

Original Paper

A Web-Based Computerized Adaptive Testing (CAT) to Assess Patient Perception in Hospitalization

Tsair-Wei Chien¹, MBA; Wen-Chung Wang², PhD; Sheng-Yun Huang², SRA; Wen-Pin Lai³, MD; Julie Chi Chow⁴, MD

¹Department of Management, Chi Mei Medical Center, Yungkuang, Taiwan

²Assessment Research Center, The Hong Kong Institute of Education, Hong Kong, China

³Department of Emergency, Chi Mei Medical Center, Yungkuang, Taiwan

⁴Department of Paediatrics, Chi Mei Medical Center, Yungkuang, Taiwan

Corresponding Author:

Julie Chi Chow, MD

Department of Paediatrics

Chi Mei Medical Center

No. 901 Junghua Rd.

Yungkuang, 710

Taiwan

Phone: 886 62812811 ext 52903

Fax: 886 62820534

Email: jchow@mail.chimei.org.tw

Abstract

Background: Many hospitals have adopted mobile nursing carts that can be easily rolled up to a patient's bedside to access charts and help nurses perform their rounds. However, few papers have reported data regarding the use of wireless computers on wheels (COW) at patients' bedsides to collect questionnaire-based information of their perception of hospitalization on discharge from the hospital.

Objective: The purpose of this study was to evaluate the relative efficiency of computerized adaptive testing (CAT) and the precision of CAT-based measures of perceptions of hospitalized patients, as compared with those of nonadaptive testing (NAT). An Excel module of our CAT multicategory assessment is provided as an example.

Method: A total of 200 patients who were discharged from the hospital responded to the CAT-based 18-item inpatient perception questionnaire on COW. The numbers of question administered were recorded and the responses were calibrated using the Rasch model. They were compared with those from NAT to show the advantage of CAT over NAT.

Results: Patient measures derived from CAT and NAT were highly correlated ($r = 0.98$) and their measurement precisions were not statistically different ($P = .14$). CAT required fewer questions than NAT (an efficiency gain of 42%), suggesting a reduced burden for patients. There were no significant differences between groups in terms of gender and other demographic characteristics.

Conclusions: CAT-based administration of surveys of patient perception substantially reduced patient burden without compromising the precision of measuring patients' perceptions of hospitalization. The Excel module of animation-CAT on the wireless COW that we developed is recommended for use in hospitals.

(*J Med Internet Res* 2011;13(3):e61) doi: [10.2196/jmir.1785](https://doi.org/10.2196/jmir.1785)

KEYWORDS

Computerized adaptive testing; computer on wheels; classic test theory; IRT; item response theory; nonadaptive testing

Introduction

As computer technology and health care become more integrated, many hospitals have adopted mobile nursing carts that can be easily rolled up to a patient's bedside to access charts

and help nurses perform their rounds [1-3]. Besides increasing efficiency by including basic functions such as billing records and decreasing the number of trips nurses need to take to the medication room [3], the carts can reduce patient burden by allowing them to answer questions on activities of daily living

using computerized adaptive testing (CAT) [1]. However, few papers have reported data regarding the bedside use of wireless computers on wheels (COW) to collect questionnaire-based information on their perception of hospitalization on discharge from the hospital. Collecting patients' feedback on their perspectives has become an important part of patient involvement and participation for health caregivers; thus, this question is important [4-6].

Gathering Feedback Efficiently From Patients

Two new modes of survey administration have been reported to make surveys more easily accessible to those who cannot read or write [7]. These include using automated telephone technology through an interactive voice response system and using Internet-like visualizations to complete questionnaires online. In medical practice, hospital staff usually hand a questionnaire to patients at the end of their visit and ask them to complete it prior to leaving hospital. At the Picker Institute Europe [5], questionnaires are sent annually to a randomized list of eligible patients who had been discharged from the hospital. Both of these methods are less prompt and efficient than using wireless COW to collect data on patients' perspectives on being discharged from the hospital.

Computer Assessment and Computer-Adaptive Testing

There is no doubt that using wireless COW at a patient's bedside is an efficient way of instantly gathering feedback from patients. Traditional paper-and-pencil or computer-based devices (nonadaptive testing [NAT]) impose a large respondent burden because patients are required to answer all the questions. In contrast, CAT-based tests developed using item response theory (IRT) [8] can achieve a similar degree of measurement precision to NAT using only about half the test length [1,9-11]. Most studies investigating IRT- and CAT-based tests have evaluated both efficiency and precision for CAT-based tests with dichotomous items. Whether CAT-based tests with polytomously scored items (CAT as defined in this study) can be incorporated with wireless COW in hospitals for gathering feedback from patients should be investigated.

Rasch Analysis

In classical test theory, raw scores (or linear transformation scores, eg, T score) are usually adopted as respondent measures. However, subsequent parametric statistical analyses, such as computing mean, variance, correlation coefficient, and Cronbach alpha [12,13], would be incorrect because raw scores are not on an additive interval scale [14].

