Physical Therapy Faculty Publications

Physical Therapy

10-2007

Patient Satisfaction in Physical Therapy: Concurrent Comparison of Two Instruments in Outpatient Settings

Susan E. Roush
University of Rhode Island, roush@uri.edu

Michelle Jones

Melissa Nassaney

Follow this and additional works at: https://digitalcommons.uri.edu/pt_facpubs

Patient Satisfaction in Physical Therapy: Concurrent Comparison of Two Instruments in Outpatient Settings

Publisher Statement

Creative Commons License

PATIENT SATISFACTION IN PHYSICAL THERAPY: CONCURRENT COMPARISON OF TWO INSTRUMENTS IN OUTPATIENT SETTINGS

Abstract

Study Design: This study compared the psychometric properties of the Physical therapy Outpatient Satisfaction Survey (PTOPS) and the Patient Satisfaction Questionnaire (PSQ). **Objective:** To compare these instruments on missing data, ceiling effect, inter-item reliability, and external validity. **Methods and Measures:** Subjects (n=156) from five outpatient clinics in New England completed both scales. The PTOPS is a 34-item Likert scale with four subscales; the PSQ is a 20-item Likert scale with no subscales. Instruments were scored; descriptive statistics, ceiling effect, and inter-item reliability were calculated. Regression analysis examined the external validity of each scale. **Results:** Missing data were found primarily on the cost-related items of the PSQ. Two PTOPS subscales and the PSQ demonstrated ceiling effects. Inter-item reliability coefficients ranged from .60 to .86 for the PTOPS subscales and was .94 for the PQS. The PTOPS accounted for twice as much validity variance as the PSQ. **Conclusion:** The single-scale PSQ is the more reliable instrument, but is limited by its substantial ceiling effect. While the ceiling effect is also a concern for two subscales of the PTOPS, it produces results that are more interpretable and demonstrates better external validity than the PSQ. Consumers of physical therapy patient satisfaction surveys need to develop greater understanding of psychometrics to critically analyze the surveys they use. [Roush SE, Jones M, Nassaney M. Patient Satisfaction in Physical Therapy: Concurrent Comparison of Two Instruments in Outpatient Settings. *HPA Resource/HPA Journal 2007*; 7(3): J1 – J8.]

KEY WORDS: patient satisfaction, physical therapy, survey instrument development

Susan E. Roush, Michelle Jones, and Melissa Nassaney

A Peer-Reviewed Publication of the Section on Health Policy & Administration

JOURNAL OF HEALTH POLICY & ADMINISTRATION

Patient satisfaction has become a critical issue in today's health care. Quality control, performance in managed care environments, and the competitive marketplace are several of the forces behind this emergence. Additionally, consumers are acting more as partners in their care than ever before,[1] increasing the importance of their perceptions of service delivery. While customer or consumer satisfaction is a concept with which most people are familiar, measuring this construct in health care settings in a reliable and valid manner is complex and difficult.

The scientific understanding of patient satisfaction in physical therapy is in its infancy. Much of the literature appears as anecdotal material in support of increasing the overall quality of health care.[2,3] There is, however, a growing recognition of the need for rigorous methodologies and empirical data if the patient satisfaction construct is going to be understood beyond a superficial level. Additionally, these methods will give health care providers evidence upon which to choose the best instrument for their work setting and guide them in appropriately using the resultant data.

An early attempt to understand patient satisfaction came from Winter and Keith[4] who used telephone surveys to determine the satisfaction level of patients with their outpatient physical therapy experience. Their results suggested that, although survey results consistently show high levels of overall satisfaction, satisfaction levels vary when separate dimensions are investigated. Their assertion suggests that patients may be satisfied with some aspects of their physical therapy care, but less satisfied with other aspects. Since 1988, authors have assessed a variety of dimensions of patient satisfaction with physical therapy services. For example, Yoshioka[5] measured satisfaction with physical therapy at the end of life by considering the dimensions of technical competence, professionalism, courtesy, and caring. Oermann et al[6] assessed satisfaction of patients with cystic fibrosis and considered the domains of efficacy, convenience, comfort, and overall satisfaction, which they identified by using factor analysis. Most recently, Beattie[7] developed the Medrisk patient satisfaction instrument with patients receiving care through Worker's Compensation insurance. This instrument measures two patient satisfaction dimensions, internal factors associated with patient/therapist interactions and external factors associated with clinic and environmental factors.

