

**Organizational readiness to change assessment (ORCA):**

**Development of an instrument based on the Promoting Action on Research in Health Services (PARiHS) framework**

Christian D. Helfrich<sup>1,2§\*</sup>, Yu-Fang Li<sup>1,3\*</sup>, Nancy D. Sharp<sup>1,2\*</sup>, Anne E. Sales<sup>4\*</sup>

<sup>1</sup>Northwest HSR&D Center of Excellence, VA Puget Sound Healthcare System, Seattle, Washington, USA

<sup>2</sup>Department of Health Services, University of Washington School of Public Health, Seattle, Washington, USA

<sup>3</sup>Department of Biobehavioral Nursing and Health Systems, University of Washington School of Nursing, Seattle, Washington, USA

<sup>4</sup>University of Alberta, Edmonton, Alberta, Canada

\*These authors contributed equally to this work

§Corresponding author

Email addresses:

CDH: [christian.helfrich@va.gov](mailto:christian.helfrich@va.gov)

YFL: [yufang.li@va.gov](mailto:yufang.li@va.gov)

NDS: [nancy.sharp@va.gov](mailto:nancy.sharp@va.gov)

AES: [anne.sales@ualberta.ca](mailto:anne.sales@ualberta.ca)

## **ABSTRACT**

### **BACKGROUND**

The Promoting Action on Research Implementation in Health Services, or PARIHS, framework is a theoretical framework widely promoted as a guide to implementation of evidence-based clinical practices. However, it has as yet no pool of validated measurement instruments that operationalize the constructs defined in the framework. The present article introduces an Organizational Readiness to Change Assessment instrument (ORCA), derived from the core elements and sub-elements of the PARIHS framework, and reports on initial validation.

### **METHODS**

We conducted scale reliability and factor analyses on cross-sectional, secondary data from three quality improvement projects ( $n = 113$ ) conducted in the Veterans Health Administration. In each project, identical 78-item ORCA instruments were administered to one or more staff involved in quality improvement efforts. Items were organized into 19 subscales and three primary scales corresponding to the core elements of the PARIHS framework: (1) Strength and extent of evidence for the clinical practice changes represented by the QI program, assessed with four subscales, (2) Quality of the organizational context for the QI program, assessed with six subscales, and (3) Capacity for internal facilitation of the QI program, assessed with nine subscales.

### **RESULTS**

Cronbach's alpha for scale reliability were 0.75, 0.85 and 0.95 for the evidence, context and facilitation scales, respectively. The evidence scale and its three constituent subscales failed to meet the conventional threshold of 0.80 for reliability, and three individual items

were eliminated from evidence subscales following reliability testing. In exploratory factor analysis, three factors were retained. Seven of the nine facilitation subscales loaded onto the first factor; five of the six context subscales loaded onto the second factor; and the three evidence subscales loaded on the third factor. Two subscales failed to load significantly on any factor. One measured resources in general (from the context scale), and one clinical champion role (from the facilitation scale).

## **CONCLUSIONS**

We find general support for the reliability and factor structure of the ORCA. However, there was poor reliability among measures of evidence, and factor analysis results for measures of general resources and clinical champion role did not conform to the PARIHS framework. The next step is criterion validation.

## **INTRODUCTION**

The Promoting Action on Research Implementation in Health Services, or PARIHS, framework is a theoretical framework widely promoted as a guide to implementation of evidence-based clinical practices [1-5]. It has been the subject of much interest and reference by implementation researchers [6-13], at a time when theoretical frameworks are needed to guide quality improvement activities and research [14-16].

However, a key challenge facing the PARIHS framework is that it has as yet no pool of validated measurement instruments that operationalize the constructs defined in the framework, and the PARIHS framers have prioritized development of diagnostic or evaluation tools.[5] Currently the only published instrument related to PARIHS is a survey from Sweden on clinical practice guideline implementation, which includes items based on PARIHS [13].

The purpose of the present article is to introduce an Organizational Readiness to Change Assessment instrument (ORCA), derived from the major dimensions and sub-dimensions of the PARIHS framework, and to report scale reliability and factor structures. The ORCA was developed by the Veterans Health Administration (VHA) Quality Enhancement Research Initiative for Ischemic Heart Disease and was initially field tested in three quality improvement projects. The scales were designed to assess organizational readiness to change in preparation for implementation of evidence-based changes in clinical practice. The scales are intended for diagnostic use, to identify hypothesized needs or conditions that can be targeted by implementation activities or resources, and to provide a prognosis of the overall readiness to change at the organizational level.

## **BACKGROUND**

### **The PARiHS framework**

The PARiHS framework was developed to represent parsimoniously the determinants of successful implementation of research into clinical practice.[1] The PARiHS framework posits three core elements that determine the success of research implementation: (1) Evidence: the strength and nature of the evidence as perceived by multiple stakeholders; (2) Context: the quality of the context or environment in which the research is implemented, and (3) Facilitation: processes by which the implementation process is facilitated. Each of the three core elements, in turn, comprises multiple, distinct components.

**Evidence** includes four components, corresponding to different sources of evidence: (1) research evidence from published sources, or participation in formal experiments, (2) evidence from clinical experience or professional knowledge, (3) evidence from patient preferences or based on patient experiences, including those of caregivers and family; and (4) routine information derived from local practice context, which differs from professional experience in that it is the domain of the collective environment and not the individual [4, 5]. While research evidence is often treated as the most heavily weighted form, the PARiHS framers emphasize that all four forms have meaning and constitute evidence from the perspective of users.