To overcome this obstacle, the IRT-based Rasch model [15], a probabilistic relationship between a person's level of a latent trait (commonly referred to as ability or measure) and an item's property (difficulty or threshold), was developed. Both person ability and item difficulty (calibrated in terms of log odds or logits) are located along the same continuum. A useful scale (or questionnaire) is usually examined by 3 important criteria

for the Rasch model, namely, unidimensionality, item fit, and item invariance (or so-called differential item functioning [16]). These criteria are detailed in Smith et al [17]. There are many published papers [1,18-21] of studies using the Rasch model to develop CAT in clinical settings, but none of them have incorporated the Internet-based polytomously scored CAT to gather feedback from patients in hospitals.

Objectives

The purpose of this study was to evaluate the relative efficiency of an Internet-based polytomously scored CAT and the precision of CAT-based measures of perceptions of hospitalized patients, as compared with those measured by NAT. An Excel (Microsoft Corporation, Redmond, WA, USA) module of our CAT multicategory assessment is provided as an example.

Methods

Data collection

Participants

The study sample was recruited from inpatients at a 1333-bed medical center in southern Taiwan. Patients who had been discharged were selected randomly by the digit code of their invoice number during each morning and afternoon interval from Monday through Friday in summer 2010.

Procedure

As an incentive for participation, patients were offered a gift card for US \$3.20 good for purchases at 7-11 convenience stores. A total of 200 patients either completed the questionnaire on COW themselves or were helped by a trained volunteer (if they were unable to personally complete the questionnaire); proxies were allowed because most of those helping patients carry out their discharge procedure were the patients' family members or friends. Time spent by each patient was recorded in Excel after they completed the questionnaire. This study was approved and monitored by the Research and Ethical Review Board of the Chi-Mei Medical Center, Tainan, Taiwan.

Tool: CAT-Format Questionnaire

We designed the 18-item CAT questionnaire in Excel based on an 18-item inpatient perception questionnaire (IPQ-18) [5]; see Table 1). Unidimensionality, local independence, item fit, and differential item functioning using the Rasch model to investigate these criteria have been previously reported [5].

Data collected from the patients included demographic characteristics (gender, treatment department, age, and person completing survey, ie, proxy or patient); perception measure in a logit unit; number of items needed to be completed; and mean square errors (MNSQ) of infit and outfit (indicators of response patterns for the IPQ-18 scale [5]) (see Table 1, Multimedia Appendix 1, and Multimedia Appendix 2).

Table 1. Items of the 18-item scale ordered by item overall difficulties

| Item number | Scale content | | Difficulty | | | | |
|-------------|-----------------------|---|------------|-------|-------|-------|-------|
| | Category ^a | Item | Overall | Step1 | Step2 | Step3 | Step4 |
| 39 | L | Did staff tell you about medication side effects when going home? | 3.78 | 0.02 | 1.87 | 5.35 | 7.89 |
| 41 | L | Did doctors or nurses give your family information needed to help you? | 2.76 | -1.00 | 0.85 | 4.33 | 6.87 |
| 27 | N | Did hospital staff talk about your worries and fears? | 2.22 | -1.54 | 0.31 | 3.79 | 6.33 |
| 11 | W | Were you ever bothered by noise at night from other patients? | 1.58 | -2.18 | -0.33 | 3.15 | 5.69 |
| 24 | N | Were you involved in decisions about your care and treatment? | 0.67 | -3.09 | -1.24 | 2.24 | 4.78 |
| 30 | N | How long was it after using the call button before you got the help you needed? | 0.42 | -3.34 | -1.49 | 1.99 | 4.53 |
| 42 | L | Did staff tell you how to contact them if worries arose after leaving? | -0.3 | -4.06 | -2.21 | 1.27 | 3.81 |
| 9 | A | Did you feel you waited a long time to get to a bed on a ward? | -0.63 | -4.39 | -2.54 | 0.94 | 3.48 |
| 44 | O | How would you rate how well the doctors and nurses worked together? | -0.71 | -4.47 | -2.62 | 0.86 | 3.4 |
| 2 | A | How organized was the care you received in the emergency room? | -0.95 | -4.71 | -2.86 | 0.62 | 3.16 |
| 5 | A | Were you given enough notice of your date of admission? | -1.08 | -4.84 | -2.99 | 0.49 | 3.03 |
| 12 | W | Were you bothered by noise at night from hospital staff? | -1.1 | -4.86 | -3.01 | 0.47 | 3.01 |
| 17 | D | Did you have confidence and trust in the doctors treating you? | -1.1 | -4.86 | -3.01 | 0.47 | 3.01 |
| 23 | N | Did staff say one thing and something quite different happened to you? | -1.1 | -4.86 | -3.01 | 0.47 | 3.01 |
| 38 | L | Did staff explain the purpose of the medicines so that you could understand? | -1.1 | -4.86 | -3.01 | 0.47 | 3.01 |
| 18 | D | Did doctors talk in front of you as if you weren't there? | -1.12 | -4.88 | -3.03 | 0.45 | 2.99 |
| 19 | N | Did you get answers that you could understand from a nurse? | -1.12 | -4.88 | -3.03 | 0.45 | 2.99 |
| 34 | P | Did hospital staff do everything they could to help you control your pain? | -1.12 | -4.88 | -3.03 | 0.45 | 2.99 |

^a Categories are A: admission to hospital; D: doctors; L: leaving hospital; N: nurses; O: overall; P: pain; W: hospital and ward.

