Résumé

La technique fondée sur une échelle d'évaluation d'un produit permettant de mieux évaluer la satisfaction d'un patient

Mary R. Lynn et Bradley J. McMillen

Dans le but de trouver une solution à la pénurie actuelle d'infirmières dans les communautés autochtones du Nord-Ouest de l'Ontario, la Direction générale de la protection de la santé des Premières nations et des Inuits à Santé Canada a commandé une étude sur la pertinence d'établir une équipe de relève constituée d'infirmières provenant des petites villes minières des environs. On a présenté un questionnaire à questions libres et à questions fermées à un échantillon aléatoire de 237 infirmières en vue d'analyser leur degré de sensibilisation, leur disposition et leur niveau de préparation à la pratique des soins infirmiers en région nordique et de déterminer quels sont les facteurs favorables et défavorables au recrutement. Les conclusions révèlent une connaissance du recoupement des dimensions professionnelles et personnelles qui caractérise cette pratique; elles justifient le bien-fondé d'un système de rotation qui chevaucheraient les compétences fédérale, provinciale et locale. Malgré sa complexité, avec du temps et de la volonté, ce type de structure de relève régionale semble viable.

Mots clés : technique fondée sur une échelle d'évaluation d'un produit, satisfaction du patient, mesurage, outil

CJNR 2004, Vol. 36 N° 3, 66–81

The Scale Product Technique as a Means of Enhancing the Measurement of Patient Satisfaction

Mary R. Lynn and Bradley J. McMillen

Measurement of patient satisfaction has long been hampered by two resolvable problems: a lack of content validity in commonly used instruments, and a lack of variability in satisfaction scores when these same instruments are used. Most patient satisfaction instruments have been developed from the perspective of the provider or institution rather than that of the patient, creating a situation of questionable content validity for these measures. Additionally, most patient satisfaction measures yield data that are invariant and consistently positively biased. Both of these problems can be addressed methodologically — through tool development using a qualitative method designed to obtain the patient's perspective, and through the use of the scale product technique to decrease the effect of acquiescence, thereby increasing variability in item responses.

Keywords: measurement, Likert, scale product, patient satisfaction, patient-centred

Patient satisfaction surveys are commonly used to evaluate health care, primarily because the patient's input into the evaluation of care is almost universally seen as essential (Committee on Quality Health Care in America, 2001; Donabedian, 1980; Walker, 1993). In several models of health care, patient satisfaction is considered *the* outcome of care (Oberst, 1984), or one of two or three outcomes of care (Ellencweig, 1992; Jelinek, 1967). Elling (1974, 1980) suggests that patient satisfaction with care is one of five secondary objectives that are antecedent to the primary objective of desirable health status.

However, as important as patient satisfaction is in the evaluation of care, the current state of its measurement is severely lacking, for two reasons. First, despite recent advances in the assessment of satisfaction among hospitalized patients (e.g., Yen, Chen, & Chou, 2003), many items commonly used in patient satisfaction tools still tend to address either "hotel services" (Louden, 1989) or the perspective of the provider rather than that of the patient (Corriher, 1994; McDaniels & Nash, 1990; Rubin, 1990). And second, these items (or any others), once placed into a instrument with the now standard five-point Likert balanced response format (strongly disagree to strongly agree), result in a positively biased or a highly acquiescent response set (Ware, Davies-Avery, & Stewart, 1977). Since it is difficult to disagree with items that vary minimally (e.g.,

"the nurses acted friendly") or have little relevance for the evaluation of care (e.g., "my food was served hot"), it is not surprising to find that patient satisfaction tools generally yield minimally useful data (McDaniels & Nash).

However, both of these restrictions on the utility of patient satisfaction measures can be addressed methodologically. First, despite the progress that has been made in this area (e.g., Hancock et al., 2003; Hilton, Budgen, Molzahn, & Attridge, 2001; Yen, Chen, & Chou, 2003), qualitative methods should still be used to determine what patients value about nursing care and what they expect when hospitalized. Only with such approaches can the validity of patient satisfaction scales be maximized. Second, the use of a response format that includes a method for enhancing variability and decreasing the positive response bias commonly found will advance the measurement of patient satisfaction. This paper reports on a method for achieving such enhancements by using items derived from qualitative interviews conducted with patients to determine what they consider important in nursing care received in the acute-care setting, combined with a method to augment traditional scaling of such items.

