

BMJ Open Development and preliminary psychometric properties of the Care Experience Feedback Improvement Tool (CEFIT)

Michelle Beattie,¹ Ashley Shepherd,² William Lauder,³ Iain Atherton,⁴ Julie Cowie,² Douglas J Murphy⁵

To cite: Beattie M, Shepherd A, Lauder W, *et al.* Development and preliminary psychometric properties of the Care Experience Feedback Improvement Tool (CEFIT). *BMJ Open* 2016;**6**:e010101. doi:10.1136/bmjopen-2015-010101

► Prepublication history for this paper is available online. To view these files please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2015-010101>).

Received 24 September 2015
Revised 7 April 2016
Accepted 26 April 2016



CrossMark

For numbered affiliations see end of article.

Correspondence to

Dr Michelle Beattie;
michelle.beattie@stir.ac.uk

ABSTRACT

Objective: To develop a structurally valid and reliable, yet brief measure of patient experience of hospital quality of care, the Care Experience Feedback Improvement Tool (CEFIT). Also, to examine aspects of utility of CEFIT.

Background: Measuring quality improvement at the clinical interface has become a necessary component of healthcare measurement and improvement plans, but the effectiveness of measuring such complexity is dependent on the purpose and utility of the instrument used.

Methods: CEFIT was designed from a theoretical model, derived from the literature and a content validity index (CVI) procedure. A telephone population surveyed 802 eligible participants (healthcare experience within the previous 12 months) to complete CEFIT. Internal consistency reliability was tested using Cronbach's α . Principal component analysis was conducted to examine the factor structure and determine structural validity. Quality criteria were applied to judge aspects of utility.

Results: CVI found a statistically significant proportion of agreement between patient and practitioner experts for CEFIT construction. 802 eligible participants answered the CEFIT questions. Cronbach's α coefficient for internal consistency indicated high reliability (0.78). Interitem (question) total correlations (0.28–0.73) were used to establish the final instrument. Principal component analysis identified one factor accounting for 57.3% variance. Quality critique rated CEFIT as fair for content validity, excellent for structural validity, good for cost, poor for acceptability and good for educational impact.

Conclusions: CEFIT offers a brief yet structurally sound measure of patient experience of quality of care. The brevity of the 5-item instrument arguably offers high utility in practice. Further studies are needed to explore the utility of CEFIT to provide a robust basis for feedback to local clinical teams and drive quality improvement in the provision of care experience for patients. Further development of aspects of utility is also required.

Strengths and limitations of this study

- The psychometric findings demonstrate the structural validity and internal consistency of Care Experience Feedback Improvement Tool (CEFIT) to quantify the patient experience of quality of healthcare.
- While the large sample (n=802) enabled exploration of the CEFIT structure, the findings are limited to an Australian community population, with a healthcare experience, as opposed to inpatients.
- Validity and reliability are not all or nothing approaches. Rather, each study with positive results adds to the psychometric strength of the instrument. Further testing of CEFIT is required to establish the utility of CEFIT to measure patient experience of hospital quality of care for quality improvement at a ward/unit level.

INTRODUCTION

Background

Sustaining and improving hospital quality of care continues to be an international challenge.^{1–3} These challenges include an ageing population, with an associated increase in comorbidities, coupled with increasingly complex care due to advances in technology, pharmacology and clinical specialisation.^{4–6} These demographic and societal changes have resulted in an increase in healthcare resource and expenditure.^{3 7} Hence, there are competing demands for limited healthcare resources. UK policymakers and healthcare organisations have responded to these trends by shifting the balance of care from an over-reliance on hospitals to community-led and home-led services.^{8–11} The aspiration has been, and is, for mutual health services; co-design and co-creation of services with patients, rather than for patients.¹¹ Despite these changes, there remains significant

pressure on hospitals to deliver high quality of care with limited resources.

Measurement is necessary to determine whether or not new interventions or approaches are indeed improving the quality of hospital care.¹² Healthcare providers translate strategic targets into local measurement plans for hospital quality of care. How hospital quality of care is measured matters, as limited resources are often directed towards what is being measured.¹³ Hence, though narratives of the patient experience can provide powerful insights into quality of healthcare, more tangible and measurable aspects of quality, such as waiting times, are used as ongoing indicators of quality, and so attract resources to address them.

Measurement of the patient perspective is important. Investigations into high profile failures of care highlight a disregard for patients' concerns, leading to calls for the patients' voice to be heard.^{14 15} Views on what constitutes quality of care differ between patients, clinicians and managers; the patient can provide an additional (and essential) perspective to professional views.^{16 17} The complexity of hospital care ensures that the only consistent person in the patient's journey is the patient; their perspective is unique. Devising a global measure of hospital quality of care, from the patient perspective, has the potential to use the patients' voice to direct quality improvement efforts within hospital care.

Many tools exist to measure patient satisfaction of hospital quality of care,^{18–22} but theoretical and methodological challenges limit their use as quality measures.²³ For example, there is evidence to show that patients over-rate satisfaction due to gratitude bias and fear of reprisal; therefore, results tend to show a high ceiling effect, which could prevent the measure from differentiating between poor and good quality of care.²⁴ Satisfaction tends to be influenced by patient expectations, which fluctuate over time, again limiting their use in measurement of hospital quality.²⁵ As an alternative, there is evidence to suggest that patient reports of their experiences of healthcare reduce these limitations.^{26 27} Measuring patient experience requires questions to be designed around what and how often care processes or behaviours occurred, as opposed to patient ratings of care. For example, a satisfaction survey may ask patients to rate the care process of medicine administration, whereas a patient experience survey may ask how often they received the right medication at the right time.

Instruments already exist to measure the patient experience, such as Patient Reported Experience Measures (PREMS). However, PREMS are designed to measure the patient experience of care of a specific condition or treatment as opposed to a global measure of hospital quality of care experience.²⁸ There are also instruments measuring aspects of the patient experience of hospital quality of care, such as the Consumer Quality Index²⁹ which measures collaboration between general practitioners and medical specialists; the Patient Evaluation of Emotional Care during hospitalisation³⁰ measuring

relational aspects of care; and the Consultation and Relational Empathy³¹ measure which quantifies the patient perception of clinicians' empathy. There remains a need for a global measure of hospital quality of care.

There is no 'best' tool when identifying an instrument to measure the patient experience of hospital quality of care. Rather, decision-making is largely dependent on the purpose and context in which the data will be used. For example, patient experience data used for national league tables would require the use of a highly reliable and valid instrument, while accepting the associated high cost of resource necessary for administration, whereas patient experience data used for local quality improvement may tolerate lower levels of reliability in favour of cost-effectiveness and user acceptability.¹² Although it is important to consider whether an instrument accurately measures a concept (validity) in a consistent manner (reliability), other factors, such as the brevity of the instrument, may also be important.

