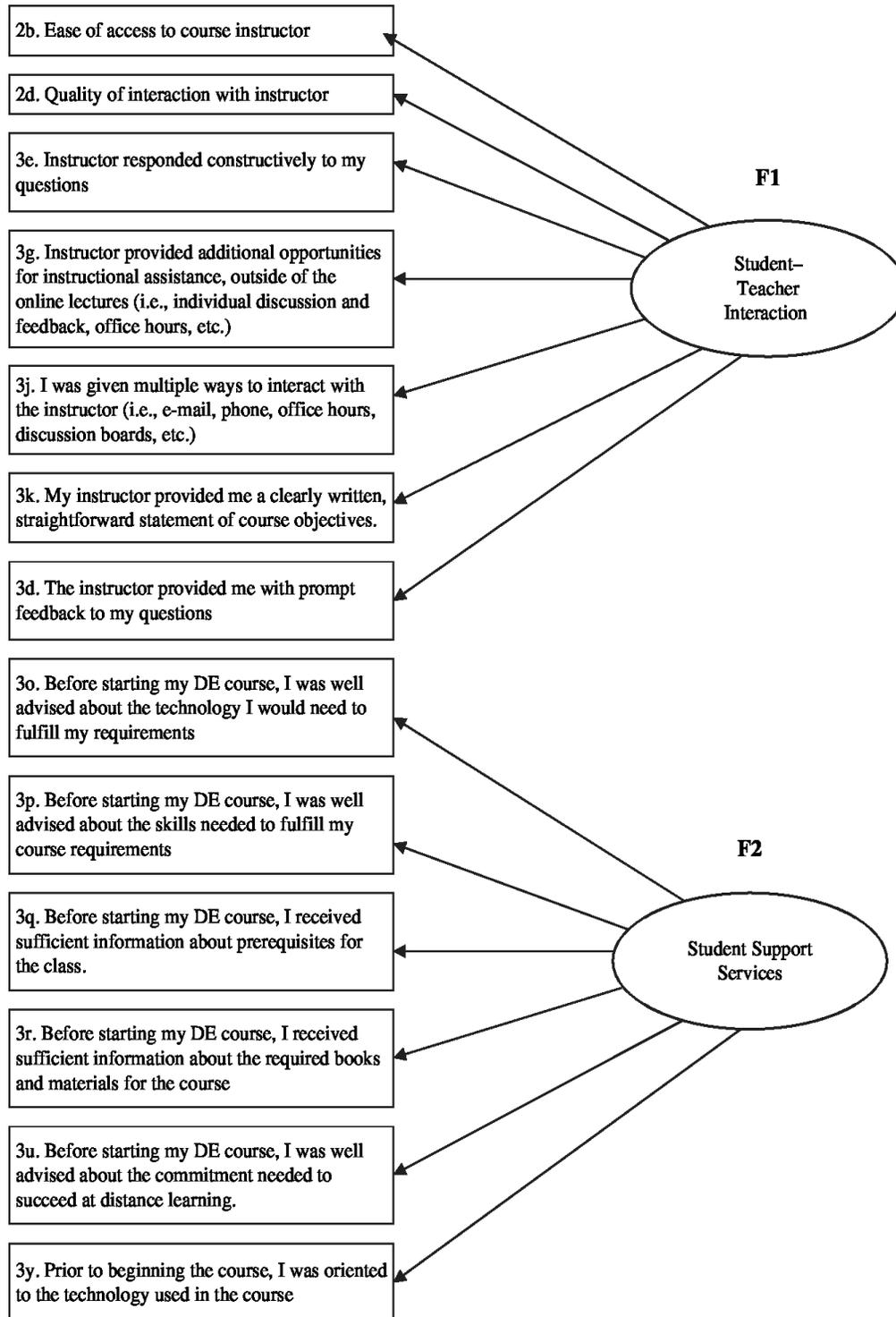


Figure 2a. Confirmatory Factor Analysis Model 2 for Factor 1 (Student- Teacher Interaction) and Factor 2 (Student Support Services).