A major limitation in the early physical therapy patient satisfaction research was the lack of a measurement tool that had been comprehensively developed based on psychometric principles as opposed to merely described with psychometric statistics. Indeed, until recently, physical therapy lacked a formally developed tool to measure overall (not disease or patient specific) patient satisfaction, relying instead on "homemade" instruments developed in restricted settings and supported, at most, by face validity.[8] Since 1999, however, other overall physical therapy-specific satisfaction instruments have been published.[9-12] The Physical Therapy Outpatient Satisfaction Survey (PTOPS)[9] and the Patient Satisfaction Questionnaire (PSQ)[10] were the only two patient satisfaction instruments in physical therapy that had extensive psychometric evidence at the time of our study. The PTOPS measures four dimensions of satisfaction: Enhancers, Detractors, Location, and Cost. The PSQ measures satisfaction as a single dimensional construct, ie, it measures patient satisfaction as a single phenomenon.

At the time of our study, no data that compared satisfaction instruments in physical therapy were available. While reliability and validity are the most obvious psychometric properties to use for comparisons, we suggest that two additional psychometric considerations — missing data and the ceiling effect — may also aid in the decision of choosing the appropriate patient satisfaction instrument for a specific setting. Given the time and expense associated with collecting patient satisfaction data, the preferred instrument should produce less missing data or missing data that are randomly distributed, ie, not associated with the content of the instrument's items. The ceiling effect, which considers the relationship between a scale's mean score and its maximum value, is defined as the number of standard deviation units between the obtained scale mean and the maximum scale value.[13] The ceiling effect is of interest because it indicates how well the scale differentiates among the responses in the upper range of a score distribution. This phenomenon is of particular interest with patient satisfaction data because the distribution of satisfaction scores is typically clustered at the top (or high satisfaction) end of the possible range of scores.[14] The amount of variation is necessarily limited when a scale produces a mean score that is too close to its maximum value, consequently providing little meaningful information. In patient satisfaction assessment, a problematic ceiling effect means that subjects with differing levels of satisfaction may produce the same high score. Since ceiling effects are a com-

Susan E. Roush, PT, PhD is Professor of Physical Therapy at the University of Rhode Island in Kingston, RI. Address all correspondence to Dr. Roush at roush@uri.edu.

Michelle Jones, PT, MS is a physical therapist with Rehab America in Memphis, TN.

Melissa Nassaney, PT, DPT is a physical therapist in Rhode Island and resides in West Kingston, RI.

mon finding in satisfaction survey results, surveys that minimize this effect are preferable to those that do not.

We wanted to know which patient satisfaction survey – PTOPS[9] or PSQ[10] – provides better psychometric results. To find out, we compared them using four psychometric properties: missing data, the ceiling effect, inter-item reliability, and external validity.

METHOD

Sample

Participants were a convenience sample of 156 physical therapy patients who were receiving treatment in one of five outpatient settings in the New England region of the United States. Each subject completed a brief demographic data form and both the PTOPS and PSQ instruments while they were in the clinic waiting room during one of their scheduled appointments. All patients scheduled for treatment were eligible for the study except those who (1) were under 18 years of age, (2) were being seen for an initial evaluation on the day of data collection, or (3) had cognitive ability incompatible with responding to the survey. Subjects were approached by one of the researchers and asked to participate in the study. Exclusion criteria 1 and 2 were verified by clinic clerical staff; exclusion criterion 3 was left to the judgment of the researchers.