We previously conducted a systematic review which found 11 instruments measuring the patient experience of hospital care.¹² Six instruments, with extensive psychometric testing, were designed for the primary purpose of data being used for national comparisons.^{32–37} The other five instruments, which were primarily used and designed for quality improvement purposes, demonstrated some degree of validity and reliability.^{38–42} However, their utility was compromised either by having too many items or by a poor fit to the UK healthcare context. Those which were considered too lengthy included the Quality from the Patients' Perspective with 68 items,³⁸ the Quality from the Patients' Perspective Shortened³⁹ instrument with 24 items and the Patient Experience Questionnaire⁴⁰ consisting of 20 items. Quality improvement measurement involves repeated data collection over time, often displayed on statistical process control charts.⁴³ Data collection may happen on a daily, weekly or monthly basis; therefore, it needs to be brief to be sustained within clinical practice. The Patient Experiences with Inpatient Care (I-PAHC)⁴¹ and Patient Perceptions of Quality (PPQ)⁴² were developed in non-western, low-income healthcare settings, which limits transferability. For example, the instruments included items around medicine availability, which are irrelevant in more affluent countries.

A concise instrument which enables rapid completion is required for improvement purposes within clinical wards or units. Lengthy instruments are a burden to patients, who are likely to be in a period of convalescence at the time of hospital discharge. Similarly, lengthy instruments could divert resources from clinicians providing care to those measuring care, which might have negative effects on the very concept we are trying to improve: quality of care. There are, of course, challenges of designing an instrument to measure patient experience of hospital quality of care for improvement purposes. For example, the brevity of the instrument risks that the tool does not fully capture the

concept of interest (threat to validity). Considering that all aspects of utility will enable a balanced consideration of the important but often competing interests for instrument development and use, Van der Vleuten⁴⁴ suggests that instrument utility comprises five aspects, namely: validity, reliability, cost-efficiency, acceptability and educational impact. Using this wide conception of utility will aid the development of a brief instrument to measure the patient experience of hospital quality of care for local quality improvement which has high practicality. The complexity inherent in achieving this requires a series of investigations, the first of which is reported in this study. We will require to conduct a future generalisability study to determine the number of patient opinions needed to obtain reliable results.

AIM

To develop a structurally valid and internally consistent brief measure of hospital care experience, the Care Experience Feedback Improvement Tool (CEFIT), by:

1. Developing items from the literature and patient experience expertise;
2. Examining the factor structure of CEFIT with those who have had previous care experience;
3. Determining the internal consistency of CEFIT in a population with care experience;
4. Critiquing utility aspects of CEFIT.

INSTRUMENT DEVELOPMENT

Stage 1: theoretical development

The instrument was theoretically informed by a literature review exploring current definitions and domains of healthcare quality. The review found that quality of care was composed of six domains, namely care which is: safe, effective, timely, caring, enables system navigation and is person-centred.¹⁶ The review informed development of a model of healthcare quality, which illustrates that person-centred care is foundational to the other five domains of quality (figure 1). For example, to ensure quality of care, effectiveness needs to be delivered in a person-centred way, that is, adjusting treatment regimes in accordance with individual circumstances and needs. Therefore, person-centred care is not a separate domain, but inherent within the five domains of quality of healthcare.

Stage 2: item construction

An item was constructed for each of the five domains. Items were worded to capture the patient experience of quality of care, as opposed to ratings of satisfaction. Behavioural or observational prompts were devised for each item to aid patient interpretation. For example, 'I received procedures and treatments within acceptable waiting times' is a prompt for the item 'I received timely care'. Prompts were designed to enable easy adaptation to differing contexts. For example, a prompt for timely care might be 'staff responded to my call bell within a reasonable time' for a hospital context or 'I waited an



Figure 1 Beattie's model of healthcare quality.

acceptable amount of time to be seen' in an outpatient setting. A five-point Likert response was devised from 'never' to 'always' to determine the frequency of care behaviours. Any response which does not indicate 'always' suggests there is room for improvement. A global question was designed to rate the patient's overall healthcare quality experience, where 1=the worst quality care and 5=the best quality of care.

Stage 3: content validity

The draft tool was subjected to a content validity index (CVI) to test for a statistically significant proportion of agreement between experts on the instrument construction.⁴⁵ This process requires between 5 and 10 experts to review the instrument and complete a four point rating scale to determine item relevance, where 1=irrelevant and 4=very relevant.⁴⁵ A section is available for reviewers to make suggestions for improvement. Content validity is achieved when items are rated as 3 or 4 by a predetermined proportion of experts.⁴⁵ The research and development manager deemed the CVI procedure to be service evaluation.

CVI was completed by two expert panels simultaneously. One panel consisted of five volunteers who had had a previous hospitalisation for more than 24 hours within the past 6 months. The public volunteer group was derived from a cardiac rehabilitation class. At the end of the class, attendees were asked by their clinician if they would be interested in giving their views on the comprehension and completeness of a draft tool designed to obtain patient feedback. Those remaining at the end of the class reviewed the draft tool and provided feedback. The other panel consisted of five 'experts' in patient experience. Professional experts met the following criteria: current or previous role providing direct patient care, and had either a specific role in patient experience

policy or practice (ie, public involvement officer), or published research in relation to patient experience or quality of care. The patient experience experts completed the CVI while attending an international conference on healthcare quality. Eight out of 10 experts were required to rate the items as 3 or 4 to achieve content validity beyond the 0.05 level of significance.⁴⁵

Both volunteer and professional experts rated all items (five domains and one overall rating question) as content valid (rating responses as 3='relevant but needs minor alterations' or 4='very relevant'), with the exception of one item (2='needing revision'). Nine out of 10 experts rated all item contents valid to measure the patient experience of hospital quality of care. Their

aggregated score gave an overall CVI as 0.90, endorsing validity beyond the 0.05 level of significance.⁴⁵ Suggestions for improvement within the open text comments section included the use of colour to visually separate sections within the questionnaire. The item rated as 'needing revision' was the overall global rating scale as there was no yardstick or parameter to help patients respond appropriately. The expert feedback and literature supported removal of the global rating scale as CEFIT was designed to highlight key areas for improvement and action as opposed to an assessment for judgement.⁴⁶ Minor alterations were made to wording and prompts. The five items, with example prompts and rating response options, are displayed in [figure 2](#).

