Development of the Patient Activation Measure (PAM): Conceptualizing and Measuring Activation in Patients and Consumers

Judith H. Hibbard, Jean Stockard, Eldon R. Mahoney, and Martin Tusler

Background. Controlling costs and achieving health care quality improvements require the participation of activated and informed consumers and patients.

Objectives. We describe a process for conceptualizing and operationalizing what it means to be "activated" and delineate the process we used to develop a measure for assessing "activation," and the psychometric properties of that measure.

Methods. We used the convergence of the findings from a national expert consensus panel and patient focus groups to define the concept and identify the domains of activation. These domains were operationalized by constructing a large item pool. Items were pilot-tested and initial psychometric analysis performed using Rasch methodology. The third stage refined and extended the measure. The fourth stage used a national probability sample to assess the measure's psychometric performance overall and within different subpopulations.

Study Sample. Convenience samples of patients with and without chronic illness, and a national probability sample ($N = 1,515$) are included at different stages in the research.

Conclusions. The Patient Activation Measure is a valid, highly reliable, unidimensional, probabilistic Guttman-like scale that reflects a developmental model of activation. Activation appears to involve four stages: (1) believing the patient role is important, (2) having the confidence and knowledge necessary to take action, (3) actually taking action to maintain and improve one's health, and (4) staying the course even under stress. The measure has good psychometric properties indicating that it can be used at the individual patient level to tailor intervention and assess changes.

Key Words. Patient activation, self-management, consumer roles in health care

Two significant emerging policy directions put patients and consumers in a key role for influencing health care quality and costs. First, consumer-directed health plans rely on informed consumer choices to contain costs and improve the quality of care. This approach assumes that consumers will make more prudent health and health care choices when they are given financial incentives along with access to comparative cost and quality information. This

approach also assumes that the combination of financial incentives and relevant information will increase their "activation" (Gabel, Lo Sasso, and Rice 2002). Second, the Chronic Illness Care Model (Bodenheimer et al. 2002) emphasizes patient-oriented care, with patients and their families integrated as members of the care team. A critical element in the model is activated patients, with the skills, knowledge, and motivation to participate as effective members of the care team (Von Korff et al. 1997).

A key health policy question is, what would it take for consumers to become effective and informed managers of their health and health care? What skills, knowledge, beliefs, and motivations do they need to become "activated" or more effectual health care actors? These are essential questions if we hope to improve the health care process, the outcomes of care, and control costs. This is true especially with regard to the 99 million Americans with a chronic disease. Because those with chronic illness need ongoing care, account for a large portion of health care costs, and must play an important role in maintaining their own functioning, encouraging their activation should be a priority.

Even though patient activation is a central concept in both the consumer driven health care approach and the chronic illness care models, it remains conceptually and empirically underdeveloped. There has been a lack of conceptual clarity regarding "activation," and thus a lack of adequate measurement. There are a number of existing methods for assessing different aspects of activation, such as health locus of control (Wallston, Stein, and Smith), self-efficacy in self-managing behaviors (Lorig et al. 1996), and readiness to change health-related behaviors (DiClemente et al. 1991; Prochaska, Redding, and Evers 1997), but these measures tend to focus on the prediction of a single behavior. Moreover, there is no existing measure that includes the broad range of elements involved in activation, including the knowledge, skills, beliefs, and behaviors that a patient needs to manage a chronic illness.

In this paper we describe the development of the Patient Activation Measure (PAM), a measure of activation that is grounded in rigorous

Address correspondence to Judith Hibbard, Dr.P.H., Professor of Health Policy, University of Oregon–1209, Department of Planning, Public Policy, and Management, 119 Hendricks Hall, Eugene, OR 97403-1209. Jean Stockard, Ph.D., is a Professor at the University of Oregon. Eldon Mahoney, Ph.D., is Director, Survey Research and Development, PeaceHealth, Bellevue, Washington. Martin Tusler, M.S., is a data analyst at the Department of Planning, Public Policy, and Management, University of Oregon.

conceptualization and appropriate psychometric methods. The PAM was developed in four stages:

> *Stage 1.* Conceptually defining activation involved a literature review, systematic consultation with experts using a "consensus method," and consultation with individuals with chronic disease using focus groups.
>
> *Stage 2.* Preliminary scale development began by building on the domains identified in stage one and operationalizing them with survey items within each domain. Steps included generating, refining, and testing a large item pool. We used Rasch psychometric methods to develop the scale and test the preliminary measure's psychometric properties.
>
> *Stage 3.* Stage three involved exploring the possibility of extending the range of the measure, refining the response categories, and testing whether the measure could be used with respondents who had no chronic illnesses.
>
> *Stage 4.* In the fourth and final stage a national probability sample was used to assess the performance of the measure across different subsamples in the population and to assess the construct validity of the measure.

STAGE 1: CONCEPTUALIZING ACTIVATION

Literature Review

Methods. A review of published articles that discuss skills and knowledge needed to successfully manage a chronic illness was conducted. Articles on self-care, self-management, doctor–patient communication, and using comparative information to inform health care choices were reviewed.

Findings. The review findings indicated that being an engaged and active participant in one's own care is linked to better health outcomes (Von Korff et al. 1997; Lorig et al. 1999; Von Korff et al. 1998; Bodenheimer et al. 2002) and measurable cost savings (Glasgow et al. 2002). Training patients with chronic diseases to self-manage their disease is effective, at least in the short term, in increasing functioning, reducing pain, and reducing health care costs (Lorig et al. 1999). Research also indicated a positive relationship between self-efficacy, preventive actions, and health outcomes (Bandura 1991; Grembowski et al. 1993; O'Leary 1985; Day, Bodmer, and Dunn 1996; Kaplan, Greenfield, and Ware 1989).