CAT Procedure

Outfit and Infit Statistics

Outfit statistics are based on unweighted sum of squared standardized residuals and are sensitive to unexpected outlying, off-target, and low-information responses; whereas infit statistics are based on weighted sum of squared standardized residuals and are sensitive to unexpected inlying patterns among informative and on-target observations [22]. Smith [23] found that Rasch outfit MNSQ approaching 1.0 [24] demonstrates a higher power than its counterpart of infit MNSQ. Outfit MNSQ of 2.0 or greater for a patient indicate a possibly aberrant response pattern [24].

CAT Procedures and Features

We programmed a Visual Basic for Applications (VBA) module in Microsoft Excel and on the Internet (<http://www.healthup.org.tw/cat.asp>, <http://www.webcitation.org/60xWv6p6d>) complying with several rules and criteria for CAT-based testing on COW (Figure 1, Figure 2). The person separation reliability (similar to Cronbach alpha) calculated from the original paper [5] was 0.94 (mean 2.64, SD 2.09). Based on this number, the CAT stop rule for measurement of standardized error was determined to be $0.51(\text{SD} \times \sqrt{1 - \alpha}) = 2.09 \times \sqrt{1 - 0.94}$.

Figure 1. Using a wireless computer on wheels (COW) to collect data on patients' perspectives on hospitalization



Figure 2. Snapshot of computerized adaptive testing (CAT)-based inpatient perception questionnaire for patients

CAT based inpatient satisfaction survey

| No | Measure | response | [I-score] |
|----|---------|----------|-----------|
| 10 | 3.14 | 3 | 0.77 |
| 5 | 2.47 | 4 | 1.88 |
| 4 | 2.1 | 3 | 0.79 |
| 1 | 1.87 | 2 | 0.26 |
| 2 | 1.67 | 3 | 1.58 |
| 3 | 1.19 | 3 | 1.22 |
| 6 | 0.64 | 4 | 1.73 |
| 9 | -4.39 | 3 | 0.62 |
| 8 | -2.21 | 1 | 3.97 |
| 7 | 0 | 2 | 2.03 |

| Item | Item | difficult |
|------|---------------------|-----------|
| 1 | Staff told you abo | 3.78 |
| 2 | Doctors or nurses | 2.76 |
| 3 | Hospital staff talk | 2.22 |
| 4 | Ever bothered by | 1.58 |
| 5 | Were you involved | 0.67 |
| 6 | How long after usi | 0.42 |
| 7 | Staff told you how | -0.3 |
| 8 | You feel wait a lo | -0.63 |
| 9 | How would you ra | -0.71 |
| 10 | How organized wa | -0.95 |
| 11 | Were you given a | -1.08 |
| 12 | Bothered by noise | -1.1 |
| 13 | Having confidence | -1.1 |
| 14 | Staff saying one t | -1.1 |
| 15 | Staff explained th | -1.1 |
| 16 | Doctors talked in f | -1.12 |
| 17 | You get answers t | -1.12 |
| 18 | Hospital staff did | -1.12 |

1.NO: 001414 + Date: 2010/10/24

email: [input field]

2.Over satisfied

Gender [dropdown] Age [dropdown] Dept [dropdown] Proxy [dropdown]

Extreme N Very NS Not Sati. Fair Satisfied Very Sati. Extreme S

How organized was the care you received in A&E?

Strongly Not Sati Fair Satisfied Strongly

Start End Scoring to_excel

Output Satisfied 3.14

SE 0.47

Inform 4.53

Outfit 3.21

Infit 3.3

Item 10

send by mail

Next Guide

We also set another stop rule so that the minimum number of questions required for completion was 10 items (10/18, 56%), because CAT achieves a similar measurement precision to NAT with only about half the test length [1,9-11]. The initial question was selected from the pool of 18 items according to the patient's overall perception of satisfaction in their hospitalization. The provisional person measure is estimated by the maximum of the log-likelihood function using an iterative Newton-Raphson procedure [1,10] (Multimedia Appendix 2) after 3 items were responded to, without all answers corresponding to either 0 or 4. The next question selected was the one with the highest information value among the remaining unanswered questions weighted against the provisional person measure. The details of CAT procedures are shown in Multimedia Appendix 2 and Multimedia Appendix 3.

Comparison of Efficiency Between NAT and CAT

Two indicators used to examine CAT efficiency in this study are (1) whether the number of questions needed was significantly less than for NAT (18 questions) and (2) whether the precision of person measures was less than that for NAT. We used paired *t* tests to make these two statistical inferences.