| The five statements below have been identified as important aspects of care by patients. We would welcome your feedback on how often you experienced these aspects of quality whilst receiving care. Examples are given to help describe each of the statements. You may have different examples or interpretations of the statements, which is OK. | | | | | |
|---|--------------------------|---------------------------|--------------------------|--------------------------|--------------------------|
| Please tick one box like this <input checked="" type="checkbox"/> with a ball point pen. If you change your mind just cross out your old response and make a new choice. Please answer every statement. | | | | | |
| | Never ¹ | Occasionally ² | Sometimes ³ | Often ⁴ | Always ⁵ |
| 1. I received safe care... | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| <i>Examples might include: Staff washed their hands before and after any direct contact with me I received the right medication dose</i> | | | | | |
| | Never ¹ | Occasionally ² | Sometimes ³ | Often ⁴ | Always ⁵ |
| 2. I received timely care... | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| <i>Examples might include: I waited an acceptable amount of time to be seen I received procedures and treatments within a reasonable time</i> | | | | | |
| | Never ¹ | Occasionally ² | Sometimes ³ | Often ⁴ | Always ⁵ |
| 3. My care met my personal needs... | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| <i>Examples might include: My care was tailored to my individual needs i.e. adjustments made for my other conditions and personal lifestyle I was involved in all decisions about my care</i> | | | | | |
| | Never ¹ | Occasionally ² | Sometimes ³ | Often ⁴ | Always ⁵ |
| 4. Staff were caring to me... | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| <i>Examples might include: I was treated pleasantly and with courtesy Staff really listened to me and took my concerns seriously</i> | | | | | |
| | Never ¹ | Occasionally ² | Sometimes ³ | Often ⁴ | Always ⁵ |
| 5. I was able to get the care I needed... | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| <i>Examples might include: I knew how to access the care I needed There were no barriers to me getting the treatment/services that I needed</i> | | | | | |

Figure 2 Care Experience Feedback Improvement Tool (CEFIT).

INSTRUMENT TESTING

Design

CEFIT was administered via a telephone population survey of people dwelling in Queensland from July to August 2014. Embedding CEFIT within a large-scale survey provided an opportunity to test the internal consistency and structural validity of CEFIT with a random sample, to determine how well items were related and to check for any possible duplication or redundancy of final content. The Queensland Social Survey is an annual statewide survey administered by the Population Research Laboratory at Central Queensland University (CQU) Australia to explore a wide range of research questions relevant to the general public. The CQU survey consists of demographic, introductory and specific research questions. We embedded the five-item CEFIT instrument within the survey. Pilot testing of CEFIT in randomly selected households (n=68) suggested that no question changes were necessary.

Sampling

A two-stage sampling procedure was used. First, two geographically proportionate samples were drawn from (1) South-East Queensland and (2) the rest of Queensland. A telephone database was used to draw a random sample of telephone numbers within postcode areas. Second, participants were selected per household on the basis of gender (to ensure an equal male and female sample) and age (≥ 18 years). Where more than one male/female met the criteria in one household, the adult with the most recent birthday was selected. The questions were preceded with a screening question to identify participants with a recent healthcare experience (<12 months). Samples of patient experience surveys are usually estimated on rates of patient discharge and include patients within 5 months of hospital discharge.¹² A wider time frame and context was required for CEFIT due to the nature of a population survey.

Data collection

The sample was loaded into a Computer-Assisted Telephone Interviewing (CATI) System held within the Population Research Laboratory. The system allocates telephone numbers and provides standardised text instructions to trained interviewers.

Analysis

Data were entered into SPSS V.19. Descriptive statistics were calculated for questionnaire characteristics to establish the range of responses. Interitem (question) total correlations were calculated in order to identify the possibility of unnecessary or redundant questions. Internal consistency of the CEFIT was calculated using Cronbach's α to determine the consistency in responses to the questions and examine the error generated by the questions asked.⁴⁷ Exploratory factor analysis was used to examine the factor structure of CEFIT and determine structural validity. The Kaiser-Meyer-Olkin test was conducted to measure the sampling adequacy and compare magnitudes of correlation. Bartlett's Test of Sphericity was used to ensure the study assumptions were met for factor analysis. Factor analysis using the principal component method (principal component analysis) and Eigenvalue >1 rule was performed. Results for structural validity are determined as positive if factor analysis explains at least 50% of the variance.⁴⁸

RESULTS

The overall CQU survey response rate was 35.9%. From the 1223 survey participants, 802 (66%) were eligible to complete CEFIT (healthcare experienced within the preceding 12 months). All eligible participants responded to the CEFIT questions (100% response). Of the 802, 50.5% (n=405) were male and 49.5% (n=397) were female. The mean age was 58 years, range 18–95 years. The range of question responses was towards the high end of the responses, indicating that the majority of quality care processes were occurring 'often' or 'always'. However, all rating responses were used highlighting necessary range (table 1).

Questions demonstrating interitem total correlations of 0.2–0.8 are considered as offering unique and useful content which is related to the overall inventory.⁴⁷ Correlations (0.28–0.74) between questions were all significant ($p \leq 0.05$), and this provided reassurance that all questions were inter-related, yet unique enough to necessitate their inclusion within the inventory (table 2). Cronbach's α for the final questionnaire was good (0.78) and confirmed that the scale as a whole was sufficient. This provided reassurance that CEFIT's overall inventory of questions did not include unnecessary questions

Table 1 Descriptives of Care Experience Feedback Improvement Tool (CEFIT) questions

| Question | Never n (%) | Occasionally n (%) | Sometimes n (%) | Often n (%) | Always n (%) | Missing n (%) |
|-------------------------------------|----------------|-----------------------|--------------------|----------------|-----------------|------------------|
| I received safe care | 10 (1.3) | 5 (0.6) | 9 (1.1) | 42 (5.3) | 724 (91.6) | 12 (1.49) |
| I received timely care | 25 (3.1) | 19 (2.4) | 49 (6.1) | 104 (13.0) | 602 (75.3) | 3 (0.37) |
| My care met my personal needs | 18 (2.3) | 8 (1.0) | 27 (3.4) | 58 (7.3) | 689 (86.1) | 2 (0.24) |
| Staff were caring towards me | 7 (0.9) | 4 (0.5) | 17 (2.1) | 49 (6.1) | 724 (90.4) | 1 (0.12) |
| I was able to get the care I needed | 13 (1.6) | 9 (1.1) | 20 (2.5) | 36 (4.5) | 723 (90.3) | 1 (0.12) |