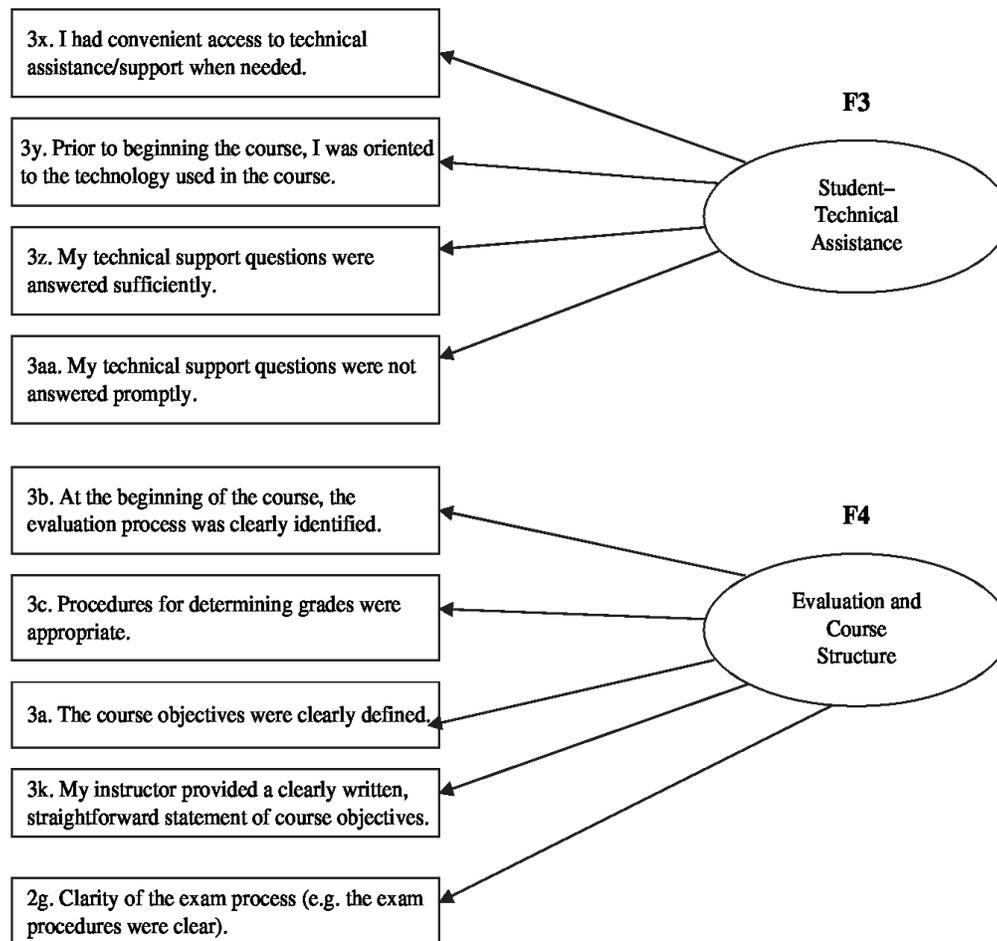


Figure 2b. Confirmatory Factor Analysis Model 2 for Factor 3 (Student–Technical Assistance/Instruction), and Factor 4 (Evaluation and Assessment/Course Structure Benchmarks)

Finally, Table 4 lists multiple *R*-square output produced by the CFA analysis in Mplus. These values are calculated for continuous latent variables (underlying continuous variables that are not observed) rather than the observed categorical/ordinal variables. It is important to understand that multiple *R*-square values for ordinal or categorical outcome variables should not be interpreted as the proportion of explained variance; therefore, parameter estimates and standard errors shed more light on model fit and appropriateness than the multiple *R*-square values (University of Texas 2000).

Predictive Validity Measures. A structural model was developed to test if the measurement model (Model 2) predicted students’ perception of overall quality of distance education and their overall learning experience in distance education courses. Overall quality of distance education was measured by one general item: “I would rate the overall quality of the distance education course as ... 1 = poor, 2 = fair, 3 = good, 4 = excellent.” The overall learning experience was measured by one general item: “Considering all factors combined, I would rate my online learning experience at TAMU as ... 1=poor, 2=fair, 3=good, 4=excellent.” Furthermore, the structural model evaluated how the four factors in the measurement model were related to the following four general items of distance education: (1) “I would rate the overall administrative process of getting started with this distance education course (registering, initial log-on, etc.) as ... 1 = poor, 2 = fair, 3 = good, and 4 = excellent”; (2) “I would rate the overall ease of use of the delivery technology (online lectures and related support resources such as remote library access) as ... 1 =poor, 2 = fair, 3 = good, and 4 = excellent”; (3) Rate the “Quality of instructional methods (online lectures, Web site, CDs, DVDs, etc.) as ... 1 =very dissatisfied, 2 = dissatisfied, 3