The Patient's Perspective in Quality Nursing Care

Uncertainty about what patients consider important in nursing care has been mentioned in a number of discussions on patient satisfaction (Bond & Thomas, 1991; Murdaugh, 1992; Oberst, 1984; Ware et al., 1977). Substantial qualitative work has clearly demonstrated that patients can and do articulate the specific aspects of care they consider important in nursing care (Fosbinder, 1994; Lynn & Sidani, 1995). In fact Lynn and Sidani (1995) identify 90 specific aspects of care that patients consider the essential components of quality nursing care. These include having a nurse who is able to determine, based on her/his knowledge, what the patient needs; instructs the patient in a clear and understandable manner; calls the patient by his/her preferred name; and encourages the patient to become involved in her/his own care.

However, a patient satisfaction tool based on these 90 aspects will be sufficiently sensitive to the subtleties in individual patient evaluations only if it considers the relative importance of each aspect. Traditional Likert scaling, the most common scaling method chosen for most tools, including patient satisfaction ones, does not allow for items with different degrees of importance; in fact, it assumes that all items are of equal intensity or strength (Likert, 1932; van Alphen, Halfens, Hasman, & Imbos, 1994). If Likert items could have varying degrees of importance, and if patient satisfaction tools incorporated this revised scoring format,

then perhaps the response bias limitation of the patient satisfaction scales could be overcome.

Oberst (1984) poses the central question regarding varying levels of patient satisfaction with nursing care: "Is there a hierarchy of satisfaction that can be identified and are certain aspects of [care] more satisfying?" (p. 2367). This hierarchy question can be answered methodologically using the judgement technique developed by Thurstone (1928). Once the judgement stage is reached, the judgements can be directly converted to weights for use in combination with Likert scaling (Likert, 1932) to enhance the sensitivity of the patient satisfaction measurement. This combination of Thurstone judging and Likert scaling is the scale product technique developed by Eysenck and Crown (1949) and used by Hulka, Zyzanski, Cassel, and Thompson (1970) and Zyzanski, Hulka, and Cassel (1974) in studies of patient evaluation of medical care.

Thurstone or Likert Scaling — or Both?

Among the many decisions made when an instrument is developed, the choice of scaling method is particularly important. The scaling method developed by Likert (1932), a summated rating method, is one of the most common methods used by those developing or revising instruments. With the Likert scale, the subject is placed on a continuum according to her/his degree of agreement or disagreement on the topic being measured (McIver & Carmines, 1981). This is done by summing the subject's responses using a balanced continuous response format (usually ranging from strongly disagree to strongly agree) for each item in the scale to derive a total score, which is assumed to represent the person's relative position on the topic. Likert developed his method of scaling in response to the extensive effort entailed in the dominant scaling technique of the time, the equal appearing interval scaling method developed by Thurstone (Thurstone, 1928; Thurstone & Chave, 1929). Perhaps a better scaling method is one that combines the best of these two approaches, the scale product method (Eysenck & Crown, 1949; Eysenck, Crown, & Shapiro, 1950). After a review of the Thurstone and Likert methods, the scale product method will be discussed, using an example, as a means of addressing the shortcomings of the two classic methods.

Thurstone Equal Appearing Interval Scales

L. L. Thurstone, a pioneer in the development of attitudinal measurement, conceived attitude measurement as the judgement concerning one stimuli in reference to or compared to another. In his method of equal appearing intervals, the researcher generates a large pool of statements

pertaining to the attribute to be measured. Ideally, these are simple, declarative statements that a person can either agree or disagree with, and represent a range of intensity or strength of opinion regarding the concept being measured.