Table 2 Correlation of Care Experience Feedback Improvement Tool (CEFIT) domains

| Domain | Safe | Timely | Effective | Caring | System navigation |
|-------------------|-------|--------|-----------|--------|-------------------|
| Safe | 1.00 | | | | |
| Timely | 0.300 | 1.00 | | | |
| Effective | 0.411 | 0.383 | 1.00 | | |
| Caring | 0.328 | 0.399 | 0.679 | 1.00 | |
| System navigation | 0.318 | 0.284 | 0.736 | 0.664 | 1.00 |

($\alpha > 0.9$), and neither was the overall inventory too narrow in its scope ($\alpha > 0.7$).⁴⁷

The Kaiser-Meyer-Olkin measure of sampling adequacy was good (0.792) and Bartlett's Test of Sphericity (1546.08, Df=10, $p=0.000$) indicated that the study met the assumptions for factor analysis. Eigenvalues identified a one-factor solution which accounted for 57.33% of total variance (table 3). The 57.33% variance of the one-factor solution was shared by five components, namely: safety 0.579, timely 0.582, effective 0.884, caring 0.845 and system navigation 0.836 (table 4).

CEFIT QUALITY CRITIQUE

Five aspects of utility were critiqued for CEFIT. First, we applied the appropriate Consensus-based Standards for the Selection of Health Measurement Instruments (COSMIN) checklist for content validity, structural validity and internal consistency to assess the quality of the study methods (criteria and results are available in tables 5–7, respectively).⁴⁹ Responses within individual checklists were given a methodological score by applying the COSMIN four-point checklist scoring system, namely: excellent, good, fair or poor. Where individual answers to checklist questions were of variable ratings (ie, some excellent, some poor), the overall score was determined by taking the lowest rating of any item. In other words, the worst score counted.⁴⁹

Second, we applied the criteria developed by Terwee⁴⁸ to determine rating for the quality of the results of each psychometric test performed on CEFIT (see figure 3). This enabled study results to be categorised as positive (+), indeterminate (?) or negative

Table 4 Component matrix for Care Experience Feedback Improvement Tool (CEFIT) single-factor solution

| | Component 1 |
|-------------------|-------------|
| Safe | 0.579 |
| Timely | 0.582 |
| Effective | 0.884 |
| Caring | 0.845 |
| System navigation | 0.836 |

Extraction method: principal component analysis.

*1 means that components were only extracted for factors with an Eigenvalue of > 1 .

(–) according to the quality criteria for each measurement property. For example, positive ratings for internal consistency are given, using Terwee *et al* criteria, if the Cronbach's α is > 0.70 . Studies with Cronbach's α results of < 0.70 would be categorised as negative, or where Cronbach's α was not determined the result would be categorised as indeterminate. A full explanation, with justification for all COSMIN criteria results, is available from Terwee.⁴⁸

Third, we applied criteria developed and tested in our previous systematic review for additional aspects of instrument utility: cost-efficiency, acceptability and educational impact (detailed in figure 4). Further explanation of the criteria and scoring is available at Beattie *et al*.¹² Results from all three steps are presented in an adaptation of the Beattie and Murphy Instrument Utility Matrix for CEFIT (table 8).

The study quality for the content validity of CEFIT was rated as fair as there was no assessment of whether all items were relevant for the study population (eg, gender, disease characteristics, country and setting). The overall rating of the content validation results was positive as the target population considered all items in the questionnaire to be relevant and complete. The quality of the structural validity was rated as excellent as there was an adequate sample size and no major flaws in the study design. Results for structural validity were categorised as positive as the one-factor solution explained more than 50% of the variance (57.3%).⁴⁸

The study quality for internal consistency was rated as excellent using the COSMIN checklist as all questions were answered positively and there were no major flaws identified in the study. The quality of the results for

Table 3 Total variance from factor analysis

| Component | Initial Eigenvalues | | | Extraction sums of squared loadings | | |
|-----------|---------------------|----------------------|---------------------|-------------------------------------|----------------------|---------------------|
| | Total | Per cent of variance | Cumulative per cent | Total | Per cent of variance | Cumulative per cent |
| 1 | 2.867 | 57.336 | 57.336 | 2.867 | 57.336 | 57.336 |
| 2 | 0.834 | 16.670 | 74.006 | | | |
| 3 | 0.710 | 14.194 | 88.200 | | | |
| 4 | 0.339 | 6.786 | 94.986 | | | |
| 5 | 0.251 | 5.014 | 100.000 | | | |

Table 5 Consensus-based Standards for the Selection of Health Measurement Instruments (COSMIN) criteria and Care Experience Feedback Improvement Tool (CEFIT) results for content validity

| COSMIN questions for content validity | Response | Rating |
|--|---------------------|-----------|
| 1. Was there an assessment of whether all items refer to relevant aspects of the construct to be measured? | Yes | Excellent |
| 2. Was there an assessment of whether all items are relevant for the study population? (eg, age, gender, disease characteristics, country, setting) | No | Fair |
| 3. Was there an assessment of whether all items are relevant for the purpose of the measurement instrument? (discriminative, evaluative and/or predictive) | Partial | Fair |
| 4. Was there an assessment of whether all items together comprehensively reflect the construct to be measured? | Yes | Excellent |
| 5. Were there any important flaws in the design or methods of the study? | No | Fair |
| 6. Was there evidence that items were theoretically informed? | Yes | Excellent |
| Total rating | Lowest score counts | Fair |

Table 6 COSMIN criteria and CEFIT results for structural validity

| COSMIN questions for structural validity | Response | Rating |
|---|-----------------------|-----------|
| 1. Does the scale consist of effect indicators, that is, is it based on a reflective model? | Yes | Excellent |
| 2. Was the percentage of missing items given? | Yes | Excellent |
| 3. Was there a description of how missing items were handled? | Not necessary as none | Excellent |
| 4. Was the sample size included in the analysis adequate? | Yes | Excellent |
| 5. Were there any important flaws in the design or methods of the study? | No | Excellent |
| 6. For CTT: Was exploratory or confirmatory factor analysis performed? | Yes | Excellent |
| 7. For IRT: Were IRT tests for determining the (uni) dimensionality of the items performed? | NA | NA |
| Total rating | Lowest score counts | Excellent |

CEFIT, Care Experience Feedback Improvement Tool; COSMIN, Consensus-based Standards for the Selection of Health Measurement Instruments; CTT, Classical Test Theory; NA, not available.