Accordingly, the perception measure based on NAT should be estimated in advance for each patient who was assumed to have answered all 18 items. The following steps were thus followed: (1) we used a standard item response-generation method [25-29] to generate responses to 18 questions for each patient with given question difficulty parameters (in Table 1) and a patient perception measure (by CAT), and (2) we applied the rectangle metric of 18 questions \times 200 persons to re-estimate NAT measures for each patient using WINSTEPS software (WINSTEPS version 3.72.0: Winsteps.com, Chicago, IL, USA)

(the 18-question difficulties are anchored in WINSTEPS with iaf file shown in Multimedia Appendix 2).

Statistical Analysis

SPSS software for Windows (Version 12, SPSS, Chicago, IL) was used for all statistical analysis.

Descriptive Statistics

Data on patient gender, age, treatment department, and proxy respondent were collected. Noncontinuous variables were reported as frequency and percentages, and were examined by chi-square tests.

Analytic statistics

For continuous variables, CAT and NAT measures were compared using the Pearson correlation coefficient. Patient perception measures obtained by CAT were compared between groups using *t* tests or analysis of variance (ANOVA). Time spent by patients was averaged by using logarithmic transformation and reported as mean (SD) by exponential function. All analyses were considered to be statistically significant at the .05 alpha level.

Results

As seen in Table 2, there were no significant associations between gender and other demographic characteristics (ie, treatment department, age, and participant). Among inpatients we approached, 13% (31/231) were unwilling to respond to the CAT questions due to personal reasons, despite the incentive we offered. CAT and NAT measures were highly correlated ($r = 0.98$).

Table 2. Demographic characteristics of the study population (N = 200)

| Variable | Male | | Female | | Total | $\chi^2_{(r-1)*(c-1)}$ | |
|---------------------------|------|----|--------|----|-------|------------------------|---------|
| | n | % | n | % | | Test | P value |
| Respondent | | | | | | 0.6 | .45 |
| Willing to participate | 100 | 50 | 100 | 50 | 200 | | |
| Unwilling to participate | 13 | 42 | 18 | 58 | 31 | | |
| Age (years) | | | | | | 0.9 | .82 |
| ≤16 | 31 | 31 | 25 | 25 | 56 | | |
| 17–40 | 27 | 27 | 30 | 30 | 57 | | |
| 41–70 | 25 | 25 | 27 | 27 | 52 | | |
| >70 | 17 | 17 | 18 | 18 | 35 | | |
| Department | | | | | | 3.9 | .42 |
| Internal medicine | 44 | 44 | 41 | 41 | 85 | | |
| Surgery | 28 | 28 | 22 | 22 | 50 | | |
| Obstetrics and gynecology | 8 | 8 | 14 | 14 | 22 | | |
| Pediatrics | 11 | 11 | 7 | 7 | 18 | | |
| Other | 12 | 12 | 16 | 16 | 28 | | |
| Participant /proxy | | | | | | 1.1 | .57 |
| Family | 75 | 75 | 81 | 81 | 156 | | |
| Friend | 15 | 15 | 12 | 12 | 27 | | |
| Patient | 10 | 10 | 7 | 7 | 17 | | |

Mean time spent by patients in CAT was 54.91 seconds (SD 8.00; maximum 76; minimum 33). As shown in Table 3, CAT required substantially fewer questions than NAT ($P < .001$). NAT required all of the 200 patients to respond to all 18 questions, and thus yielded a total of 3600 responses. In CAT,

a total of 2083 responses were required, meaning that on average a patient answered 10.42 questions. Thus, as compared with NAT, CAT received an efficiency gain in test length of 0.42 (defined as $1 - \text{ratio of total responses by CAT and NAT: } 1 - 2084/3600$).

Table 3. Comparison of computerized adaptive testing (CAT) versus nonadaptive testing (NAT) (all questions having to be answered) in efficiency as assessed by paired t test

| | Mean | Variance | Response | Maximum | Minimum | Paired t_{199} test | P value |
|---|--------------------|--------------------|-------------------|-----------------|-----------------|-----------------------|---------|
| Test length (number of questions answered) | | | | | | | |
| NAT | 18 | 0.00 | 3600 ^b | 18 | 18 | -476.72 | <.001 |
| CAT | 10.42 | 0.25 | 2084 ^b | 12 | 10 | | |
| Estimated measures(mean and variance) | | | | | | | |
| NAT | 0.69 | 2.66 | 3600 | 4.16 | -2.69 | 1.10 | .14 |
| CAT | 0.71 | 2.62 | 2084 | 4.00 | -2.56 | | |
| Time spent (seconds) | | | | | | | |
| CAT | 54.91 ^c | 64.04 ^c | 2084 | 76 ³ | 33 ³ | | |

^aEfficiency = $(1 - 2084/3600) = 0.58$.

^b3600 = 200×18 ; 2084 = 200×10.42 .

^cOn second unit.

Regarding precision of measurement, person measures from CAT did not statistically differ from those from NAT ($P = .14$). ANOVA revealed that patient perception measures did not differ

between groups across elements; t test analyses showed that they also did not differ between genders (Table 4).