The statements are then written on cards and given to a group of individuals who serve as judges. The judges rate each statement according to its placement along a continuum, reflecting the extent to which it represents a positive or negative attitude towards the concept being measured. Items are typically rated on an 11-interval scale, from A = unfavourable to K = favourable. Only the middle (neutral) and the two extremes are defined for the judges (Edwards, 1957). After each item is rated by the judges, its median ranking is determined using the centile formula to obtain the 50th percentile, which becomes the scale score associated with that item (Stevens, 1946). Only the items that have stable median rankings (those with relatively low semi-interquartile ranges) are retained for use in the scale (Edwards). The desired outcome of this process is a scale with approximately two items for each of the 11 equal appearing intervals. When there are more than two items per interval, the items chosen are those with the lowest semi-interquartile ranges.

The "correct" number of judges to employ in the Thurstone method depends on the number needed to ensure substantial agreement on the placement of items along the continuum. Thurstone's original study used 300 judges, although subsequent research using far fewer judges has produced sufficient agreement on item placement along the continuum (Crocker & Algina, 1986). When a Thurstone scale is administered, respondents indicate whether they agree or disagree with each item. The average scale score (0–10) of the items that the person agrees with becomes his/her score.

One of the drawbacks of the Thurstone method is that it is quite laborious (Likert, 1932; Murphy & Likert, 1938). A substantial amount of time is required to find the judges and to generate the dozens of items needed to ensure an 11-point range. Also, it can be difficult to find items that fall into the moderate intervals of the scale (Edwards, 1957; Ferguson, 1939) and to secure agreement on the scale value of an item, which, by definition, falls into the grey area between quite favourable and quite unfavourable. The Thurstone method is also criticized because it assumes that the concept being measured is unidimensional, which is not a reasonable assumption for many affective phenomena.

Likert Summated Rating Scales

The Likert method of "summated ratings" came about because of the logistical difficulties of the Thurstone method. The items in a Likert scale are split between those that are positively worded (favourable) and those

that are negatively worded (unfavourable). Unlike items in the Thurstone method, those in the Likert scale do not have to be distributed across a continuum; in fact, all of the items are assumed to be of the same magnitude in favouring (or not favouring) the measurement objective of the instrument. It is this last assumption, that all items are of approximately the same strength or intensity, that mathematically allows for the items to be summed, with equal weighting, to derive the total score. Interestingly, Likert items, when rated according to the Thurstone method, tend to cluster at the top and bottom of the 11-point continuum (Ferguson, 1939).

The Likert scale has a more precise response format than the Thurstone scale. Its typical response format has five options, ranging from strongly disapprove/disagree to strongly approve/agree, with an undecided/neutral central point (Likert, 1932). Scores are obtained by assigning integer values to item responses and summing them to derive a total score (e.g., strongly disagree = 1, disagree = 2, neither clearly agree nor clearly disagree = 3, agree = 4, strongly agree = 5). It should be noted, however, that a Likert response format can have from 3 to 21 response categories (although 3 to 7 is the usual range), and a neutral response option is not a requirement. Many different labels have been used for Likert response formats; no one response format is required in a Likert scale; the only requirement is that the response options be balanced, with equal numbers of agree and disagree options.

The Likert method, however, also has its shortcomings. All items count equally in the total score. Therefore, agreeing (or disagreeing) with any one item contributes just as much to a person's score as agreeing (or disagreeing) with any other item. As stated above, the equal value of each item in the total score is based on the assumption that each item is as important as the next in terms of its centrality to the concept being measured. Therefore, a Likert scale lacks the scale values that are generated by Thurstone judging, which can provide valuable information about the items.

Another limitation of the Likert method is what can be called the "ocular test" of item quality. While the Thurstone method allows one to weed out items of suspect quality by eliminating those that receive ambiguous rankings during the judging, the Likert method does not allow for direct assessment of item quality.