Collaborating on care and engaging in shared clinical decision making are also linked with better health outcomes (Von Korff et al. 1997; Kaplan, Greenfield, and Ware 1989; Glasgow 2002). Coaching patients to be more involved and to have more control in the medical encounter has been shown to produce better health and functioning in patients (Wasson et al. 1999; Greenfield, Kaplan, and Ware 1985; Greenfield et al. 1988).

Several studies document the problems consumers have in understanding and navigating the health care system, which may lead to reduced access to appropriate and timely care (Isaacs 1996; Hibbard et al. 1998, 2001). Similarly, because of the documented variability in the quality of different health care providers and hospitals, it is hypothesized that consumers who use comparative quality information to choose health care providers will receive higher-quality medical care (Marshall et al. 2000).

To summarize, the review of the literature indicates that patients who are able to: (1) self-manage symptoms/problems; (2) engage in activities that maintain functioning and reduce health declines; (3) be involved in treatment and diagnostic choices; (4) collaborate with providers; (5) select providers and provider organizations based on performance or quality; and (6) navigate the health care system, are likely to have better health outcomes. We used these six domains as a starting point for an expert consensus process and for patient focus groups.

Expert Consensus

Methods. The expert consensus process was adapted from Kahn et al. (1997) and Thorndike and Hagen (1991) and was designed to identify consensus among experts who view the issue of activation from a wide range of perspectives. Twenty-one panelists were chosen, in part, because they had demonstrated national prominence in their area of expertise. To limit the influence of one respondent on another, we gathered input through mailed surveys.

The process involved two rounds of contact, and 18 of the 21 experts completed both rounds. The key question posed to the experts in each round was, "What are the knowledge, beliefs, and skills that a consumer needs to successfully manage when living with a chronic disease?"

The first round was designed to elicit a broad range of ideas about domains to be included. We began with the six "domains" developed from the literature review (listed above) and elaborated these to include patients' beliefs, knowledge, and skills associated with each of these areas. Thus we had

Figure 1a: Example of Subdomains, under the Domain

Has the Skills for Maintaining Function, Prevention, and Making Lifestyle Changes

Subdomains
1. Has the skills to reduce the impact of symptoms on ability to function.
2. Can self-monitor and make lifestyle and environmental changes that improve functioning and well-being.
3. Maintains recommended diet and exercise regimens or other lifestyle / environmental recommendations.
4. Maintains activities when experiencing some pain or fatigue.
5. Can use evidence-based strategy for managing their primary risks.
6. Ability to "come back" after a behavioral lapse.

Subdomains 1–3 were rated as important or very important to defining the domain by the expert panel; Subdomains 4–6 were rated as less important.

a total of 18 possible domains: beliefs, knowledge, and skills for each of the six domains. Within each of these 18 domains we listed a number of subdomains involving specific characteristics, attributes, or behaviors. (Figure 1a shows examples of the subdomains.) We asked the experts to edit these subdomains, add any new subdomains, and rate the importance of each subdomain and each general domain for its importance to the construct.

Findings. The results of the first round indicated considerable consensus in conceptualizing activation with many of the experts providing similar comments and additions. On the basis of this expert feedback our original classification of beliefs, knowledge, and skills was altered slightly to include "accessing emotional support."

In the second expert consensus round the expert respondents were given an expanded set of subdomains that more clearly defined the larger domain and included subdomains suggested by the experts. The experts rated and rank-ordered the importance of each domain and each subdomain. The domains where there was expert consensus are identified in Figure 1b.

Patient Focus Groups

Methods. Two focus groups explored the same potential domains with a convenience sample of chronic disease patients. One focus group had ten participants; the other group had nine participants. The domains that were explored with the experts were revised and reworded in layman's terms and were used as the basis for a discussion on the key components of successful management of chronic disease. Participants, like the expert panel, could also edit or add to the list. The participants were recruited with ads in the local newspaper and were paid $35 for their participation. The average age was 55

Figure 1b: Domains Endorsed by Expert Consensus and Patients

	Self-Manage	Collaborate with Provider	Maintain Function / Prevent Declines	Access Appropriate and High-Quality Care
Believes patient is important in:	▨	▨	▨	
Has the knowledge to:	▨		▨	
Has the skills to:	▨	▨	▨	▦
Can access emotional supports to:	▤		▤	

▨ Identified by experts and patients as a key component and retained in later stages of scale development

▤ Identified only by experts as a key component and omitted in later stages of scale development

▦ Identified by experts as a key component and identified by patients as a secondary component and retained for preliminary scale development

(range 39 to 78). Sixty-eight percent of the respondents were female. Ninety percent had more than one chronic condition.

Findings. The expert panel and focus group participants were in agreement regarding most of the domains (Figure 1b). However, the focus group participants were much less likely than the experts to view emotional support of family and friends as important in successful management of chronic disease.

Based on results from the expert panel and the consumer focus groups we derived a conceptual definition of health activation in patients and consumers: Those who are activated *believe* patients have important roles to play in self-managing care, collaborating with providers, and maintaining their health. They *know* how to manage their condition and maintain functioning and prevent health declines; and they have the *skills and behavioral repertoire* to manage their condition, collaborate with their health providers, maintain their health functioning, and access appropriate and high-quality care. We used this definition as the basis for developing the measure.

STAGE 2: PRELIMINARY SCALE DEVELOPMENT

Methods

To operationalize the domains in Figure 1b, an 80-item pool was constructed by selecting questions from existing instruments and creating new ones where

none existed. The items in the pool were categorized under the domains they were intended to measure and were reviewed by a subset of the expert panel for face and content validity.

All 80 items were further refined with three rounds of face-to-face cognitive testing with 20 respondents with chronic conditions. Items were evaluated in terms of how well they were understood, the degree to which there was variability in responses, and the adequacy of the response categories. Seventy-five items were retained after the cognitive interviews and used for the pilot study.

