
Reliability and Validity of a Student Scale for Assessing the Quality of Internet-Based Distance Learning

Craig L. Scanlan, EdD, RRT, FAARC*
Professor and Director
MS and PhD in Health Sciences Programs
University of Medicine and Dentistry of New Jersey
School of Health Related Professions
scanlan@umdnj.edu

**I wish to acknowledge the assistance provided by the UMDNJ-SHRP Technology Task Force in this effort: Drs. Laura Nelson, Joyce O'Connor, Julie O'Sullivan-Maillet, Carlos Pratt, Ann Tucker, and Gail Tuzman.*

Introduction

About 85% of all U.S. universities and colleges offer distance education courses, up from 62% since 1998 (National Governors' Association, 2001). Also since 1998, distance education enrollments in credit-bearing courses have more than doubled, from about 1,364,000 to over 2,870,000 in 2001. (National Center for Education Statistics, 1997; 2003). By 2004, it is expected that distance learners will constitute about 14% of all those enrolled in degree programs (International Data Corporation, 1999). Most of this growth is attributed to the dramatic upsurge in Internet-based education, which now represents the primary means by which colleges and universities provide distance learning (National Center for Education Statistics, 2003).

Concomitant with the rapid growth in Internet-based education has come a flurry of guidelines and standards designed to help assure its quality (American Council on Education, (n.d.); American Distance Education Consortium, (n.d.); American Federation of Teachers, 2000; Council of Graduate Schools, 1998; Council of Regional Accrediting Commissions, 2000; Institute for Higher Education Policy, 2000; Quality Assurance Agency for Higher Education, 1999; Western Cooperative for Educational Telecommunication, 1997). Although there is a remarkable degree of congruence among these standards (Twigg, 2001), and while most include specifications regarding the evaluation of Internet-based education, none provide the actual measurement tools needed to conduct quality assessment. Indeed, in its preliminary review of distance learning, the Institute for Higher Education Policy (1998) emphasized the need for reliable and valid performance measurements.

The situation at the UMDNJ School of Health-Related Professions (SHRP) mirrors the national picture. SHRP has been offering Internet-based education since 1997, with the number of courses and enrollments growing substantially each year. Although individual course evaluation has always been applied, no overall assessment of the School's distance learning program had ever taken place, nor was there a strategy to do so. To that end, a Technology Task Force (TTF) was created and charged with evaluating the School's overall distance learning program.

Methods

Based on a review of the literature, the TTF selected a benchmarking model as the basis for its evaluation. Benchmarking is a quality improvement process that compares actual program or institutional performance to exemplary or best practices (McGregor & Attinasi, 1998). Existing and/or prospectively gathered data are used to determine the actual performance level, with the

best practices set as the achievement ideal. Any observed discrepancy between the actual and ideal performance points to a specific need for quality improvement - the greater the discrepancy, the greater the need.

In terms of best practices, after a review of the published guidelines and standards cited above, the TTF decided to adopt the 24 benchmarks developed and promulgated by the Institute for Higher Education Policy or IHEP (2000). This decision was based on the strong content validation process employed by the IHEP to derive their benchmarks (discussed subsequently under findings).

The TTF reviewed each IHEP benchmark to determine (a) the feasibility and appropriateness of its review; (b) applicable data sources; and (c) data collection methods. Because it became clear that multiple perspectives and approaches would be needed to determine actual performance levels, the TTF adopted a mixed-method strategy (Greene & Caracelli, 1997) that focused on triangulating data obtained from students, faculty and support personnel (Fielding & Fielding, 1986). Student data were gathered primarily via a quantitative survey questionnaire, while faculty and support personnel data were obtained, respectively, via focus group and individual interviews. The remainder of this paper focuses on the psychometric properties of the student survey questionnaire, in particular its benchmark scale.

The student survey questionnaire consisted of four parts: (1) selected IHEP benchmark statements (hereafter referred to as the benchmark scale); (2) four global rating of students' online experience; (3) an open-ended request for recommendations for improvement; and (4) demographic information.*

The TTF initially identified 10 of the 24 IHEP benchmarks as applicable for student assessment. To enhance the specificity of data collection and analysis, the TTF broke 3 of the IHEP benchmarks into two each. For example, the single technical assistance benchmark was separated into two, one focusing on availability and the other focusing on the quality of support. Last, based on prior concerns, the TTF added one of its own benchmarks, addressing enrollment and registration processes.