Table 7 COSMIN criteria and CEFIT results for internal consistency

| COSMIN questions for internal consistency | Response | Rating |
|---|-----------------------|-----------|
| 1. Does the scale consist of effect indicators, that is, is it based on a reflective model? | Yes | Excellent |
| 2. Was the percentage of missing items given? | Yes | Excellent |
| 3. Was there a description of how missing items were handled? | Not necessary as none | Excellent |
| 4. Was the sample size included in the internal consistency analysis adequate? | Yes | Excellent |
| 5. Was the unidimensionality of the scale checked? That is, was factor analysis or IRT model applied? | Yes | Excellent |
| 6. Was the sample size included in the unidimensionality analysis adequate? | Yes | Excellent |
| 7. Was an internal consistency statistic calculated for each (unidimensional) (sub)scale separately? | NA | NA |
| 8. Were there any important flaws in the design or methods of the study? | No | Excellent |
| 9. For CTT: Was Cronbach's α calculated? | Yes | Excellent |
| 10. For dichotomous scores: Was Cronbach's α or KR-20 calculated? | NA | NA |
| 11. For IRT: Was a goodness of fit statistic at a global level calculated? For example, χ^2 , reliability coefficient of estimated latent trait value (index of subject or item) | NA | NA |
| Total rating | Lowest score counts | Good |

CEFIT, Care Experience Feedback Improvement Tool; COSMIN, Consensus-based Standards for the Selection of Health Measurement Instruments; CTT, Classical Test Theory; IRT, Item Response Theory; KR-20, Kuder-Richardson Formula 20; NA, not available.

internal consistency was rated as positive as the quality criteria for measurement properties suggest that positive ratings are applied when Cronbach's α is ≥ 0.70 (Cronbach's α for CEFIT was 0.78).⁴⁸

Applying the additional aspects of utility scoring criteria found that CEFIT was rated as 'good' for cost-efficiency. While the majority of responses were good to excellent, it is not yet known how many CEFIT

Figure 3 Quality criteria for measurement properties (Terwee).⁴⁸

| Property | Rating [†] | Quality Criteria |
|---|---------------------|--|
| Reliability | | |
| Internal consistency | + | Cronbach's alpha(s) ≥ 0.70 |
| | ? | Cronbach's alpha not determined or dimensionality unknown |
| | - | Cronbach's alpha(s) < 0.70 |
| Reliability | + | ICC / weighted Kappa ≥ 0.70 OR Pearson's $r \geq 0.80$ |
| | ? | Neither ICC / weighted Kappa, nor Pearson's r determined |
| | - | ICC / weighted Kappa < 0.70 OR Pearson's $r < 0.80$ |
| Measurement error | + | MIC $>$ SDC OR MIC outside the LOA |
| | ? | MIC not defined |
| | - | MIC \leq SDC OR MIC equals or inside LOA |
| Validity | | |
| Content validity | + | All items are considered to be relevant for the construct to be measured, for the target population, and for the purpose of the measurement AND the questionnaire is considered to be comprehensive |
| | ? | Not enough information available |
| | - | Not all items are considered to be relevant for the construct to be measured, for the target population, and for the purpose of the measurement OR the questionnaire is considered not to be comprehensive |
| Construct validity - Structural validity | + | Factors should explain at least 50% of the variance |
| | ? | Explained variance not mentioned |
| | - | Factors explain $<$ 50% of the variance |
| - Hypothesis testing | + | Correlations with instruments measuring the same construct ≥ 0.50 OR at least 75% of the results are in accordance with the hypotheses AND correlations with related constructs are higher than with unrelated constructs |
| | ? | Solely correlations determined with unrelated constructs |
| | - | Correlations with instruments measuring the same construct < 0.50 OR $<$ 75% of the results are in accordance with the hypotheses OR correlations with related constructs are lower than with unrelated constructs |
| - Cross-cultural validity | + | No differences in factor structure OR no important DIF between language versions |
| | ? | Multiple group factor analysis not applied AND DIF not assessed |
| | - | Differences in factor structure OR important DIF between language versions |
| Criterion validity | + | Convincing arguments that gold standard is "gold" AND correlation with gold standard ≥ 0.70 |
| | ? | No convincing arguments that gold standard is "gold" |
| | - | Correlation with gold standard < 0.70 |
| Responsiveness | | |
| Responsiveness | + | Correlation with changes on instruments measuring the same construct ≥ 0.50 OR at least 75% of the results are in accordance with the hypotheses OR AUC ≥ 0.70 AND correlations with changes in related constructs are higher than with unrelated constructs |
| | ? | Solely correlations determined with unrelated constructs |
| | - | Correlations with changes on instruments measuring the same construct < 0.50 OR $<$ 75% of the results are in accordance with the hypotheses OR AUC < 0.70 OR correlations with changes in related constructs are lower than with unrelated constructs |

MIC = minimal important change, SDC = smallest detectable change, LoA = limits of agreement, ICC = intraclass correlation coefficient, DIF = differential item functioning, AUC = area under the curve

[†] + = positive rating, ? = indeterminate rating, - = negative rating

questionnaires will be necessary to ensure reliable data to differentiate between those experiencing different care experiences, which resulted in an overall score of good as opposed to excellent for cost-efficiency. Acceptability of CEFIT was scored as 'poor' as it has not yet been tested in the context in which it is intended (hospital). Educational impact was scored as 'good' as CEFIT was determined as easy to calculate the score and use results, but there is not yet evidence of it being used for quality improvement purposes.

DISCUSSION

Our psychometric findings support CEFIT as a structurally valid instrument with positive internal consistency to measure patient care experience within an Australian

community population. The uniqueness of CEFIT is that it provides a brief yet structurally valid and reliable tool. The brevity of the instrument indicates its potential use as a quality improvement measure of patient experience. Quality improvement advocates ongoing measurement over time, as opposed to snapshot audits or before and after measures.⁴³ The sustainable use of CEFIT within busy hospital wards depends on its simplicity and brevity. However, there are ongoing challenges to ensure that such a brief measure captures the complexity of the patient experience of hospital quality of care (validity), that a small number of items within an instrument sufficiently captures the concept of interest. This study found that the CEFIT structure is measuring the patient experience of quality of care sufficiently (validity) and is doing so in a consistent manner (internal consistency