Study Sample

The pilot study was conducted with a convenience sample of 100 respondents. Participants were recruited through newspaper advertisements and were paid for their participation. Respondents ranged in age from 19 to 79 and reported a wide range of chronic conditions. Items were administered through a telephone interview that included the 75-item pool and a limited set of demographic and health status questions.

Psychometric Analysis

The initial set of items constituting the PAM were selected using Rasch analysis (Rasch 1960; Wright and Masters 1982; Wright and Stone 1979; Massof 2002). Rasch measurement can be used to create interval-level, unidimensional, probabilistic Guttman-like scales from ordinal data such as rating scale responses to survey questions. The measurement model calibrates the "difficulty" of the items in terms of response probabilities. The calibration of an item on the measurement scale indicates how much of the measured variable a respondent must exhibit to be able to endorse the item.

Once the measure is constructed, individuals are measured as to where they fall on the scale, and their location represents how much of the variable each respondent possesses. In the case of the PAM, an individual's location indicates how activated the person is. Both the people who are measured and the items doing the measurement are located on the same equal interval scale, yet these two parameters are statistically independent of each other. This concept of parameter separation means that the calibration of the items is independent of the activation levels of the particular respondents measured.

The precision with which an item's scale location, or calibration, has been estimated is represented by the item's standard error of measurement.

Likewise, the precision of each individual respondent's estimated scale location is specified by the standard error of measurement of that person.

Item selection is based on item fit statistics representing how much responses to an item deviate from the model's expectations. A fit value of 1.0 indicates perfect fit to model expectations. Fit values >1.0 indicate more stochastic variability in responses than expected (e.g., persons with low measured activation endorsing items requiring a high level of activation) and fit values <1.0 indicate that responses to the item by persons of different activation levels do not vary as much as the model expects.

Two item fit statistics are calculated. Infit is an information-weighted residual and is most sensitive to item fit when the item's scale location is close to the respondent's scale location. Outfit is more sensitive to item fit for items with a scale location that is distant from the respondent's scale location. Simulation studies and experience suggest that item fit values between .5 and 1.5 produce sufficient unidimensionality and expected response variability for useful rating scale measurement (Smith 1996). All analyses were conducted with the *Winsteps* Rasch models software application (Linacre 2002).

Findings

Table 1 shows the 21 items constituting the preliminary activation measure, the calibrated scale location (difficulty) of each item, and the fit and item discrimination statistics. Item difficulty calibration on the "calibration" shown in Table 1 indicates how much activation is required for a patient to have .5 probability of responding "agree" to an item. Item scale locations have been transformed from the original logit metric to a user-friendly 0–100 metric where $0 =$ the lowest possible activation and $100 =$ the highest possible activation as measured by this set of items. While the metric allows for a potential range of 0–100, the items included in the measure only covered the range from 40–60, not tapping what would be theoretically the lowest or highest ranges of the construct.

All the domains derived through the conceptualization stage (Figure 1b) are reflected in the 21 items, except for the domain of accessing appropriate and high-quality care. While items addressing this domain correlate with the 21-item measure, fit statistics revealed these items tap a different construct than activation.

Most importantly, this analysis indicates that the items form a unidimensional, probabilistic Guttman-like scale. Close inspection of the

Table 1: Preliminary (from Stage 2) 21-Item Activation Measure with Calibrations

Item	Calibration	SEM	Infit	Outfit
How much do you know about why you are supposed to take each of your prescribed medicines?	40.3	1.4	1.12	1.15
Taking an active role in my own care is the most important factor in determining my health and ability to function.	41.0	1.5	1.15	1.11
How much do you know about the lifestyle changes, like diet and exercise, that are recommended for your condition?	42.4	1.4	1.33	1.14
How much do you know about the nature and causes of your health condition(s)?	44.3	1.4	1.28	1.28
How confident are you that you can tell your health care provider concerns you have even when he/she does not ask?	45.9	1.3	1.40	1.33
How much do you know about how to prevent further problems with your condition?	46.2	1.3	0.90	0.82
Even if I make the changes in diet and exercise recommended for my condition, it won't make any difference to my health.	47.0	1.3	1.13	1.13
How much do you know about self-treatment approaches for your condition?	47.9	1.3	1.20	1.06
How much do you know about the medical treatment options available for your condition?	48.9	1.2	1.20	1.12
How confident are you that you can find trustworthy sources of information about your health condition and your health choices?	48.9	1.2	1.10	1.03
How confident are you that you can follow through on medical treatments you need to do at home?	50.0	1.2	0.87	0.81
How confident are you that you can identify when it is necessary to get medical care and when you can handle the problem yourself?	50.2	1.2	0.92	1.10
How confident are you that you can take actions that will help prevent or minimize some symptoms or problems associated with your condition?	51.2	1.2	0.92	0.88
How confident are you that you can follow through on medical recommendations your health care provider makes such as changing your diet or doing regular exercise?	52.9	1.2	0.88	0.90
To what extent are you able to handle symptoms on your own at home?	54.4	1.2	1.02	1.01
How well have you been able to maintain these lifestyle changes?	55.2	1.2	0.73	0.74
To what extent have you made the changes in your lifestyle, like diet and exercise, that are recommended for your condition?	56.4	1.2	0.74	0.73
Maintaining the lifestyle changes that have been recommended for my condition is too hard to do on a daily basis.	57.0	1.1	0.76	0.76
Even if I'm dissatisfied, it is usually too much of a hassle to change health care providers.	57.7	1.1	1.04	1.12
How confident are you that you can figure out solutions when new situations or problems arise with your condition?	57.7	1.1	0.74	0.73
How confident are you that you can keep the symptoms of your disease from interfering with the things you want to do?	59.5	1.1	1.02	1.04

Ordering is by difficulty calibration.

SEM: SEM is the standard error of measurement in estimation of the item difficulty. SEM is the precision of the item difficulty estimation and is shown in 0–100 units.

