

Measuring Online Course Design: A Comparative Analysis

Peter You
Sue Hochberg
Phoebe Ballard
Mingli Xiao
Tony Walters

The University of Toledo

Abstract: This paper investigates the differences between students' and QM peer reviewers' perspectives of essential QM standards in three online courses. Rasch analysis is used to validate the instrument and nonparametric Mann-Whitney U test is used to evaluate a difference between the two. The primary goal of this study is to provide an explanation of the gaps between students' perceptions and QM peer reviewers' regarding the online courses reviewed by QM peer reviewers.

Introduction

The growth of online learning programs is tremendous in the past ten years. Best practices for the implementation of online programs have been employed by universities offering online programs. Best practices and standards for online courses also have been developed and implemented in higher education. To ensure the quality of online courses, it is critical that online courses are designed according to a set of best practices or standards before it is delivered to students. Quality Matters is a faculty-driven, peer-review process that is collaborative, collegial, continuous, and centered in national standards of best practices, and research findings in online and blended learning to promote student learning. It has been widely adopted by higher education in the nation as a process and a rubric to continuously improve online course quality.

This study attempts to (1) validate the instrument design based on QM Standards to measure online course design; (2) investigate to what degree the selected courses meet QM standards from a student's perspective, and (3) identify gaps between student's perspective and QM certified reviewers' perspective about QM essential standards.

The results of this study indicate that most of the items in the instrument designed according to the Quality Matters standards work well to measure the design perspective of online courses. The results also show there are three tiers (Tier I: to a great extent, Tier II: to a moderate extent, and Tier III: to little or some extent) in regard to meeting the standards in the three courses.

The results on most of the standards evaluated in this study provided by both reviewers and students are the same, indicating that both peer reviewers and students take the same point view in regard to evidences of the standards. However, they differed significantly regarding three of the

essential standards regarding course objectives, unit learning objectives, and grading policy. One factor that might cause the discrepancy is that reviewers look for solid evidences for measurable learning outcomes while students look for clearly articulated objectives.

Literature Review

Student Perspectives

Several studies have been conducted about student perspectives. These studies can be separated into two categories: a) student perceptions of the value of QM features in an online course, b) student opinions about whether a course meets QM standards or not. Ralston-Berg and Nath (2008) stated that students valued the same standards marked as essential “3” and very important “2” by QM, but valued significantly less on standards marked as important “1” by QM. They further noted that students who claimed high satisfaction in online courses also valued all QM features than those who claimed low satisfaction. Ralston (2011) also found out in another study that the results by rank of importance to students for success correlated with QM standards. Knowles and Kalata (2010) as cited in Shattuck (2012) stated that there might be a discrepancy in expectations between students and experienced QM master reviewers. They further offered possible explanations about this discrepancy: a) that students simply completed the survey without thinking about the standards and the course content, or b) many of the design aspects were clarified by the instructors during the course were being taught via channels that are not available to the peer reviewers.

Quality Matters Standards and Review Process

Quality Matters (QM) is a process and a rubric to continuously improve online course quality (Shattuck, 2012). It is a faculty-driven, peer-review process that is collaborative, collegial, continuous, and centered in national standards of best practices, and research findings in online and blended learning to promote student learning. QM is a leader in quality assurance for online education and has received national recognition for its peer-based approach and continuous improvement in online education and student learning. The research-based rubric is designed to evaluate only course design and not course delivery or content. The rubric consists of eight broad categories broken down into 41 individual standards. These 41 standards can be used in a variety of ways ranging from providing guidelines for course development to the evaluation and certification of courses through an internal or external review process.

The goal of the QM review process is to continuously improve online course quality. According to Shattuck (2012) the process begins with a mature course, which means the course had been offered for at least 2 semesters and the course instructor has revised it based on previous experiences. A review team with three certified QM reviewers who have online teaching experiences will review the course and provide feedback to the course developer. To conduct formal reviews one of the review team members must be a subject matter expert in the field of the course being reviewed and one member must be a master reviewer. In the event that a course does not meet the required 85% (81 of 96 points, including all 21 3-point essential specific standards) constructive recommendations will be sent to the instructor/course developer. The instructor/course developer can schedule a meeting

with instructional designers to revise the course according to the recommendations. All courses reviewed by QM review team are expected to meet the standards after necessary design improvements.