= satisfied, 4 = very satisfied, 5 = not applicable”; and (4) Rate the “Quality of the course materials as ... 1 =very dissatisfied, 2 = dissatisfied, 3 = satisfied, 4 = very satisfied, 5 = not applicable.”

Table 3. Parameter Estimates of Confirmatory Factor Analysis Model 2

	Estimates	SE	Est./SE
Factor 1 (Student–Teacher Interaction)			
Item 2b.	1.000	0.000	0.000
Item 2d.	2.161	0.431	5.012
Item 3e.	3.765	0.709	5.312
Item 3g.	0.890	0.269	3.315
Item 3j.	–0.361	0.222	–1.625
Item 3k.	–0.056	0.211	–0.264
Item 3d.	3.569	0.677	5.273
Factor 2 (Student Support Services)			
Item 3o.	1.000	0.000	0.000
Item 3p.	1.081	0.043	25.150
Item 3q.	0.962	0.044	21.714
Item 3r.	1.005	0.048	21.088
Item 3u.	0.959	0.048	19.905
Item 3y.	0.638	0.055	11.509
Factor 3 (Student–Technical Assistance)			
Item 3x.	1.000	0.000	0.000
Item 3y.	0.392	0.077	5.109
Item 3z.	1.608	0.096	17.347
Item 3aa.	1.579	0.091	17.347
Factor 4 (Evaluation/Assessment and Course Structure)			
Item 3b.	1.000	0.000	0.000
Item 3c.	0.865	0.038	22.929
Item 3a.	1.089	0.047	23.394
Item 3k.	1.064	0.049	21.914
Item 2g.	0.353	0.069	5.127

Table 4. Multiple R-square Values for Confirmatory Factor Analysis Model 2

R-square for items	Residual variance	R-square
Item 2a.	0.678	0.322
Item 2d.	0.689	0.311
Item 3a.	0.313	0.687
Item 3b.	0.420	0.580
Item 3c.	0.566	0.434
Item 3d.	0.152	0.848
Item 3e.	0.056	0.944
Item 3j.	0.640	0.360
Item 3k.	0.353	0.647
Item 3o.	0.491	0.509
Item 3p.	0.405	0.595
Item 3q.	0.529	0.471
Item 3r.	0.486	0.514
Item 3u.	0.531	0.469
Item 3x.	0.628	0.372
Item 3y.	0.695	0.305
Item 3z.	0.069	0.961
Item 3aa.	0.073	0.927

Fit indexes for the structural model indicate that the model provides a satisfactory fit for these data. The chi-square goodness-of-fit ($\chi^2 = 473.405$, $d.f. = 81$, $p = .000$) was statistically significant; however, other fit indexes were examined to further investigate model fit. CFI (0.936) and TLI (0.963) are acceptable and provide evidence of good model fit. Additionally, RMSEA (0.107), SRMR (0.072), and WRMR (1.668) were approximately satisfactory numbers (Figure 3).

The parameter estimates and standard errors of the estimates are in Table 5. Parameter estimate to standard error ratios for the model reveal that Factor 1 (student–teacher interaction) and Factor 4 (evaluation and assessment) helps to explain quality of instructional methods. Factor 2 (student support/services) did not significantly explain any of the general distance education constructs, whereas Factor 3 (student–technical assistance/instruction) helped to explain the overall ease of use of distance education technology and the quality of the course materials. However, the relationship between the quality indicator involving student–technical assistance/instruction (Factor 3) was negatively correlated with overall ease of use of distance education technology. This negative relationship could be due to the fact that students who needed more technical assistance probably did not find the distance education technology easy to use. Factor 4 (evaluation and assessment/course objectives) helped explain all four of the general distance education constructs. Finally, the four general distance education items helped explain (and predict) overall quality of distance education and learning experience of students in distance education courses, with statistically significant parameter estimate to standard error ratios for each construct (Table 5).

