Family Practice, 2021, 339–345
doi:10.1093/fampra/cmaa102
Advance Access publication 24 September 2020

Research Methods

The validity, reliability, sensitivity and responsiveness of a modified Patient Enablement Instrument (PEI-2) as a tool for serial measurements of health enablement

Laura Elizabeth Bedford[a,*,†], Maegan Hon Yan Yeung[a,†], Chi Ho Au[a], Emily Tsui Yee Tse[a], Wing Yee Yim[a], Esther Yee Tak Yu[a], Carlos King Ho Wong[a,b] and Cindy Lo Kuen Lam[a]

[a]Department of Family Medicine and Primary Care and [b]Department of Pharmacology and Pharmacy, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong SAR, China
[†]These authors contributed equally.

*Correspondence to Laura Elizabeth Bedford, Department of Family Medicine and Primary Care, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Rm 1-01, 1/F, Jockey Club Building for Interdisciplinary Research, 5 Sassoon Road, Pokfulam, Hong Kong SAR, China; E-mail: lbedford@hku.hk

Abstract

Background: Patient enablement is a core tenet of patient-centred and holistic primary care. The Patient Enablement Instrument (PEI) is a transitional measure limited in its ability to measure changes over time. A modified version, PEI-2, has been developed to measure enablement at a given time-point without comparison to a recalled baseline.
Objective: To assess the validity, reliability, sensitivity and responsiveness of PEI-2.
Methods: PEI-2 was modified from the Chinese PEI to assess enablement over 4 weeks in a prospective cohort study nested within a community support programme [Trekkers Family Enhancement Scheme (TFES)] in Hong Kong. Construct validity was assessed by factor analysis and convergent validity by Spearman's correlations with health-related quality of life and depressive symptoms. Internal reliability was assessed using Cronbach's alpha. Test–retest reliability was assessed by intraclass correlation (ICC), responsiveness by 12–24-month change in PEI-2 score and sensitivity by differences in change of PEI-2 score between TFES participants and a control group.
Results: PEI-2 demonstrated construct validity with all items loading on one factor (factor loadings >0.7). Convergent validity was confirmed by significant correlations with 12-item Short Form Questionnaire, version 2 ($r = 0.1089$–0.1919) and Patient Health Questionnaire-9 ($r = -0.2030$). Internal reliability was high (Cronbach's alpha = 0.9095) and test–retest reliability moderate (ICC = 0.520, $P = 0.506$). Significant improvements in PEI-2 scores among the TFES group suggested good responsiveness ($P < 0.001$). The difference in change of PEI-2 scores between TFES and control was significant ($P = 0.008$), indicating good sensitivity.
Conclusions: This study supports the validity, reliability, sensitivity and responsiveness of PEI-2 in measuring changes in enablement, making it a promising tool for evaluating enablement in cohort and intervention studies.

Key words: Health enablement, patient-centred care, primary care, questionnaire, reliability, validity

Key Messages

- Patient enablement is a core tenet of patient-centred and holistic primary care.
- We have developed a modified Patient Enablement Instrument (PEI-2).
- This tool can evaluate care in cohort and intervention studies.
- Our results support its validity, reliability, sensitivity and responsiveness.

Introduction

Patient-centred and holistic primary care is conducive to patient empowerment (1–3), whereby patients are motivated to gain greater control over decisions affecting their health through collaborative efforts with health care providers (4). A core tenet of empowerment is enablement (5), which refers to self-perceived ability to understand and cope with illnesses and health issues (6). The Patient Enablement Instrument (PEI) (7) is a commonly used measure of enablement in a primary care setting. It includes six items with three response options to assess the degree of improvement in enablement as a result of the GP consultation ('much better/better/same or less'). The PEI (7) has been widely translated and implemented across different populations (6,8–16), including the Chinese population in Hong Kong (17). It demonstrates moderate to good performance in internal consistency [Cronbach's alphas: 0.85–0.922 (9,17,18)], construct validity [$r > 0.4$ (17); $\beta = 2.09$ (19)], test–retest reliability [intraclass correlation (ICC): 0.65–0.75 (9,17)] and sensitivity (17).