**Context** comprises three components: (1) organizational culture, (2) leadership, and (3) evaluation [3, 5]. Culture refers to the values, beliefs, and attitudes shared by members of the organization, and can emerge at the macro organizational-level, as well as among sub-units within the organization. Leadership includes elements of teamwork,

control, decision making, effectiveness of organizational structures, and issues related to empowerment. Evaluation relates to how the organization measures its performance, and how (or whether) feedback is provided to people within the organization, as well as the quality of measurement and feedback.

**Facilitation** is defined as a "technique by which one person makes things easier for others" which is achieved through "support to help people change their attitudes, habits, skills, ways of thinking, and working" [1]. Facilitation is a human activity, enacted through roles. Its function is to help individuals and teams understand what they need to change and how to go about it [2, 10]. That role may encompass a range of conventional interventions, such as education, feedback and marketing [10], though two factors appear to distinguish facilitation from conventional multifaceted interventions. First, as its name implies, facilitation emphasizes enabling (as opposed to doing for others) through critical reflection, empathy, and counsel. Second, facilitation is expressly responsive and interactive, whereas conventional multi-faceted interventions do not necessarily involve two-way communication. Stetler and colleagues provide a pithy illustration from an interview[10]:

On the site visit, I came in with a PowerPoint presentation. That is education. When they called me for help ... that was different. It was facilitation.

Harvey and colleagues propose that facilitation is an appointed role, as opposed to an opinion leader who is defined by virtue of his or her standing among peers [2]. Prior publications have also distinguished facilitation roles filled by individuals internal versus external to the team or organization implementing the evidenced-based practice [2, 10]. Internal facilitators are local to the implementation team or organization, and are directly

involved in the implementation, usually in an assigned role. They serve as a major point of interface with external facilitators [10].

This distinction between internal and external facilitation is particularly important in the context of assessing organizational readiness to change. Most prior publications on the PARIHS framework focused on external, rather than internal facilitation. (Stetler and colleagues even make the point of referring to internal facilitators by another name entirely: internal change agents [10]). However, for the purposes of assessing organizational readiness to change, internal facilitation is most pertinent, because it is a function of the organization. Assessing the organization or team's initial state becomes the first step in external facilitation, guiding subsequent facilitation activities [5].

The framers of PARIHS propose that the three core elements of evidence, context and facilitation have a cumulative effect [6], and that none is inherently more important than the others [1]. We further assert that the components of each of the three core elements are, in turn, implicitly compensatory and equifinal. By this, we mean that individual components (operationalized as subscales in the ORCA) are generally neither necessary nor sufficient to achieve implementation; different combinations of components might achieve similar implementation outcomes (i.e., are equifinal); and weakness in some components may be compensated for by strengths in others [17]. For example, the "evidence" may be overall favorable for a new pharmacist-managed, telemonitoring system for hypertension, even in the absence of published research evidence. The lack of research evidence may be compensated for by favorable practice experience with pharmacists managing patients with hypertension, and perceived patient

preferences to increase provider contact but limit travel to clinics. Accordingly, the overall evidence in favor of the telemonitoring system may be high.

## **METHODS**

We conducted two sets of psychometric analyses on cross-sectional, secondary data from three quality improvement projects conducted in the Veterans Health Administration. The survey instrument was developed in consultation with other researchers and is described below.

### **Data and Setting**

Data came from surveys (n = 113) conducted for three quality improvement (QI) projects: 1) the Cardiac Care Initiative (n = 65 from 49 facilities); 2) the Lipids Clinical Reminders project (n = 12 from 1 facility) [18]; and 3) an intensive care unit quality improvement project (n = 36 from 9 facilities). In each project, identical 78-item surveys were administered to one or more staff involved in quality improvement efforts; the only survey element that varied was the initial question stem which stated what the evidence-based practice change was.

In each project, respondents addressed issues related to that specific project (implementing changes in care for acute coronary syndrome, intensive care, or reminders for secondary prevention of ischemic heart disease). The projects were conducted between 2002 and 2006.

### **Survey Instrument**

The survey instrument (Appendix A) comprised three major scales corresponding to the core elements of the PARIHS framework: (1) Strength and extent of evidence for the clinical practice changes represented by the QI program, assessed with four subscales,

(2) Quality of the organizational context for the QI program, assessed with six subscales, and (3) Capacity for internal facilitation of the QI program, assessed with nine subscales. Each subscale comprised between three and six items assessing a common dimension of the given scale. Each item measures the extent to which a respondent agrees or disagrees with the item statement on a 5-point Likert-type scale (1=strongly disagree; 5=strongly agree). Below, we briefly introduce and describe each of the 19 subscales.

**Evidence.** The evidence scale comprised four subscales. The first scale consists of two items which are meant to measure discord within the practice team about the evidence, that is, the extent to which the respondent sees his or her colleagues concluding a weaker or stronger evidence base than the respondent. The other three subscales correspond to the three hypothesized components of evidence in the PARIHS framework: research evidence, clinical experience and patient preferences.

The instrument omits items measuring the fourth hypothesized component of evidence, that of routine information. We argue that routine information is a function of the existence and use of data gathering and reporting systems and is a contextual factor related to the place where the evidence-based practice or technology is being implemented rather than a characteristic of the evidence-based practice itself and how it is perceived by users.