Statement of Problem

Research indicates that there are many factors that can affect online course quality. The factors include course design, course delivery, infrastructures, learning management system, faculty readiness, student readiness, etc. Course design is one of the critical pieces in the quality control process as it will affect the course delivery and the overall success of online programs. Quality Matters (QM) is a process and a tool to continuously improve online course quality (Shattuck, 2012). The 2011-2013 edition of QM rubric standards for higher education include eight general categories with 41 specific standards addressing different aspects of online course design. Each of the standards is designed based on rigorous research in the field of online and blended learning. A team of three certified QM peer reviewers will review online courses according to QM annotated standards and provide constructive feedback to course developers. Although QM peer reviewers are asked to take a student's point of view when reviewing online courses there still might be a gap between the reviewers and the students who take the online course. Therefore, it is necessary to collect feedback from students about the course design.

This study attempts to achieve three objectives. First it attempts to validate the instrument design based on QM Standards to measure online course design. Second it attempts to analyze the data and understand to what degree the selected courses meet QM standards from a student's perspective. Third it attempts to identify if there is a gap between student's perspective and QM certified reviewers' perspective about QM essential standards.

Method

Instrument

Online course design evaluation, a questionnaire with 27 questions in Likert scale (to little or no extent 1-5 to a great extent) and three open-ended questions was designed based on Quality Matters standards by the instructional design team. Feedback was also obtained from a professor in the field of research and measurement. The instrument focuses on design aspect of online courses.

Data Collection

Student Data

The instrument has been used at the university to collect feedback from students about the design aspect of online courses since fall 2011. The project team identified three online courses for this project. One course was offered in fall 2011 and 35 students completed the course design evaluation survey and two courses were offered in spring 2012, 18 students completed the course design

evaluation in the first one, and 20 students completed the course design evaluation survey in the second course.

Reviewer Data

Three reviewer reports have been collected on each of the three courses. All three reviewers are Quality Matters certified reviewers and they were trained to review online courses from a student's point of view. However, this review is not an official review and none of the reviewers are subject matter experts in the field of study of these courses and no master reviewers are in the review team.

Data Coding and Analysis

To achieve the first two objectives of the project data collected from three online courses were analyzed separately with Winsteps. Winsteps is a Windows-based software that assists with many applications of the Rasch model, particularly in the areas of educational testing, attitude surveys and rating scale analysis (Linacre, J. M., 2009).

To achieve the third objective of this project the data were treated. Students' results are converted into a measure that is comparable to the reviewers' rating. Student responses of 'To a great extent "4" or To a very great extent "5" are used as at or above 85% level and coded as "1". Student responses of 'To a moderate extent "3", To some extent "2" and To little or no extent "1" are used as below 85% level and coded as "0". According to the majority rule principle if 2/3 of the students selects 'To a great extent "4" or To a very great extent "5" for an item in the survey then the course meets that specific standard from a student's perspective. See Tables 1, 2, and 3.

Three QM certified peer reviewers reviewed the three courses according to QM standards and recorded their scores in a spreadsheet. If a standard is met, "1" is recorded for the standard. If a standard is not met, "0" is recorded for the standard. If two (2/3) of the peer reviewers assigned a score to a specific standard then the course meets that standard from a peer reviewer's perspective. See Tables 1, 2, and 3.

The data were treated in a spreadsheet and analyzed with SPSS. A Nonparametric Mann-Whitney U test (2 independent samples) was used to evaluate a difference in medians between the two groups (students and peer reviewers). The two groups are different even though peer reviewers are asked to take a student's view when completing course reviews. They are independent of each other.