The PEI is a transitional measure, such that the respondent reports their perceived level of change in enablement when compared with their recalled status before the consultation (20–22). Therefore, a baseline measurement of enablement is not required when using the PEI as it depends solely on a respondent's recollection (21). In addition to being administered immediately after a consultation (2,13,18), the PEI has also been used to assess both short- (2 weeks) and longer-term (up to 2 years) changes following an intervention (23–26). However, while the PEI is considered the 'gold-standard' (10) measure of enablement, a key problem with transitional scales is that respondents are not always able to accurately recall their baseline state (22), an issue that will become more pertinent if administering the measure long after the intervention or consultation. Evidence has shown that, in cases where respondents are unable to recall their prior health state, that they will complete a transitional measure according to their current health state, thereby rendering it an inaccurate measure of change (20,27). A further limitation with the PEI is that it may be susceptible to hypothesis guessing. For example, a study testing patient understanding of the PEI found that some respondents answered items based on their evaluation of the GPs performance as opposed to perceived change in their own enablement (21).

In order to overcome such limitations, a measure that can capture health enablement at a given point in time would be a more effective means by which to track enablement when used serially in a repeated measures design, for example, as part of the longer-term evaluation of an intervention. Such a measure could be administered at baseline and then at follow-ups with the differences in scores between timepoints used to identify any changes in enablement. We therefore developed a modified PEI (PEI-2) where respondents are asked to rate their level of enablement over the past 4 weeks. The aim of this study was to evaluate its validity, reliability, responsiveness and sensitivity.

Methods

Development of the Chinese (Hong Kong) PEI-2

The Chinese PEI-2 was modified from the Chinese PEI, which has good translation equivalence, validity, reliability and sensitivity (17). The PEI-2 utilizes the same stem questions as the PEI, which assess coping, understanding and self-care (Supplementary Figure 1). In order to assess the magnitude of perceived enablement, each item is rated on a 5-point Likert scale, ranging from 1 (not at all) to 5 (extremely well). The item scores are summed to give a total score (range: 6–30), with higher scores indicating better enablement. The increase in the number of response options to five confers higher sensitivity to discriminate between varying levels of enablement.

Study population and setting

This study included participants from a prospective cohort study nested within the Trekkers Family Enhancement Scheme (TFES) (17,28). The TFES was established in 2012 and is funded by a local philanthropic organization. The objective is to empower low-income families by providing support and opportunities in health, education, employment and environmental harmony (29). A health empowerment program is delivered, which consists of annual health examinations and various programs to enhance health literacy and enable self-care. Participants were identified through local non-government organizations and sent an invitation letter with inclusion criteria: (i) ≥ one family member working full-time or part-time; (ii) ≥ one dependent children aged 6–11 years; (iii) a monthly income less than 75% of Hong Kong's median monthly household income and (iv) ability to consent to take part (29). TFES participants were invited to take part in the cohort study that aimed to evaluate the impact of the health empowerment program on a range of outcomes, the primary outcome being health enablement, measured by the PEI-2 (29). Control families were recruited to act as a comparison group. Their inclusion criteria were similar to the TFES families and they also received a health examination at baseline and 5 years. They were not provided with the other health empowerment programmes. All TFES and control participants were invited to complete a questionnaire on morbidity, service utilization, lifestyle, health enablement, health-related quality of life (HRQoL) and mental health at baseline and 12–24-month follow-up.

Data collection

The data of adult participants who completed the PEI-2 at baseline and follow-up (12–24 months) were extracted from the cohort study database. For the TFES group, baseline records were collected between 2012 and 2013. As control participants were recruited at a later date, baseline data were collected between 2014 and 2016. The PEI-2 was administered by an annual telephone health survey or during annual health assessments. Lam *et al.* previously demonstrated similar results obtained from telephone interviews compared with face-to-face surveys (30).