**Context.** Context comprises six subscales. Two subscales assess dimensions of organizational culture: one for senior leadership or clinical management, and one for staff members. Two subscales assess leadership practice: one focused on formal leadership, particularly in terms of teambuilding, and one focused on attitudes of opinion leaders for practice change in general (as a measure of informal leadership practice). One subscale

assesses evaluation in terms of setting goals, and tracking and communicating performance.

In addition, we included a subscale measuring resources to support practice changes, in general, once they had been made an organizational priority. We added this because research on organizational innovation suggests that slack resources, such as funds, staff time, facilities and equipment, are important determinants of successful implementation [19].

**Facilitation.** As explained above, we felt the organization's capacity for internal facilitation was of chief interest for prospective measurement of organizational readiness. Facilitation comprises nine elements: (1) senior leadership management characteristics, such as proposing feasible projects and providing clear goals; (2) clinical champion characteristics, such as assuming responsibility for the success of the project and having authority to carry it out; (3) senior leadership or opinion leader roles, such as being informed and involved in implementation and agreeing on adequate resources to accomplish it; (4) implementation team member roles, such as having clearly defined responsibilities within the team and having release time to work on implementation; (5) implementation plan, such as having explicitly delineated roles and responsibilities, and obtaining staff input and opinions; (6) communication, such as having regular meetings with the implementation team, and providing feedback on implementation progress to clinical managers; (7) implementation progress, such as collecting feedback from patients and staff; (8) implementation resources, such as adequate equipment and materials, and incentives; and (9) implementation evaluation, such as staff and/or patient satisfaction, and review of findings by clinical leadership.

## **Analyses**

We conducted two sets of psychometric analyses: (1) item analysis to determine if items within scales correlate as predicted [20] and (2) exploratory factor analyses of the aggregated subscales to determine how many underlying “factors” might be present, and their relationships to each other [21].

The item analysis consisted of two measures of item correlation with a given subscale: (1) Cronbach's alpha was calculated for reliability; and (2) item-rest correlations were calculated as an indicator of convergent validity and to identify items that do not correlate well with a given subscale and could be dropped for parsimony. We used a minimum threshold of 0.80 for Cronbach's alpha [20], and assessed how dropping an item from its given subscale would affect the Cronbach's alpha for the subscale. We considered the minimum threshold 0.20 for item-rest correlation [20]. We also calculated Cronbach's alpha for the overall scales (e.g., evidence) as a function of the constituent subscales.

We conducted exploratory factor analysis with promax rotation to examine the emergent factor structure of the aggregated subscales, and to determine if the data supported alternative factor solutions other than the three core elements hypothesized by the PARiHS framework. We chose oblique rotation to allow the factors to correlate [22]. This is consistent with both the conceptual underpinnings of the framework which supposes that core elements are interrelated (e.g., facilitation may be influenced by context), and with the items used to operationalize the framework, which include common themes across scales (e.g., leadership culture and leadership implementation role). We factor analyzed the subscales (i.e., the aggregate of constituent items) rather

than the constituent items due to the small number of observations (113) relative to the number of items (78). Data were analyzed using STATA version 9.2.

We retained factors with (1) eigenvalues  $\geq 1.0$ ; (2) eigenvalues greater than the point at which the slope of decreasing eigenvalues approaches zero on a scree plot; and (3) with two or more items loaded with factor loading  $\geq 0.60$  [22]. We only retained factors that met all three criteria. Conversely, we dropped subscales that failed to load on any factor at  $\geq 0.60$ . The threshold of 0.60 was chosen, as opposed to a lower factor loading such as 0.40, to ensure stable factor solutions derived from a small sample size (e.g.,  $\leq 150$  observations) [22].

## **RESULTS**

### **Descriptive Statistics**

Mean scores on the subscales ranged from 2.25 (general resources subscale in the Lipids Reminders project sample) to 4.19 (research evidence subscale in the Lipids Reminders project sample) on a 5-point scale (Table 1). Across the three samples, clinical experience favoring the evidence-based practice changes was rated marginally lower, on average, than was the perceived research evidence, and the evidence in terms of patient preferences was rated lowest of the three evidence subscales. Among the subscales measuring context, staff culture was the highest rated in the Lipids Reminders and Cardiac Care Initiative projects, and opinion leaders was highest in the ICU QI Intervention. Across the three samples, the general resources subscale was the lowest rated of all subscales. Among the subscales measuring facilitation, leaders' practices was rated highest in the Lipids Reminders and Cardiac Care Initiative projects, and project evaluation was highest in the ICU QI Intervention; project evaluation was the single

highest rated subscale in the ICU QI Intervention sample, although it was among the lower half of subscales in the Lipids Reminders and Cardiac Care Initiative projects. Across the three samples, the project resources subscale was the lowest rated of the facilitation subscales.

### **Item Analysis**

Cronbach's alpha for scale reliability for the overall scales were 0.75, 0.85 and 0.95 for the evidence, context and facilitation scales, respectively. Cronbach's alpha for the constituent subscales ranged from 0.70 for the patient experience subscale of the evidence scale to 0.94 for both the leadership behavior subscale of the context scale and the implementation plan subscale of the facilitation scale (Table 1).