Infit: Infit mean square error is one of two quality control fit statistics assessing item dimensionality (the degree to which the item falls on the same single, real number line as the rest of the items). Infit is an information-weighted residual of observed responses from model expected responses and is most sensitive to item fit when the item is located near the person's scale location.

Outfit: Outfit mean square error fit statistic is most sensitive to item dimensionality when the item scale location is distant from the person's scale location.

difficulty order of items on the scale suggests that they reflect a *developmental model of activation* (Bond and Fox 2001). Beliefs about the patient role and basic knowledge about one's condition and treatment appear to be important early developmental steps. Items in this early stage involve areas such as knowledge of medications and needed lifestyle changes as well as a belief that active involvement in one's health care is important. Only a small amount of activation is required to be able to endorse these items. Skills and confidence appear to come at later developmental steps. Items at the midpoint of the scale involve confidence that one can identify when medical care is needed, and that one can follow through on medical recommendations and handle symptoms on one's own. Items at the top of the activation continuum, indicating greatest activation, include maintaining needed lifestyle changes, having the confidence to handle new situations or problems, and keeping chronic illness from interfering with one's life.

Reliability Assessments

Rasch person reliability is the proportion of the total sample variability in measured activation that is not measurement error. Rasch person reliability provides upper and lower bounds to the estimate of the "true score" reliability of a measure. Real person reliability is calculated under the assumption that all of the misfit in the responses is due to departure of the data from the model's expectations. This is the lower bound reliability of the measurement of persons in this sample with this set of items. Model person reliability is based on the assumption that the data fit model expectations and that the misfit in the data is due to the probabilistic nature of the model. This is the upper-bound reliability. The true reliability of the measure lies somewhere between these lower and upper bounds. The Rasch person reliability for the preliminary 21-item measure was between .85 (real) and .87 (model). Cronbach's alpha was .87.

We also conducted a test–retest reliability assessment. Thirty respondents from the pilot survey were reinterviewed two weeks after the initial interview with the same protocol. For each person we calculated the precision of their measured activation at test and again at retest, measured by the standard error of measurement (SEM) for each person's estimated activation at each time point. The SEM times 1.96 provides the 95 percent confidence interval (CI) for each person's measured (estimated) activation. Twenty-eight of the 30 respondents had a retest activation estimate within the 95 percent CI of their test activation estimate.

Criterion Validity

To assess criterion validity, we interviewed 10 respondents from the pilot study: five who scored at the lowest end of the activation scale, and five who scored at the highest. An in-depth, open-ended, semistructured interview protocol was used to elicit elaborated explanations of how respondents dealt with common problems and challenges associated with managing their conditions, such as handling a situation with a physician who did not answer questions well, their responses to recommendations to change their lifestyle, and handling self-treatments on their own. The interviews were transcribed and three judges, blinded to the person's measured activation, reviewed and independently categorized each transcript as that of a person "low" or "high" in activation.

The three independent judges' classification of respondents agreed with their measured activation level (high or low) 83 percent of the time (or 25 of the 30 classifications were correct). Cohen's kappa for measured activation and each judge's classification were .80, .90, and .90 ($p < .001$ for all three kappas). No one respondent was misclassified by all three judges. These findings suggested that the preliminary measure had criterion validity when evaluated using the key criterion of self-described behavior.

STAGE 3: EXTENSION AND REFINEMENT OF THE PAM

Our goals for the third phase of scale development were to refine the measure and extend the range of activation assessed by the items. First, because the items in the preliminary scale calibrated only the midrange of activation (40–60), we tested items for possible inclusion that might extend the item difficulty. Second, because the items in the preliminary survey used several different response scales we tested the items using the same response scale for all items. Thus, the items were changed from questions to statements with the respondent indicating degree of agreement (four categories of degrees of agreement). Third, we wanted to assess how well the instrument would perform with a population that did not have a chronic disease. Fourth, we wanted to collect data from a larger sample to further assess the psychometric properties of the measure. Finally, we evaluated the use of a self-administered questionnaire.

Methods

A convenience sample of 486 respondents was recruited from among cardiac rehabilitation patients ($n = 120$) and employees of a large health system in a

second community $(n = 366)$. The employee sample responded to a web-based version of the survey; the clinic sample responded to a self-administered paper questionnaire. Twenty-four percent of the sample reported no chronic disease $(n = 118)$ and the remainder reported from 1 to 8 chronic illnesses.
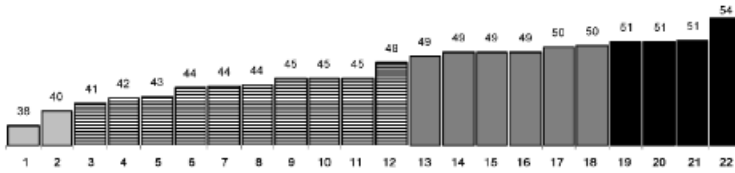
Findings

A Rasch rating scale model (Andrich 1978; Wright and Stone 1979) analysis yielded a 22-item measure (Figure 2).[1] Importantly, despite the slight change in item content, response categories, and the two different modes of administration, the findings confirm the item hierarchy observed in the preliminary 21-item scale. These results strongly suggest that activation is *developmental* in nature: the different elements of knowledge, belief, and skill that constitute activation have a hierarchical order, as shown in Figure 2.

In comparing this refined measure to our conceptual definition of activation, it appears that activation has four stages: The first involves beliefs about the importance of the patient role. The second involves the confidence and knowledge necessary to take action, including knowledge of medications and lifestyle changes, confidence in talking to health care providers and knowing when to seek help, and (at slightly higher levels of activation) confidence in following through on recommendations, knowing the nature and causes of the health condition, and different medical treatment options. The third stage involves actually taking action, including maintaining lifestyle changes, knowing how to prevent further problems, and handling symptoms on one's own. The fourth stage involves actually staying the course even when under stress. Patients who endorse these items are confident they can maintain lifestyle changes when under stress, that they can handle problems (rather than simply symptoms) on their own at home, and that they can keep their health problems from interfering with their life.