Additional measures were administered as part of the cohort study, which included the Chinese Patient Health Questionnaire (PHQ)-9 (31) to measure depression and the Chinese (Hong Kong) adaptation of the 12-item Short Form Questionnaire, version 2 (SF-12v2) (17) to measure HRQoL. Their data were extracted to assess convergent validity of the PEI-2.

In order to evaluate test–retest reliability, we invited TFES participants to complete two telephone surveys with PEI-2 at 2 weeks apart.

Analyses

Descriptive statistics [mean, standard error, response distributions] were calculated by assessment time. Floor and ceiling effects of PEI-2 scores were considered to be present if >15% of subjects reported the minimum or maximum possible scores (32).

Validity. We hypothesized that the six PEI-2 items measure one single construct of enablement. Exploratory factor analysis (EFA) and confirmatory factor analysis (CFA) were used to assess construct validity. EFA using a principal components method with Varimax was first used to determine the underlying factor structure. CFA was used to test whether the relationship between the observed variables and their underlying latent factor(s) exists. A model chi-square test, root mean square error of approximation (RMSEA), comparative fit index (CFI), non-normed fit Tucker–Lewis index (TLI) and standardized root mean square residual (SRMR) were used to assess the model goodness-of-fit. These were considered adequate if: (i) chi-square test ($P \geq 0.05$); (ii) RMSEA <0.08 (33); (iii) CFI ≥0.90 (34); (iv) TLI ≥0.90 (35); (v) SRMR <0.08 (36).

Convergent validity is established when theoretically corresponding constructs are observed to be correlated. We hypothesized that health enablement should have a correlation with mental health and HRQoL, although they do not measure identical constructs. Spearman's correlations between PEI-2 scores and those of PHQ-9, SF-12v2 domain and summary scores were tested.

Reliability was first assessed by internal consistency and Cronbach's alpha coefficient was used to determine internal consistency relative to the expected standard of ≥0.7 (37). The effect of imputed data substitutions (missing values) on internal consistency was undertaken in a sensitivity analysis. Test–retest reliability was calculated between the total PEI-2 scores at baseline and the 2-week follow-up. Paired *t*-tests on the difference in means and ICC were employed as indices of test–retest reliability. Inter-rater reliability of each individual PEI-2 item was assessed using (i) Gwet's agreement coefficient (Gwet's AC) and (ii) percent agreement calculations between test and retest. Gwet's AC is generally the statistic of choice for the case of two raters (test–retest) as it does not depend upon the assumption of independence between raters (38).

The changes in PEI-2 scores within groups (TFES and control) at 12–24 months were used to evaluate responsiveness. We hypothesized that TFES participants would have significant improvements in enablement and the PEI-2 total score would detect this. The difference in PEI-2 scores between baseline and follow-up was evaluated using paired *t*-tests and effect size. The minimally clinical important difference (MCID) of the new PEI-2 is not known, we set the threshold MCID at ≥0.5 SD of the mean baseline score, taking reference of the MCID of other patient-reported outcome measures (39). Sensitivity was measured by ability to detect a difference between groups. We hypothesized that the change in PEI-2 scores would be higher among the TFES group than that among the control group. The significance of the difference in change was assessed by the two-sample *t*-test.

Analyses were conducted by STATA version 13.1. Ethical approval was obtained from the Institutional Review Board of the University of Hong Kong—the Hospital Authority Hong Kong West Cluster (reference no: UW 12-517).

Results

Subject characteristics

In total, 360 cohort study participants were eligible for inclusion, of whom 285 completed the 12–24-month follow-up. Their mean age was 41.9 years, 70.8% were females, 47.7% were employed and 46.1% had an income below the Hong Kong median of $13 500. No statistically significant differences were found between subjects with and without follow-up data (Table 1). For the test–retest study, 60 TFES subjects completed the baseline survey, of which 53 completed the follow-up with no missing data.

Response distribution

The response distribution of PEI-2 items showed that there was no floor effect at baseline and follow-up (Supplementary Table 1). A ceiling effect was identified for three items at baseline and all six items at follow-up. The total PEI-2 scores were lowest (6) in 5 (1.4%) and 2 (0.7%) subjects at baseline and follow-up, respectively, and were highest (30) in 13 (3.6%) and 16 (5.6%) subjects at baseline and follow-up, respectively.

