Journal of Nursing Measurement, Volume 16, Number 1, 2008

Field Testing, Refinement, and Psychometric Evaluation of a New Measure of Quality of Care for Assisted Living

Marilyn J. Rantz, PhD, RN, FAAN
Sinclair School of Nursing and Family and Community Medicine, School of Medicine, University Hospital, University of Missouri, Columbia, MO

Myra A. Aud, PhD, RN
Sinclair School of Nursing, University of Missouri

Mary Zwygart-Stauffacher, PhD, RN, FAAN
College of Nursing and Health Sciences, University of Wisconsin–Eau Claire

David R. Mehr, MD, MS
Family and Community Medicine, School of Medicine, University of Missouri

Gregory F. Petroski, PhD
School of Medicine, University of Missouri

Steven V. Owen, PhD
Department of Pediatrics, School of Medicine, University of Texas Health Science Center at San Antonio

Richard W. Madsen, PhD
Department of Statistics, University of Missouri

Marcia Flesner, RN, PhD
Sinclair School of Nursing, University of Missouri

Vicki Conn, PhD, RN
Sinclair School of Nursing, University of Missouri

Meridean Maas, RN, PhD, FAAN
College of Nursing, University of Iowa

Field test results are reported for the Observable Indicators of Nursing Home Care Quality Instrument–Assisted Living Version, an instrument designed to measure the quality of care

DOI: 10.1891/1061-3749.16.1.16

in assisted living facilities after a brief 30-minute walk-through. The OIQ-AL was tested in 207 assisted-living facilities in two states using classical test theory, generalizability theory, and exploratory factor analysis. The 34-item scale has a coherent six-factor structure that conceptually describes the multidimensional concept of care quality in assisted living. The six factors can be logically clustered into process (Homelike and Caring, 21 items) and structure (Access and Choice; Lighting; Plants and Pets; Outdoor Spaces) subscales and for a total quality score. Classical test theory results indicate most subscales and the total quality score from the OIQ-AL have acceptable interrater, test–retest, and strong internal consistency reliabilities. Generalizability theory analyses reveal that dependability of scores from the instrument are strong, particularly by including a second observer who conducts a site visit and independently completes an instrument, or by a single observer conducting two site visits and completing instruments during each visit. Scoring guidelines based on the total sample of observations ($N = 358$) help guide those who want to use the measure to interpret both subscale and total scores. Content validity was supported by two expert panels of people experienced in the assisted-living field, and a content validity index calculated for the first version of the scale is high (3.43 on a four-point scale). The OIQ-AL gives reliable and valid scores for researchers, and may be useful for consumers, providers, and others interested in measuring quality of care in assisted-living facilities.

Keywords: assisted living; residential care; nursing homes; quality of care; generalizability theory; reliability and validity

The number of assisted living facilities in the United States increased rapidly during the 1990s with an estimated 33,000 facilities caring for approximately 800,000 residents (NCAL, 2001). While the concept of assisted living is not new, having its roots in congregant housing in the 1960s (Pruchno & Rose, 2000), today's assisted-living facilities provide housing and supportive services with some capacity to meet both scheduled and unscheduled health care needs of an older and more frail adult population (Allen, 1999; Brummett, 1997). Today's assisted-living facilities also emphasize the social model of care rather than the medical model of care. Frail older adults and their families are attracted to homelike environments that do not resemble hospitals or traditional nursing homes and that stress independence, autonomy, and continuation of prior lifestyles.

Quality of care or life has been only minimally addressed in the slowly growing body of research literature related to assisted-living facilities. Quality of life in assisted-living facilities has been linked to the facility's social environment and whether that environment encourages social participation and family involvement (Mitchell & Kemp, 2000) and to positively rated resident–staff communication and staff involvement in care planning (Zimmerman et al., 2005). Surveys of resident satisfaction related to staff, direct care, services, activities, amenities, and environment in assisted-living facilities have been conducted in four states as elements of state-sponsored quality-improvement initiatives (Lowe, Lucas, Castle, Robinson, & Crystal, 2003). Quality of care in assisted living has had some exploration that resulted in a preliminary measure of quality of care in this long-term care environment (Aud, Rantz, Zwygart-Stauffacher, & Manion, 2004).

The question for prospective residents, their families, and health care professionals is this: When you walk into one of these assisted-living facilities, how do you know that the care is of high quality? Although the assisted-living facility philosophy places a high value on individualized services provided to older adult residents in a homelike environment, there is great variation in operation of assisted-living facilities across the United States. This is due in part to the absence of federal regulations that delineate minimal standards of practice and by the absence of a national consensus on the definition of an assisted-living facility. Instead, several states have promulgated their own regulations for the assisted-living segment of institutional long-term care while other states have not, resulting in wide variation across states. Variation among states begins with the naming of such facilities—approximately 26 different names are in use—then continues with variations in services allowed or disallowed and in admission/discharge criteria (Mollica, 2003).