Instruments

The PTOPS[9] and the PSQ[10] are both Likert-type survey instruments that were developed using factor analytical techniques to measure patient satisfaction in outpatient settings. Table 1 summarizes the key developmental features of each instrument. The PTOPS was developed in three phases based on data from 607 subjects and measures four dimensions of patient satisfaction: Enhancers, Detractors, Cost, and Location. In contrast, the PSQ was developed in one phase, is based on data from 289 subjects, and measures one dimension of satisfaction. Both instruments offer respondents a 5-point, agree-disagree scale with the addition of a "no experience" option for each PSQ item.

Principal components factor analyses identified the PTOPS' four dimensions and the PSQ single dimension.[9,10] Principal components analysis is a descriptive technique that does not consider sampling error or assess the fit of data to the identified factor structure. Repeated identification of the factor structure on new samples adds evidence to the factor structure's credibility, as does the inferential technique of confirmatory factor analysis.[15] The PSQ single-factor structure was identified from one sample of 289 subjects,[10] while the PTOPS four-factor structure was identified and refined from three samples totaling 607 subjects.[9] The PTOPS factor structure was confirmed in another study that had a sample size of 1,175.[16] Finally, confirmatory factor analysis of the PTOPS[9] exceeded fit index standards for four accepted criteria.[17-19]

Developers of both the PTOPS and the PSQ calculated inter-item reliability of their respective factors.[9,10] The reliability of the PSQ's single factor demonstrated greater reliability than any of the PTOPS subscales. As expected, (sub)scales with more items demonstrated greater reliability. Additionally, the PTOPS was designed to control for the effects of two types of response bias: acquiescence and social desirability. Response bias is the "systematic tendency to respond to … items on some basis other than the specific item content."[20 (p17)] Acquiescence and social desirability are two prominent response biases.[20] Acquiescence was minimized in the PTOPS by including both positively- and negatively-worded items in all subscales. The PTOPS was also constructed to be independent of two aspects of social desirability: Impression Management and Self-deception Enhancement as identified by Paulhus.[21] Social desirability is a characteristic of survey items that relates to the likelihood that respondents will answer in a socially acceptable way, as opposed to how they truly feel about the item.[21] PTOPS items that correlated highly with these measures were deleted from the scale during development, leaving items that were only minimally correlated.[9]

Validity evidence for the PSQ and PTOPS was gathered in different ways. Concurrent validity for the PSQ[10] was demonstrated with high correlations (r >.95) between three of its items, which were judged by PSQ authors to measure overall satisfaction, and the remaining 17 items. Construct validity was addressed through non-empirical consideration of a multitrait-multimethod matrix. The multitrait-multimethod approach to validity[22] uses two or more methods to measure two or more traits and considers the relative strengths of the correlations among results obtained from measuring the different traits with the different methods.[13] Goldstein et al's[10] multitrait-multimethod analysis considered each PSQ item as a different satisfaction trait, although the nature of their second measurement "method" was unclear. The external validity of the PTOPS[9], on the other hand, was assessed through discriminate function analysis, which accurately differentiated 94% of the time between patients who generically reported high satisfaction with physical therapy from those who generically reported low satisfaction. Also, the PTOPS was found to differentiate between patients who had attended their scheduled physical therapy appointments less than 50% of the time and those who attended 90% or more of their appointments. This differentiation was accurate 88% of the time.[9]

Procedure

Participants were approached during non-treatment time and asked to participate in the study. Informed consent was obtained consistent with the standards of the University of Rhode Island Institutional Review Board and included assurance that the survey results would remain anonymous. Participants completed a brief demographic form and the PSQ and PTOPS satisfaction surveys. The satisfaction surveys were given in alternating order, ie, half of the subjects received the PTOPS first and the other half received the PSQ first. Participants placed their completed surveys in a slotted collection box and received a token gift (note paper, playing cards, or scented soap) for their participation.

Data Analysis

Across all five clinics, 54% of eligible participants completed the surveys with the response rate by clinic ranging from 34% to 80%. Data collection was more successful in the smaller clinics. Data were analyzed using version 11.0 of SPSS statistical software.[23]

Frequencies were calculated for the categorical variables of gender, method of payment, and ethnicity. Descriptive statistics were calculated for the continuous variables of age and years of education. The data were then analyzed to address the psychometric properties of interest: (1) missing data, (2) ceiling effect, (3) inter-item reliability, and (4) external validity.