Three subscales, the three comprising the evidence scale, failed to meet the conventional threshold of 0.80 for reliability [20]. Reliability statistics were initially 0.51, 0.62 and to 0.70, for the research evidence, clinical experience and patient preference subscales, respectively. One item from the research evidence subscale, q3e (the practice change will fail to improve patient outcomes), had an item-rest correlation of 0.18, failing to meet the threshold of 0.20. Eliminating this item improved the Cronbach's alpha to 0.55, but decreased the item-rest correlation for item q3d (the practice change will improve patient outcomes, even though it is experimental) to 0.17. Dropping q3d further improved the Cronbach's alpha for the research evidence subscale to 0.71.

For the clinical experience subscale, item q4d (the practice change has not been previously attempted in the facility) had the lowest item-rest correlation at 0.27. Although it met the minimum threshold for item-rest correlations, the Cronbach's alpha for the subscale improved from 0.62 to 0.75 when item q4d was dropped from the subscale.

These three items (q3e, q3d, and q4d) were excluded in subsequent analyses. This decision was based on the reliability results, and because of the items appeared to address potentially distinct concepts, such as predicting the effect of the practice change on patient outcomes (this is further explained in the Discussion). The figures in Table 1 were calculated without these three items.

The patient preferences subscale failed to meet the 0.80 threshold for reliability, but item-rest correlations for all four items ranged from 0.45 to 0.51, well above the minimum threshold of 0.20. Eliminating any item decreased the Cronbach's alpha for the subscale. Although the subscales comprising the evidence scale failed to meet the minimum threshold for reliability, we elected to retain them for the factor analysis because the scale represented concepts central to the PARiHS model.

### **Factor Analysis**

Based on the three criteria discussed in the methods section, three factors were retained (Table 2). Following promax rotation, based on the criterion of factor loading  $\geq 0.60$ , seven of the nine facilitation subscales loaded onto the first factor; five of the six context subscales loaded onto the second factor; and the three evidence subscales loaded on the third factor. No subscales cross-loaded on multiple factors, and all subscales, except the leaders' practices subscale from the facilitation scale, loaded primarily on factors corresponding to the core element they were intended to measure. The subscale measuring leader practices had a factor loading of 0.76 on the second factor, which the majority of the context subscales loaded on.

Two subscales failed to load significantly on any of the factors. One measured resources in general (from the context scale), and one clinical champion role (from the

facilitation scale), although both loaded primarily on the first factor, with the majority of facilitation subscales. The factor loadings were 0.41 and 0.49, respectively.

Two subscales had uniqueness values in excess of 0.50, which indicates the variance in the subscales left unexplained by the three emergent factors taken together. The general resources subscale had a uniqueness value of 0.71, and the patient preferences subscales had uniqueness value of 0.67.

## **DISCUSSION**

We find some statistical support, in terms of reliability and factor analyses, for aggregation of survey items and subscales into three scales of organizational readiness-to-change based on the core elements of the PARiHS framework: evidence, context and facilitation. Reliability statistics met conventional thresholds for the majority of subscales, indicating that the subscales intended to measure the individual components of the main elements of the framework (e.g., the six components of the context scale) held together reasonably well. Exploratory factor analysis applied to the aggregated subscale scores supports three underlying factors, with the majority of subscale scores clustered corresponding to the core elements of the PARiHS framework.

However, there were three results that raised concerns and suggest need for further revision to the instrument and further research on its stability: (1) reliability was poor for the three evidence subscales; (2) the subscales measuring clinical champion (as part of the facilitation scale), and availability of general resources (as part of the context scale) failed to load significantly on any factor; and (3) the leadership practices subscale loaded on the second factor with most of the context subscales. We discuss each of these in turn.

### **Reliability of evidence subscales**

Reliability, as measured by Cronbach's alpha, was mediocre for the evidence scale and the three constituent subscales. Poor reliability could be a function of too few items (alpha coefficients are highly sensitive to the number of items in a scale [20]; could indicate that the items are deficient measures of the evidence construct; or could signal that the subscales are not uni-dimensional, i.e., they reflect multiple underlying factors with none measured reliably or well.

There is some evidence for the latter given the observed improvement in reliability statistics after dropping three items: q3d and q3e from the research evidence subscale, and q4d from the practice experience subscale. These items had some important conceptual differences from other items in their respective subscales. Both q3d and q3e are about anticipating the effect of the practice change on patient outcomes, whereas the other items in the subscale (q3a - q3c) are about the scientific evidence for the practice change. The former require respondents to make a prediction about a future state, not just an assessment of a current one (i.e., the state of the research evidence). Item q4d, on the other hand, is about whether the practice change has previously been attempted in the respondent's clinical setting, which was unlikely given the context was quality improvement projects introducing new practices. However, factor analysis generally supported a common factor solution for the three subscales, suggesting that overall the subscales may tap into a common latent variable.

The patient preferences subscale requires further consideration, and we feel remains an open question as to how it fits with the model and with the survey. It had high uniqueness, indicating that the majority of variance in the items was not accounted for by

the three factors. Furthermore, past research appears to undermine the contention that patient preferences are a significant factor in how favorably clinicians evaluate a given practice or course of treatment. For example, there is little or no correlation between patient preferences and what clinicians do [23, 24], and even after interventions to increase shared decision making (a practice intended to better incorporate patient preferences into health care practice), the actual effects on clinical choices appear limited, although providers and patients may perceive greater participation [25]. Patient preferences *should* be a major driver of implementation of evidence-based practices, but we suspect that in our current health care system it is generally not. It remains unclear what this means for assessing patient preferences as a distinct component of organizational readiness to change, but additional exploratory research would seem to be in order.