Results

Course A

Thirty-five out of the 44 students completed the course design evaluation survey with a response rate of 79.55%. The person reliability is 0.83 and the item reliability is 0.48.

The item statistics indicate that Item 1 (MNSQ =3.31) will need to be revised and Item 16 (MNSQ=3.13) will need to be revised or dropped if the instrument is used in the future.

The Item Map (Fig. 1) indicates that there are three tiers regarding course quality from a student's perspective. Tier I contains the items that students strongly agreed with, which indicates the course meets the Standards 2.1 (item 4), 3.1 (item 14), 3.2 (item 15), 6.1 (item 20), 3.3 (item 7) to a great extent. Tier II contains the items that students agreed with, which indicates that course meets those standards to a moderate extent. Tier III contains items that students agreed with to some extent, which indicates that the course does not meet the Standards 7.1 (item 24), 5.1 (item 13), and 8.1 (item 25), or just meets the standards.

Fig. 1 Course A Item Map



Course B

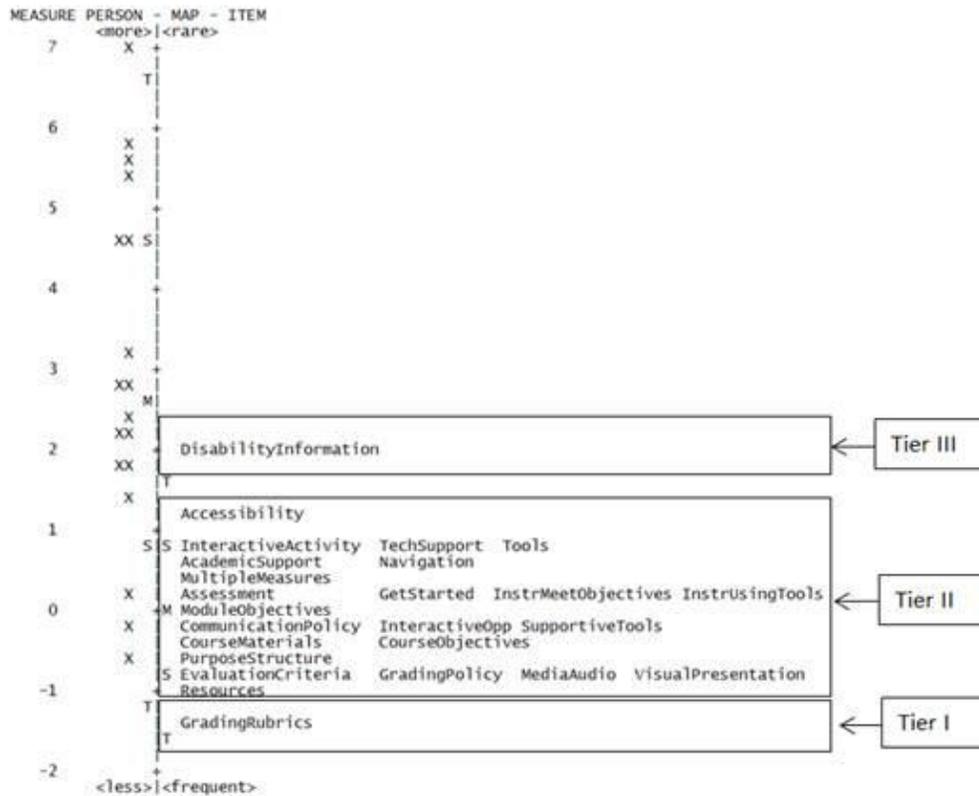
Eighteen out of the 38 students completed the course design evaluation survey with a response rate of 47.37%. The person reliability is 0.95 and the item reliability is 0.63.

The item statistics indicate that Item 14 (MNSQ = 2.29) needs to be revised if the instrument is used in the future.