Table 4. Comparison of inpatient perception by demographic characteristic

| Variable | Male | | Female | | ANOVA ^a | |
|-----------------------------|------|------|--------|------|--------------------|---------|
| | Mean | SD | Mean | SD | Test | P value |
| Proportion | 0.77 | 1.59 | 0.65 | 1.66 | $t_{398} = 0.55$ | .59 |
| Age (years) | | | | | $F_{3,196} = 0.71$ | .55 |
| ≤16 | 0.77 | 1.72 | 0.83 | 1.81 | -0.12 | .89 |
| 17–40 | 1.23 | 1.54 | 0.58 | 1.40 | 1.68 | .09 |
| 41–70 | 0.72 | 1.48 | 0.53 | 1.74 | 0.42 | .67 |
| >70 | 0.13 | 1.45 | 0.69 | 1.83 | -1.00 | .32 |
| Department | | | | | $F_{4,195} = 0.92$ | .45 |
| Internal medicine | 0.65 | 1.53 | 0.49 | 1.48 | 0.47 | .63 |
| Surgery | 0.61 | 1.56 | 0.9 | 1.77 | -0.77 | .44 |
| Obstetrics and gynecology | 1.00 | 1.91 | 0.77 | 1.70 | 0.28 | .77 |
| Pediatrics | 0.45 | 1.79 | 0.19 | 2.00 | 0.30 | .78 |
| Other | 1.73 | 1.29 | 0.68 | 1.85 | 1.67 | .11 |
| P participant/p roxy | | | | | $F_{2,197} = 0.36$ | .69 |
| Family | 0.90 | 1.58 | 0.60 | 1.62 | 1.14 | .25 |
| Friend | 0.58 | 1.60 | 0.93 | 2.10 | -0.49 | .62 |
| Patient | 0.16 | 1.62 | 0.72 | 1.43 | -0.73 | .47 |

^a Analysis of variance.

Total person mean 0.71 logits (SD 1.62); median 0.59; coefficient of skewness 0.103 ($P = .54$); coefficient of kurtosis -0.89 ($P = .03$); D'Agostino-Pearson test for normal distribution accept normality ($P = .09$).

Discussion

Key Finding

The results from this study indicate that CAT-based testing using COW can reduce patient (or proxy) burdens. It is up to 42% more efficient in answering questions and achieves a very similar degree of measurement precision to NAT.

What This Adds to What Was Known

Consistent with the literature [1,9-11,30], the efficiency of CAT was supported. We confirmed that the CAT-based IPQ-18 on COW requires significantly fewer questions to measure patient perception than NAT, but does not compromise precision of measurement.

What is the Implication, What Should be Changed

Using an Excel module of animation for CAT on COW as a tool that can help hospital staff efficiently and precisely gather feedback from patients is technically feasible. Outfit MNSQ of 2.0 or greater can be used to examine whether patient responses are distorted or abnormal—that is, many more responses unexpectedly did not fit the model's requirement and were deemed to be very likely to be careless, mistaken, or awkward [1,5,6,24]. Thus, CAT makes detection of problematic responses easier—normally, inspecting problematic feedback from patients using classical test theory is rather difficult.

Strength of This Study

There are 2 major forms of standardized assessments in clinical settings [31]: (1) a lengthy questionnaire to achieve a precise assessment that requires significant amounts of time and training to administer, and (2) a rapid short-form scale that briefly screens for the most common symptoms using cut-off points to identify degrees of impairment [32,33]. CAT has the advantages of both forms: precision and efficiency. This paper reports several achievements, including using the Rasch rating scale model [34] (instead of dichotomy) to design CAT in a perception survey, applying CAT on a COW, and offering an Excel module with an animation prototype (demonstrated in [Multimedia Appendix 2](#) or <http://www.healthup.org.tw/cat.asp>). This module and available files can complement the limited uses for interactive voice response or Internet-like visualization online [7] to complete questionnaires and put them into clinical practice.

We conducted an actual CAT-based survey instead of CAT with simulations. This study demonstrates the whole procedure of a CAT-based survey, from the beginning of data collection ([Figure 2](#) and [Multimedia Appendix 3](#)) through the end of the evaluation report ([Table 4](#)), and fulfills the goal of creating a Web-CAT with graphs and animations, as stated in our previous paper [35].

Limitations of the Study

Several issues should be considered more thoroughly in further studies. First, a total of 200 patients were surveyed on the IPQ-18. The generalizability of this study needs to be investigated with more studies on different samples and different

instruments. Second, there is a potential sampling bias in this study. Those who completed the IPQ-18 CAT on COW tended to be younger; and proxies were used to represent patients to complete the discharge procedure from hospital, because they were selected randomly by the digit code of their invoice number on the patient's discharge. The proportion of proxies, who are assumed to be healthier and more capable of filling out a questionnaire, was very high (183/200, 91.5%; see [Table 2](#)). This sample therefore does not reflect mostly the patients' perspective on hospitalization, which possibly affects the study results shown in [Table 4](#). Third, the patient burden was determined by the number of questions administered in this study. Other indices may be collected where available, such as time and effort required for test administration, and accessibility of the hospital [[33](#), [34](#)].

In addition, we set at least 10 items in CAT to be completed as one of the stop rules, which might inflate the test length to some extent. As a result, the test length of CAT was about 58% that of NAT, a little higher than in previous studies with about half the test length [[1](#),[9-11](#)].