Scale Product Technique

In light of these shortcomings, the Thurstone and Likert methods were combined for the purpose of creating the scale product technique, which capitalizes on the strengths of both methods while ameliorating some of their weaknesses. The scale product technique was pioneered by Eysenck

and Crown (1949) and modified by Hulka, Zyzanski, Cassel, and Thompson (1970) and by Zyzanski, Hulka, and Cassel (1974). In essence, it combines the Thurstone scale's position anchoring of an item with the Likert scale's integer item response scores. It has been shown to produce reliabilities superior to those that either the Thurstone method or the Likert method can produce separately (Castle, 1953; McNemar, 1946; Zyzanski et al.).

The scale product method entails the creation of a pool of items that are rated by judges, as described above for the Thurstone method. This judgement results in median ratings that are used as weights for each item. When the scale is administered, the subject responds to items using a balanced Likert format. The integer values assigned to the individual responses are then multiplied by the corresponding Thurstone weight to create a "weighted item score" that is summed to derive the total score. Therefore, a respondent's score for any one item is based on two pieces of information — the extent to which s/he agrees or disagrees with the statement, and the rated importance of that statement in relation to the concept being measured. Total scores can be summed for the entire instrument or for factors or other component scores, depending on the dimensionality of the instrument. The ability to use the weights within factors or other dimensions of an instrument avoids the unidimensional limitation of the Thurstone method.

The scale product method is not the only means of weighting an item. Other weighting methods include item response modelling (Hambleton, Swaminathan, & Rogers, 1991), whereby items in an affective measure are weighted according to their ability (least difficult to most difficult) to be agreed with at evaluation (Beck & Gable, 2001), and an explicit approach such as giving more weight to items that correlate with an external criterion (Rudner, 2001). Both of these approaches have advantages (e.g., item response theory allows for estimation of the "difficulty" [here, "agree attractiveness"] of items so that a developer can strive to have items that cover the range of "difficulties"), but their basis for weighting is different from that of the scale product method. The scale product method's basis for weighting — the importance of the items, as explicitly stated by the respondents, in reference to the concept being assessed — has both conceptual and cognitive appeal in the assessment of patient satisfaction. This is not to suggest that this method is superior to the aforementioned approaches to weighting; it is simply different.

The disadvantages of the scale product method stem primarily from its foundation in classical measurement (test) theory. In this theory, all items in a scale are assumed to be of equal value and are therefore summed without regard to differences in intensity or importance. Weighting of these items does not necessarily alter any statistical opera-

tions, leading some measurement experts to suggest it is not worth the effort (Gulliksen, 1950). However, knowing that the items are not of equal importance/intensity may well be enough to make the effort worthwhile.

Use of the Scale Product Technique

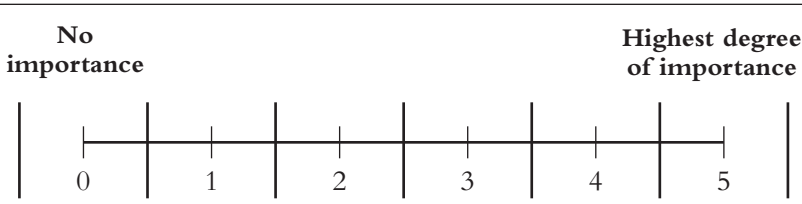
In the current study, the scale product technique was used to derive item weights for the 90 aspects of quality nursing care derived from qualitative interviews with 29 patients at two medical centres in the southwestern United States. All patients were asked the same grand tour question: "How would you describe or define good nursing care?" Further questioning depended on the response to the previous question(s) and continued until there were no unexplored areas. At the conclusion of the 29 interviews, data saturation had been achieved (Lynn & Sidani, 1991, 1995). The interviews were transcribed verbatim and the analysis began with the first interview. Ninety distinct data bits, representing attitudes ("likes being a nurse"),[1] actions ("responds to my calls promptly"), characteristics ("looks professional"), and physical environment ("my room is not noisy"), were identified in the interview data. These 90 data bits were translated into items (e.g., "looks professional" became "the nurse has a professional appearance") and presented to a panel of six patients not previously interviewed, to determine the extent to which the items matched their experience. The list was deemed accurate and comprehensive by the panel, thereby supporting the content validity of the items (Lynn, 1986). From the patients' perspective, the conceptual categories of quality nursing care were "responsiveness," "attitude," "knowing me as a person," and "respecting me as a person."