The Item Map (Fig. 2) indicates that there are three tiers regarding course quality from a student's perspective. Tier I contains the items that students strongly agreed with, which indicates the course meets the Standards 3.3 (item 7), 3.2 (item 15), 3.2 (item 15) to a great extent. Tier II contains the items that students agreed with, which indicates that course meets those standards to a moderate

extent. Tier III contains items that students agreed with to some extent, which indicates that the course does not meet the Standards 7.2 (item 24) and 8.1 (item 25), or just meets the standards.

Fig. 2 – Course B Item Map



Course C

Twenty out of the 22 students completed the course design evaluation survey with a response rate of 90.91%. The person reliability is 0.96 and the item reliability is 0.78.

The item statistics indicate that Item 10 (MNSQ=2.83) will be dropped and Item 12 (MNSQ=2.64) and Item 6 (MNSQ=2.60) will need to be revised if the instrument is used in the future.

The Item Map (Fig. 3) indicates that there are three tiers regarding course quality from a student's perspective. Tier I contains the item that students strongly agreed with, which indicates the course meets the Standard 2.2 (item 5) to a great extent. Tier II contains the items that students agreed with, which indicates that course meets those standards to a moderate extent. Tier III contains the item that students agreed to some extent, which indicates that the course does not meet the Standard (item 10) a non-essential standard.

Fig. 3 – Course C Item Map



To achieve the third objective of this project the data were treated as follows:

- Students' results are converted into a measure that is comparable to the reviewers' rating. Student responses of To a great extent "4" or To a very great extent "5" are used as at or above 85% level and coded as "1". Student responses of To a moderate extent "3", To some extent "2" and To little or no extent "1" are used as below 85% level and coded as "0". According to the majority rule principle if 2/3 of the students selects To a great extent "4" or To a very great extent "5" for an item in the survey then the course meets that specific standard from a student's perspective. See Tables 1, 2, and 3.
- Three QM certified peer reviewers reviewed the three courses according to QM standards and recorded their scores in a spreadsheet. If a standard is met, "1" is recorded for the standard. If a standard is not met, "0" is recorded for the standard. If two (2/3) of the peer reviewers assigned a score to a specific standard then the course meets that standard from a peer reviewer's perspective. See Tables 1, 2, and 3.

The data were treated in a spreadsheet and analyzed with SPSS. A Nonparametric Mann-Whitney U test (2 independent samples) was used to evaluate a difference in medians between the two groups

(students and peer reviewers). The two groups are different even though peer reviewers are asked to take a student's view when completing course reviews. They are independent of each other.

Course A

Students reported that the course meets all the essential standards except Standards 7.1, 7.2, and 8.1 as measured by the instrument developed by the research team. The review results conducted by the three certified peer reviewers (none of the peer reviewers served as an SME on this review) also indicate that the course meets most of the essential standards except Standards 2.1, 2.2, 2.4, 3.3, and 8.1.

Table 1

Essential Standards	Student Results	Peer Reviewer Results	Items
1.1	YES	YES	2
1.2	YES	YES	1
2.1	YES	NO	4
2.2	YES	NO	5
2.4	YES	NO	6
3.1	YES	YES	14
3.2	YES	YES	15
3.3	YES/YES	NO	7, 16
4.1	YES	YES	8
5.1	YES	YES	13
5.2	YES	YES	12
6.1	YES	YES	20
6.3	YES	YES	21
7.1	NO	YES	22
7.2	NO	YES	24
8.1	NO	NO	25

No statistical differences were detected regarding those standards except Standards 2.1, 2.4, and 3.3. The two groups differed significantly regarding Standard 2.1 $U = 1.000$, $Z = -5.192$, $P = .000$, Standard 2.4 $U = 7.500$, $Z = -3.393$, $P = .001$, and Standard 3.3 (item 7) $U = 22.000$, $Z = -2.819$, $P = .005$. See Fig. 4.