Unlike federally regulated skilled nursing facilities, quality indicators have not been developed for assisted living facilities (Aud & Rantz, 2004). Consumers and health care professionals alike are perplexed by the absence of a single published standard of practice. The absence of a single standard also impedes the establishment of a reliable and valid method to measure care quality in assisted-living facilities.

To rectify the absence of a reliable and valid method of care quality, Rantz and colleagues adapted her Observable Indicators of Nursing Home Care Quality Instrument (OIQ-NH; Rantz et al., 2000, 2006; Rantz & Mehr, 2001; Rantz & Zwygart-Stauffacher, 2005; Rantz, Zwygart-Stauffacher, & Flesner, 2005), creating a 34-item tool to measure assisted-living facility care quality. Two panels of experts assessed the validity of this initial adaptation. Interrater and test–retest reliability were assessed with a modest sample of 35 facilities with promising results (Aud et al., 2004). This article describes the subsequent revision and more extensive psychometric testing of the Observable Indicators of Nursing Home Care Quality–Assisted Living Version (OIQ-AL).

PREVIOUS CONTENT VALIDITY EVIDENCE FOR THE OIQ-AL

A panel of experts reviewed the initial assisted-living/residential-care facility version for content validity: Do the instrument's items cover the domain we want to measure? Relevance of each item was rated on a four-point scale: not relevant, somewhat relevant, quite relevant, very relevant (Waltz, Strickland, & Lenz, 2005). The content validity index for the total scale was 3.43 with only five items having average ratings less than 3.00, none of which were less than 2.0 (Aud et al., 2004).

Development of the OIQ-AL

After the initial field-test in 35 facilities, the instrument was revised but retained its overall structure of a five-point rating scale for each question with anchoring descriptors for each point. Similar to the nursing home version, there is a user's guide that provides directions about how to perform a 20–30 minute tour of a facility prior to answering the questions. Items include topics of resident grooming, personalization and cleanliness of resident rooms, odors, conversations between staff and residents, and others.

After field-testing version 1 in the 35 residential care facilities, we carefully reviewed the results of the reliability and validity studies for possible item revisions to improve item and subscale performance. Two of the original 34 items were dropped. Two items asking separately about pets and plants were combined. The statistician expressed concern that responses to several items were clustered at one end of the five-point scale rather than distributed among the range of responses. Therefore, the anchors of these items with poor distribution of response were revised.

A new panel of five assisted-living/residential-care facility experts met as a focus group to discuss the revised instrument's items and whether additional items were indicated. The experts suggested inclusion of items related to food choices, availability of snacks, access to telephones, and access to e-mail and computers. The research team drafted new items and a second version with 41 items was ready for evaluation.

Study Purpose

The primary aim of the study was to advance the development of the OIQ-AL for use by researchers studying the dimensions of care quality in assisted living. A secondary study aim was to assess the psychometric properties of the OIQ-AL when used by nurse observers.

METHODS

Sample

A major field-study of the OIQ-NH was funded by the National Institute of Nursing Research (NINR) of the National Institutes of Health (NIH) to further develop the nursing-home version of the quality of care measurement instrument (Rantz et al., 2005, 2006; Rantz & Zwygart-Stauffacher, 2005). The nursing-home study facilitated data collection in assisted-living facilities in two states, Missouri and Wisconsin. Both freestanding and assisted-living facilities located on campuses with nursing homes that in many cases included independent living were contacted to participate in a field test of the OIQ-AL. Facilities were required to have more than 12 residents and have a primary population of older adults to be included in the study. Registered nurse (RN) observers made initial and return visits to tour assisted-living facilities and completed the instrument following the directions in the user's guide developed for the OIQ-AL.

A total of 207 assisted-living facilities were visited, 189 in Missouri and 18 in Wisconsin; of these, observations from 207 different assisted-living facilities were complete and usable for analyses. To measure interrater reliability and test–retest stability, simultaneous visits by two RNs were made to 73 assisted-living facilities, and return visits by one of the RNs were made to 77 assisted-living facilities. One return visit was excluded from the analysis due to excessive time between initial and return visit, resulting in 76 usable test–retest visits. Return visits followed the initial visit by an average of 8 days (SD = 3) with a minimum of 6 days and maximum of 20 days. Table 1 summarizes the visits made in each state by RN observers.