### **Two subscales with low factor loadings**

Two subscales failed to load significantly on any of the three factors: One measured dimensions of facilitation related to the clinical champion role, the other measured dimensions of context related to the availability of general resources, such as facilities and staffing. There are at least three ways to interpret this finding, with different attendant implications.

First, the failure of the two subscales to load on any of the three factors may indicate that overall availability of resources and clinical champion roles are functions of unique factors, distinct from evidence, context and facilitation (as framed in this instrument). Empirically and conceptually, we believe this may be the case for the general resource availability, but not for the clinical champion role.

In the case of general resource availability, the subscale had high uniqueness, indicating that a majority of variance of the items was not accounted for by any of the three factors. Conceptually, this subscale was not part of the original PARiHS framework; it was added to the ORCA based on other organizational research supporting the powerful influence of resource availability as an initial state that often sets boundaries in planning and execution. Although this seems to fit logically within the domain of the context scale, general resources may be a function of factors at other levels. This is consistent with the observed subscale scores, which were lowest for the general resources subscale across the three study samples. General resource availability may be less a function of the organization (in this case individual VHA facilities), and more a function of the broader resource environment in the VHA, or in the US health care system generally. The period covered in these three quality improvement projects has been one of high demand on Veterans Health Administration services [26], and cost containment was (and continues to be) a major and pervasive issue in healthcare [27]. We still believe that resource availability is an important factor in the determination of organizational readiness to change, however, we believe it may be distinct from the three factors hypothesized in the PARiHS model, appearing different from the other dimension of context. We propose that additional conceptual work is needed on this subscale and that more items are likely needed to reliably measure it.

Second, the failure of the two subscales to load on any of the three factors may indicate that organizational readiness to change is more properly understood as a formative, rather than a reflective scale. Principal factors analysis assumes that the individual items are reflective of common, latent variables (or factors) that cause the item

responses [28, 29]; when a scale is reflective, it corresponds to a given latent variable. However, organizational readiness to change may be more aptly understood as a formative scale, meaning that the constituent pieces (items or subscales) are the determinants and the latent variable organizational readiness to change is the intermediary outcome [28]. In the former case, the constituent parts are necessarily correlated (see Howell et al 2007 for a comparison of the mathematical assumptions underlying formative and reflective scales). For example, a scale meant to measure native athletic ability should register high correlations among constituent components meant to assess speed, strength, and agility; i.e., the physiological factors that determine speed, are also thought to determine strength and agility, and therefore a person scoring "high" on one component should score relatively high on the others. Conversely, a scale meant to measure how good a baseball player is, might assess their throwing, fielding, and batting to create a composite score. Throwing, fielding and batting may often be related—being in part a function of native athletic ability—but they're also a function of specific training activities and experience, and skill developed in one does not parlay into skill in the others. Rigorous training in pitching will not make you a good batter. Currently, it is not clear if organizational readiness to change should properly be understood as a formative or a reflective scale.

Third, the distinctiveness of the two subscales may indicate measurement error. General resource availability and clinical champion role might be appropriately understood as distinct reflections of the favorability of the context in the organization. However, the items, and their component subscales, may simply be inaccurate measures of the latent variables, or the number of observations in this analysis may have been

insufficient for a stable estimate of the factors. We believe the latter is the case for the clinical champion subscale, which had a relatively low uniqueness value (0.34), and relatively high factor loading (0.49). Although the factor loading did not meet the threshold (0.60), we set an unusually high threshold for this analysis because the relatively small number of observations needed to be balanced with high factor loadings in order to achieve stable estimates [22]. We expect that repeating the analysis with a larger sample will confirm that the clinical champion subscale loads onto the same factor as the other facilitation subscales.

### **The leadership practices subscale loaded on the context factor**

The subscale measuring leaders' practices (from the facilitation scale) loaded on the second factor with context subscales. The leaders' practice subscale addressed whether senior leaders or clinical managers propose an appropriate, feasible project; provide clear goals; establish a project schedule; and designate a clinical champion. The high loading on the second factor could indicate that the leaders' practices subscale is properly understood as part of context, or it could signal poor discriminant validity between the context and facilitation scales. However, in this case, we believe the overlap may be a function measurement error related to item wording. Two of the items refer to "a project," which put the respondent in mind of a generic change more consonant to the questions in the context scale, whereas many of the facilitation items in the subsequent subscales refer to "this project" or "the intervention" implying the specific implementation project named in the question stem at the opening of the survey.

We believe that this unintended discrepancy in the pattern of wording cued respondents to answer the leader practices questions in a different frame of mind,

conceiving of them in terms of projects in general rather than their estimate of leadership practices in the project they were actively engaged upon. This will be a revision to explore in future use of the survey.

### **Limitations**

There are five major limitations to our work. First, this analysis does not address the validity of the instrument as a predictor of evidence-based clinical practice, or even as a correlate of theoretically relevant covariates, such as implementation activities. Our objective with the present analysis was confined to assessing correlations among items within respondents to determine if the items cluster into scales and subscales as predicted. Criterion validation using implementation and quality-of-care outcomes is the next phase of our work.