Fig.4

	S2.1	S2.2	S2.4	S3.3A	S4.1
Mann-Whitney U	1.500	23.500	7.500	22.000	43.500
Wilcoxon W	7.500	29.500	13.500	28.000	673.500
Z	-5.192	-2.485	-3.393	-2.819	-.771
Asymp. Sig. (2-tailed)	.000	.013	.001	.005	.441
Exact Sig. [2*(1-tailed Sig.)]	.000 ^a	.122 ^a	.008 ^a	.109 ^a	.644 ^a

a. Not corrected for ties.

b. Grouping Variable: Role

Course B

Students reported that the course meets all of the essential standards except Standards 7.1, 7.2 and 8.1 as measured by the instrument developed by the research team. However, the review results conducted by the three certified peer reviewers (none of the peer reviewers served as an SME on this review) indicate that at least five course essential Standards 1.1, 1.2, 2.2, 2.4 and 3.2 do not meet the standards. Meanwhile, the peer reviewers' results indicate that the course does meet Standards 7.1, 7.2, and 8.1, but the students do not agree with that decision.

Table 2

Essential Standards	Student Results	Peer Reviewer Results	Items
1.1	YES	NO	2
1.2	YES	NO	1
2.1	YES	YES	4
2.2	YES	NO	5
2.4	YES	NO	6
3.1	YES	YES	14
3.2	YES	NO	15
3.3	YES/YES	YES	7, 16
4.1	YES	YES	8
5.1	YES	YES	13
5.2	YES	YES	12
6.1	YES	YES	20
6.3	YES	YES	21
7.1	NO	YES	22
7.2	NO	YES	24
8.1	NO	YES	25

No statistical differences were detected regarding those standards except Standards 2.2, and 3.2. The two groups differed significantly regarding Standard 2.2 $U = 4.500$, $Z = -2.887$, $P = .004$, and Standard 3.2 $U = 3.000$, $Z = -3.266$, $P = .001$. See Fig. 5

Fig. 5

	S1.2	S5.2	S5.1	S3.1	S3.2	S3.3B
Mann-Whitney U	12.000	21.000	19.500	25.500	3.000	22.500
Wilcoxon W	18.000	192.000	172.500	31.500	9.000	193.500
Z	-2.214	-.886	-.916	-.192	-3.266	-.745
Asymp. Sig. (2-tailed)	.027	.376	.360	.847	.001	.456
Exact Sig. [2*(1-tailed Sig.)]	.153 ^a	.600 ^a	.546 ^a	.887 ^a	.011 ^a	.669 ^a

a. Not corrected for ties.

b. Grouping Variable: Role

Test Statistics^b

	S2.1	S2.2	S2.4	S3.3A	S4.1
Mann-Whitney U	24.000	4.500	16.500	24.000	19.500
Wilcoxon W	195.000	10.500	22.500	195.000	190.500
Z	-.592	-2.887	-1.291	-.592	-1.021
Asymp. Sig. (2-tailed)	.554	.004	.197	.554	.307
Exact Sig. [2*(1-tailed Sig.)]	.814 ^a	.017 ^a	.307 ^a	.814 ^a	.471 ^a

a. Not corrected for ties.

b. Grouping Variable: Role

Course C

Students reported that the course meet only a few essential standards 1.2, 2.1, 2.2, 2.4, 3.2, 4.1, and 5.2 as measured by the instrument developed by the research team. However, the review results conducted by the three certified peer reviewers (none of the peer reviewers served as an SME on this review) indicate that only three of the essential standards are not met, 2.2, 7.2 and 8.1. The peer reviewers' results for standards 7.2 and 8.1 are in conformity with the students' results that the course does not meet these standards.