The structure of this probabilistic hierarchy of item difficulty implies that what is needed to increase activation depends on where the person is on the activation continuum. For example, those at the low end of activation may lack the belief that they have an important role to play in their health and lack elementary knowledge about their condition and their care. Respondents scoring in the mid range of the scale tend to have the necessary knowledge for self-care, but appear to lack some of the skills and confidence needed to carry through on all that is required for effective self-care. Those scoring at the higher end of the scale largely possess the necessary knowledge, skills, and

Figure 2:   Final Patient Activation Measure (PAM) with Item Calibrations

	Believes Active Role Important	SEM	Infit	Outfit
1	When all is said and done, I am the person who is responsible for managing my health condition	0.39	0.92	0.88
2	Taking an active role in my own health care is the most important factor in determining my health and ability to function	0.39	1.08	1.06
	Confidence and Knowledge to Take Action			
3	I know what each of my prescribed medications do	0.44	1.02	0.99
4	I am confident I can tell my health care provider concerns I have even when he or she does not ask	0.39	0.96	0.91
5	I am confident that I can tell when I need to go get medical care and when I can handle a health problem myself	0.39	0.86	0.82
6	I know the lifestyle changes like diet and exercise that are recommended for my health condition	0.39	0.71	0.67
7	I am confident that I can follow through on medical treatments I need to do at home	0.39	0.92	0.94
8	I am confident that I can take actions that will help prevent or minimize some symptoms or problems associated with my health condition	0.39	0.91	0.86
9	I am confident that I can find trustworthy sources of information about my health condition and my health choices	0.38	0.85	0.81
10	I am confident that I can follow through on medical recommendations my health care provider makes, such as changing my diet or doing regular exercise	0.38	0.92	0.88
11	I understand the nature and causes of my health condition(s)	0.40	0.95	0.91
12	I know the different medical treatment options available for my health condition	0.39	0.84	0.80
	Taking Action			
13	I have been able to maintain the lifestyle changes for my health that I have made	0.37	0.91	0.90
14	I know how to prevent further problems with my health condition	0.36	0.95	0.92
15	I know about the self-treatments for my health condition	0.37	0.87	0.85
16	I have made the changes in my lifestyle like diet and exercise that are recommended for my health condition	0.37	1.30	1.34
17	I am confident I can figure out solutions when new situations or problems arise with my health condition	0.36	1.00	1.02
18	I am able to handle symptoms of my health condition on my own at home	0.36	1.13	1.15
	Staying the Course under Stress			
19	I am confident that I can maintain lifestyle changes like diet and exercise even during times of stress	0.35	0.90	0.91
20	I am able to handle problems of my health condition on my own at home	0.35	1.07	1.13
21	I am confident I can keep my health problems from interfering with the things I want to do	0.35	1.16	1.25
22	Maintaining the lifestyle changes that are recommended for my health condition is too hard to do on a daily basis	0.33	1.44	1.64

Ordering is by difficulty calibration.

SEM: The standard error of measurement in estimation of the item difficulty. SEM is the precision of the item difficulty estimation and is shown in 0–100 units.

Infit: Infit mean square error is one of two quality control fit statistics assessing item dimensionality (the degree to which the item falls on the same single, real number line as the rest of the items). Infit is an information-weighted residual of observed responses from model expected responses and is most sensitive to item fit when the item is located near the person's scale location.

Outfit: Outfit mean square error fit statistic is most sensitive to item dimensionality when the item scale location is distant from the person's scale location.

confidence, but may be derailed from their course when they are under stress or encounter unexpected health events.[2]

The items have infit values between .76 and 1.32, well within the range required for a unidimensional measure. The Rasch person reliability for the 22-item measure was between .85 (real) and .88 (model). Cronbach's alpha was .91. Reliability statistics for those with and without chronic conditions are comparable.

In addition, an analysis to determine whether there were any observable mode effects was conducted. The log-odds equivalent of a Mantel-Haenszel differential item function analysis was conducted in *Winsteps* (2002) comparing web-based questionnaire and paper questionnaire item calibrations. No significant differences in item calibrations could be attributed to administration method.

STAGE 4: TESTING WITH A NATIONAL SAMPLE

Methods

This stage of the research evaluated the measure in a heterogeneous national probability sample to evaluate the performance of the measure across diverse groups and assess the construct validity of the measure.

Study Sample

A national probability sample ($N = 1,515$) of people 45 years and older was included in the telephone survey. Respondents were selected via a random digit dial selection and a screening question to determine age eligibility. No other eligibility requirement was employed. A 48 percent response rate was achieved with a protocol of a minimum of 12 call-backs. Many "no answer" or "busy" numbers had in excess of 20 attempts. Respondents ranged in age from 45 to 97, with 66 percent of the sample under the age of 65. Half the sample had a high school education or less and 32 percent had a household income of less than $25,000. Seventy-nine percent of the sample reported at least one chronic disease.[3] Among those with a chronic condition, 73 percent reported 2 or more conditions. Table 2 shows the distribution of the sample on other health and demographic characteristics.