Applications

A large variety of behavior-change techniques and other methods to promote exposure to interventions have been used [[36](#)]. There are concerns about how to entice patients (or proxies) to complete surveys before they are discharged from the hospital. Offering reward points or coupons good for credits toward another service is recommended because perception surveys are not similar to other clinical scales conducted by clinicians, where patients themselves consider the benefits to their health.

A telephone survey with CAT-based administration or patient self-report on the Internet (demonstrated at

<http://www.healthup.org.tw/cat.asp>) can be combined with the CAT on COW for gathering feedback from patients easily, quickly, and efficiently.

There are many issues that should be addressed in the future, including studies that address the limitations noted above. For example, using CAT on COW at patients' bedsides to gather their feedback before discharge from the hospital can solve the problem of sampling bias (eg, when proxies constitute a high proportion of respondents) and warrants further study. Surveying perceptions of hospital service via the Internet by CAT-type telephone or self-report is encouraged to complement CAT on COW and questionnaires delivered by mail to discharged patients, such as the Picker Institute Europe's annual survey.

One of the important advantages of CAT scoring is that the item pool can be expanded without changing the metric [[37](#)]. CAT administrators may expand the IPQ-18 item pool or replace items with other kinds of questions as presented in the Excel spreadsheet example. It must be noted that (1) overall item and step (threshold) difficulties of the questionnaire must be calibrated in advance using Rasch analysis (eg, the IPQ-18 of this study was examined by Rasch analysis in a previous paper [[5](#)]), and (2) picture and voice files for each question should be well prepared in an appropriate folder that can be shown simultaneously with the corresponding question in an animation module of CAT.

Conclusion

CAT-based administration of surveys of patient perception reduces patient burden without compromising measurement precision. The Excel module for animation-CAT on COW connected to a mainframe computer is recommended for assessing patients' perceptions of their experience in the hospital.

Acknowledgments

This study was supported by Grant 98cm-kmu-18 from the Chi Mei Medical Center, Taiwan.

Conflicts of Interest

None declared

Authors' Contributions

Chien,Lai and Chou collected all data, generated the database, designed and performed the statistical analysis and wrote the manuscript. Wang and Huang contributed to the development of the study design and advised on the performance of the statistical analysis. The analysis and results were discussed by all authors together. Chien contributed to the Excel programming, helped to interpret the results and helped to draft the manuscript. All authors read and approved the final manuscript.

Multimedia Appendix 1

Excel VBA module for CAT delivering results to the website through an Internet address

[\[ZIP file \(Zip Archive\), 1141 KB-Multimedia Appendix 1\]](#)

Multimedia Appendix 2

Comprehensive overview of Rasch models and the CAT process

[\[PDF file \(Adobe PDF File\), 405 KB-Multimedia Appendix 2\]](#)

Multimedia Appendix 3

Screenshot of the module with an animation-CAT design

[\[WMV file \(Windows Media Video File\), 16.264 KB-Multimedia Appendix 3\]](#)