Table 3

Essential Standards	Student Results	Peer Reviewer Results	Items
1.1	NO	YES	2
1.2	YES	YES	1
2.1	YES	YES	4
2.2	YES	NO	5
2.4	YES	YES	6
3.1	NO	YES	14
3.2	YES	YES	15
3.3	NO/YES	YES	7, 16
4.1	YES	YES	8
5.1	NO	YES	13
5.2	YES	YES	12
6.1	NO	YES	20
6.3	NO	YES	21
7.1	NO	YES	22
7.2	NO	NO	24
8.1	NO	NO	25

No statictical differences were detected regarding those standards except Standard 2.2. The two groups differed significantly regarding Standard 2.2 $U = 11.500$, $Z = -2.892$, $P = .004$. See Fig. 6

Fig. 6

Test Statistics^b

	S2.1	S2.2	S2.4	S3.3A	S4.1
Mann-Whitney U	22.500	11.500	21.000	19.500	21.000
Wilcoxon W	232.500	17.500	231.000	229.500	231.000
Z	-.957	-2.892	-1.079	-1.202	-1.079
Asymp. Sig. (2-tailed)	.338	.004	.280	.230	.280
Exact Sig. [2*(1-tailed Sig.)]	.514 ^a	.094 ^a	.457 ^a	.355 ^a	.457 ^a

a. Not corrected for ties.

b. Grouping Variable: Role

Discussions

We investigated the differences between students and peer reviewers regarding the essential standards in three online courses. When the courses were approved for design the faculty course developers were provided a copy of the Quality Matters rubrics in the beginning of the course development process. Instructional designers, who are certified QM peer reviewers, are available for individual consultations in the design and development. The faculty course developers are fully aware of the standards and the standards need to be incorporated into the course design.

As reported in the results section, in course A the results reported by students and peer reviewers differed significantly in regards to Standard 2.1 (The course learning objectives describe outcomes that are measurable.) and Standard 2.4 (Instructions to students on how to meet the learning objectives are adequate and stated clearly.) For Standard 2.1, the students were asked to report whether course objectives are clearly presented in the course syllabus. While the reviewers look for solid evidences for measurable learning objectives. For Standard 2.4, both reviewers and students were asked to report whether clear instructions on how students should meet the learning objectives are articulated in the course. Students reported that the instructions are available. However, reviewers do not agree and one reviewer state:

Standard 2.4 calls for clear instructions on how students should meet the learning objectives. The course design does a good job in providing students with a brief introduction to each Chapter topic; however, it is somewhat difficult to understand which learning activities, resources, assignments, and assessments support the learning objectives for each unit week. It is important to help students connect the dots between chapter level objectives and the assigned activities and assessment for the week.

Apparently peer reviewers are looking for above average at approximately 85%. Students might think the brief introduction to each chapter provides instructions on how to achieve the learning objectives. The overall satisfaction of the course might also affect students' rating on the standards

as the majority of the students rate the course as excellent. A third factor that might contribute to the difference is that student satisfaction of the teacher. The answers to the open ended questions indicate the professor is excellent and cares about their learning, as one student stated:

The professor always leads a very informative, fun, and creative class and this one was not an exception. I learned a plethora of new things from the reading, assignments, and independent studies throughout the semester.

In course B the results reported by students and peer reviewers differed significantly in regards to Standard 2.2 (The module/unit learning objectives describe outcomes that are measurable and consistent with the course-level objectives.) and Standard 3.2 (The course grading policy is stated clearly.) For Standard 2.2, the students were asked to report whether module/unit objectives are clearly stated in each unit. While the reviewers look for solid evidences for measurable learning objectives and alignment to course level objectives. For Standard 3.2, both reviewers and students were asked to report whether grading policy is clearly articulated in the course. Students reported that the grading policy is available. However, the majority of the reviewers think that the policy is not clear enough. One reviewer states:

Standard 3.2 asks for a clear, written description on how student's grades will be calculated, for instance, the total points for each assignment, the percentages or weights for each component of the course grade. It would be helpful to provide an overall list of assignments, points, percentages or weights in the syllabus so that students are acknowledged upfront on how they will be evaluated without digging deeper in the Unit content pages.

Again in this course, the overall satisfaction of the course and the instructor might also affect students' rating on the standards as students stated:

Overall, this course has given me a lot of valuable information that I can use in the classroom.

I appreciate all the help given to me throughout the years. This was not an easy thing to accomplish, but I have and I will always remember all those that have helped me succeed.