After approval had been received from the Institutional Review Board, the scale product technique was initiated by employing the Thurstone judgement stage with patients in seven hospitals in the southeastern United States — an academic medical centre, a Veterans' Administration Medical Center, three urban community hospitals, and two rural community hospitals. These hospitals were participating in a study on the perceptions of quality of care held by patients and nurses and the influences on those perceptions. The 90 aspects identified by patients as important in good nursing care were presented to 448 patients in these hospitals. The patients ranged in age from 18 to 90 years (*mean* = 53, *SD* = 16.4), were almost evenly divided between females (48%) and males (52%), and were primarily Caucasian (64%), with 65% having no more than a high-school education. Most (77%) had been hospital-

[1] Each phrase in parentheses is one of the 90 distinct data bits obtained from the interviews.

ized at least twice previously and had been hospitalized for at least 48 hours before participating in the study. All patients spoke English but were not required to read English; the data collector could read the card aloud so that the patient could then place the card on the numbered board according to her/his evaluation of the intensity of the item. See Figure 1 for the board used in this procedure.

Figure 1 *Item Sorting Board*



The participants were told that information written on each card had come from interviews with patients who were asked to describe or define good nursing care. The participants were then asked to rate the aspect described on each card in terms of its importance to the overall concept of good nursing care. The participants were given as much time as necessary to perform the card sorting and were allowed to re-evaluate any card (or stack of cards) if they chose. They took between 15 and 25 minutes to perform the card sorting.

The Thurstone method was modified in the current study by eliminating the unfavourable end of the judging continuum, since the items were all generated from interviews relating to good care. Therefore a six-point scale was used in the judgement, with panellists placing each item in one of six categories ranging from 0 (no importance) to 5 (highest degree of importance) according to their perception of quality nursing care (Figure 1). Sorting was continued until stable median rankings were achieved.

Median ratings for the 90 items ranged from 2.88 to 4.81. Aspects of care pertaining to the competence of the nurses and the adequacy of resources ranked the highest and aspects pertaining to the patient's physical environment ranked the lowest. Applying the Thurstone technique to weight these aspects of good nursing care will add to the variability and, it is hoped, the precision of the scale. Items will contribute to the total score of the scale or subscale(s) in direct proportion to their overall relevance to quality nursing care. See Table 1 for a sample of the items and their associated median ratings.

Table 1 ***Sample PPQS-ACV Items with Median Rating***

I wish the view from my room was more interesting	2.88
The nurse uses touch to reassure or support me	3.88
The nurse knows who I am as a person	3.97
The nurse makes sure that I have plenty of time to talk to her/him	4.12
The nurse shows me that I am her/his first concern	4.34
The nurse helps me take care of my daily physical needs	4.42
The nurse frequently checks on me	4.47
The nurse is patient	4.52
The nurse is able to talk to me	4.55
The nurses see me as an individual, a real person	4.56
The nurse is attentive and responsive to my needs	4.59
My room is kept clean	4.61
The nurse is clear when teaching me about my care	4.66
The nurse gives me my medications on time	4.74
The nurse knows what she/he is doing	4.81

The scale in question has five factors — Communication, Professionalism, Individualization, Timeliness, and Environment. Weighted and unweighted means and standard deviations for the factors are shown in Table 2. As might be expected, the means and standard deviations of the factors increased in the weighting procedure, by approximately the same magnitude. However, it should be noted that, as is true with any linear transformation of scores, the weighting of items has little or no effect on statistical operations, specifically those based on covariances/correlations. Only in some instances have the resulting correlations between weighted factors and outcome of interest been different from those of the original, unweighted, factor scores (Jansen, Stigglebout, Nooij, & Kievit, 2000). A few correlations in this examination were different, but by only .01 or .02. Therefore, correlations with other variables will be almost identical using weighted or unweighted factor scores, and reliability of the factors should be similar across the two scoring methods. While there were differences in the reliability estimates for the factors using the weighted and unweighted items, these differences occurred at the 3rd and 4th decimal points and therefore also are of no consequence. There are some differences in the skewness and kurtosis of the distributions of the factor scores caused by the wider dispersion of the scores created by the weighting. Despite the lack of statistical differences with the scale product technique, its primary advantage is its role in enhancing the variability of the scores.