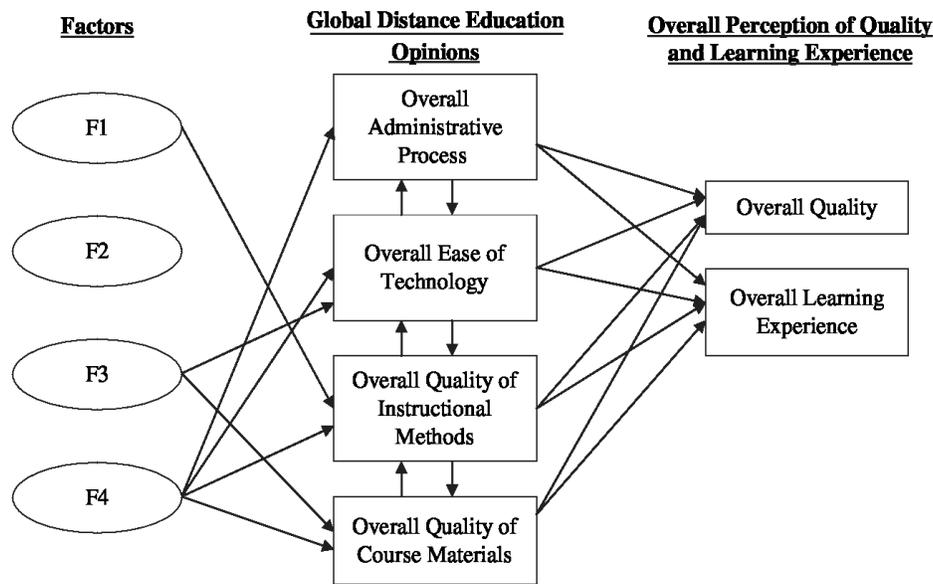


Figure 3. Final Structural Model.

Reliability Measures. Cronbach’s (1984) alpha was assessed for the four quality indicator scales, and all reliability measures were above the acceptable 0.70 alpha coefficient (Gable and Wolf 1993) (Table 6). Cronbach’s alpha was also assessed for each scale by eliminating one item at a time to see if reliability improved by deleting items; however, no deletion improved the alpha coefficient significantly (improvement fell between 0.0012 and 0.0183). Therefore, no items were deleted from the reliability analysis.

Table 5. Parameter Estimates of Final Structural Model

	Estimates	SE	Est./SE
Overall Administrative Process			
Factor 4	0.686	0.105	6.553
Overall Ease of Technology			
Factor 3	0.210	0.112	1.884
Factor 4	0.610	0.110	5.555
Overall Quality of Instructional Methods			
Factor 1	0.500	0.266	1.880
Factor 4	0.575	0.098	5.886
Overall Quality of Course Materials			
Factor 3	0.213	0.098	2.168
Factor 4	0.510	0.093	5.483
Overall Quality of Distance Education			
Overall Administrative Process	0.285	0.045	6.287
Overall Ease of Technology	0.117	0.052	2.238
Overall Quality of Instructional Methods	0.285	0.045	6.375
Overall Quality of Course Materials	0.226	0.052	4.356
Overall Learning Experience			
Overall Administrative Process	0.342	0.038	9.075
Overall Ease of Technology	0.193	0.042	4.612
Overall Quality of Instructional Methods	0.199	0.045	4.450
Overall Quality of Course Materials	0.173	0.044	3.929

Table 6. Reliability Measures of Quality Indicator Scales

Quality Indicator Scales	Cronbach's Alpha
Student-Teacher Interaction Scale	0.718
Student Support/Services	0.785
Student-Technical Assistance/Instruction	0.718
Evaluation and Assessment/Course Structure	0.711

Stage 4: A Final Check. Did We Do Something Silly?**Methods**

In this final step, test developers should ask a few people who have had no part in the development process to answer the questions and check for problems (Dillman 2000). In this study, three additional people were asked to review the survey for wording or content problems.

Results

Stage 4 did not result in additional changes or edits to the final version of SASODE.