Missing data were examined in two ways: per item and across (sub)scales. Several issues arose as to how to manage the missing data in this study. We decided not to delete all cases with missing data, because this action may have had an adverse effect on the sample size. Another strategy would have been to replace the missing data with a "best guess", ie, the mean score of that scale or the mean score of the item. The missing data point could also have been replaced with a generated random score that takes into consideration the variance of the responses. Neither of these solutions was acceptable because

much of the PSQ data were not technically missing. Rather, respondents used the "No Experience with Item" option instead of leaving the item unanswered. The PSQ developers, however, did not provide guidance on how to score the "no experience" responses. To further complicate the missing data issue, the PTOPS does not have a comparable "no experience" response category, and applying a "best guess" treatment to the PSQ "no experience" responses could have distorted any subsequent analyses. Given the fundamental difference between the interpretation of missing data in the PSQ and the PTOPS, and the adequate sample size of 156 subjects, we decided to omit cases for each analysis that had missing data in any variable using listwise deletion. Analysis, therefore, proceeded with cases that provided complete data for the PSQ and for each of the PTOPS subscales. The results reported below identify the number of valid cases upon which each analysis was based.

To consider the ceiling effect, descriptive statistics were calculated for each PTOPS subscale and the PSQ. Then the number of standard deviation units separating the mean and maximum (or highest satisfaction) response was calculated and compared to the recommended minimum two standard deviation separation. Since two of the PTOPS subscales (Detractors and Cost) are negative scales, ie,

Table 1. Developmental Features of the PTOPS[a] and PSQ[b]

Development Feature	PTOPS[9]	PSQ[10]
Sample	3 independent samples N = 607 Outpatients	1 sample N = 289 Outpatients
Factor structure[c] (# of items/factor)	Enhancers: (10) Detractors: (10) Location: (7) Cost: (7)	One factor (20)
Response options	Likert-type scale: 1=strongly disagree, 2=disagree, 3=uncertain, 4=agree, 5=strongly agree	Likert-type scale: Same as PTOPS with added option: "no experience with the item"
Inter-item reliability (Cronbach α)	Enhancers: $\alpha = 0.86$ Detractors: $\alpha = 0.82$ Location: $\alpha = 0.87$ Cost: $\alpha = 0.71$	$\alpha = 0.99$
Control for response bias	Yes, both positively and negatively worded items Minimal correlation with social desirability	No, all items positively worded
Confirmatory assessment of factor structure	Confirmatory factor analysis (Empirical technique)	None
Construct validity	Discriminate function analysis	3 inter-item correlations Non-empirical consideration of multimethod-multitrait matrix

[a] Physical Therapy Outpatient Satisfaction Survey
[b] Patient Satisfaction Questionnaire
[c] Identified through Principal Component Factor Analysis

higher scores are associated with lower satisfaction, the ceiling effects for these scales became known as floor effects.

The inter-item reliability of the (sub)scales was quantified by using the Cronbach alpha statistic. Finally, regression was used to examine the relationship between the PTOPS and an overall measure of patient satisfaction, and between the PSQ and that same measure. Specifically, multiple linear regression was used with the PTOPS since it is composed of four subscales. The four subscales were the independent variables which were used to predict the dependent variable of overall satisfaction. Simple linear regression was used with the PSQ data, which was considered the sole independent variable used to predict the dependent variable of overall satisfaction.

RESULTS
Participants
Approximately two-thirds of the participants were female and over 85% were Caucasian. The average subject was 51 years old (SD = 15.5) and had 13.9 years (SD = 3.4) of formal education. Thirty-seven percent of the subjects obtained their physical therapy through private insurance, 25% through managed care, 13% through Medicare, and 12% through Worker's Compensation.