References

1. Chien TW, Wu HM, Wang WC, Castillo RV, Chou W. Reduction in patient burdens with graphical computerized adaptive testing on the ADL scale: tool development and simulation. *Health Qual Life Outcomes* 2009;7:39 [FREE Full text] [doi: [10.1186/1477-7525-7-39](https://doi.org/10.1186/1477-7525-7-39)] [Medline: [19416521](https://pubmed.ncbi.nlm.nih.gov/19416521/)]
2. Briggs B. Point of care on a roll. *Health Data Manag* 2003 Dec;11(12):48-50. [Medline: [14682255](https://pubmed.ncbi.nlm.nih.gov/14682255/)]
3. Lavin M, Sierzega G, Pucklavage D, Kleinbach D, Gogal C, Bokovoy J. Carts and care. Roll out safer medication delivery and smoother workflow with mobile technology. *Nurs Manage* 2007 Nov;Suppl Pharmacy:16-18. [doi: [10.1097/01.NUMA.0000298279.92200.8c](https://doi.org/10.1097/01.NUMA.0000298279.92200.8c)] [Medline: [18176103](https://pubmed.ncbi.nlm.nih.gov/18176103/)]
4. Davies AR, Ware JE. Involving consumers in quality of care assessment. *Health Aff (Millwood)* 1988;7(1):33-48 [FREE Full text] [Medline: [3360392](https://pubmed.ncbi.nlm.nih.gov/3360392/)]
5. Chien TW, Wang WC, Wang HY, Lin HJ. Online assessment of patients' views on hospital performances using Rasch model's KIDMAP diagram. *BMC Health Serv Res* 2009;9:135 [FREE Full text] [doi: [10.1186/1472-6963-9-135](https://doi.org/10.1186/1472-6963-9-135)] [Medline: [19646267](https://pubmed.ncbi.nlm.nih.gov/19646267/)]
6. Chien TW, Wang WC, Lin SB, Lin CY, Guo HR, Su SB. KIDMAP, a web based system for gathering patients' feedback on their doctors. *BMC Med Res Methodol* 2009;9:38 [FREE Full text] [doi: [10.1186/1471-2288-9-38](https://doi.org/10.1186/1471-2288-9-38)] [Medline: [19534773](https://pubmed.ncbi.nlm.nih.gov/19534773/)]
7. Ritter P, Lorig K, Laurent D, Matthews K. Internet versus mailed questionnaires: a randomized comparison. *J Med Internet Res* 2004 Sep 15;6(3):e29 [FREE Full text] [doi: [10.2196/jmir.6.3.e29](https://doi.org/10.2196/jmir.6.3.e29)] [Medline: [15471755](https://pubmed.ncbi.nlm.nih.gov/15471755/)]
8. Lord FM. Applications of item response theory to practical testing problems. Hillsdale, NJ: Lawrence Erlbaum; 1980.
9. Wainer HW, Dorans NJ, Flaugh R, Green BF, Mislevy RJ, Steinberg L, et al. Computerized adaptive testing: a primer. Hillsdale, N.J.: L. Erlbaum Associates; 1990.
10. Embretson S, Reise S. Item response theory for psychologists. Mahwah, N.J.: L. Erlbaum Associates; 2000:158-186.
11. Weiss DJ. Improvement measurement quality and efficiency with adaptive testing. *Applied Psychological Measurement* 1982;6:473-492. [doi: [10.1177/014662168200600408](https://doi.org/10.1177/014662168200600408)]
12. Crocker L, Algina J. Introduction to classical and modern test theory. New York: Holt, Rinehart, and Winston; 1986.
13. Nunnally JC, Bernstein JH. Psychometric theory. New York: McGraw-Hill; 1994.
14. Wright BD, Linacre JM. Observations are always ordinal; measurements, however, must be interval. *Arch Phys Med Rehabil* 1989 Nov;70(12):857-860. [Medline: [2818162](https://pubmed.ncbi.nlm.nih.gov/2818162/)]
15. Rasch G. Probabilistic models for some intelligence and attainment tests. Chicago: University of Chicago Press; 1980.
16. Holland PW, Wainer H. Differential item functioning. Hillsdale: Lawrence Erlbaum Associates; 1993.
17. Smith AB, Wright P, Selby PJ, Velikova G. A Rasch and factor analysis of the Functional Assessment of Cancer Therapy-General (FACT-G). *Health Qual Life Outcomes* 2007;5:19 [FREE Full text] [doi: [10.1186/1477-7525-5-19](https://doi.org/10.1186/1477-7525-5-19)] [Medline: [17448239](https://pubmed.ncbi.nlm.nih.gov/17448239/)]
18. Velozo CA, Wang Y, Lehman L, Wang JH. Utilizing Rasch measurement models to develop a computer adaptive self-report of walking, climbing, and running. *Disabil Rehabil* 2008;30(6):458-467. [doi: [10.1080/09638280701617317](https://doi.org/10.1080/09638280701617317)] [Medline: [18297500](https://pubmed.ncbi.nlm.nih.gov/18297500/)]
19. Lehman LA, Woodbury M, Shechtman O, Wang YC, Pomeranz J, Gray DB, et al. Development of an item bank for a computerised adaptive test of upper-extremity function. *Disabil Rehabil* 2011 Mar 14. [doi: [10.3109/09638288.2011.560336](https://doi.org/10.3109/09638288.2011.560336)] [Medline: [21401332](https://pubmed.ncbi.nlm.nih.gov/21401332/)]
20. Öztuna D, Elhan AH, Küçükdeveci AA, Kutlay S, Tennant A. An application of computerised adaptive testing for measuring health status in patients with knee osteoarthritis. *Disabil Rehabil* 2010;32(23):1928-1938. [doi: [10.3109/09638281003777572](https://doi.org/10.3109/09638281003777572)] [Medline: [20384449](https://pubmed.ncbi.nlm.nih.gov/20384449/)]
21. Elhan AH, Oztuna D, Kutlay S, Küçükdeveci AA, Tennant A. An initial application of computerized adaptive testing (CAT) for measuring disability in patients with low back pain. *BMC Musculoskelet Disord* 2008;9:166 [FREE Full text] [doi: [10.1186/1471-2474-9-166](https://doi.org/10.1186/1471-2474-9-166)] [Medline: [19094219](https://pubmed.ncbi.nlm.nih.gov/19094219/)]
22. Linacre JM, Wright BD. Chi-square fit statistics. *Rasch Meas Trans* 1994;8(2):360.
23. Smith RM. Fit analysis in latent trait measurement models. *J Appl Meas* 2000;1(2):199-218. [Medline: [12029178](https://pubmed.ncbi.nlm.nih.gov/12029178/)]
24. Linacre JM. Optimizing rating scale category effectiveness. *J Appl Meas* 2002;3(1):85-106. [Medline: [11997586](https://pubmed.ncbi.nlm.nih.gov/11997586/)]
25. Kieffer KM, Reese RJ. A reliability generalization study of the ceriatric scale. *Educational and Psychological Measurement* 2002;62(6):969-994. [doi: [10.1177/0013164402238085](https://doi.org/10.1177/0013164402238085)]
26. Harwell M, Stone CA, Hsu TC, Kirisci L. Monte Carlo studies in item response theory. *Applied Psychological Measurement* 1996, 20, 101-125 1996. [doi: [10.1177/014662169602000201](https://doi.org/10.1177/014662169602000201)]
27. Macdonald P, Paunonen SV. A monte carlo comparison of item and person statistics based on item response theory versus classical test theory. *Educational and Psychological Measurement* 2002;62:921-943. [doi: [10.1177/0013164402238082](https://doi.org/10.1177/0013164402238082)]