As depicted in Figure 1, each of these 14 benchmarks was directly translated into a questionnaire statement applicable for student response and presented together on a 4-point Likert scale. The questionnaire was then pilot tested for clarity among the members of the Task Force, with the final edited version administered anonymously in April 2002 to all 115 students enrolled in the School's Internet-based courses. After two follow-ups, 77 students completed the questionnaire, for a response rate of 67%.

Figure 1. Example translation of IHEP benchmark into student scale item.

IHEP Benchmark:
Feedback to student is constructive and provided in a timely manner

<p>In my online course(s), I always received constructive and timely feedback on my assignments and questions</p>			
Strongly Agree	Agree Somewhat	Disagree Somewhat	Strongly Disagree

Findings

All quantitative survey responses were coded, entered into the data editor and analyzed using the Statistical Package for the Social Sciences (SPSS), Release 9.0.0 (1998).

Scale Reliability

Reliability of the benchmark scale was determined by computation of Chronbach's alpha. The standardized alpha for the 14 item scale was 0.94, indicating a high degree of internal consistency (Thorndike, 1996). As depicted in Table 1, individual scale item statistics confirmed this finding, with all items exhibiting both high item-to-scale correlations and high multiple r² values. Moreover, because deleting any item would have lowered overall scale reliability, all 14 items were justified for retention.

Table 1 – Benchmark Scale Item Statistics (items presented in the order they appeared).

Scale Items	Item-Scale	Multiple	Alpha if
	Corr	r ²	Deleted
I found it easy to get the information I needed about my online course(s) and complete the enrollment & registration process	.642	.644	.936
I was always able to gain access to my online course(s) and the applicable UMDNJ network resources (library, e-mail, etc) when needed.	.609	.525	.937
I was given multiple ways to interact with the teacher and other students (e.g., e-mail, discussion) in all online course(s)	.812	.793	.932
In my online course(s), I always received constructive and timely feedback on my assignments and questions	.738	.712	.934
Before starting my online course(s), I was well advised about the self-motivation and commitment needed to succeed at distance learning	.774	.830	.933
Before starting my online course(s), I was well advised about the technology and skills I would need to fulfill my course requirements	.727	.789	.934
My online instructor(s) always provided a clearly written, straightforward statement of course objectives and learning outcomes or expectations	.742	.785	.934
I had sufficient access to the online library resources I needed to fulfill my course objectives and complete all my assignments	.6416	.594	.937

Before starting my online course(s), I received sufficient information about admission requirements or prerequisites, tuition and fees, books and materials, test proctoring or phone conferencing requirements, and student support services	.614	.549	.937
My course(s) provided me with the skills I needed to secure outside course materials through electronic databases, interlibrary loans, government archives, news services, and other sources	.703	.758	.935
Prior to the beginning my online course(s), I was orientated to WebCT and had the opportunity to practice using it	.636	.546	.937
I had convenient access to technical assistance/support whenever needed	.692	.716	.935
My technical support questions or problems were answered accurately or solved quickly	.775	.773	.933
There is a structured system in place to address student complaints about online learning	.755	.639	.933

Scale Validity. Validity of the benchmark scale was approached in three ways: content validity, construct validity and criterion-referenced validity (Thorndike, 1996).

Content Validity. For purposes of this study, content validity was defined as the degree to which the scale properly reflected student-related dimensions of quality in Internet-based distance education. The content validity of the benchmarks themselves were separately established by the Institute for Higher Education Policy (2000). To do so, IHEP first compiled a set of 45 guidelines on Internet-based educational quality via a comprehensive literature search. It then assessed the degree to which colleges and universities were applying these guidelines by visiting six institutions identified as leaders in Internet-based distance learning. At each site, faculty, administrators and students were surveyed to determine (a) the importance of these guidelines, (b) the extent to which they were being followed, and (c) whether or not they made a difference in academic quality. After eliminating statements for which consensus could not be established, deleting those not considered mandatory, and combining standards that addressed the same issue(s), the IHEP derived its final set of 24 benchmarks. As previously described, the content validity of the scale described here was assured by carefully translating each applicable IHEP benchmark into a statement for student response.