Second, this study relied on secondary data from quality improvement projects, which did not employ some standard practices for survey development intended to mitigate threats to internal validity. We note two specific examples. First, the items were organized according to the predicted scales and subscales, rather than being presented to respondents in a random order. Item ordering can influence item scoring, and introduces the danger that reliability statistics may be inflated because items were organized according to the predicted subscales. However, this is not an uncommon practice in health services research survey instruments. Second, two of the quality improvement projects (Cardiac Care Initiative, and the intensive care unit quality improvement project) entailed multiple evidence-based practice changes, each of which could conceivably elicit different responses in terms of evidence, context and facilitation. The surveys assessed these practice changes as a whole, and therefore may have introduced measurement error

to the extent that respondents perceived evidence, context and facilitation differently for different components. However, the danger here is less significant than for the item ordering, as the measurement error would tend to inflate item variance within scales, and therefore bias results towards the null (i.e., toward an undifferentiated mass of items rather than distinct scales), which we did not observe.

Third, the survey instrument is quite long (78 items), and to be practicable will need to be considerably shorter. Despite the length, we note that most respondents are able to complete the survey in about 15 minutes. However, the research presented included only preliminary item reduction based on scale reliability. Although scale reliability statistics often serve as a basis for excluding items [20], we believe that item reduction is best done as a function of criterion validation, i.e., that items are retained as a function of how much variance they account for in some theoretically meaningful outcome. We suggest this is a priority for the next stage of research.

Fourth, the sample size was small (113) relative to the number of survey items (78). This led us to factor analyze the aggregated subscales rather than the constituent items. This assumed that the subscales were unidimensional. While Cronbach's alpha findings generally supported the reliability of the subscales, high average correlations can still occur among items that reflect multiple factors [30], and high reliability is no guarantee that the subscales were unidimensional. This limitation will be corrected with time when additional data become available and the analysis can be repeated with a larger sample.

Fifth, the ORCA was fielded a single time in each project, which leaves unanswered questions both about the proper timing of the assessment and how variable

subscales and scales are over time. In terms of timing, in the Lipids Clinical Reminders project, and the intensive care unit quality improvement project the instrument was fielded before any work related to the specific change was undertaken. In the case of the Cardiac Care Initiative, some work had already begun at some sites. It is possible that administering the instrument at more than one time point might yield different factor structures.

Other limitations include questions of external validity, for example, in terms of the setting in the VHA and these particular evidence-based practices; and questions of internal validity, in terms of the sensitivity of the measures to changes in wording or format. These limitations are all important topics for future research on the instrument.

## **CONCLUSIONS**

We find general support for the reliability and factor structure of an organizational readiness to change assessment based on the PARiHS framework. We find some discrepant results, specifically poor reliability among subscales intended to measure distinct dimensions of evidence, and factor analysis results for measures of general resources and clinical champion role that do not conform to the PARiHS framework.

The next critical step is to use outcomes from implementation activities for criterion validation. This should provide more information about which items and scales are the most promising candidates for a revised readiness to change instrument.

## **Abbreviations**

ORCA – Organizational Readiness to Change Assessment; PARiHS – Promoting Action on Research Implementation in Health Services; VHA – Veterans Health Administration;

## **Competing interests**

The authors declare that they have no competing interests.

### **Authors' contributions**

CDH conceived of the study and framed the research design, carried out the analyses, interpreted findings, and drafted the manuscript. YFL collaborated on study design, advised on the analyses, interpreted findings, and helped draft the manuscript. NDS led the development of the ORCA, helped frame the study, interpreted findings, and helped draft the manuscript. AES was a co-developer of the ORCA, helped frame the study, advised on the analyses, interpreted findings and helped draft the manuscript. All authors read and approved the final manuscript.

### **Acknowledgements**

The research reported here was supported by Department of Veterans Affairs, Veterans Health Administration, Health Services Research and Development Service, project grant number RRP 07-280. Drs. Helfrich, Li and Sharp were supported by the VA Northwest HSR&D Center of Excellence.

We wish to thank Mary McDonnell for overall project management, and Rachel Smith and Liza Mathias for project support for this research study. We also wish to thank Jennie Bowen who completed early reliability analyses for the instrument.

The views expressed in this article are the authors' and do not necessarily reflect the position or policy of the Department of Veterans Affairs.

### **REFERENCES**

1. Kitson A, Harvey G, McCormack B: **Enabling the implementation of evidence based practice: a conceptual framework.** *Quality in Health Care* 1998, **7**:149-158.
2. Harvey G, Loftus-Hills A, Rycroft-Malone J, Titchen A, Kitson A, McCormack B, Seers K: **Getting evidence into practice: the role and function of facilitation.** *Journal of Advanced Nursing* 2002, **37**:577-588.
3. McCormack B, Kitson A, Harvey G, Rycroft-Malone J, Titchen A, Seers K: **Getting evidence into practice: the meaning of 'context'.** *J Adv Nurs* 2002, **38**:94-104.
4. Rycroft-Malone J, Seers K, Titchen A, Harvey G, Kitson A, McCormack B: **What counts as evidence in evidence-based practice?** *J Adv Nurs* 2004, **47**:81-90.
5. Kitson A, Rycroft-Malone J, Harvey G, McCormack B, Seers K, Titchen A: **Evaluating the successful implementation of evidence into practice using the PARIHS framework: theoretical and practical challenges.** *Implementation Science* 2008, **3**:1.
6. Rycroft-Malone J, Kitson A, Harvey G, McCormack B, Seers K, Titchen A, Estabrooks C: **Ingredients for change: revisiting a conceptual framework.** *Quality & Safety in Health Care* 2002, **11**:174-180.
7. Rycroft-Malone J: **The PARIHS framework--a framework for guiding the implementation of evidence-based practice.** *J Nurs Care Qual* 2004, **19**:297-304.
8. Rycroft-Malone J, Harvey G, Seers K, Kitson A, McCormack B, Titchen A: **An exploration of the factors that influence the implementation of evidence into practice.** *J Clin Nurs* 2004, **13**:913-924.
9. Brown D, McCormack B: **Developing Postoperative Pain Management: Utilising the Promoting Action on Research Implementation in Health Services (PARIHS) Framework.** *Worldviews on Evidence-Based Nursing* 2005, **2**:131-141.
10. Stetler C, Legro M, Rycroft-Malone J, Bowman C, Curran G, Guihan M, Hagedorn H, Pineros S, Wallace C: **Role of "external facilitation" in**