Missing data
All of the 34 PTOPS items and 16 of 20 (80%) PSQ items were missing five or fewer data points across the 156 cases. This means that, at most, only 3% of subjects did not respond to any one of these items. The percentage of subjects who did not respond to the four remaining PSQ items, however, was larger than 3%. Specifically, twelve participants (8%) did not respond to item 17 (I am satisfied with the services provided by my physical therapist assistant); 14 (9%) did not respond to item 25 (If I had to I would pay these physical therapy services myself) and 34 (22%) did not respond to either item 16 (My bills were accurate) or 24 (The cost of the physical therapy treatment received was reasonable).

Ceiling effect
Descriptive statistics of each PTOPS subscale and the PSQ are presented in Table 2 along with the calculated ceiling (or floor) effect. Three ceiling effects were found. The PTOPS Enhancer subscale and the PSQ demonstrated a ceiling effect with 1.41 and 1.04 standard deviations (SDs), respectively that separated their mean and their scales' maximum values. The PTOPS Detractor subscale, a negatively-worded scale where low scores are associated with high satisfaction, demonstrated a floor effect with 1.27 SDs separating its mean from its minimum value. Whether titled a ceiling effect (with the positive scales) or a floor effect (with the negative scales), the PTOPS Enhancer and Detractor subscales and the PSQ were compromised in their ability to differentiate among subjects at the high satisfaction end of the distribution.

Inter-item reliability
The PTOPS subscale inter-item reliability coefficients (Cronbach a) for the four subscales were as follows: Enhancers

.80, Detractors .86, Location .60, and Cost .62. The PQS inter-item reliability coefficient was 0.94. The reliability of the PQS and the PTOPS Enhancers and Detractors subscales is considered good, while the reliability of the PTOPS Location and Cost subscales is moderate.[24]

External validity
The external validity of the two satisfaction measures was assessed by examining their relationship with an external criterion measure, ie, responses to one five-point (agree-disagree) Likert-like item measuring overall satisfaction. The mean of this overall satisfaction item was 1.22 (SD = 4.6) with higher numbers indicating lower satisfaction. The PTOPS accounted for 12.7% of the variance in this measure (p<0.001), while the PSQ accounted for 6.3% (p<0.01).

DISCUSSION
The PTOPS demonstrated less missing data than the PSQ. While neither instrument demonstrated missing data sufficient to discard entire cases, four items on the PSQ that related to finance issues did not provide meaningful information except that many respondents had "no experience" with these items. The PTOPS does not have a comparable "no experience" response option, although it seemed likely that respondents who did not have experience with a PTOPS item responded with "uncertain." Consideration of the endorsement of the "uncertain" option on the PTOPS Cost items, for example, lends support to this assertion. Indeed, "uncertain" responses on the seven PTOPS finance-related items were high. It is likely that some of these responses represent uncertainly due to inexperience with financial issues rather than uncertainty related to satisfaction, but the PTOPS did not capture the distinction between these two types of uncertainty. The practical difference between the two approaches used by the PSQ and the PTOPS is unknown.

The wide variety of payment options and processes used in physical therapy make assessing satisfaction in relationship to financial matters difficult. The lack of a consistent frame of reference from which to write items results in confusion when subjects encounter an item that is outside of their experience. For example, both the PTOPS and the PSQ have an item related to the reasonableness of cost. Depending on the circumstances, some patients may not know what the cost is, making a judgment about its reasonableness inappropriate. For those who know the cost of treatment, however, this item provides valuable information. Capturing this information without compromising the psychometric properties of a survey, however, is difficult.

The PSQ and two PTOPS subscales (Enhancers and Detractors) demonstrated ceiling effects. A high ceiling effect is a serious limitation as it reflects an inherent minimizing of the range of obtained scores. Without adequate variance, a survey item provides no useful information. While it may be affirming to report high satisfaction scores, high scores that are associated with ceiling effects are likely to be an artifact of the survey items themselves and would be obtained in any clinical setting.