The national sample largely mirrors census data for this age group. Differences between the sample and the census data are in gender distribution (census data 54 percent female, our sample 63 percent) and in the distribution on race (census data 83 percent white, our sample 88 percent)

Table 2: Reliability of 22-Item PAM (National Sample)

	N	%	Rasch Person Real	Rasch Person Model
Sample	1,515	100%	.87	.91
Gender				
Male	557	37%	.87	.90
Female	958	63%	.89	.91
Age Group				
45–54	560	38%	.88	.91
55–64	410	28%	.88	.91
65–74	300	20%	.89	.91
75–84	186	13%	.87	.90
85 or older	34	2%	.76	.82
Self-Rated Health				
Poor	109	7%	.83	.87
Fair	240	16%	.84	.88
Good	422	28%	.84	.87
Very Good	476	31%	.87	.90
Excellent	268	18%	.89	.91
Race				
White	1,326	88%	.88	.91
Black	114	8%	.84	.88
Other	68	5%	.89	.92
Education				
High school graduate or less	647	43%	.84	.88
Some college or trade school	391	26%	.89	.91
College graduate or more	469	31%	.89	.91
Household Income				
Less than $15,000	216	16%	.86	.89
$15,000 to $24,999	209	16%	.87	.90
$25,000 to $34,999	164	12%	.84	.88
$35,000 to $49,999	230	17%	.87	.90
$50,000 to $74,999	229	17%	.88	.91
$75,000 or more	286	21%	.88	.90
Chronic Condition				
None	323	21%	.87	.90
Angina/heart problem	191	13%	.88	.90
Arthritis	575	38%	.89	.91
Chronic pain	374	25%	.88	.91
Depression	219	15%	.87	.89
Diabetes	172	11%	.88	.91
Hypertension	510	34%	.88	.91
Lung disease	184	12%	.87	.91
Cancer	80	5%	.89	.91
High cholesterol	458	30%	.89	.91

Findings

The Rasch analysis of items from the national survey replicated the results obtained with the stage three pilot survey, showing the same developmental hierarchy of items and that the items maintain this same difficulty structure for both those with and without chronic illness.

Reliability

Assessments of the 22-item PAM using national sample data show a high level of reliability with infit values ranging from .71 to 1.44. All but one of the outfit statistics are between .80 and 1.34.

    The Rasch person reliability statistics for the measure are shown in Table 2 for the entire sample and meaningful subsamples. The consistency of performance of the measure is apparent in the reliability coefficients across subsamples. The high-reliability estimates indicate that the measure is appropriate for individual-level use, such as designing a care plan for an individual patient.

    Some other notable characteristics of the measure are apparent in Table 2. First, the measure performs well for both those with a chronic condition as for those with no chronic condition. It is also stable across differing levels of health status. Reliability is also stable across gender and different age groups with a slight decline in the oldest group (85+years). Finally, the measurement precision is stable across the several different chronic illnesses represented in the sample. This suggests that the measure can be reliably used to assess activation across a variety of subgroups in the population.

Validity

To assess construct and criterion validity, the 22-item PAM variables believed to be conceptually related to activation were examined for their relationship to measured activation. In addition, outcomes that are hypothesized to be a result of activation levels were examined, such as health behaviors and health functioning. Validity was assessed for the sample as a whole and for those with specific chronic illnesses. It was hypothesized that those with higher activation would be more likely to engage in specific self-care and preventive behaviors. Further, those with higher activation who have a specific chronic disease should be more likely to engage in the self-care behaviors specific to their condition (e.g., exercising to control arthritis pain). Similarly, it was hypothesized that those with higher measured activation should engage in other health "consumeristic" behaviors, such as seeking relevant health care

information, being persistent in getting clear answers from providers, and using comparative performance information to make health care choices. We further expected that those with greater activation would have better health and functioning and lower rates of health care utilization. Finally, because being activated implies having a sense of control over one's health, an item that is intended to measure "health fatalism"[4] was included. We hypothesized that those with more activation would indicate less fatalism about their future health.

The results indicate considerable evidence for the construct validity of the PAM. Those with higher activation report significantly better health as measured by the SF 8 ($r = .38$, $p < .001$), and have significantly lower rates of doctor office visits, emergency room visits, and hospital nights ($r = -.07$, $p < .01$). Those with higher activation are significantly more likely to exercise regularly, follow a low-fat diet, eat more fruits and vegetables, and not smoke (Table 3). In addition, those with higher activation are significantly more likely to engage in consumeristic health behaviors, such as finding out about a new provider's qualifications. Self-management behaviors associated with specific conditions are also significantly associated with measured activation levels. For instance, diabetics with higher activation are more likely to keep a glucose journal, more-activated arthritics are more likely to exercise, and among those with high cholesterol, those with higher activation are more likely to follow a low-fat diet. Finally those with higher activation indicate a lower degree of fatalism about their health.

These findings indicate that the measure has a high degree of construct and criterion validity. Future work is needed to determine the predictive validity of the measure, its sensitivity to detect changes in underlying behavior, and the types of interventions that help people move up the activation scale. Research is underway to assess predictive validity and sensitivity to changes in self-care behaviors.

CONCLUSIONS

There is wide agreement that engaging patients to be an active part of the care process is an essential element of the quality of care. Any serious attempts to improve this aspect of care will require three essential steps: (1) The development of a measure to assess patient activation; (2) The identification and use of evidenced-based interventions to increase patient activation; and (3) A method to hold providers and delivery systems accountable for