Table 2 *Means, Standard Deviations, Skewness, and Kurtosis for Unweighted and Weighted Scale Factors*

Scale Factor (Alpha)	Unweighted				Weighted			
	Mean	*SD*	*Skewness*	*Kurtosis*	*Mean*	*SD*	*Skewness*	*Kurtosis*
Communication (.93)	4.17	.51	-.451	1.777	18.99	2.31	-.444	1.765
Professionalism (.79)	4.27	.51	-.641	2.192	19.46	2.31	-.645	2.214
Individualization (.84)	4.00	.55	-.465	1.256	16.93	2.34	-.469	1.260
Timeliness (.84)	3.99	.63	-.916	2.243	17.95	2.80	-.917	2.234
Environment (.68)	3.62	.69	-.497	.480	13.73	2.54	-.557	.620

Discussion

In essence, the scale product technique allows for subjects to consider the importance of each item when responding to it, so that the total score indicates the extent to which they agree or disagree with the item as well as the importance of the item with respect to the concept being measured. Since the score on an instrument using the scale product technique is directly proportional to the mathematical combination of the item weight and the respondent's selection, the overall importance of selected aspects is integral to the final score without the need for importance to be assessed each time the scale is used. Such an approach is particularly suited to the measurement of patient satisfaction, because it combines the evaluation of the extent to which the aspect of patient care occurred and the importance of that aspect to the patient's overall experience. Certainly there are instruments that incorporate an "importance" response scale and a "presence" or "agreement" response scale. When completing such scales, respondents are asked to provide two responses to each statement, one indicating the importance they give to the item and the other the extent of their agreement or disagreement with the statement. There are two problems with this response format: it places a large response burden on the respondent, especially when the scale is fairly long, and the importance ratings are not generalizable. Importance in such instruments is idiosyncratic.

With the genuine Thurstone scale, the items are expected to fall across the entire judgement range. In the case of the scale product technique, they are not expected to be distributed in this manner, since they already represent "important" aspects of quality nursing care. However, the fact that they varied at all suggests that even items that appear to be similar in their relative strength towards the measurement objective of an instrument may not be similar at all. It may not be reasonable, or even desirable, to assume that items in a Likert scale have the same strength, intensity, or degree of favourability towards the measurement objective. While it is conventional to treat Likert items as if they were all of equal intensity, perhaps the time has come to examine their intensity as part of the instrument's development or revision. The advantages of being able to weight items according to their strength or intensity will likely increase the sensitivity of the instrument, thereby enhancing both its validity and its reliability.

Use of a qualitative method to derive the content for the Thurstone judging method addresses a significant problem with traditional satisfaction surveys — the lack of clarity on what patients value about the care they receive (Ware et al., 1977). Weighting also helps to resolve the persistent problem of score invariance in traditional satisfaction assessments

(McDaniels & Nash, 1990; Ware, 1977). When an item is weighted according to its judged priority, the total score reflects the patient's evaluation of that aspect of nursing care, weighted for the importance of that component of care to patients in general. A respondent's agreement with an item contributes more if it is a highly weighted rather than a lesser-weighted item. Conversely, if a respondent disagrees with a highly weighted item, more points are subtracted from the score than would be the case with a lesser-weighted item. These subtle differences in the effect of items on the total score make some scales more sensitive to patient evaluation of the quality of nursing care and provide considerably more variance than traditional patient quality or satisfaction inventories.