The high ceiling effect means that two PTOPS subscales and the PSQ were unable to adequately differentiate in the upper ranges of satisfaction resulting in different groups (eg, different clinics) getting roughly the same "satisfaction grade", even though there may be differing levels of satisfaction among their patients. In patient satisfaction assessment, a problematic ceiling effect means that, for example, therapists who have patients with differing levels of satisfaction may get the same satisfaction score. In other words, therapists with highly satisfied patients will obtain the same satisfaction scores as therapists with only moderately satisfied patients. As a result, managers will not be able to differentiate between the work of two therapists based on subscales that demonstrate ceiling effects.

may have contributed to its high reliability and suggests the presence of an acquiescence response bias, ie, a tendency for the respondent to give each item the same response without reading each one carefully.[20] Indeed, test developers routinely employ mixed item polarity to counter acquiescence response bias.[25, 26] In addition, the negative effect of the consistent item polarity of the PSQ may have contributed to the survey's high ceiling effect.

The reliability of the PTOPS subscales was found to be good to adequate. One factor that may contribute to the differences between the PTOPS and PSQ reliability scores is the smaller number of items associated with each PTOPS subscale. Inter-item reliability scores are sensitive to the number of items such

Table 2. Descriptive data and ceiling/floor effect for the PTOPS[9] subscales and the PSQ[10] scale

(Sub)scale (# of items)	N	Mean	Standard Deviation	Possible Range	Obtained Range	Ceiling effect [b]	Floor effect [c]
PTOPS							
Enhancers (10)	147	4.33	0.47	1 - 5	2.30 - 5.00	1.41[d]	
Detractors[a] (10)	148	1.81	0.64	1 - 5	1.00 - 4.10		1.25[d]
Location (7)	150	3.88	0.52	1 - 5	2.00 - 4.71	2.15	
Costa (7)	149	2.97	0.29	1 - 5	2.14 - 3.86		6.79
PSQ (20)	106	4.57	0.42	1 - 5	3.45 - 5.00	1.04[d]	

[a] Negative scale (after recoding) i.e., high scores were associated with lower satisfaction.
[b] Number of standard deviation units between obtained mean and scale's maximum value, which on positive scales is associated with the highest satisfaction.
[c] Number of standard deviation units between obtained mean and scale's minimum value, which on negative scales is associated with the highest satisfaction.
[d] Demonstrated the highest level of a ceiling (or floor) effect.

When selecting instrumentation to measure patient satisfaction, or any other construct for that matter, surveys that minimize the ceiling effect are preferable to those that do not. The results of our study suggest that the PTOPS was less affected by ceiling effect than the PSQ. Greater variation is possible with the PTOPS subscales that have a ceiling (or floor) effect than with the PSQ, because sufficient variation is obtainable through the PTOPS's other subscales. Finally, those using physical therapy satisfaction surveys must interpret their results in the context of known ceiling effects. This means that high satisfaction survey results do not necessarily indicate that patients' needs are maximally being met or that there is not room for improvement in a particular aspect of care. High satisfaction scores may be as reflective of the inherent limitations associated with the present 'state of the art' in measuring patient satisfaction as high quality care. The psychometric evidence only supports relative comparisons.

In general, the PSQ demonstrated higher inter-item reliability than the PTOPS and could be considered to be a more "patient-friendly" survey because it contains fewer items that are all positively-worded. The PSQ's lack of item polarity, however,

that scales with fewer items are associated with lower reliability.[13] Another interesting aspect of the PTOPS reliability results is that the reliability of the Cost and Location subscales are lower than previously reported,[9] although the reason for this is unclear.

No external "gold standard" exists with which patient satisfaction surveys can be compared. Since an unambiguous external criterion does not exist, validation of patient satisfaction measures becomes a multifaceted process. Consideration needs to be given to a variety of types of evidence, gathered from multiple studies, with judgments about validity relying on the preponderance of evidence. To that end, the evidence generated in this study suggests that the individual relationships between the PSQ or the PTOPS and an overall measure of satisfaction should be considered only as a supplement to previous validity studies.

The external criterion measure used in this study demonstrated a serious ceiling effect and, as a result, the magnitude of the correlations is likely to be spuriously low and should not be taken as true measures of shared variance. A calculated correla-

tion can be artificially deflated when there is a limited range of responses as was found in our results. A higher correlation may have been found if a full range of responses were available.[13] The correlations between each of the satisfaction surveys and the external criterion measure, however, does provide initial evidence of the relative strength of each measure in predicting overall satisfaction. Since we found that the PTOPS accounted for just over twice as much validity variance as the PSQ, the PTOPS may be the better measure when based on this criterion alone. The PTOPS's higher validity variance is consistent with the more rigorous development history of the PTOPS when compared to the PSQ.