A

| Questions for Cost Efficiency | Excellent (****) | Good (***) | Fair (**) | Poor (*) |
|--|--|---|--|--|
| 1. What are the number of assessments (observations, questionnaires) needed to ensure reliable data? | Only a small sample needed (<30) | A moderate sample size (30-49) | Not explicit but can be assumed or (50-99 assessments needed) | No details given or (>100 assessments needed) |
| 2. How long does an assessment take to complete | ≤15 mins | ≤30mins | 30-60min | >60mins |
| 3. What are the administrative costs of completing the assessment? | Easily embedded within existing resource. Little additional support required. | Some administrative resource but no specialist resource required. | Large amount of resource required to assess and administer | Significant specialist expertise and administrative time required to assess and administer |
| 4. What is the cost to complete a reliable sample? | Minimal | Moderate | Considerable | Extensive |
| Questions for Acceptability | Excellent (****) | Good (***) | Fair (**) | Poor (*) |
| 1. Is there evidence of subjects understanding of the instrument/assessment? | Investigations of subjects understanding (i.e. cognitive testing of instruments) | Estimated evidence of subjects understanding (i.e. high number of questions missed) | Subject understanding not explicitly stated but some can be assumed (i.e. student guide to OSCE) | No evidence of subject understanding |

B

| 2. How many assessments are not completed? | There are low numbers of missing items (<10%) and adequate response rates (≥40%) | There are a high number of missing items (≥10%) and an adequate response rates (≥40%) | There are low numbers of missing items or poor (<10%) and an inadequate response rate (<40%) | There are high numbers of missing items (≥10%) and poor response rates (<40%) |
|--|--|---|--|--|
| 3. Has the instrument/assessment been tested in an appropriate context? | Evidence of successful administration/use within an appropriate setting | Tested in vivo and changes recommended would be achievable | Testing in vivo and changes recommended would be difficult or only partially tested in vivo | Testing has only been conducted in vitro/simulation |
| Questions for Educational Impact | Excellent (****) | Good (***) | Fair (**) | Poor (*) |
| 1. There is evidence of the instruments intended purpose being achieved (i.e. if aim is to enable hospital ranking for patient selection, is there evidence that the results are actually influencing patient choice?) | Clear evidence of intended purpose being fulfilled. | Explanatory or theoretical link between intended and actual use, but no clear evidence. | Evidence of theoretical work but relationship between intended and actual purpose poorly or not described. | No evidence of intended purpose becoming actual. |
| 2. The scoring system is easily translated or available in an easy to use format? | Explicitly stated and easy to calculate | Explicitly stated but not easy to calculate | Scoring only calculated by resource with statistical knowledge | Scoring not explained well enough to calculate. |
| 3. The feedback from the results can be readily used for action where necessary? | Feedback is readily available in a format that enables necessary action. | Feedback is readily available but not drilled down enough to enable targeted action. | Minimal feedback available or delay results in limited impact. | No explanation to determine adequacy of feedback. No direct feedback could be readily used without additional expertise. |

Figure 4 Additional aspects of utility scoring criteria. OSCE, Objective Structured Clinical Examination.

reliability). Testing of internal consistency reliability and interitem correlations confirmed the need for five components, which were inter-related yet unique enough to require their own component. A Cronbach's α of 0.78 provides reassurance of the reliability of the instrument and confirms that each of the five items is valuable. In addition, factor analysis revealed CEFIT as measuring a single factor, which we have called healthcare quality. We have therefore taken the first steps to ensure that CEFIT is structurally sound. This is an initial but essential step in instrument development.⁵⁰ Of course, validity and reliability are not all or nothing approaches. Rather, each study, with positive results, adds to the psychometric strength of the instrument.

Many instruments are criticised for not being derived from a theoretical model, which is an essential step in instrument development and subsequent validation.²⁷ CEFIT was derived from our theoretically informed model with factor analysis identifying one factor (quality of care), composed of five domains: safety, effectiveness, caring, system navigation and timeliness. To ensure that

patients experience high quality of care, these domains need to be expressed in a person-centred way. Hence, person-centred care is foundational and necessary for all other domains of healthcare quality. Our one-factor solution enables a brief yet structurally valid measure.

While there are patient experience instruments to measure national-level performance, the data are not timely, nor specific enough, to direct or measure local improvement efforts at the clinical ward level.³⁵ For example, when data are used for national comparisons, there are robust and lengthy procedures to ensure a reliable sample. While such criteria are necessary, the delay between sampling and analysis often creates a significant delay (up to 1 year) between data collection and results.¹² Given that much change can occur over that time within a hospital, such data would be of limited use for improvement purposes at a ward level. This is not to suggest that brief instruments are superior to lengthy evaluations, but rather that they serve different purposes.

Further, if hospital-level or national-level data identified a deterioration in patient experience around

Table 8 CEFIT results of Beattie and Murphy Instrument Utility Matrix

| Aspect of utility | Measurement property/criteria checklist | Result | Rating of study quality | Rating of quality of results | Combined score of methods and results |
|--------------------|--|---|-------------------------|------------------------------|---------------------------------------|
| Validity | Content validity (COSMIN checklist ⁴⁹ and Terwee) ⁴⁸ | Items derived from integrative review of the literature and associated theoretical model. Content validity index with a patient and expert group found items were relevant and comprehensive. | Fair | Positive | *** |
| | Structural validity (COSMIN checklist ⁴⁹ and Terwee) ⁴⁸ | Principal component analysis confirmed a unidimensional scale (one-factor solution) accounting for 57.3% variance. | Excellent | Positive | ***** |
| Reliability | Internal consistency (COSMIN checklist ⁴⁹ and Terwee) ⁴⁸ | Cronbach's α 0.78 | Excellent | Positive | ***** |
| Cost | Additional aspects of utility scoring criteria in figure 4 | Number of CEFIT questionnaires needed to ensure reliable data is not yet known. Completion of CEFIT <15 min. Some administrative resource but no specialist resource required. Overall cost-efficiency calculated as moderate | NA | Good | *** |
| Acceptability | Additional aspects of utility scoring criteria in figure 4 | Investigations of participants' understanding. There are low numbers of missing items (<10%) and adequate response rates (>40%). Testing has not yet been conducted in context (ie, hospital setting). | NA | Poor | * |
| Educational impact | Additional aspects of utility scoring criteria in figure 4 | Explanatory or theoretical link between intended and actual use, but no clear evidence yet. Scoring stated and easy to calculate. Feedback is readily available in a format that enables necessary action. | NA | Good | *** |

Ratings of study quality *poor, **fair, ***good, ****excellent ratings of quality of results: (+) positive rating, (–) negative rating, (?) indeterminate rating, mixed (+?).

CEFIT, Care Experience Feedback Improvement Tool; COSMIN, Consensus-based Standards for the Selection of Health Measurement Instruments.

privacy and dignity, for example, there is no way of knowing from which ward or unit within the hospital this problem originates. Similarly, episodes of positive patient experiences cannot be linked to specific wards or teams, thus limiting the ability to spread good practice. Since the intention of CEFIT is to gather data per ward, results will be directly applicable to that area. CEFIT will most likely be a useful indicator of areas of problem identification or to demonstrate improvement in key aspects of quality of care from the patient experience. It is also likely that the addition of an open question to CEFIT will help direct improvement activity given the known strengths of patient narratives to motivate clinical teams to improve.⁵¹ The ability of CEFIT to drive improvement remains untested and is a matter for future work.