Validity

The EFA overall Kaiser–Meyer–Olkin measure (0.8761) and Bartlett's test of sphericity (chi-square test = 109.691, $P < 0.001$) confirmed the sampling adequacy and variability (Table 2). Using the iterated principal-factor method, a one-factor solution was found, which explained 86.9% of the total variance. All six items loaded significantly on this single factor (factor loadings ≥0.7). CFA confirmed the one-factor structure with all items loading strongly and significantly on the factor. The CFI (0.9281), TLI (0.8802) and SRMR (0.0397) supported adequate model fit, but the model chi-square test was significant ($P < 0.001$) and the RMSEA was relatively large (0.1768).

Convergent validity was confirmed by a significant negative correlation between the total PEI-2 and PHQ-9 score indicating that subjects with better enablement were less likely to report depressive symptoms (Table 3). There were significant, although weak, correlations between the total PEI-2 and the Mental Component Summary score as well as seven of the eight domain scores of the SF-12v2.

Reliability

Internal consistency was strong with an overall Cronbach's alphas of 0.9095. There was no significant difference in the total PEI-2 score between test and retest (Table 4). The total PEI-2 score demonstrated moderate reliability (ICC = 0.520, $P = 0.506$). The reliability of individual items assessed inter-rater reliability and showed moderate to good reproducibility (Gwet's AC: 0.382–0.637, 49.1–69.8%).

Responsiveness and sensitivity

As hypothesized, for the TFES group, PEI-2 total scores showed a statistically significant improvement over time (difference = 3.05, $P < 0.001$) with an effect size of 0.522 that was greater than the MCID standard (39) (Table 5). There was only slight improvement for the control group (difference = 0.90, $P = 0.050$) with an effect size of 0.167 SD that was unlikely to be clinically important. When the changes of the two groups were compared, the TFES group showed significantly greater improvements in mean total PEI-2 scores (difference between the mean difference = 2.15, $P = 0.001$).

Table 1. Sociodemographic and clinical characteristics for cohort study participants (TFES and control combined) and the test–retest subjects recruited from the TFES

	Cohort study participants				Test–retest subjects[c]
	Baseline[a]	Completion of 12–24-month follow-up[b]			
	Total (*N* = 360)	Yes (*n* = 285)	No (*n* = 75)	*P*-value	Total (*N* = 60)
Gender (%, *n*)	*N* = 360	*N* = 285	*N* = 75	0.544	*N* = 60
Female	70.8% (255)	71.6% (204)	68.0% (51)		71.7% (43)
Male	29.2% (105)	28.4% (81)	32.0% (24)		28.3% (17)
Age (mean ± SE), year	41.9 (0.4)	41.9 (0.4)	41.8 (0.8)	0.931	48.2 (1.0)
Marital status (%, *n*)	*N* = 360	*N* = 285	*N* = 75	0.543	*N* = 60
Married	18.9% (68)	18.3% (52)	78.7% (59)		80.0% (48)
Unmarried	81.1% (292)	81.8% (233)	21.3% (16)		20.0% (12)
Employment status (%, *n*)	*N* = 329	*N* = 279	*N* = 50	0.379	*N* = 54
Working	47.7% (157)	48.8% (136)	42.0% (21)		70.4% (38)
Not working	52.3% (172)	51.3% (143)	58.0% (29)		29.6% (16)
Chronic diseases (%, *n*)	*N* = 360	*N* = 285	*N* = 75		*N* = 60
No chronic disease	70.8% (255)	74.4% (212)	57.3% (43)		68.3% (41)
One or more chronic diseases	29.2% (105)	25.6% (73)	42.7% (32)		31.7% (19)
Household income	*N* = 360				*N* = 45
Median	$13 000				$16 000
Above population median	46.1% (166)				51.1% (23)
Below population median	53.9% (194)				48.9% (22)

SE, standard error.

[a]Baseline data collection period: TFES group (January 2012–September 2014); control group (January 2014–December 2016).