Overview of Data Analysis

Analysis of the field-testing of the 40-item second version of the assisted-living/residential-care facility version of the OIQ included the following: item analysis with

TABLE 1. Summary of Visits Made by State

State	Observer (*n*)	Facilities Visited	Total Visits	Facilities Used in Interrater Reliability	Facilities in Test–Retest Reliability
Missouri	12	189	308	57	61
Wisconsin	4	18	50	16	16
Total	16	207	358	73	77

review of Spearman Rho and weighted Kappa statistics for interrater and test–retest reliability for each item, exploratory factor analysis, recalculation of interrater and test–retest reliability for the full scale and subscales, calculation of Cronbach's alpha for the full scale and subscales, generalizability theory (GT) analysis of sources of error, and a decision study (D-study) to estimate the numbers of raters and visits necessary to maximize dependability of the instrument. The outcome of these analyses is a 34-item instrument that can reliably assess care quality in assisted-living facilities. Because the 34-item version of the instrument was closely linked to version 10 of the nursing home OIQ-NH instrument, it was officially denoted as OIQ-AL Version 10AL. The following describes the steps of the analysis and results of each step.

Psychometric Analyses

Items with insufficient variability do not discriminate between facilities, so the initial step in the psychometric analysis was to examine the frequency distribution of each item. Spearman's rank correlation coefficient and weighted Kappa coefficients were calculated to evaluate the interrater and test–retest stability of each item.

Factor Analysis Procedures. Following the item analysis, a number of exploratory factor analyses (EFAs) were performed to evaluate the dimensionality of the instrument. The total sample size for the EFA was 207, with each facility used only once in the analysis. In the case of multiple observers or visits, the first visit from only one observer was used for the factor analysis. This sample size was judged as adequate, as Gorsuch (1983) suggested that sample size in EFA is somewhat uncertain because one cannot know in advance the strength of the latent variables. In any case, he proposed an "absolute minimum ratio of five individuals to every variable, but not less than 100 individuals" (p. 332). The lower boundary of $N = 100$ might work, he said, if item communalities are high and each factor has "many" indicators.

Factor extraction used the iterated principal factor method followed by an oblique Promax rotation. Factor solutions consisting of four to nine factors were considered. Choosing the number of factors was guided primarily by the interpretability and simple structure of the factor solutions. A variety of parallel analysis (Horn, 1965) was also performed to contrast expert judgment with a data-driven approach to identifying the number of factors.

The idea behind parallel analysis is that estimated eigenvalues differ from their population values because of sampling error and that simulation methods can be used to estimate this error. The eigenvalues of correlation matrices for randomly generated data should be close to 1.0 and exceed unity only because of sample variation. The implementation of parallel analysis used here was to simulate uncorrelated item scores for the same number of items and with the same marginal frequency distributions as in the study data. Correlation matrices were calculated and eigenvalues extracted for 100 simulated data sets each with the same number of observations ($n = 207$) as in the study data. The mean and 95th percentiles of the simulated eigenvalues were overlaid on a scree plot with the eigenvalues from the OIQ-AL data. A guide to the number of factors to retain is that point at which the plot of real eigenvalues rises above the plot for the mean or 95th percentile of the eigenvalues from random data (Hayton, Allen, & Scarpello, 2004). In this study the choice to use the mean or 95th percentile made no substantive difference. Although there is literature supporting the use of parallel analysis to choose the number of factors (e.g., Hayton, Allen, & Scarpello, 2004; Zwick & Velicer, 1986) we stress that in this study parallel analysis was used as an adjunct to expert judgment and not as the sole tool for identifying the number of factors.

Generalizability Theory (GT) Analysis Procedures. Observed scores usually include multiple sources of error. Variability across facilities, occasions, raters, and the sampled items are all sources of error. In classical test theory (CTT) these are addressed by examining test–retest and interrater reliability. A shortcoming of CTT is that there is no overall evaluation of error in which all the sources of error can be evaluated simultaneously. Generalizability theory (GT; Brennan, 2001) represents an extension of CTT that allows for a more comprehensive evaluation of measurement error. With GT, observed score variation is partitioned into variation due to differences in the objects of measurement (assisted-living facilities), plus differences among raters (RNs), occasions (visits), and variation due to items. Furthermore, with GT any specific set of items is regarded as a sample from a conceptually infinite pool of items and thus also represent a source of variation.

Because the full data set for this study contained the multiple items of the OIQ-AL collected by two observers on a single visit and by one observer on two occasions, a GT study was conducted to combine these three sources of error (items, raters, occasions, and their interactions) into a single coefficient. Once calculated, the G-coefficient has similar properties as the conventional reliability estimate; it ranges from 0.00 to 1.00, with larger values indicating greater reliability. If measurements were error free, then all variability in observed scores would be attributable to differences in the objects of measurement, the assisted-living facilities in this current study. If most of the variability is attributable to error, that is, sources other than differences in assisted-living facilities, then the G-coefficient would be close to zero. Thus the G-coefficient represents the proportion of total variation in observed scores attributable to differences in the facility, sometimes also called the proportion of dependable variance.