As well as the ongoing threat to validity, brief measures also present psychometric challenges. For example, the five-point rating response options assume that there are equidistant differences between each rating and that each of the five questions has equal importance in the patient experience of hospital quality of care.⁵² These issues will be investigated further using statistical modelling in future studies. However, this needs to be balanced with brevity and simplicity. Use of the instrument by frontline staff should not require sophisticated statistics.

Application of the utility matrix and clarity of the instrument's primary purpose will continue to help inform the development of CEFIT. As evidence of instrument utility develops over time, it is important not to dismiss newer instruments with poorer scores.

For example, the content validity of CEFIT was rated as fair, as there is not yet evidence of the instrument being tested with hospitalised patients. We are also aware that the criteria for additional aspects of utility will need further development, but they offer a useful starting point to consider these other important aspects of instrument utility. Conducting a utility critique will help ensure the usefulness of CEFIT for frontline care.

A limitation of the study is that CEFIT was embedded within a telephone survey and tested with people who had had a healthcare experience within the preceding 12 months. While this enabled exploration of the CEFIT structure with a large sample, the findings are limited to an Australian community population, with a healthcare experience, as opposed to inpatients. However, the promising psychometric results suggest that it would be worthwhile to test CEFIT with recently hospitalised patients to determine validity in this population, as well as identifying the numbers needed to differentiate between good and not so good patient experience of hospital quality of care.

We recognise the potential limitation of range due to CEFIT responses being mostly towards the high end of response options. There is a risk that data grouped towards the high end of option responses limit the ability to distinguish between 'good' and 'poor' care. However, although CEFIT responses were mostly towards the high end of the scale, all responses were used, indicating the potential range necessary to differentiate between varying care experiences. Other instruments with high ceiling effects have been able to differentiate between aspects of good and not so good care.³¹ There also remains debate as to the accuracy of using Cronbach's α and factor analysis with non-normally distributed data, although large samples can reduce the effect.^{53 54} We will remain vigilant of the potential threat, but will not know whether CEFIT can differentiate between different experiences of quality of care until we conduct our future generalisability study.

CONCLUSIONS

Measuring the quality of hospital care from the patient perspective is a vital component of healthcare measurement and improvement plans, but the effectiveness of the data collected is dependent on a balanced consideration of all aspects of instrument utility. This study has established a structurally valid and internally consistent measure of the patient experience of hospital quality of care, namely the CEFIT. The next steps are to validate within an inpatient context and establish its reliability to discriminate between different patient experiences at ward/unit level to direct improvement efforts in clinical practice.

Author affiliations

¹School of Health Sciences, University of Stirling, Centre for Health Science, Inverness, UK

²School of Health Sciences, University of Stirling, Stirling, UK

³College of Nursing, University of South Florida, South Florida, Florida, USA

⁴School of Nursing, Midwifery and Social Care, Edinburgh Napier University, Edinburgh, UK

⁵Quality, Safety and Informatics Research Group, University of Dundee, Dundee, UK

Twitter Follow Michelle Beattie at @BeattieQi

Acknowledgements The authors thank their colleagues at Population Research Laboratory, Central Queensland University, Australia for embedding the CEFIT questions within their Annual Social Survey.

Contributors MB and DJM conceived and designed the CEFIT instrument. MB designed the theoretical model of healthcare quality. WL and IA contributed to the thinking and development of the work in their role as MB supervisors. MB and DJM designed the study. AS facilitated acquisition of data via the Queensland survey. MB and DJM designed and collected data for the CVI. WL and IA conducted statistical analysis and interpretation. JC helped in result interpretation and statistical revision. MB drafted the manuscript which was critically revised by all authors before agreeing on the final version of the manuscript.

Funding The University of Stirling provided funding to embed the CEFIT questions within the Queensland Annual Social Survey.

Competing interests None declared.

Ethics approval Human Ethics Review Panel Central Queensland University (Reference H13/06-120).

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement The CEFIT measure is available free of charge for NHS and University staff for the purpose of research or improvement. Anyone wishing to use the measure should contact and register their request with MB (michelle.beattie28@gmail.com) and/or DJM (douglasmurphy27@gmail.com).

Open Access This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>

REFERENCES

1. Institute of Medicine. *Crossing the quality chasm: a new health system for 21st century*. Washington DC: National Academy Press, 2001.
2. Department of Health. High Quality Care for All. Gateway Ref, 2008:10106.
3. Scottish Government. *Achieving sustainable quality in Scotland's healthcare: a '20:20' Vision; 2012*. <http://www.gov.scot/Topics/Health/Policy/2020-Vision/Strategic-Narrative> (accessed 19 Jun 2015).
4. World Health Organisation. *Innovative care for chronic conditions. Building blocks for action*. Geneva: WHO, 2002.
5. Department of Health. *Equity and excellence: liberating the NHS*. London: Stationary Office, 2010.
6. Ham C, Dixon A, Brooke B. *Transforming the delivery of health and social care: the case for fundamental change*. London: The Kings Fund, 2012.
7. World Health Organisation. *World Health Day 2012—ageing and health-toolkit for event organizers*. WHO, 2012. http://apps.who.int/iris/bitstream/10665/70840/1/WHO_DCO_WHD_2012.1_eng.pdf (accessed 26 Aug 2015).
8. Department of Health. *Our health, our care, our say: a new direction for community services*. Her Majesty Government, 2006.
9. Scottish Executive. *Better health, better care: discussion document*. Edinburgh: Scottish Executive, 2007.
10. Scottish Executive. *Building a health service fit for the future*. Edinburgh: Scottish Executive, 2005.
11. Scottish Government. *The healthcare quality strategy for NHS Scotland*. Scottish Government, 2010.
12. Beattie M, Murphy D, Atherton I, et al. Instruments to measure patient experience of health care quality in hospitals: a systematic review. *Syst Rev* 2015;4:97.