Construct Validity. For purposes of this study, construct validity was assessed by identifying the concepts underlying students' scores on this scale. To determine if the scale had a meaningful component structure, it was factor analyzed. In addition, factor scores were derived on the identified components and were compared to students' global ratings of quality, obtained via Part 2 of the survey questionnaire.

All 14 scale items were included in an exploratory factor analysis. The initial components solution was rotated using the Varimax procedure, with an Eigenvalue > 1.0 used as the criterion for factor retention. As depicted in Table 2, after three iterations and using a minimum factor loading of 0.60 (Nunnally and Bernstein, 1994), a meaningful two-factor solution emerged.

Table 2 – Factor Loadings of Benchmark Scale Items.

Scale Item	Factor	Factor
------------	--------	--------

	1	2
In my online course(s), I always received constructive and timely feedback on my assignments and questions	.825	.259
Before starting my online course(s), I was well advised about the self-motivation and commitment needed to succeed at distance learning	.793	.345
My online instructor(s) always provided a clearly written, straightforward statement of course objectives and learning outcomes or expectations	.791	.315
I was given multiple ways to interact with the teacher and other students (e.g., e-mail, discussion) in all online course(s)	.749	.437
There is a structured system in place to address student complaints about online learning	.747	.409
My course(s) provided me with the skills I needed to secure outside course materials through electronic databases, interlibrary loans, government archives, news services, and other sources	.745	.329
Before starting my online course(s), I was well advised about the technology and skills I would need to fulfill my course requirements	.716	.377
Prior to the beginning my online course(s), I was orientated to WebCT and had the opportunity to practice using it	.265	.797
I found it easy to get the information I needed about my online course(s) and complete the enrollment & registration process	.245	.754
I had convenient access to technical assistance/support whenever needed	.388	.743
My technical support questions or problems were answered accurately or solved quickly	.537	.692
Before starting my online course(s), I received sufficient information about admission requirements or prerequisites, tuition and fees, books and materials, test proctoring or phone conferencing requirements, and student support services	.296	.674
I was always able to gain access to my online course(s) and the applicable UMDNJ network resources (library, e-mail, etc) when needed	.338	.608
I had sufficient access to the online library resources I needed to fulfill my course objectives and complete all my assignments	.452	.537

In combination, the two factors accounted for 66.7% of the benchmark scale variance. Factor 1 consisted of seven items and accounted for 36.4% of the scale variance. Since Factor 1 items related mainly to the teaching/learning process (prerequisite skills, learning outcomes, teacher interaction and provision of two-way feedback), it was labeled Teaching-Learning Process. Factor 2 consisted of six items and accounted for 30.3% of the scale variance. Since Factor 2 items related mainly to administrative and support services (admissions and enrollment information, registration, orientation, network access, technical support), it was labeled Administrative Support.

To further explore the concepts underlying students' scores on the benchmark scale, standardized factor scores were computed for each case in the analysis. A Pearson Product-Moment bivariate correlation coefficient (r) was then computed between the students' factor scores and their global ratings (Excellent – Good – Fair – Poor) on the following four statements:

Global Rating Statements (Survey Questionnaire, Part 2)

- I would rate the overall administrative process of getting online (registering, initial logon, etc.) as: (Administrative Rating)

- I would rate the overall quality of online instruction I received as: (Instruction Rating)
- I would rate the overall ease of use of the delivery technology (WebCT and related support resources such as remote library access) as: (Ease of Use Rating)
- Considering all factors combined, I would rate my overall online learning experience at UMDNJ-SHRP as: (Overall Rating)

As evident in Table 3, both factor scores correlated positively and significantly with all global ratings. The highest correlations were observed between each factor and its most equivalent global rating, e.g., $r = 0.811$ for the Teaching-Learning Process factor versus the rating of instructional quality. On the other hand, the lowest correlations were observed between each factor and its complement, e.g., $r = 0.394$ for the Administrative Support factor versus rating of instructional quality. This was taken to indicate good convergent validity (Trochim, 2001).