GT also makes a distinction between the dependability of relative judgments versus absolute judgments and provides coefficients for both circumstances. Relative judgments focus on consistent rank ordering of measured objects. Absolute decisions are those that interpret the magnitude of individual scores without regard to their rank ordering. In educational testing, this distinction is usually termed *norm-referenced* versus *criterion-referenced.* Norm-referenced scores take meaning by giving an object an ordered position relative to others. Criterion referenced scores are used to establish a passing grade or to demonstrate mastery, without consideration of the performance of others. GT provides reliability-like coefficients for both relative and absolute decisions. It was decided that absolute decisions are not meaningful in the context of this study, and so only the G-coefficient for relative decisions is provided. We treated the number of items in each scale as fixed and estimated the G-coefficient for relative decision for each combination of one to five raters and one to five visits to an assisted-living facility.

RESULTS

Exploratory Factor Models

The review of item frequency response distributions revealed that all items showed sufficient variability to be retained for further analysis. One item (about staff seeming to know the residents so they are able to provide care) was excluded during the initial analysis because of low interrater agreement ($\rho = 0.09$; $\kappa = 0.13$). All other items were retained for the exploratory factor analysis. Factor solutions with oblique rotations consisting of four to nine factors were considered. Both expert judgment and the parallel analysis pointed to a six-factor solution as representing the highest degree of interpretability and consistency with clinical judgment. Two items were deleted because they lacked sufficient loading (≥ 0.3) on any factor, and an additional four items were discarded because of strong and roughly equivalent cross-loadings on two factors. Other items that loaded on more than one factor were assigned to the factor with the higher loading if the difference between factor loadings was greater than 0.10. Table 2 is a summary of the six factors and the range of factor loadings. Details of factor loadings for the individual items are available from the authors.

Interrater Reliability

After the factor analyses, the final 34-item instrument was analyzed using the simultaneous interrater observations made by two RNs to 73 assisted-living/residential-care facilities. Table 3 displays the Spearman's Rho correlations with upper and lower confidence intervals. Five of the six subscales and the total scale had very good interrater reliability, while the plants and pets subscale showed substandard interrater reliability.

TABLE 2. Six Factors of the OIQ-AL With Range of Factor Loadings

Factor Number	Number of Items	Factor Name	Range of Factor Loadings
1	12	Homelike	0.55–0.79
2	9	Caring	0.42–0.87
3	7	Access and choice	0.36–0.65
4	2	Lighting	0.67–0.73
5	2	Pets and plants	0.72–0.81
6	2	Outdoor spaces	0.38–0.76
Total scale	34		

TABLE 3. Interrater Reliability (Spearman's Rho) (73 Assisted-Living Facilities)

		Spearman's Rho		
Subscales	Number of Items	Correlation	Lower CI	Upper CI
1 Homelike	12	0.94	0.90	0.96
2 Caring	9	0.56	0.38	0.70
3 Access and choice	7	0.68	0.52	0.78
4 Lighting	2	0.70	0.56	0.80
5 Plants and pets	2	0.33	0.10	0.52
6 Outdoor spaces	2	0.73	0.59	0.82
Total scale	34	0.76	0.64	0.84

Note. CI = confidence interval.

Test–Retest Reliability

There were return visits to 77 assisted-living/residential-care facilities. One observation was excluded because of an extended period between observations. The remaining 76 return visit observations were used in the test–retest analysis; the average time between visits was 8 days (*SD* = 3) with a range of 6 to 20 days. Table 4 displays the test–retest results, which demonstrate generally good test–retest reliability of scores from the instrument. As before, the pets and plants subscale was the lowest correlation.

Internal Consistency

Items remaining after the factor analysis were also used in calculation of Cronbach's coefficient alpha for the six subscales and the full scale. The sample was the complete sample of first visits to 207 different assisted-living/residential-care facilities in Missouri and Wisconsin. Table 5 displays the Cronbach's coefficient alpha that reveals excellent internal consistency for five of the six subscales and the total scale. The outdoor spaces subscale has poor internal consistency.

Generalizability Theory

Because different forms of reliability estimates in CTT (the above interrater, test–retest, and internal consistency) use different definitions of score consistency, there is no way to combine the various sources of error into an "overall" value. Generalizability theory

TABLE 4. Test–Retest Results (Spearman's Rho) (76 Assisted-Living Facilities)

Subscales	Number of Items	Spearman's Rho		
		Correlation	Lower CI	Upper CI
1 Homelike	12	0.91	0.86	0.94
2 Caring	9	0.64	0.48	0.75
3 Access and choice	7	0.85	0.78	0.90
4 Lighting	2	0.77	0.66	0.85
5 Plants and pets	2	0.60	0.43	0.73
6 Outdoor spaces	2	0.71	0.57	0.80
Full scale	34	0.82	0.72	0.88

Note. CI = confidence interval.