In course C the results reported by students and peer reviewers differed significantly in regards to Standard 2.2 (The module/unit learning objectives describe outcomes that are measurable and consistent with the course-level objectives). The students were asked to report whether module/unit objectives are clearly stated in each unit. While the reviewers look for solid evidences for measurable learning objectives. One reviewer states:

Standard 2.2 requires that the module/unit learning objectives describe outcomes that are measurable and consistent with the course level objectives. Many of the module level learning objectives are overlapping. It is suggested that you develop unique learning objectives for each module based on Bloom's taxonomy.

The peer reviewers expect that the course meets this standard at or above 85% level and use the opportunity for the faculty developer to improve their course and ultimately meets the standards.

Conclusions



Most of the items in the instrument designed according to the Quality Matters standards work well to measure the design perspective of online courses. However, the misfit items will be dropped (item 10) or revised (items 1, 6, 14, and 16) according to the analysis results. The results from students indicate three tiers (Tier I: to a great extent, Tier II: to a moderate extent, and Tier III: to little or some extent) in regard to meeting the standards in the three courses.

The results on most of the standards evaluated in this study provided by both reviewers and students are the same, indicating that both peer reviewers and students take the same point view in regard to evidences of the standards. However, they differed significantly regarding three of the essential standards. One factor that might cause the discrepancy is that reviewers look for solid evidences for measurable learning outcomes while students look for clearly articulated objectives. The second factor might be that instructors clarified those unclear design aspects via email while the course was delivered and not available to the reviewers. The third factor might be that the reviewers look for above average approximately 85%, which students look for the basic elements regarding the standards. The reviewers also think that the overall satisfaction of the course and the instructor might also affect students' rating about the essential standards. Further study is needed to investigate the causes of discrepancy.

References:

- Chickering, A., & Gamson, Z. (Eds.). (1987). Seven principles for good practice in undergraduate education. *AAHE Bulletin*, 38(7) 3-7.
- Higher Learning Commission. (2007). Statement of commitment by the regional accrediting commissions for the evaluation of electronically offered degree and certificate programs. Retrieved from <http://www.ncahlc.org/index.php?&Itemid=236>
- Holmberg, B. (1995). *Theory and practice of distance education*. (2nd ed.). London and New York: Routledge.
- Knowles, E. E., & Kalata, K. (2010, June). The impact of QM standards on learning experiences in online courses. [2009 QM Research Grant]. Presentation at the 2nd Annual Quality Matters Conference, Oak Brook, IL.
- Linacre, J.M. (2009). *Winsteps* (Version 3.68.0) [Computer Software]. Chicago: Winsteps.com.

Quality Matters (2011). Quality Matters rubric standards 2011-2013 edition Retrieved from <https://www.qmprogram.org/rubric>

Ralston-Berg, P. & Nath, L. (2008). What makes a quality online course? The student perspective. 24th Annual Conference on Distance Teaching & Learning. Retrieved from http://www.uwex.edu/disted/conference/Resource_library/proceedings/08_12876.pdf

Ralston-Berg, P. (2011). What makes a quality online course? The student perspective. Presentation at the 3rd Annual Quality Matters Conference, Baltimore, MD.

Rosner, B. and Grove, D. (1999). Use of the Mann Whitney U-Test for clustered data. *Statist. Med.* 18, 1387-1400.

Shattuck, K. (2012)  What we're learning from Quality Matters-focused research: Research, practice, continuous improvement. Retrieved from https://www.qmprogram.org/files/Learning%20from%20QM%20Focused%20Research%20Paper_0.pdf

Shattuck, K. (2003)  Quality matters: Collaborative program planning at a state level. *Online Journal of Distance Learning Administration*. Retrieved from <http://www.westga.edu/~distance/ojdl/fall103/shattuck103.htm>

The Higher Education Program and Policy Council (2000). Distance education: Guidelines for good practice. Retrieved from http://www.aft.org/pubs-reports/higher_ed/distance.pdf