[b]Follow-up data collection period: TFES group (August 2013–March 2016); control group (September 2014–May 2017).

[c]Test–retest data collection dates: December 2019–January 2020.

Table 2. Exploratory and confirmatory factor loadings of PEI-2 items for the cohort study participants (TFES and control combined)

	EFA	CFA	
	Factor loading	Factor loading	Variance explained
PEI-2 items (*N* = 360)[a]			
Able to cope with life	0.7544	0.7446	0.5544
Able to understand your illness	0.7247	0.7219	0.5211
Able to cope with your illness	0.8898	0.8743	0.7645
Able to keep yourself healthy	0.8281	0.8172	0.6678
Confident about your health	0.7962	0.7840	0.6146
Able to help yourself	0.8347	0.8178	0.6687
Overall KMO	0.8761		
Chi-square test		109.691 ($P < 0.001$)	
P-value		<0.001*	
RMSEA		0.1768	
CFI		0.9281	
TLI		0.8802	
SRMR		0.0397	

KMO, Kaiser–Meyer–Olkin.

*Statistically significant.

[a]Baseline data collection period: TFES group (January 2012–September 2014); control group (January 2014–December 2016).

Table 3. Spearman's correlations between PEI-2 total scores, PHQ-9 scores and SF-12v2 summary scores for the cohort study participants (TFES and control combined)

	PEI-2 total score (*N* = 358)[a]	
	Correlation coefficient (*r*)	*P*-value
PHQ-9 score	**–0.2030**	**0.014***
SF-12v2		
Physical functioning	0.1089	0.040*
Role physical	0.1094	0.039*
Bodily pain	0.1919	<0.001*
General health	–0.0747	0.159
Vitality	0.1838	<0.001*
Social functioning	0.1703	0.001*
Role emotional	0.1118	0.035*
Mental health	0.1827	<0.001*
Physical component summary	**0.0663**	**0.212**
Mental component summary	**0.1720**	**0.001***

*Statistically significant.

[a]Baseline data collection period: TFES group (January 2012–September 2014); control group (January 2014–December 2016).

Discussion

Summary

Our results support the validity, reliability, responsiveness and sensitivity of the PEI-2 in Chinese adults recruited from the general population. No floor or ceiling effects were found for total PEI-2 score implying that the PEI-2 is potentially sensitive in detecting improvement or deterioration in health enablement. A ceiling effect

Table 4. Test–retest reliability of the PEI-2 for subjects recruited from the TFES group

	Baseline (N = 53)[a]			Follow-up (N = 53)[b]			P-value[†]	ICC	Gwet's AC	Agreement, %
	Mean (SD)	Floor (%)	Ceiling (%)	Mean (SD)	Floor (%)	Ceiling (%)				
Able to cope with life	4.09 (0.95)	3.3	40.0	3.94 (0.77)	0.0	26.4	N.A.	N.A.	0.478	56.6
Able to understand your illness	3.55 (0.99)	5.0	15.0	3.68 (0.80)	0.0	17.0	N.A.	N.A.	0.477	56.6
Able to cope with your illness	3.75 (0.96)	0.0	21.7	3.77 (0.78)	0.0	20.8	N.A.	N.A.	0.412	54.7
Able to keep yourself healthy	3.68 (0.85)	0.0	18.3	3.62 (0.79)	0.0	15.1	N.A.	N.A.	0.491	60.4
Confident about your health	3.74 (1.06)	3.3	28.3	3.57 (0.89)	1.9	17.0	N.A.	N.A.	0.382	49.1
Able to help yourself	4.02 (0.95)	3.3	36.7	3.94 (0.84)	1.9	24.5	N.A.	N.A.	0.638	69.8
PEI-2 total score	22.83 (4.70)	0.0	5.0	22.53 (4.21)	0.0	7.5	0.506	0.520	N.A.	N.A.

N.A., not applicable.
[a]Baseline data collected in December 2019.
[b]Follow-up data collected between December 2019 and January 2020.
[†]Tested by paired t-test.