TABLE 5. Cronbach's Coefficient Alpha (207 Assisted-Living Facilities)

Subscales	Number of Items	Cronbach Coefficient Alpha
1 Homelike	12	0.94
2 Caring	9	0.92
3 Access and choice	7	0.78
4 Lighting	2	0.86
5 Plants and pets	2	0.88
6 Outdoor spaces	2	0.47
Full scale	34	0.94

(GT), however, provides such a framework, since it allows examining and summarizing multiple sources of measurement error that make up the total variation in a score (Brennan, 2001). This GT analysis used 230 observations made by more than one observer from 79 different assisted-living/residential-care facilities; facilities that were visited only once by one observer were not used. Table 6 displays the G-coefficients for the subscales and the total scale for different numbers of raters and different numbers of visits. The value of this analysis is that one can project how the dependability of the instrument's scores might be improved by increasing the numbers of observations or the number of raters. What is learned from this analysis is that increasing the number of raters from one to two does increase the dependability of the scores. For example, the total score for one rater going for one visit has a *G* estimate of 0.73; this is estimated to improve to 0.83 by going a second time, but little is gained by going three times (0.88). Similarly, increasing the visits of one rater from one to two will improve the dependability estimate from 0.73 to 0.83; but little is gained by sending a third rater (0.87). So before making a judgment about quality, someone using this instrument may want to visit the facility twice or have someone accompany him or her on one visit. It should be noted that with multiple raters or multiple visits the reliability of most subscales and the total score is very reliable with coefficients over 0.90. Figure 1 is a graphical display of the improvement in scores by increasing numbers of rater or numbers of visits, improving the dependability of the instrument.

Both the GT and the internal consistency results argue for the deletion of the outdoor spaces subscale; however, both interrater and test–retest correlations are good for this subscale. The raw data reveals very little variation in our sample of facilities for the scores of this subscale. Another conflicting result is the pets and plants subscale; it has good GT results and poor interrater results. The research team made the decision to retain the items of the outdoor spaces and the pets and plants subscales until the instrument is used more extensively in assisted living. Experts in assisted living who participated in our content validity panels believe outdoor spaces as well as pets and plants are important to quality of care.

Scoring Guidelines

Interpreting performance is an essential step for assisted-living facilities to take to determine actions to improve quality of care and for consumers to compare facilities as they make crucial choices for long-term care. To guide those interested in using the Observable Indicators Instrument, we used all data from all raters in both states to construct ranges of scores from which to guide interpretation of scores from the instrument. Examining the distribution of scores for all observations ($n = 358$), the research team decided to use the score at the 20th percentile as suggestive of quality problems in a facility, the score at the 80th percentile as suggestive of a quality facility, and the scores between as typical of most assisted-living facilities. This approach of selecting the 20th and 80th percentiles as benchmarks for interpretation of quality is similar to upper and lower control limit setting used by others (Dearmin, Brenner, & Migliori, 1995; Katz & Green, 1997; Kiefe, Woolley, Allison, Box, & Craig, 1994) and was the approach used in developing guidelines for the nursing home version, OIQ-NH (Rantz et al., 2006). The range of scores for the OIQ-AL observations is 57 to 167, with 116 at the 20th percentile and 151 at the 80th percentile. Therefore, the guidelines for interpretation of the OIQ total score are:

- A score above 151 suggests a quality facility.
- Scores between 116 and 151 suggest the typical assisted-living facility care quality.
- A score below 116 suggests a facility with quality problems.

TABLE 6. OIQ-AL Generalizability Coefficients (Relative Decisions) for 1–5 Raters and 1–5 Visits (230 Observations; 79 Assisted-Living Facilities)

			Number of Visits				
Scale	*n* Items	*n* Raters	Visits 1	Visits 2	Visits 3	Visits 4	Visits 5
Access & choice	7	1	0.618	0.723	0.766	0.790	0.805
	7	2	0.708	0.781	0.809	0.824	0.833
	7	3	0.745	0.803	0.824	0.836	0.842
	7	4	0.765	0.814	0.832	0.842	0.847
	7	5	0.777	0.821	0.837	0.845	0.850
Caring	9	1	0.680	0.786	0.829	0.852	0.867
	9	2	0.775	0.846	0.872	0.886	0.895
	9	3	0.812	0.868	0.888	0.898	0.904
	9	4	0.833	0.879	0.896	0.904	0.909
	9	5	0.845	0.886	0.900	0.908	0.912
Homelike	12	1	0.893	0.938	0.954	0.962	0.967
	12	2	0.933	0.960	0.969	0.973	0.976
	12	3	0.947	0.967	0.974	0.977	0.979
	12	4	0.955	0.971	0.976	0.979	0.981
	12	5	0.959	0.973	0.978	0.980	0.982
Lighting	2	1	0.837	0.902	0.926	0.938	0.946
	2	2	0.893	0.933	0.948	0.955	0.959
	2	3	0.914	0.945	0.955	0.961	0.964
	2	4	0.924	0.950	0.959	0.964	0.966
	2	5	0.931	0.954	0.961	0.965	0.968