- implementation of research findings: a qualitative evaluation of facilitation experiences in the Veterans Health Administration.** *Implementation Science* 2006, **1**:23.
11. Cummings GG, Estabrooks CA, Midodzi WK, Wallin L, Hayduk L: **Influence of organizational characteristics and context on research utilization.** *Nurs Res* 2007, **56**:S24-39.
  12. Estabrooks CA, Midodzi WK, Cummings GG, Wallin L: **Predicting research use in nursing organizations: a multilevel analysis.** *Nurs Res* 2007, **56**:S7-23.
  13. Bahtsevani C, Willman A, Khalaf A, Östman M: **Developing an instrument for evaluating implementation of clinical practice guidelines: a test-retest study.** *Journal of Evaluation in Clinical Practice* 2008, **Electronic edition ahead of publication.**
  14. Eccles M, Grimshaw J, Walker A, Johnston M, Pitts N: **Changing the behavior of healthcare professionals: the use of theory in promoting the uptake of research findings.** *Journal of Clinical Epidemiology* 2005, **58**:107-112.
  15. (ICEBeRG) TICeTBRG: **Designing theoretically-informed implementation interventions.** *Implementation Science* 2006, **1**:4.
  16. Grol RPTM, Bosch MC, Hulscher MEJL, Eccles MP, Wensing M: **Planning and Studying Improvement in Patient Care: The Use of Theoretical Perspectives.** *The Milbank Quarterly* 2007, **85**:93-138.
  17. Klein KJ, Conn AB, Sorra JS: **Implementing computerized technology: An organizational analysis.** *Journal of Applied Psychology* 2001, **86**:811-824.
  18. Sales A, Helfrich C, Ho PM, Hedeem A, Plomondon ME, Li Y-F, Connors A, Rumsfeld JS: **Implementing Electronic Clinical Reminders for Lipid Management in Patients with Ischemic Heart Disease in the Veterans Health Administration.** *Implementation Science* 2008, **3**.
  19. Bourgeois LJ: **On the measurement of organizational slack.** *Academy of Management Review* 1981, **6**:29-39.
  20. Bernard HR: *Social Research Methods: Qualitative and Quantitative Approaches.* Thousand Oaks, CA: Sage; 2000.

21. Nunnally JC, Bernstein IH: *Psychometric Theory*. 3rd edn. New York, NY: McGraw-Hill Inc.; 1994.
22. Floyd F, Widaman K: **Factor Analysis in the Development and Refinement of Clinical Assessment Instruments**. *Psychological Assessment* 1995, **7**:286-299.
23. Sanchez-Menegay C, Stalder H: **Do physicians take into account patients' expectations?** *J Gen Intern Med* 1994, **9**:404-406.
24. Montgomery AA, Fahey T: **How do patients' treatment preferences compare with those of clinicians?** *Qual Saf Health Care* 2001, **10**:i39-43.
25. Davis RE, Dolan G, Thomas S, Atwell C, Mead D, Nehammer S, Moseley L, Edwards A, Elwyn G: **Exploring doctor and patient views about risk communication and shared decision-making in the consultation**. *Health Expect* 2003, **6**:198-207.
26. Getzan C: **VA Funding Fails to Meet Increased Demand for Services, Groups Say; As Congress and the President haggle over future Veterans Administration funding, a New England Journal of Medicine study shows an increased risk of mental health disorders among Middle East veterans**. In *The New Standard*. Syracuse, NY; 2004.
27. Mays GP, Claxton G, White J: **Managed care rebound? Recent changes in health plans' cost containment strategies**. *Health Aff (Millwood)* 2004, **Suppl Web Exclusives**:W4-427-436.
28. Howell RD, Breivik E, Wilcox JB: **Reconsidering formative measurement**. *Psychological Methods* 2007, **12**:205-218.
29. Edwards JR, Bagozzi RP: **On the nature and direction of relationships between constructs and measures**. *Psychological Methods* 2000, **5**:155-174.
30. Shevlin M, Hunt N, Robbins I: **A confirmatory factor analysis of the Impact of Event Scale using a sample of World War II and Korean War veterans**. *Psychol Assess* 2000, **12**:414-417.