Table 3 - Correlation of Factor Scores with Students' Global Ratings

Factor	Administrative Rating	Instruction Rating	Ease of Use Rating	Overall Rating
1 <i>Teaching-Learning Process</i>	.423*	.811**	.580**	.677**
2 <i>Administrative Support</i>	.564**	.394*	.479**	.606**

**p < 0.001

*p < 0.01

Criterion-Referenced Validity. For the benchmark scale to have criterion-related validity, it should explain or predict students' perceptions of the quality of their online experience and/or distinguish between groups of students with differing conceptions of that experience (Trochim, 2001). Although Table 3 provides some perspective on this component of validity, additional analysis was performed. Specifically, (1) students' summed benchmark scale scores (all 14 items) were correlated with their four global ratings, (2) students' factor scores were regressed on their overall ratings of their online learning experience, and (3) students' factor scores were used to predict their overall satisfaction/dissatisfaction with their online learning experience (logistic regression).

Table 4 provides the Pearson correlation matrix among students' overall benchmark scale scores and the four global ratings of their online experience. As is evident, all correlations between the overall benchmark scale scores and the four global ratings were significant and ranked from moderately strong (Admin Rating) to very strong (Overall Rating). In the latter case, students' summed benchmark scores explained nearly 80% of the variance in their overall ratings of their online experience.

Table 4 - Pearson Correlation Matrix among Benchmark Scale Scores and Global Ratings

	Scale Score	Admin Rating	Instruct Rating	Ease Rating	Overall Rating
Scale Score	1.00	.661**	.857**	.738**	.893**
Admin Rating		1.00	.460**	.608**	.630**

Instruct Rating			1.00	.595**	.802**
Ease Rating				1.00	.690**
Overall Rating					1.00

**p
<
0.001

To ascertain the ability of the scale factors to predict students' overall rating of their online learning experience, stepwise linear regression was performed, with independent variable entry and removal criteria set to $F < 0.5$ and $F > 0.10$ respectively. Table 5 summarizes the results of this analysis. In combination, the two factors accounted for almost 80% of the variance in the overall rating of students' online learning experience (adjusted R²), comparable to that explained by the summed scale score (above). Moreover, the R² and F change statistics for the steps and the model's standardized beta weights for the factors indicated that both components contribute significantly and are roughly equivalent in predicting the overall rating.

Table 5 - Regression of Factor Scores on Overall Rating

Factor	Adjusted R ²	R ² Change	F Change	Stand Beta+
1 <i>Teaching-Learning Process</i>	.450	.459	50.03**	.662
2 <i>Administrative Support</i>	.799	.347	103.71**	.589

**
p
<
0.001
+model
constant
=
3.04
?
.060

Last, to determine whether the benchmark scale factors could distinguish between groups of students with differing conceptions of their online experiences, respondents were divided according to their overall ratings into two groups: satisfied (rating of excellent or good) and less than satisfied (ratings of fair or poor). The factor scores were entered as a block into a binary logistic regression on these two categories. The resulting model correctly predicted group membership in 95.1% of the cases (chi-square = 48.57, df = 2).

Conclusions and Recommendations

Psychometric analysis of this benchmark scale indicates high reliability (internal consistency) and good content, construct and criterion-related validity. Results also reveal two different dimensions underlying student conceptions of the quality of their Internet-based distance

education – one related to teaching-learning processes and the other associated with provision of administrative support. Both appear to have a significant influence on students' overall perception of their online learning experiences.

Based on these findings and the immensely practical value found by the UMDNJ School of Health-Related Professions Technology Task Force in using scale data in its the quality improvement efforts, it is recommended that this tool be incorporated or adapted by colleges and universities as one of the means used to assess the quality of their Internet-based distance learning efforts. Of course, distance education providers should recognize that any comprehensive quality improvement effort in this area will need to tap additional sources of data (other than students) and explore other dimensions of the quality construct, including actual learning outcomes.

References

American Council on Education. (n.d.). Guiding principles for distance learning in a learning society. Retrieved May 12, 2003, from http://www.acenet.edu/calec/dist_learning/dl_principlesIntro.cfm

American Distance Education Consortium. (n.d.). ADEC guiding principles for distance teaching and learning. Retrieved May 12, 2003, from http://www.adec.edu/admin/papers/distance-teaching_principles.html.