Table 6. continued

Scale	*n* Items	*n* Raters	Number of Visits				
			Visits 1	Visits 2	Visits 3	Visits 4	Visits 5
Outdoor spaces	2	1	0.047	0.064	0.074	0.080	0.084
	2	2	0.068	0.088	0.098	0.104	0.108
	2	3	0.080	0.101	0.110	0.116	0.119
	2	4	0.088	0.109	0.118	0.123	0.126
	2	5	0.094	0.114	0.122	0.127	0.130
Plants & pets	2	1	0.817	0.858	0.873	0.880	0.885
	2	2	0.873	0.903	0.913	0.919	0.922
	2	3	0.893	0.919	0.928	0.933	0.935
	2	4	0.903	0.927	0.935	0.939	0.942
	2	5	0.910	0.932	0.940	0.944	0.946
Total	34	1	0.726	0.834	0.877	0.901	0.915
	34	2	0.829	0.898	0.923	0.937	0.945
	34	3	0.870	0.921	0.940	0.949	0.955
	34	4	0.892	0.934	0.948	0.956	0.961
	34	5	0.906	0.941	0.954	0.960	0.964

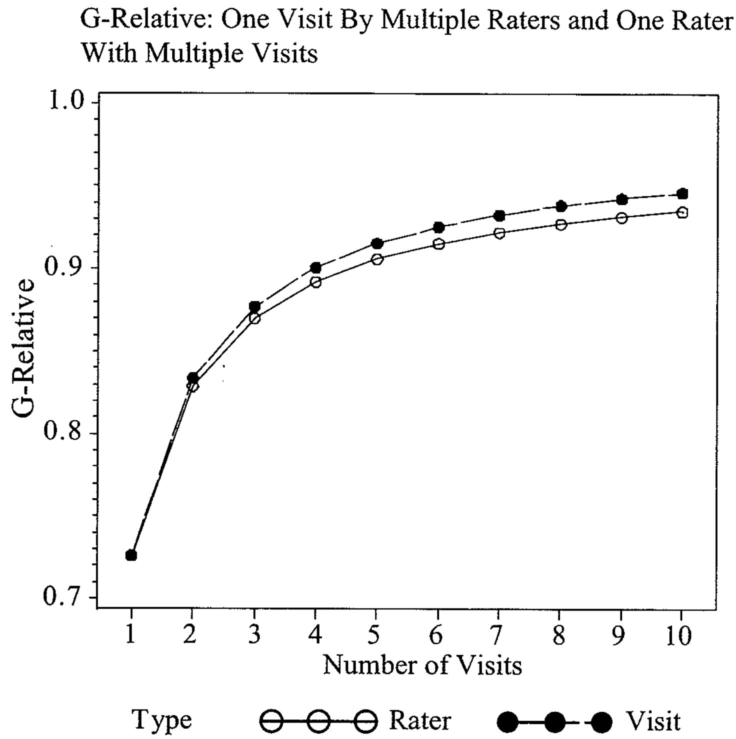

Figure 1. Display of G-relative results.

Detailed scoring guidelines for each subscale are available from the first author, as is the complete instrument and user's guide. The instrument and user's guide can be ordered at no charge via the Internet at www.nursinghomehelp.org, a Web site maintained by our research team for providers, consumers, regulators, and other researchers.

DISCUSSION

After field-testing in 207 assisted-living facilities, the OIQ-AL has been reduced from 41 to 34 reliable and discriminating items for researchers or others interested in directly observing and quickly evaluating (within 30 minutes of observation) quality of care in assisted-living facilities. The scale has a coherent six-factor structure that conceptually describes the multidimensional concept of care quality in this long-term care setting. These six factors can be logically clustered into process (Homelike and Caring, 21 items) and structure (Access and Choice; Lighting; Plants and Pets; Outdoor Spaces) subscales and for a total quality score that are the classic constructs of quality (Donabedian, 1969, 1988).

CTT results indicate the total quality score from the OIQ-AL has acceptable interrater and test–retest reliabilities (0.76 and 0.82, respectively), and strong internal consistency (0.94). Most subscales demonstrate strong internal consistency as well, and two subscales (Outdoor Spaces, Access and Choice) demonstrate weak internal consistency. GT analyses reveal that dependability of scores from the instrument can be improved by including a second observer who conducts a site visit and independently completes an instrument, or by a single observer conducting two site visits and completing instruments during each visit. GT analyses also indicate little additional benefit from increasing either observers or visits by more than two. Other researchers such as Hintze and Matthews (2003) have found similar effects in other instruments. Importantly, the results of the GT analysis reveal that with multiple observers or multiple visits (or a combination of these), the OIQ-AL is a dependable measure. Scoring guidelines based on the total sample of observations ($n = 358$) help to guide those who want to use the measure to interpret both subscale and total scores.