Table 3: Association between Scale Calibration and Construct Validity Indicators

Variable Answer Category (n)	Mean Score on Measure	F; df	P
General Preventive Behaviors			
Follow a low-fat diet: *Always or almost always (798)*	61.1	72.8; 1,1512	.001
Sometimes or never (716)	56.2		
Follow regular exercise schedule: *Yes (884)*	61.3	116.3; 1,1511	.001
No (629)	55.2		
Five servings of fruits or vegetables per day		68.1; 1,1512	.001
At least four days per week (755)	61.2		
Three days per week or less (759)	56.4		
Smoke tobacco: *Yes (262)*	56.3	15.1; 1,1512	.001
No (1,252)	59.3		
Disease-Specific Behaviors			
Diabetes			
Use glucose journal: *Always or almost always (110)*	57.7	4.9; 1,157	.05
Sometimes or never (49)	53.8		
Arthritis			
Arthritis exercise: *Always or almost always (254)*	60.6	60.2; 1,571	.001
Sometimes or never (319)	53.9		
High Cholesterol			
Follow a low-fat diet: *Always or almost always (256)*	60.4	37.6; 1,456	.001
Sometimes or never (202)	54.2		
Consumeristic Behaviors			
Before I go to a new health care provider, I find out as much as I can about his or her qualifications.		182.2; 2,1416	.001
Disagree or strongly disagree (232)	54.0		
Agree (880)	56.5		
Strongly agree (307)	68.2		
When I do not understand, I am persistent in asking my health care provider to explain something until I understand it.		322.9; 2,1491	.001
Disagree or strongly disagree (79)	50.4		
Agree (992)	55.3		
Strongly agree (423)	68.5		
As far as I know medical science has developed guidelines for treating my condition.*		150.4; 2,175	.001
Disagree or strongly disagree (54)	48.4		
Agree (572)	55.3		
Strongly agree (102)	69.7		
Health Fatalism			
In the next 5 years how likely do you think it is that you will develop a new or additional health condition that requires ongoing medical care?		81.3; 1,1431	.001
Very likely or likely (526)	55.6		
Unlikely or very unlikely (907)	61.1		

*Asked only of respondents with a condition for which there are guidelines.

supporting and increasing patient activation. The first step of developing a measure is necessary before the other two steps can be attempted.

The Patient Activation Measure (PAM) appears to be a valid and reliable instrument to measure activation. The measure has strong psychometric properties and appears to tap into the developmental nature of activation. Because the measure is highly reliable at the person level, it is possible to use it on an individual patient basis to diagnose activation and individualize care plans. Moreover, because the measure maintains precision across different demographic and health status groups, it can also be used at the aggregate level to evaluate and compare the efficacy of interventions and health care delivery systems.

It is not unreasonable to expect that providers delivering high-quality care would have, over time, more-activated patients. Changes in the activation levels of patient populations might be used as an indicator of the performance of providers or delivery systems, and be employed for quality assessment and public accountability purposes. Consumers will likely want to know which providers and systems are performing well in this area and comparative data could drive purchaser and consumer choices.

The PAM may be useful for both designing interventions and in evaluating them. The measure can be used in a clinical setting to assess individual patients and to develop care plans tailored to that patient and integrated into the processes of their care. Because the measure is developmental, interventions could be tailored to the individual's stage of activation. For example, those at early stages of activation would need interventions designed to increase knowledge about their condition and their treatments. Patients at later stages would need interventions designed to increase their skills and confidence in the different self-management tasks. As patients advance in activation, the type of interventions that will be helpful to them will also change. The approach is economical because it is targeted rather than omnibus. Employers could also use the measure to assess interventions designed to increase engagement and activation among their employees. In summary, wide use of a precise, valid, and useful measure is the first step toward the goal of informed and engaged patients and ultimately to more effective and efficient delivery systems. The measurement properties of the Patient Activation Measure (PAM) when assessed using the stringent Rasch model suggest that it could fulfill that role.

Having a valid and reliable measure is the very first step in understanding patient activation and its role in health care quality, outcomes,

and cost containment. Of course, the validity of the measure is limited by our current level of understanding of activation. As our understanding of the construct increases through the use of the measure, it should be anticipated that refinement of the measure will be necessary.

ACKNOWLEDGMENTS

NOTES

1. For the changes in items between phase 2 and phase 3, see online-only Appendix A, Note 1. The Appendix is available at http://www.blackwell-synergy.com.
2. The PAM can be scored using a Rasch score table that converts curvilinear summated raw scores to linear, interval scores. This is essential to obtain accurate scores. To obtain a copy of the score table and instructions, contact the first author.
3. For explanation of differences in item wording depending on chronic disease status, see online-only Appendix A, Note 2.
4. For exact wording of fatalism item, see online-only Appendix A, Note 3.

REFERENCES

Andrich, D. 1978. "A Rating Formulation for Ordered Response Categories." *Psychometrica* 43 (4): 561–73.

Bandura, A. 1991. "Self-efficacy Mechanism in Physiological Activation and Health-Promoting Behavior." In *Adaption, Learning and Affect*, edited by J. Madden, S. Matthysse, and J. Barchas, pp. 226–69. New York: Raven Press.

Bodenheimer, T., K. Lorig, H. Holman, and K. Grumbach. 2002. "Patient Self-Management of Chronic Disease in Primary Care." *Journal of the American Medical Association* 288 (19): 2469–75.

Bond, T., and C. Fox. 2001. *Applying the Rasch Model: Fundamental Measurement in the Human Sciences*. Mahwah, NJ: Erlbaum.

Day, J., C. W. Bodmer, and O. M. Dunn. 1996. "Development of a Questionnaire Identifying Factors Responsible for Successful Self-management of Insulin-Treated Diabetes." *Diabetic Medicine* 13 (6): 564–73.

DiClemente, C. C., J. O. Prochaska, S. K. Fairhurst, W. F. Velicer, M. M. Velasquez, and J. S. Rossi. 1991. "The Process of Smoking Cessation: An Analysis of Precontemplation, Contemplation and Preparation Stages of Change." *Journal of Consulting and Clinical Psychology* 59 (2): 295–304.

Gabel, J. R., A. T. Lo Sasso, and T. Rice. 2002. "Consumer-Driven Health Plans: Are They More Than Talk Now?" *Health Affairs* Web exclusive, available at http://www.healthaffairs.org/WebExclusives/2201Gabel.pdf.

Glasgow, R. E., M. M. Funnell, A. E. Bonomi, C. Davis, V. Beckham, and E. H. Wagner. 2002. "Self-management Aspects of the Improving Chronic Illness Care Breakthrough Series: Implementation with Diabetes and Heart Failure Teams." *Annals of Behavioral Medicine* 24 (2): 80–7.