28. Chien TW, Lin SJ, Wang WC, Leung HW, Lai WP, Chan AL. Reliability of 95% confidence interval revealed by expected quality-of-life scores: an example of nasopharyngeal carcinoma patients after radiotherapy using EORTC QLQ-C 30. *Health Qual Life Outcomes* 2010;8:68 [FREE Full text] [doi: [10.1186/1477-7525-8-68](https://doi.org/10.1186/1477-7525-8-68)] [Medline: [20626903](https://pubmed.ncbi.nlm.nih.gov/20626903/)]
29. Linacre JM. How to simulate Rasch data. *Rasch Meas Trans* 2007;21(3):1125.
30. Ware JE, Kosinski M, Bjorner JB, Bayliss MS, Batenhorst A, Dahlöf CG, et al. Applications of computerized adaptive testing (CAT) to the assessment of headache impact. *Qual Life Res* 2003 Dec;12(8):935-952. [Medline: [14651413](https://pubmed.ncbi.nlm.nih.gov/14651413/)]
31. Eack SM, Singer JB, Greeno CG. Screening for anxiety and depression in community mental health: the beck anxiety and depression inventories. *Community Ment Health J* 2008 Dec;44(6):465-474. [doi: [10.1007/s10597-008-9150-y](https://doi.org/10.1007/s10597-008-9150-y)] [Medline: [18516678](https://pubmed.ncbi.nlm.nih.gov/18516678/)]
32. Ramirez Basco M, Bostic JQ, Davies D, Rush AJ, Witte B, Hendrickse W, et al. Methods to improve diagnostic accuracy in a community mental health setting. *Am J Psychiatry* 2000 Oct;157(10):1599-1605 [FREE Full text] [Medline: [11007713](https://pubmed.ncbi.nlm.nih.gov/11007713/)]
33. Shear MK, Greeno C, Kang J, Ludewig D, Frank E, Swartz HA, et al. Diagnosis of nonpsychotic patients in community clinics. *Am J Psychiatry* 2000 Apr;157(4):581-587 [FREE Full text] [Medline: [10739417](https://pubmed.ncbi.nlm.nih.gov/10739417/)]
34. Andrich D. A rating formulation for ordered response categories. *Psychometrika* 1978;43:561-573.
35. Chien TW, Lai WP, Lu CW, Wang WC, Chen SC, Wang HY, et al. Web-based computer adaptive assessment of individual perceptions of job satisfaction for hospital workplace employees. *BMC Med Res Methodol* 2011;11:47 [FREE Full text] [doi: [10.1186/1471-2288-11-47](https://doi.org/10.1186/1471-2288-11-47)] [Medline: [21496311](https://pubmed.ncbi.nlm.nih.gov/21496311/)]
36. Brouwer W, Kroeze W, Crutzen R, de Nooijer J, de Vries NK, Brug J, et al. Which intervention characteristics are related to more exposure to internet-delivered healthy lifestyle promotion interventions? A systematic review. *J Med Internet Res* 2011;13(1):e2 [FREE Full text] [doi: [10.2196/jmir.1639](https://doi.org/10.2196/jmir.1639)] [Medline: [21212045](https://pubmed.ncbi.nlm.nih.gov/21212045/)]
37. Aday LA. *Designing and conducting health surveys: a comprehensive guide*. San Francisco: Jossey-Bass Publishers; 1996.

Abbreviations

- ANOVA:** analysis of variance
- CAT:** computerized adaptive testing
- COW:** computers on wheels
- IPQ:** inpatient perception questionnaire
- IRT:** item response theory
- MNSQ:** mean square errors
- NAT:** nonadaptive testing
- VBA:** Visual Basic for Applications

Edited by G Eysenbach; submitted 26.02.11; peer-reviewed by T Bond, CL Hsieh, A Smith; comments to author 13.04.11; revised version received 05.05.11; accepted 09.05.11; published 15.08.11

Please cite as:

Chien TW, Wang WC, Huang SY, Lai WP, Chow JC
A Web-Based Computerized Adaptive Testing (CAT) to Assess Patient Perception in Hospitalization
J Med Internet Res 2011;13(3):e61
URL: <http://www.jmir.org/2011/3/e61/>
doi: [10.2196/jmir.1785](https://doi.org/10.2196/jmir.1785)
PMID: [21844001](https://pubmed.ncbi.nlm.nih.gov/21844001/)

©Tsair-Wei Chien, Wen-Chung Wang, Sheng-Yun Huang, Wen-Pin Lai, Julie Chi Chow. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 15.08.2011. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.