Determining the reliably and validily of instruments that measure a construct such as physical therapy patient satisfaction is a complex process. Numerous issues contribute to this complexity, including (1) variation in the construct's perceived meaning and dimensionality, (2) a strong tendency to obtain negatively skewed data, ie, patients are unlikely to endorse negative statements about their health care providers, (3) lack of a 'gold standard' criterion measure, and (4) a historical over-reliance on face validity. The last point refers to relying on satisfaction surveys that 'look' like they measure satisfaction, but lack psychometric support.

At the same time, practitioners, managers, managed care organizations, and patients are demanding and expecting much from patient satisfaction data. At the policy level, provider profilers, to date, have not included satisfaction as a quality outcome indicator, but it is likely to be included in the future.[27] Eventually, reimbursement is likely to be tied to quality improvement initiatives. As these eventualities unfold, physical therapy practices must ensure that patient satisfaction is measured with instruments that are psychometrically sound so that important decisions are based on good data. For example, since the PTOPS accounts for twice as much variance in an external satisfaction measure as does the PSQ, the PTOPS may be considered the better choice as a quality indicator. At the individual clinic level, many managers are interested in rewarding clinicians whose patients are highly satisfied. Differentiating between good and excellent patient satisfaction has clinical relevance, but can only be approached if the appropriate survey instrument is selected. The PSQ's troublesome ceiling effect limits its effectiveness in doing exactly that. Of course, the PTOPS also has a ceiling effect, but it is less severe and its subscales offer multiple sources of variance.

Our study's results have added new psychometric evidence for two patient satisfaction instruments by comparing the PSQ and the PTOPS on missing data, ceiling effect, reliability, and validity. Much work remains to be done in physical therapy patient satisfaction research. Comparisons between and among the PSQ, PTOPS, and other developing physical therapy satisfaction instruments are indicated. Specific challenges include refining items that measure cost and identifying appropriate external criterion measures. Both of these challenges will undoubtedly be addressed through the next generation of test construction techniques, eg, item response theory.[28-30] Given that

patient satisfaction is likely to become an even more important outcome indicator in physical therapy in the future and given that it is unlikely a psychometric 'gold standard' will ever be developed, anyone who uses the results of patient satisfaction surveys must be informed, analytical, and critical in their consumption. Just as patient satisfaction research in physical therapy has developed beyond 'face validity' to a more rigorous dimension, so must consumers of patient satisfaction literature.

The results of this study need to be considered in light of its limitations. Our sample, for example, was almost exclusively Caucasian and was highly educated. They may not be representative of typical physical therapy patients. The sample was clearly not representative of the US population. Also, participants had a wide range of physical therapy experience, including those who had only one physical therapy session (an initial evaluation) prior to completing the satisfaction surveys. Minimal experience with physical therapy services may not have been adequate to provide meaningfully informed responses to the surveys.

CONCLUSION
Measuring patient satisfaction is a complex and difficult task. Since no 'gold standard' in physical therapy exits, those who wish to measure the patient satisfaction construct need to be able to critically evaluate the available instruments. Our direct comparison of psychometric properties of the PTOPS and PSQ may assist physical therapists in choosing the one that works best in a particular setting and, more importantly, should be helpful in interpreting and using the results appropriately.