American Federation of Teachers, Higher Education Program and Policy Council. (2000). Distance education: Guidelines for good practice. Washington, D.C.: American Federation of Teachers. Retrieved May 12, 2003, from http://www.aft.org/higher_ed/downloadable/distance.pdf

Council of Graduate Schools, Task Force on Distance Graduate Education. (1998). Distance graduate education: Opportunities and challenges for the 21st century (policy statement). Washington, D.C.: Council of Graduate Schools. Retrieved May 12, 2003, from <http://www.cgsnet.org/pdf/DistanceGraduateEducation.pdf>

Council of Regional Accrediting Commissions. (2000). Best practices for electronically offered degree and certificate programs. Washington, DC: Council of Regional Accrediting Commissions. Retrieved May 12, 2003, from http://www.ncahigherlearningcommission.org/resources/electronic_degrees/Best_Pract_DEd.pdf

Fielding, N.G. & Fielding, J.L. (1986). Linking data (Vol 4). Beverly Hills, CA: Sage Publications.

Greene, J.C., & Caracelli V.J. (1997). Advances in mixed-method evaluation: The challenges and benefits of integrating diverse paradigms. San Francisco: Jossey-Bass.

Institute for Higher Education Policy. (1998). Assuring quality in distance learning: A preliminary review. Washington, DC: Council for Higher Education Accreditation. Retrieved May 12, 2003, from <http://www.chea.org/Events/QualityAssurance/98May.html>

Institute for Higher Education Policy. (2000). Quality on the line: benchmarks for success in Internet-base distance education. Washington, DC: Institute for Higher Education Policy. Retrieved May 12, 2003, from <http://www.ihep.org/Pubs/PDF/Quality.pdf>

International Data Corporation. (1999). Online distance learning in higher education, 1998-2002. Cited in CHEA Update, Number 2. Washington, DC: Council For Higher Education Accreditation.

National Governors' Association. (2001). State of e-learning in the states. Washington, DC: National Governors' Association. Retrieved May 12, 2003, from <http://www.nga.org/cda/files/060601ELEARNING.pdf>

Nunnally, J.C. & Bernstein, I.H. (1994). Psychometric theory, 3rd ed. New York: McGraw Hill.

McGregor, E.N. & Attinasi, L.C. (1998). The craft of benchmarking: finding and utilizing district-level, campus-level, and program-level standards. Paper presented at the Rocky Mountain Association for Institutional Research Annual Meeting, October, Bozeman, MT. ED423014

National Center for Education Statistics. (1999). Distance education at postsecondary education institutions: 1997-98. NCES 2000-013. Washington, DC: U.S. Department of Education. Retrieved May 12, 2003, from <http://nces.ed.gov/pubs2000/2000013.pdf>

U.S. Department of Education, National Center for Education Statistics. (2003). Distance education at degree-granting postsecondary institutions: 2000-2001. NCES 2003-017. Washington, DC: U.S. Department of Education. Retrieved August 1, 2003, from <http://nces.ed.gov/pubs2000/2000013.pdf>

Quality Assurance Agency for Higher Education. (1999). Guidelines on the quality assurance of distance learning. Gloucester, UK: Quality Assurance Agency for Higher Education. Retrieved May 12, 2003, from <http://www.qaa.ac.uk/public/dlg/contents.htm>

SPSS. (1998). Statistical package for the social sciences. Release 9.0.0 for Windows. Chicago, IL: SPSS, Inc.

Thorndike, R.M. (1996). Measurement and evaluation in psychology and education, 6th ed. Upper Saddle River, NJ: Prentice Hall.

Trochim, W.M.K. (2001). The research methods knowledge base, 2nd ed. Cincinnati: Atomic Dog Publishing.

Twigg, C.A. (2001). Quality assurance for whom? Providers and consumers in today's distributed learning environment. Troy, NY: The Pew Learning and Technology Program, Center for Academic Transformation, Rensselaer Polytechnic Institute. Retrieved May 12, 2003, from <http://www.center.rpi.edu/PewSym/mono3.html>

Western Cooperative for Educational Telecommunication. (1997). Principles of Good Practice for Electronically Offered Academic Degree and Certificate Programs (Pub. 2A299) Boulder, CO. Retrieved May 12, 2003, from <http://www.wcet.info/projects/balancing/principles.asp>

Online Journal of Distance Learning Administration, Volume VI, Number III, Fall 2003
State University of West Georgia, Distance Education Center

[Back to the Online Journal of Distance Learning Administration Contents](#)