Table 5. Changes in PEI-2 total scores for the TFES group and the control group

	Study time-point		Difference between baseline and follow-up		Effect size of within group change	P value for paired t-test on within group change	P value for two-sample t-test on difference of changes between groups
	Baseline[a] (mean score, SD)	Follow-up[b] (mean score, SD)	Mean, SD	(95% CI)			
Total (N = 285)	18.7 (5.4)	20.7 (5.4)	1.99 (5.71)	(1.32, 2.66)	0.348	<0.001*	—
TFES group (51.2%, N = 140)	15.6 (4.0)	18.6 (5.4)**	3.05 (5.84)	(2.09, 4.01)	0.522	<0.001*	0.001*
Control group (48.8%, N = 145)	22.0 (4.7)	22.9 (4.4)	0.90 (5.39)	(0.00, 1.80)	0.167	0.050	

CI, confidence interval. *Notes*: Analyses adjusted for age, gender, household income (<HK$20 000 versus ≥HK$20 000) and comorbidities.
[a]Baseline data collection period: TFES group (January 2012–September 2014); control group (January 2014–December 2016).
[b]Follow-up data collection period: TFES group (August 2013–March 2016); control group (September 2014–May 2017).
*Statistically significant.
**Clinically meaningful change based on an increase of more than one half a SD (39).

was observed in individual PEI-2 items, especially on follow-up assessment, which was expected from relatively healthy adults. Indeed, a much higher proportion of subjects scored the highest in individual PEI-2 items at follow-up, suggesting a positive effect from the TFES. Such a ceiling effect would be less likely among patient populations.

For construct validity, we confirmed the one-factor structure of the PEI-2 by EFA and CFA inferring that PEI-2 items are all valid indicators of the construct of health enablement. The adequate CFI, TLI and SRMR indicated that the one-factor model fit was acceptable at an absolute level although the significant model chi-square test and the relatively large RMSEA (0.1768) suggested that additional factors or more complex factors were possible. However, one factor explained 86.9% of the variance in the EFA and the factor loadings in both EFA and CFA were very strong, while the effect of other factors, if any, would be small. It should be pointed out that the P-value is a function of sample size, a statistically significant model chi-square test in a large sample may not necessarily imply the model inadequacy. As there is currently no gold-standard measure for health enablement, we used the PHQ-9 and SF-12v2 to evaluate the convergent validity of the PEI-2 based on the hypothesis that there is a mutual effect between enablement, depression and HRQoL. Although significant, the correlations were weak as they do not measure the same construct. Further studies using measures of more similar constructs (e.g. self-efficacy) may show stronger convergent validity.

The Cronbach's alpha on internal consistency was high (α = 0.9095), although the test–retest reliability was moderate. It is important to consider though that test–retest and inter-rater reliability assumes that the construct remains stable across time. The results could have been affected by variance introduced by different interviewers who administered the PEI-2 at different time-points, by inconsistencies in the participant's self-evaluation and by a real change in health enablement.

Responsiveness and sensitivity are important psychometric properties of a measure that is administered to monitor changes over time or to evaluate the effectiveness of an intervention. As hypothesized, the PEI-2 demonstrated good responsiveness with its ability to detect improvements in health enablement within the TFES group. The sensitivity was strong as the PEI-2 could discriminate the improvement in enablement between control and TFES subjects, demonstrating the effectiveness of the health empowerment intervention.

Strengths and limitations

This study has established the psychometric properties of a modified PEI. The PEI-2 could track changes in health enablement over time

without the need to compare to a recalled baseline state, making it a promising tool for cohort and intervention studies. Indeed, given that '*measurement is the first step that leads to control and eventually to improvement*' (H. James Harrington), the PEI-2 may contribute to the long-term advancement of primary care practice. However, the generalizability of this study is limited due to sampling from within an established cohort study and results may not translate to other contexts or populations.