From a validity perspective, two expert panels of people experienced in the assisted-living field reviewed and advised on additions made to the version of the instrument that was field-tested in this study. The initial version of the OIQ-AL was derived from the OIQ-NH, which had extensive qualitative and quantitative development (Rantz et al., 1998, 1999, 2000, 2005, 2006; Rantz & Mehr, 2001; Rantz & Zwygart-Stauffacher, 2005), and a content-validity index calculated for that version was high (3.4 out of a maximum of 4) (Aud et al., 2004). Because there are no national assisted-living resident-assessment data such as the Minimum Data Set assessment instrument that is used in nursing homes, known group analysis for additional construct validation was not attempted. These analyses would likely be helpful and were done with the OIQ-NH (Rantz et al., 2006) as other comparative quality data are available for nursing homes. As other quality measure data become available for assisted living, additional validity testing should be conducted with the revised OIQ-AL.

The field-testing in this study was done with RN observers using the OIQ-AL. Therefore, we know the instrument is highly reliable when used by these trained observers. Additional field-testing needs to be done with other groups of observers, such as consumers or regulators, as was undertaken with the large-scale study of the OIQ-NH. Still, it is likely that the OIQ-AL will be helpful to those who are interested in judging quality of care, such as consumers and providers of assisted-living services. Prospective assisted-living facility residents and their families may want to use the OIQ-AL while comparing assisted-living facilities for themselves or their loved ones. Based on the success of the use of the OIQ-NH for consumers of nursing home services, it is likely the OIQ-AL will help consumers of assisted-living services as well.

Assisted-living facility administrators and staff may want to include the instrument in quality-improvement programs. Although in some states assisted-living facilities are not highly regulated, in those states that do license and survey assisted-living facilities, the state surveyors may find the instrument a useful supplement to their inspection protocols and assessments of regulatory compliance. Using the OIQ-AL in addition to other inspection or quality-improvement tools can bring to light quality of care issues of great interest to residents and families as well as long-term care providers and regulators. The OIQ-AL should be of interest to researchers who wish to measure outcomes of interventions targeting any or all of the instrument's six dimensions of care quality.

REFERENCES

Allen, J. E. (1999). *Assisted living administration: The knowledge base.* New York: Springer Publishing.

Aud, M. A., & Rantz, M. J. (2004). Quality concerns in assisted living facilities. *Journal of Nursing Care Quality, 19*(1), 8–9.

Aud, M. A., Rantz, M. J., Zwygart-Stauffacher, M., & Manion, P. (2004). Developing a residential care facility version of the Observable Indicators of Nursing Home Care Quality Instrument. *Journal of Nursing Care Quality, 19*(1), 48–57.

Brennan, R. L. (2001). *Generalizability theory.* New York: Springer Publishing.

Brummett, W. J. (1997). *The essence of home: Design solutions for assisted-living housing*. New York: Van Nostrand Reinhold.

Dearmin, J., Brenner, J., & Migliori, R. (1995). Reporting on QI efforts for internal and external customers. *Joint Commission Journal on Quality Improvement, 21*, 227–288.

Donabedian, A. (1969). Some issues in evaluating the quality of nursing care. *American Journal of Public Health, 59*, 1833–1836.

Donabedian, A. (1988). The quality of care: How can it be assessed? *Journal of the American Medical Association, 260*, 1743–1748.

Gorsuch, R. L. (1983). *Factor analysis* (2nd ed.). Hillsdale, NJ: Erlbaum.

Hayton, J. C., Allen, D. G., & Scarpello, V. (2004). Factor retention decisions in exploratory factor analysis: A tutorial on parallel analysis. *Organizational Research Methods, 7*, 191–205.

Hintze, J., & Matthews, W. J. (2003, April). *Generalizability of systematic direct observation data across occasion and time*. Paper presented at the annual meeting of the National Association of School Psychologists, Toronto, CA. Summary available at http://www.umass.edu/education/schoolpsychology/Presentations/G-theoryNASP.pdf

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika, 30*, 179–185.

Katz, J. M., & Green, E. (1997). *Managing quality: A guide to system-wide performance management in health care* (2nd ed.). St. Louis, MO: Mosby.

Kiefe, C., Woolley, T. W., Allison, J. J., Box, J. B., & Craig, A. S. (1994). Determining benchmarks: A data-driven search for the best achievable performance. *Clinical Performance and Quality Health Care, 2*, 190–194.

Lowe, T. J., Lucas, J. A., Castle, N. G., Robinson, J. P., & Crystal, S. (2003). Consumer satisfaction in long-term care: State initiatives in nursing homes and assisted living facilities. *The Gerontologist, 43*, 883–896.