REFERENCES
1. Purtillo R. *Ethical dimensions in the health professions. 3rd Ed.* Philadelphia: WB Saunders; 1999.
2. Cohn R. Patient satisfaction surveys. *PT Mag*. 1997;5(12):30-31.
3. Courtney J, Ritchie J, Schunk C. CQI in the outpatient clinic. *PT Mag*. 1998; 6(11):51-58.
4. Winter P, Keith R. A model of outpatient satisfaction in rehabilitation. *Rehab Psych*. 1988; 33(3):131-142.
5. Yoshioka H. Rehabilitation for the terminal cancer patient. *Am J Phys Med Rehab*. 1994; 73(3):199-206.
6. Oermann C, Swank P, Sockrider M. Validation of an instrument measuring satisfaction with chest physiotherapy techniques in cystic fibrosis. *Chest*. 2000; 118(1):92-97.
7. Beattie PF. The Medrisk instrument for measuring patient satisfaction with physical therapy care: A psychometric analysis. *JOSPT*. 2005; 35(1):24-32.
8. American Physical therapy Association. Patient satisfaction instruments: A compendium. Alexandria VA: American Physical Therapy Association; 1995.
9. Roush SE, Sonstroem RJ. Development of the Physical therapy Outpatient Satisfaction Survey (PTOPS). *Phys Ther*. 1999; 79(2):159-170.
10. Goldstein MS, Elliot SD, Guccione AA. The development of an instrument to measure satisfaction in physical therapy. *Phys Ther*. 2000; 80(9):853-863.

11. Monnin D, Perneger TV. Scale to measure patient satisfaction with physical therapy. *Phys Ther*. 2002; 82(7):682-691.

12. Beattie PF, Pinto MB, Nelson MK, Nelson R. Patient satisfaction with outpatient physical therapy: Instrument validation. *Phys Ther*. 2002; 82(6):557-565.

13. Allen MJ, Yen WM. Introduction to measurement theory. Monterey CA: Brooks/Cole Publishing; 1979.

14. Hudak P, Wright J. The characteristics of patient satisfactions measures. *Spine*. 2000; 25(24):3167-3177.

15. Bentler, PM. EQS: Structural Equations Manual. Los Angeles, CA: BMDP Statistical Software; 1980

16. Scott CK. *Determinants of patient satisfaction with outpatient satisfaction services in the southern United States as measured by the Physical therapy Outpatient Satisfaction Survey: Precision by replication [dissertation]*. Jackson MS; University of MS Medical Center; 2001.

17. Hudak LA. *Structural equation modeling with LISREL*. Baltimore MD: The John Hopkins University Press; 1987.

18. Bentler PM, Bonnett DG. Significance tests and goodness of fit in the analysis of covariance structures. *Psych Bull*. 1980; 88(3):588-606.

19. Bentler PM. On the fit of models to covariances and methodology. *Psych Bull*. 1992; 112(3):400-404.

20. Paulhus DL. Measurement and control of response bias. In: Robinson JP, Shaver PR Wrightsman LS, Eds. *Measures of personality and social psychological attitudes*. New York, NY: Academic Press; 1991:17-60.

21.   Paulhus DL. *BIDR Reference Manual for Version 6*. Vancouver, British Columbia, University of British Columbia Press; 1991.

22. Campbell DT, Fiske DW. Convergent and discriminant validation by the multitrait-mulitmethod matrix. *Psych Bull*. 1959; 56(2):81-105.

23. Norusis MJ. *SPSS base system user's guide*. SPSS, Inc. Chicago IL; 1990.

24. Portney LG, Watkins MP. Foundations of clinical research. 2nd Ed. Upper Saddle River, NJ: Prentice Hall Health; 2000.

25. Nunnally JC. *Psychometric theory*. 2nd Ed. New York: McGraw-Hill; 1978.

26. Goldstein G, Hersen M. *Handbook of psychological assessment*. New York, NY: Pergamon Press; 1990.

27. Resnik L, Hart DL. Using clinical outcomes to identify expert physical therapists. *Phys Ther*. 2003; 83(11):990-1002.

28. Cella D. Chang CH. A discussion of item response theory and its applications in health status assessment. *Med Care*. 2000; 38(9): II-66 – II-72.

29. Hambleton RK. Principles and selected applications of item response theory. In: Linn RL ed. *Educational Measurememt* 3rd Ed. New York, NY: MacMillian Publishing Co; 1989; 147-200.

30. Hays RD, Morales LS, Reise SP. Item response theory and health outcomes measurement in the 21st century. *Med Care*. 2000; 38(9): II-28 – II-42. ■