Comparison with existing literature

As the PEI-2 is a new measure, no studies have explored its validity and reliability. The wider literature on the original PEI, however, has also reported on the instability of enablement. For example, Röösт *et al.* (9) identified a rapid decline in mean PEI scores 2 days after a consultation (kappa values: 0.54–0.65 between baseline and 2 days after) and further maturation effects have been found at 3 weeks after initial measurements [ICC = 0.62 (6)]. Enablement could therefore be viewed as being influenced by external factors, which could explain why test–retest reliability was only moderate. Widespread social unrest was prevalent in Hong Kong when the test–retest work took place, which could have had detrimental effects on perceived coping (40). It is therefore plausible that our results partly reflect actual deteriorations in enablement as opposed to measurement errors of PEI-2.

Conclusions

The results of this study support the validity, reliability, responsiveness and sensitivity of the PEI-2 in Chinese adults recruited from the general population. The PEI-2 can be used to measure changes in enablement, making it a promising tool for evaluating enablement in cohort and intervention studies. Further studies with larger samples, and self-administration of the PEI-2, should be conducted to establish its test–retest reliability and to differentiate intra- and inter-rater variability.

Supplementary material

Supplementary material is available at *Family Practice* online.

Acknowledgements

The authors thank KGKF for their funding and the Neighborhood Advice-Action Council and Yat Tung (I/II) Estate Property Management for providing a venue for health assessments. Thank you to Versitech Ltd for granting a complementary license to use the Opine Software for data collection. The completion of telephone surveys by the University of Hong Kong Social Science Research Center is much appreciated. Thank you to all our research staff for support with data collection.

Declaration

Funding: this study was supported by a research donation to the University of Hong Kong from the Kerry Group Kuok Foundation (Hong Kong) Limited (KGKF). No funding organization had any role in the design and conduct of the study; collection, management, analysis and interpretation of the data; or manuscript preparation.

Ethical approval: all study procedures were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. Ethics approval was granted by the Institutional Review Board of the University of Hong Kong/Hospital Authority Hong Kong West Cluster (Ref. UW 12-517).

Conflict of interest: Prof. Cindy LK Lam is an associate editor for *Family Practice*. There are no other conflicts of interest to declare.

Data availability

The data underlying this article cannot be shared publicly in order to protect the privacy of the individuals that participated in the study.