Table 1 : Descriptive Statistics and Reliability for Organizational Readiness to Change Assessment Subscales

	Number of items retained	Lipids-reminders (n = 12)		ICU QI-intervention (n = 36)		Cardiac Care Initiative (n = 65)		p-value <sup>‡</sup>	Overall (n = 113)		Cronbach's Alpha
		Mean	SD	Mean	SD	Mean	SD		Mean	SD	
Evidence Scale <sup>†</sup>	10	3.96	0.24	3.88	0.40	4.00	0.45	0.37	3.96	0.41	0.75 <sup>††</sup>
Research	3	4.19	0.48	4.04	0.52	4.16	0.51	<b>0.02</b>	4.12	0.51	0.71
Clinical experience	3	4.08	0.35	3.96	0.51	4.12	0.55	0.19	4.07	0.52	0.75
Patient preferences	4	3.60	0.43	3.70	0.51	3.73	0.49	0.62	3.71	0.49	0.70
Context Scale <sup>†</sup>	23	3.24	0.44	3.54	0.45	3.80	0.67	0.07	3.65	0.61	0.85 <sup>††</sup>
Leader culture	3	2.92	0.93	3.54	0.81	3.79	0.99	0.08	3.62	0.96	0.90
Staff culture	4	4.00	0.60	3.69	0.60	4.05	0.78	<b>0.02</b>	3.93	0.72	0.91
Leadership behavior	4	2.92	0.90	3.56	0.68	3.79	0.96	0.10	3.62	0.91	0.94
Measurement (leadership feedback)	4	3.63	0.73	3.44	0.67	3.99	0.77	<b>0.03</b>	3.78	0.77	0.87
Opinion leaders	4	3.73	0.38	3.81	0.42	4.04	0.64	<b>0.02</b>	3.94	0.56	0.89
General resources	4	2.25	0.61	3.13	0.81	2.88	0.86	0.18	2.89	0.85	0.88
Facilitation Scale <sup>†</sup>	40	3.14	0.50	3.85	0.31	3.51	0.69	0.76	3.55	0.63	0.95
Leaders practices	4	3.42	0.64	3.59	0.37	3.74	0.72	0.14	3.66	0.63	0.85
Clinical champion	4	3.19	0.67	3.74	0.57	3.62	0.86	0.14	3.61	0.78	0.93
Leadership implementation roles	4	2.94	0.68	3.87	0.56	3.66	0.66	<b>0.01</b>	3.64	0.69	0.87
Implementation team roles	4	2.92	0.71	3.66	0.56	3.33	0.81	0.10	3.38	0.76	0.85
Implementation plan	4	3.17	0.77	4.00	0.50	3.65	0.82	0.39	3.70	0.77	0.94
Project communication	4	3.25	0.65	4.04	0.56	3.66	0.81	0.97	3.72	0.76	0.91
Project progress tracking	4	3.25	0.51	3.95	0.48	3.40	0.73	0.22	3.54	0.69	0.84
Project resources and context	7	2.86	0.63	3.63	0.51	3.23	0.75	0.10	3.30	0.71	0.87
Project evaluation	5	3.30	0.67	4.11	0.42	3.49	0.62	<b>0.03</b>	3.65	0.64	0.86

† The three major scales (evidence, context, facilitation) are averages of their constituent subscales, thus subscales are equally weighted.

‡ p-value is for F-statistic from ANOVA.

†† Cronbach's alpha for a revised evidence scale based on just the research evidence and clinical experience subscales was 0.83.

‡‡ Cronbach's alpha for a revised context scale based dropping the general resources subscale was 0.87.

Table 2: Exploratory factor analysis of Organizational Readiness to Change Subscales

Retained Factors	Eigenvalue	Proportion		
Factor1	7.61	0.59		
Factor2	7.12	0.55		
Factor3	3.23	0.25		
Principal factors with promax rotation				
	Factor 1	Factor 2	Factor 3	Uniqueness
Evidence Scale				
Research	-0.10	0.11	<b>0.74</b>	0.42
Clinical experience	0.04	0.01	<b>0.83</b>	0.27
Patient preferences	0.06	-0.24	<b>0.62</b>	0.67 <sup>†</sup>
Context Scale				
Leader culture	0.07	<b>0.83</b>	-0.08	0.29
Staff culture	-0.17	<b>0.67</b>	0.26	0.48
Leadership behavior	0.08	<b>0.88</b>	-0.05	0.18
Measurement (leadership feedback)	0.07	<b>0.72</b>	0.01	0.41
Opinion leaders	0.04	<b>0.69</b>	0.12	0.41
General resources	0.41	0.10	0.13	0.71 <sup>†</sup>
Facilitation Scale				
Leaders practices	0.24	<b>0.74</b>	-0.02	0.19
Clinical champion	0.49	0.35	0.15	0.34
Leadership implementation roles	<b>0.65</b>	0.33	-0.08	0.28
Implementation team roles	<b>0.67</b>	0.23	0.02	0.30
Implementation plan	<b>0.73</b>	0.34	-0.10	0.13
Project communication	<b>0.80</b>	0.12	0.07	0.20
Project progress tracking	<b>0.92</b>	-0.09	-0.02	0.25
Project resources and context	<b>0.86</b>	0.01	0.00	0.24
Project evaluation	<b>0.88</b>	-0.14	0.02	0.34

Factor loadings  $\geq 0.60$ , our threshold, are bolded

† Indicates subscale for which factors failed account for  $\geq 50\%$  of variance.

Additional Files

Appendix A

File name: ORCA\_annotated\_08-01-08.pdf

File format: PDF

Title of data: Annotated copy of the Organizational Readiness to Change Assessment  
(ORCA)

Description of data: This is an annotated copy of the Organizational Readiness to Change  
Assessment (ORCA).

**Additional files provided with this submission:**

Additional file 1: orca\_annotated\_08-29-08.pdf, 128K

<http://www.implementationscience.com/imedia/1031121556219754/supp1.pdf>