The scale product technique has three additional advantages in terms of satisfaction scales. First, it provides a statistical check on the quality of the items. If an item receives very low rankings from the judges, it probably does not belong on the scale, as the judges have said it has little importance with respect to the concept being measured. Therefore, the weights can be used to help eliminate "bad" items. Second, giving patients the opportunity to comment on the importance of scale items can enhance the validity of the scale. Since a scale is intended to capture the perspective of patients with respect to quality of care, if patients think an item is irrelevant, then it *is* irrelevant, and is not a valid measure of quality of care from the perspective of patients. Finally, the determination of the relative importance of the items used to assess patient satisfaction offers clinicians and researchers insight into patients' relative valuing of care that otherwise is not available to them without overt investigation in this area. Therefore the scale product technique, when used with qualitatively derived items, should be useful in clinical, research, and quality improvement projects that require a patient-centred method of evaluating care.

References

Beck, C. T., & Gable, R. K. (2001). Item response theory in affective instrument development: An illustration. *Journal of Nursing Measurement, 9*(1), 5–22.

Bond, S., & Thomas, L. H. (1991). Issues in measuring outcomes of nursing. *Journal of Advanced Nursing, 16,* 1492–1502.

Castle, P. F. C. (1953). A note on the scale-product technique of attitude scale construction. *Occupational Psychology, 27,* 104–108.

Committee on Quality Health Care in America. (2001). *Crossing the quality chasm: A new health system for the 21st century.* Institute of Medicine. Washington: National Academy Press.

Corriher, J. B. (1994). *Critical care patients' satisfaction with nursing care.* Unpublished master's thesis, University of North Carolina School of Nursing.

Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory.* New York: Holt, Rinehart & Winston.

Donabedian, A. (1980). *Explorations in quality assessment and monitoring. Vol 1: The definition of quality and approaches to its assessment.* Ann Arbor, MI: Health Administration Press.

Edwards, A. L. (1957). *Techniques of attitude scale construction.* New York: Appleton-Century-Crofts.

Ellencweig, A. Y. (1992). *Analyzing health systems: A modular approach.* New York: Oxford University Press.

Elling, R. H. (1974). Case studies of contrasting approaches to organizing for health: An introduction to a framework. *Social Science and Medicine, 8,* 263–270.

Elling, R. H. (1980). *Cross-national study of health systems — political economics and health care.* New Brunswick, NJ: Transaction Books.

Eysenck, H. J., & Crown, S. (1949). An experimental study in opinion-attitude methodology. *International Journal of Opinion and Attitude Research, 3,* 47–86.

Eysenck, H. J., Crown, S., & Shapiro, M. (1950). Symposium on primary social attitudes. In *Proceedings and papers of the Twelfth International Congress of Psychology.* Edinburgh: Oliver & Boyd.

Ferguson, L. W. (1939). The requirements of an adequate attitude scale. *Psychological Bulletin, 36,* 664–673.

Fosbinder, D. (1994). Patient perceptions of nursing care: An emerging theory of interpersonal competence. *Journal of Advanced Nursing, 20,* 1085–1093.

Gulliksen, H. O. (1950). *Theory of mental tests.* New York: John Wiley.

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory.* Newbury Park, CA: Sage.

Hancock, K., Chang, E., Chenoweth, L., Clarke, M., Carroll, A., & Jeon, Y. (2003). Nursing needs of acutely ill older people. *Journal of Advanced Nursing, 44,* 507–516.

Hilton, B. A., Budgen, C., Molzahn, A. E., & Attridge, C. B. (2001). Developing and testing instruments to measure client outcomes at the Comox Valley Nursing Center. *Public Health Nursing, 18,* 327–339.

Hulka, B. S., Zyzanski, S. J., Cassel, J. C., & Thompson, S. J. (1970). Scale for the measurement of attitudes toward physicians and primary care. *Medical Care, 8,* 429–436.

Jansen, S. J., Stiggelbout, A. M., Nooij, M. A., & Kievit, J. (2000). The effect of individually assessed preference weights on the relationship between holistic utilities and nonpreference-based assessment. *Quality of Life Research, 9,* 541–547.

Jelinek, R. C. (1967). A structural model for the patient care operation. *Health Services Research, 2,* 226–242.

Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology, 22,* 5–55.

Louden, T. L. (1989). Customer perception counts in quality assurance. *Hospitals, 63*(2), 84.