Mitchell, J. M., & Kemp, B. J. (2000). Quality of life in assisted living homes: A multidimensional analysis. *Journal of Gerontology: Psychological Sciences, 55B*, P117–P127.

Mollica, R. (2003). Coordinating services across the continuum of health, housing, and supportive services. *Journal of Aging and Health, 15*, 165–168.

National Center for Assisted Living (NCAL). (2001). *Facts and trends: The assisted living sourcebook*. Washington, DC: Author.

Pruchno, R. A., & Rose, M. S. (2000). The effect of long-term care environments on health outcomes. *The Gerontologist, 40*, 422–428.

Rantz, M. J., & Mehr, D. R. (2001). A quest to understand and measure nursing home quality of care. *Long-Term Care Interface*, 2(7), 34–38.

Rantz, M. J., Mehr, D. R., Petroksi, G. F., Madsen, R. W., Popejoy, L., Hicks, L. L., et al. (2000). Initial field-testing of an instrument to measure "Observable indicators of nursing home care quality." *Journal of Nursing Care Quality, 14*(3), 1–12.

Rantz, M. J., Mehr, D., Popejoy, L., Zwygart-Stauffacher, M., Hicks, L., Grando, V., et al. (1998). Nursing home care quality: A multidimensional theoretical model. *Journal of Nursing Care Quality, 12*, 30–46.

Rantz, M. J., & Zwygart-Stauffacher, M. (2005). A new reliable tool for researchers, regulators, providers, and consumers for measuring quality of care in nursing homes. *Nursing Outlook, 53*(4), 212–213.

Rantz, M. J., Zwygart-Stauffacher, M., & Flesner, M. (2005). Advances in measuring quality of care in nursing homes: A new tool of providers, consumers, regulators, and researchers. *Journal of Nursing Care Quality, 20*(4), 293–296.

Rantz, M. J., Zwygart-Stauffacher, M., Popejoy, L., Grando, V., Mehr, D., Hicks, L., et al. (1999). Nursing home care quality: A multidimensional theoretical model integrating the views of consumers and providers. *Journal of Nursing Care Quality, 14*(1), 16–37.

Rantz, M., Zwygart-Staffacher, M., Mehr, D., Petroski, G., Owen, S., & Madsen, R. (2006). Field testing, refinement and psychometric evaluation of a new measure of nursing home care quality. *Journal of Nursing Measurement, 14*(2), 129–148.

Waltz, C. F., Strickland, O. L., & Lenz, E. R. (2005). *Measurement in nursing research* (3rd ed.). New York: Springer Publishing.

Zimmerman, S., Sloane, P. D., Williams, C. S., Reed, P. S., Preisser, J. S., Eckert, J. K., et al. (2005). Dementia care and quality of life in assisted living and nursing homes. *The Gerontologist, 45*(Special Issue 1), 133–146.

Zwick, W. R., & Velicer, W. F. (1986). Comparison of five rules for determining the number of components to retain. *Psychological Bulletin, 99,* 432–442.

Acknowledgments. The authors wish to acknowledge the contributions of other University of Missouri–Columbia MDS and Nursing Home Quality Research Team: Rose Porter, RN, PhD, dean; Deidre D. Wipke-Tevis, PhD, RNC, CVN, associate professor; Myra Aud, RN, PhD, assistant professor; Jill Scott-Cawiezell, RN, PhD, assistant professor; Jane Bostick, RN, PhD, clinical instructor; Donna Minner, Carol Seim, Clara Boland, Amy Vogelsmeier, Katy Nguyen, Elizabeth Sutherland, and Margie Diekemper, quality improvement nurses, Sinclair School of Nursing; Jessica Mueller, data support staff; Debra Oliver, PhD, assistant professor; consultants to our team: Dr. Robin Froman, RN, PhD, FAAN, dean, School of Nursing, University of Texas Health Science Center at San Antonio; Dr. Gwen Uman, founding partner, Vital Research, Los Angeles, California; Dr. David Zimmerman, director, Health Systems Research and Analysis, University of Wisconsin–Madison; and Dr. Francois Sainfort, professor, Georgia Institute of Technology. The members of the MU MDS and Nursing Home Quality Research Team gratefully acknowledge the ongoing support of the Missouri Department of Health and Senior Services staff, the Missouri Healthcare Association, and the Missouri Association of Homes and Services for the Aged; they are truly committed to helping homes embrace quality improvement. Research activities were partially supported by NINR 1R01NR/AG05287–01A2. Opinions are those of the authors and do not represent NINR.

Correspondence regarding this article should be directed to Marilyn J. Rantz, PhD, RN, FAAN, S406 Sinclair School of Nursing and Family and Community Medicine, School of Medicine, University Hospital Professor of Nursing, University of Missouri, Columbia, MO 65211. E-mail: rantzm@missouri.edu