Final Form of Instrument

The instrument design framework and results of the statistical analyses helped refine the instrument to sixty items. These items measure global or general distance education opinions, four quality indicators (factors 1–4), background information, and demographic information. The final form of SASODE is available, free of charge, for educational use at <http://ohi.tamu.edu/survey.html>

Discussion and Conclusion

Results of the study reveal that utilizing the instrument design framework, adapted from the *Standards* (1999) and Dillman's (2000) four stages of pretesting, creates a culturally sensitive instrument, SASODE, that produces valid and reliable scores. SASODE can be used to assess student perceptions of quality of distance education courses and provides rich data for evaluation purposes. The final version consists of five parts. Part I includes four items measuring global or general distance education opinions. The second part consists of thirty-five items measuring identified quality indicators and three items measuring perceptions of overall quality. Part III consists of ten background information questions regarding distance education, and Part IV includes four open-ended questions on strengths/weaknesses of the course and recommendations for improvement in quality. Finally, Part V contains seven demographic questions.

References

- Bentler, P. M. 1990. Comparative fit indices in structural models. *Psychological Bulletin*, 107: 238–246.
- Chaney, B. H. 2006. History, theory, and quality indicators of distance education: A literature review. Available online at <http://ohi.tamu.edu/distanced.pdf>
- Cronbach, L. J. 1984. *Essentials of psychological testing*. San Francisco: Harper & Row.
- Dillman, D. A. 2000. *Mail and Internet surveys: The tailored design method*. New York: Wiley.
- Flora, D. B., and P. J. Curran. 2004. An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. *Psychological Methods* 9 (4): 466–491.
- Gable, R. K., and M. B. Wolf. 1993. *Instrument development in the affective domain*. Boston: Kluwer.
- Gomm, R., G. Needham, and A. Bullman, eds. 2000. *Evaluating research in health and social care*. London: Sage.
- Heubeck, B. G., and J. T. Neill. 2000. Internal validity and reliability of the 30 item Mental Health Inventory for Australian Adolescents. *Psychological Reports* 87: 431–440.
- McLeroy, K. R., D. Bibeau, A. Steckler, and K. Glanz. 1988. An ecological perspective on health promotion programs. *Health Education Quarterly* 15 (4): 351–377.
- Muthén, L. K., and B. O. Muthén. 2002. *Mplus user's guide*. Available online at <http://www.statmodel.com/download/usersguide/Mplus%20Users%20Guide%20v41.pdf>
- Potthast, M. J. 1993. Confirmatory factor analysis of ordered categorical variables with large models. *British Journal of Mathematical and Statistical Psychology* 46: 273–286.
- Saba, F. 2003. Distance education theory, methodology, and epistemology: A pragmatic paradigm. In *Handbook of distance education*, ed. M. G. Moore and W. G. Anderson, 3–20. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Scanlan, C. L. 2003. Reliability and validity of a student scale for assessing the quality of Internet-based distance learning. Available online at <http://www.westga.edu/~distance/ojdl/fall63/scanlan63.htm>
- Schafer, J. L. 1999. NORM: Multiple imputation of incomplete multivariate data under a normal model, Ver. 2.03, software for Windows 95/98/NT. Available online at <http://www.stat.psu.edu/~jls/misoftwa.html>
- Sherry, A. C. 2003. Quality and its measurement in distance education. In *Handbook of distance education*, ed. M. G. Moore and W. G. Anderson, 435–459. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Standards for Educational and Psychological Testing*. 1999. Washington, DC: American Educational Research Association, American Psychological Association, and National Council on Measurement in Education.
- Stewart, I., E. Hong, and N. Strudler. 2004. Development and validation of an instrument for student evaluation of the quality of Web-based instruction. *The American Journal of Distance Education* 18 (3): 131–150.
- Thompson, B. 2004. *Exploratory and confirmatory factor analysis: Understanding concepts and applications*. Washington, DC: American Psychological Association.

Tucker, L. R., and C. Lewis. 1973. A reliability coefficient for maximum likelihood factor analysis. *Psychometrika* 38 (1): 1–10.

University of Texas, Austin, Informational Technology Services. 2000. Mplus for Windows: An introduction. Available online at <http://www.utexas.edu/its/rc/tutorials/stat/mplus/>