Lynn, M. R. (1986). Quantification and determination of content validity. *Nursing Research, 34,* 254–256.

Lynn, M. R., & Sidani, S. (1991, May 2–4). *The next step: The patient's perception of quality scale.* Paper presented at the 24th WSRN Annual Communicating Nursing Research Conference, Albuquerque, New Mexico.

Lynn, M. R., & Sidani, S. (1995). Patients' and family members' perceptions of what makes quality patient care. In S. G. Funk, E. M. Tornquist, M. T. Champagne, & R. A. Wiese (Eds.), *Key aspects of caring for the acutely ill: Technological aspects of patient education and quality of life* (pp. 292–298). New York: Springer.

McDaniels, C., & Nash, J. (1990). Compendium of instruments measuring patient satisfaction with nursing care. *Quality Review Bulletin, 16,* 182–188.

McIver, J. P., & Carmines, E. G. (1981). *Unidimensional scaling.* Beverly Hills, CA: Sage.

McNemar, Q. (1946). Opinion-attitude methodology. *Psychological Bulletin, 43,* 289–374.

Murdaugh, C. (1992). Quality of life, functional status, patient satisfaction. In *Patient outcomes research: Examining the effectiveness of nursing practice. Proceedings of the State of the Science Conference* (pp. 91–96). Bethesda, MD: Department of Health and Human Services, Public Health Service, National Institutes of Health.

Murphy, G., & Likert, R. (1938). *Public opinion and the individual.* New York: Harper.

Oberst, M. T. (1984). Patients' perceptions of care: Measurement of quality and satisfaction. *Cancer, 53,* 2366–2373.

Rubin, H. R. (1990). Patient judgments of hospital quality, patient evaluations of hospital care: A review of the literature. *Medical Care, 28,* S3–S9.

Rudner, L. M. (2001). Informed test component weighting. *Educational measurement: Issues and practice, 20*(1), 16–19.

Stevens, S. (1946). On the theory of scales of measurement. *Science, 103,* 677–680.

Thurstone, L. L. (1928). Attitudes can be measured. *American Journal of Sociology, 23,* 529–554.

Thurstone, L. L., & Chave, E. J. (1929). *The measurement of attitude.* Chicago: University of Chicago Press.

Van Alphen, A., Halfens, R., Hasman, A., & Imbos, T. (1994). Likert or Rasch? Nothing is more applicable than good theory. *Journal of Advanced Nursing, 20,* 196–201.

Walker, J. D. (1993). Enhancing physical comfort. In M. Gerteis, S. Edgman-Levitan, J. Daley, & T. L. Delbanco (Eds.), *Through the patient's eyes* (pp. 119–153). San Francisco: Jossey-Bass.

Ware, J. E. (1977). *Effects of acquiescent response set on patient satisfaction ratings.* Report #P-5675. Washington: Department of Health, Education and Welfare.

Ware, J. E., Davies-Avery, A., & Stewart, A. L. (1977). *The measurement and meaning of patient satisfaction: A review of the literature.* Washington: Department of Health, Education and Welfare.

Yen, M., Chen, C., & Chou, S. (2003). Concerns of hospitalized care from patients' perspectives. *Journal of Nursing Research, 10*(2), 121–128.

Zyzanski, S. J., Hulka, B. S., & Cassel, J. C. (1974). Scale for the measurement of "satisfaction" with medical care: Modifications in content, format and scaling. *Medical Care, 12,* 611–620.

Authors' Note

This study was funded by the National Institute for Nursing Research, R01 NR03606.

Comments or inquiries may be directed to Mary R. Lynn, School of Nursing, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599-7460 USA. Telephone: 919-966-5450. Fax: 919-966-2286. E-mail: Mary_Lynn@unc.edu

---

Mary R. Lynn, PhD, is Associate Professor, School of Nursing, University of North Carolina at Chapel Hill, USA. Bradley J. McMillen, PhD, is Senior Research and Evaluation Consultant, North Carolina Department of Public Instruction, Raleigh.