13. Berry N, Gardner T, Anderson I. *On targets: how targets can be most effective in the English NHS*. London: Health Foundation, 2015.
14. Francis R. *Report of the Mid Staffordshire NHS Foundation Trust Public Inquiry*. London: Stationary Office, 2013.
15. MacLean L. *The Vale of Leven Hospital Inquiry Report*. Scotland: The Vale of Leven Hospital Inquiry Team APS Group, 2014.
16. Beattie M, Shepherd A, Howieson B, et al. Do the Institute of Medicines' (IOM) dimensions of quality capture the current meaning of quality in health care?—An integrative review. *J Res Nurs* 2012;18:288–304.
17. Donabedian A. *Explorations in quality assessment and monitoring, vol. 1, the definition of quality and approaches to its assessment 1980*. Ann Arbor, MI: Health Admin Press.
18. Rubin H, Ware J, Nelson E, et al. The Patient Judgements of Hospital Quality (PJHQ) questionnaire. *Med Care* 1990;28:S17–18.
19. Parasuraman A, Berry L, Zeithaml V. Refinement and reassessment of the SERVQUAL Scale. *J Retail* 1991;67:420–50.
20. Mountzoglu A, Dafogianni C, Karra V, et al. Development and application of a questionnaire for assessing parent satisfaction with care. *Int J Qual Health Care* 2000;12:331–7.
21. Hendriks A, Oort F J, Vrieliink M R, et al. Reliability and validity of the Satisfaction with Hospital Care Questionnaire. *Int J Qual Health Care* 2002;14:471–82.
22. Smith LF. Postnatal care: development of a psychometric multidimensional satisfaction questionnaire (the WOMBPNSQ) to assess women's views. *Br J Gen Pract* 2011;61:628–37.
23. Williams B. Patient satisfaction: a valid concept? *Soc Sci Med* 1994;38:509–16.
24. Moret L, Nguyen J, Pillet N, et al. Improvement of psychometric properties of a scale measuring inpatient satisfaction with care: a better response rate and a reduction of the ceiling effect. *BMC Health Serv Res* 2007;7:197.
25. Williams B, Coyle J, Healy D. The meaning of patient satisfaction: an explanation of high reported levels. *Soc Sci Med* 1998;47:1351–9.
26. Luxford K. What does the patient know about quality? *Int J Qual Health Care* 2012; 24:439–40.
27. Foundation H. *Measuring patient experience: no. 18, evidence scan*. England: Health Foundation, 2013.
28. Black N, Varaganum M, Hutchings A. Relationship between patient reported experience (PREMs) and patient reported outcomes (PROMs) in elective surgery. *BMJ Qual Saf* 2014;23:534–42.
29. Ikkersheim DE, Koolman X. Dutch healthcare reform: did it result in better patient experiences in hospitals? A comparison of the consumer quality index over time. *BMC Health Serv Res* 2012;12:76.
30. Murrells T, Robert G, Adams M, et al. Measuring relational aspects if hospital care in England with the Patient Evaluation of Emotional Care during Hospitalisation (PEECH) survey questionnaire. *BMJ Open* 2013;3:e002211.
31. Mercer S, Murphy D. Validity and reliability of the CARE measure in secondary care. *Clin Govern Int J* 2008;13:261–83.
32. O'Malley AJ, Zaslavsky AM, Hays RD, et al. Exploratory factor analyses of the CAHPS hospital pilot survey responses across and within medical, surgical and obstetric services. *Health Serv Res* 2005;40(6 Pt 2):2078–94.
33. Jenkinson C, Coulter A, Reeves R, et al. Properties of the Picker Patient Experience questionnaire in a randomized controlled trial of long versus short form survey instruments. *J Public Health Med* 2003;25:197–201.
34. Picker Institute Europe. *Guidance manual for the NHS Adult Inpatient Survey 2012*. Oxford: Picker Institute Europe, 2012.
35. Scottish Government. *Scottish Inpatient Patient Experience Survey 2012, volume 2: technical report*. <http://www.scotland.gov.uk/Publications/2012/08/5616/0>.
36. Wong EL, Coulter A, Cheung AW, et al. Validation of inpatient experience questionnaire. *Int J Qual Health Care* 2013;25:443–51.
37. Oltedal S, Garrat A, Bjertnaes O, et al. The NORPEQ patient experience questionnaire: data quality, internal consistency and validity following a Norwegian inpatient survey. *Scand J Caring Sci* 2007;35:540–7.
38. Wilde B, Larsson G, Larsson M, et al. Quality of care: development of a patient-centred questionnaire based on a grounded theory model. *Scand J Caring Sci* 1994;8:39–48.
39. Larsson BW, Larsson G. Development of a short form of the Quality from the Patient's Perspective (QPP) questionnaire. *J Clin Nurs* 2002;11:681–7.
40. Pettersen KL, Veenstra M, Guldvog B, et al. The Patient Experiences Questionnaire: development. *Int J Qual Health Care* 2004;16:453–63.
41. Webster TR, Mantopoulos J, Jackson E, et al. A brief questionnaire for assessing patient healthcare experiences in low-income settings. *Int J Qual Health Care* 2011;23:258–68.
42. Rao KD, Peters DH, Banteen-Roche K. Towards patient-centered health services in India—a scale to measure patient perceptions of quality. *Int J Qual Health Care* 2006;18:414–21.
43. Benneyan JC, Llyod R, Plesk P. Statistical process control as a tool for research and healthcare improvement. *Qual Saf Health Care* 2003;12:458–64.
44. Van der Vleuten C. The assessment of professional competence: developments, research and practical implications. *Adv Health Sci Educ* 1996;1:41–67.
45. Lynn MR. Determination and quantification of content validity. *Nurs Res* 1986;35:382–5.
46. Freeman T. Using performance indicators to improve health care quality in the public sector: a review of the literature. *Health Serv Manag Res* 2002;15:126–37.
47. Streiner DL, Norman GR. *Health measurement scales*. 3rd edn. Oxford Medical Publications, 2003.
48. Terwee CB. *Protocol for systematic reviews of measurement properties*. KnowledgeCenter Measurement Instruments, 2011:13. <http://www.cosmin.nl/images/upload/files/Protocol%20klinimetrische%20review%20version%20nov%202011.pdf>
49. Mokkink L, Terwee C, Patrick D, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. *Qual Life Res* 2010;19:539–49.
50. Coaley K. *An introduction to psychological assessment and psychometrics*. London: Sage publications, 2014.
51. Sargeant J. Reflecting upon multisource feedback as 'assessment for learning.' *Perspect Med Educ* 2015;4:55–6.
52. Larsson G, Larsson B W. Quality improvement measures based in patient data: some psychometric issues. *Int J Nurs Pract* 2003;9:294–9.
53. Sheng Sheng Y, Sheng Z. Is coefficient alpha robust to non-normal data? *Front Psychol* 2012;3:34.
54. Norman G. Likert scales, levels of measurement and the "laws" of statistics. *Adv Health Sci Educ Theory Pract* 2010;15:625–32.