Glasgow, R. 2002. "Technology and Chronic Care." Paper presented at the Congress on Improving Chronic Care: Innovations in Research and Practice. September 8–10, Seattle, Washington.

Greenfield, S., S. Kaplan, and J. E. Ware. 1985. "Expanding Patient Involvement in Care. Effects on Patient Outcomes." *Annals of Internal Medicine* 102 (4): 520–8.

Greenfield, S., S. Kaplan, J. E. Ware, E. M. Yano, and H. J. Frank. 1988. "Patients' Participation in Medical Care: Effects on Blood Sugar Control and Quality of Life in Diabetes." *Journal of General Internal Medicine* 3 (5): 448–57.

Grembowski, D. E., D. L. Patrick, P. Diehr, M. Durham, S. Beresford, E. Kay, and J. Hecht. 1993. "Self-Efficacy and Health Behavior among Older Adults." *Journal of Health and Social Behavior* 34 (2): 89–104.

Hibbard, J. H., J. J. Jewett, S. Engelmann, and M. Tusler. 1998. "Can Medicare Beneficiaries Make Informed Choices?" *Health Affairs* 17 (6): 181–93.

Hibbard, J. H., M. Greenlick, H. Jimison, J. Capizzi, and L. Kunkel. 2001. "The Impact of a Community-Wide Self-Care Information Project on Self-Care and Medical Care Utilization." *Evaluation and the Health Professions* 24 (4): 404–23.

Isaacs, S. L. 1996. "Consumer's Information Needs: Results of a National Survey." *Health Affairs* 15 (4): 31–41.

Kahn, D. A., J. P. Docherty, D. Carpenter, and A. Frances. 1997. "Consensus Methods in Practice Guideline Development: A Review and Description of a New Method." *Psychopharmacology Bulletin* 33 (4): 631–9.

Kaplan, S., S. Greenfield, and J. E. Ware. 1989. "Assessing the Effects of Physician–Patient Interactions on the Outcomes of Chronic Disease." *Medical Care* 27 (3, supplement): S110–27.

Linacre, J. M. 2002. *Winsteps Manual.* Chicago: Winsteps.

Lorig, K. 1996. *Outcome Measures for Health Education and Other Health Care Interventions.* Thousand Oaks, CA: Sage.

Lorig, K. R., D. S. Sobel, A. L. Stewart, B. W. Brown, A. Bandura, P. Ritter, V. M. Gonzalez, D. D. Laurent, and H. R. Holman. 1999. "Evidence Suggesting That a Chronic Disease Self-Management Program Can Improve Health Status While Reducing Hospitalization: A Randomized Trial." *Medical Care* 37 (1): 5–14.

Marshall, M. N., P. G. Shekelle, R. H. Brook, and S. Leatherman. 2000. "Use of Performance Data to Change Physician Behavior." *Journal of the American Medical Association* 284 (9): 1079.

Massof, R. W. 2002. "The Measurement of Vision Disability." *Optometry and Vision Science* 79 (8): 516–52.

O'Leary, A. 1985. "Self-Efficacy and Health." *Behaviour Research and Therapy* 23 (4): 437–51.

Prochaska, J. O., C. A. Redding, and K. E. Evers. 1997. "The Transtheoretical Model and Stages of Change." In *Health Behavior and Health Education*, 2d ed., edited by K. Glanz, F. M. Lewis, and B. K. Rimer, pp. 60–84. San Francisco: Jossey-Bass.

Rasch, G. 1960. *Probabilistic Models for Some Intelligence and Attainment Tests* (reprint, with Foreword and Afterword by B. D. Wright, Chicago: University of Chicago Press, 1980). Copenhagen, Denmark: Danmarks Paedogogiske Institut.

Smith, R. M. 1996. "Polytomous Mean-Square Fit Statistics." *Rasch Measurement Transactions* 10 (3): 516–7.

Thorndike, R. M., and E. P. Hagen. 1991. *Measurement and Evaluation in Psychology and Education.* New York: Macmillan.

Von Korff, M., J. E. Moore, K. Lorig, D. C. Cherkin, K. Saunders, V. M. Gonzalez, D. Laurent, C. Rutter, and F. Comite. 1998. "A Randomized Trial of a Lay Person-Led Self-Management Group Intervention for Back Pain Patients in Primary Care." *Spine* 23 (23): 2608–15.

Von Korff, M., J. Gruman, J. Schaefer, S. J. Curry, and E. H. Wagner. 1997. "Collaborative Management of Chronic Illness." *Annals of Internal Medicine* 127 (12): 1097–102.

Von Korff, M., W. Katon, T. Bush, E. H. Lin, G. E. Simon, K. Saunders, E. Ludman, E. Walker, and J. Unutzer. 1998. "Treatment Costs, Cost Offset, and Cost-Effectiveness of Collaborative Management of Depression." *Psychosomatic Medicine* 60 (2): 143–9.

Wallston, K. A., M. J. Stein, and C. A. Smith. 1994. "Form C of the MHLC Scales: A Condition-Specific Measure of Locus of Control." *Journal of Personality Assessment* 63 (3): 534–53.

Wasson, J. H., T. A. Stukel, J. E. Weiss, R. D. Hays, A. M. Jette, and E. C. Nelson. 1999. "A Randomized Trial of the Use of Patient Self-Assessment Data to Improve Community Practices." *Effective Clinical Practice* 2 (1): 1–10.

Winsteps. 2002. *Winsteps: Rasch Model Statistical Software* (Version 337). Chicago: Winsteps.

Wright, B. D., and G. Masters. 1982. *Rating Scale Analysis.* Chicago: Mesa Press.

Wright, B. D., and M. H. Stone. 1979. *Best Test Design.* Chicago: Mesa Press.