References

1. Stewart M, Brown JB, Weston WW *et al. Patient-Centered Medicine: Transforming the Clinical Method*. Thousand Oaks, CA: Sage Publications, Inc., 1995, xxiv, 267 p.
2. Little P, Everitt H, Williamson I *et al.* Observational study of effect of patient centredness and positive approach on outcomes of general practice consultations. *BMJ* 2001; **323** (7318)**:** 908–11.
3. Howie JG, Heaney D, Maxwell M. Quality, core values and the general practice consultation: issues of definition, measurement and delivery. *Fam Pract* 2004; **21** (4)**:** 458–68.
4. Nutbeam D. Evaluating health promotion—progress, problems and solutions. *Health Promot Int* 1998; **13** (1)**:** 27–44.
5. Fumagalli LP, Radaelli G, Lettieri E, Bertele' P, Masella C. Patient Empowerment and its neighbours: clarifying the boundaries and their mutual relationships. *Health Policy* 2015; **119** (3)**:** 384–94.
6. Hudon C, Fortin M, Rossignol F, Bernier S, Poitras ME. The Patient Enablement Instrument-French version in a family practice setting: a reliability study. *BMC Fam Pract* 2011; **12**: 71.
7. Howie JG, Heaney DJ, Maxwell M, Walker JJ. A comparison of a Patient Enablement Instrument (PEI) against two established satisfaction scales as an outcome measure of primary care consultations. *Fam Pract* 1998; **15** (2)**:** 165–71.
8. Pawlikowska TR, Walker JJ, Nowak PR, Szumilo-Grzesik W. Patient involvement in assessing consultation quality: a quantitative study of the Patient Enablement Instrument in Poland. *Health Expect* 2010; **13** (1)**:** 13–23.
9. Röösт M, Zielinski A, Petersson C, Strandberg EL. Reliability and applicability of the Patient Enablement Instrument (PEI) in a Swedish general practice setting. *BMC Fam Pract* 2015; **16**: 31.
10. Tolvanen E, Koskela TH, Kosunen E. Comparison of the Patient Enablement Instrument (PEI) with two single-item measures among Finnish Health care centre patients. *BMC Health Serv Res* 2019; **19** (1)**:** 376.
11. Enthoven P, Peolsson A, Ludvigsson ML *et al.* Validity, internal consistency and self-rated change of the patient enablement instrument in patients with chronic musculoskeletal pain. *J Rehabil Med* 2019; **51** (8)**:** 587–97.
12. Remelhe M, Teixeira PM, Lopes I, Silva L, Correia de Sousa J. The modified patient enablement instrument: a Portuguese cross-cultural adaptation, validity and reliability study. *NPJ Prim Care Respir Med* 2017; **27**: 16087.
13. Ozvacić Adzić Z, Katić M, Kern J *et al.* Patient, physician, and practice characteristics related to patient enablement in general practice in Croatia: cross-sectional survey study. *Croat Med J* 2008; **49** (6)**:** 813–23.
14. Al Momen RK, Alotaibi MM, Abdelhay O. Patient enablement in chronic diseases in primary health care, Riyadh City, KSA. *Int J Clin Med* 2015; **6** (9)**:** 615–22.
15. Birhanu Z, Assefa T, Woldie M, Morankar S. Determinants of satisfaction with health care provider interactions at health centres in central Ethiopia: a cross sectional study. *BMC Health Serv Res* 2010; **10**: 78.
16. Azizam NA, Maon SN, Abdul Aziz NIS, Hamid NZA. Association of patient centered communication and patient enablement (conference paper). In: International Conference on Business and Economics, Shah Alam, Malaysia, 2016, pp. 235–42.
17. Lam CL, Yuen NY, Mercer SW, Wong W. A pilot study on the validity and reliability of the Patient Enablement Instrument (PEI) in a Chinese population. *Fam Pract* 2010; **27** (4)**:** 395–403.
18. Howie JG, Heaney DJ, Maxwell M *et al.* Quality at general practice consultations: cross sectional survey. *BMJ* 1999; **319** (7212)**:** 738–43.

19. Small N, Bower P, Chew-Graham CA, Whalley D, Protheroe J. Patient empowerment in long-term conditions: development and preliminary testing of a new measure. *BMC Health Serv Res* 2013; **13**: 263.
20. Kamper SJ, Maher CG, Mackay G. Global rating of change scales: a review of strengths and weaknesses and considerations for design. *J Man Manip Ther* 2009; **17** (3)**:** 163–70.
21. Murphy M, Hollinghurst S, Salisbury C. Patient understanding of two commonly used patient reported outcome measures for primary care: a cognitive interview study. *BMC Fam Pract* 2018; **19** (1)**:** 162.
22. Streiner DL, Norman GR, Cairney J. *Health Measurement Scales: A Practical Guide to Their Development and Use*. 5th edn. Oxford, UK: Oxford University Press, 2015.
23. Ahmed N, Hughes P, Winslow M *et al*. A pilot randomized controlled trial of a holistic needs assessment questionnaire in a supportive and palliative care service. *J Pain Symptom Manage* 2015; **50** (5)**:** 587–98.
24. Chen JY, Wan EY, Choi EP *et al*. Clinical and patient-reported outcomes of Chinese patients undergoing haemodialysis in hospital or in the community: a 1-year longitudinal study. *Nephrology (Carlton)* 2016; **21** (7)**:** 617–23.
25. Everitt HA, Landau S, O'Reilly G *et al*. Cognitive behavioural therapy for irritable bowel syndrome: 24-month follow-up of participants in the ACTIB randomised trial. *Lancet Gastroenterol Hepatol* 2019; **4** (11)**:** 863–72.
26. Bikker AP, Mercer SW, Reilly D. A pilot prospective study on the consultation and relational empathy, patient enablement, and health changes over 12 months in patients going to the Glasgow Homoeopathic Hospital. *J Altern Complement Med* 2005; **11** (4)**:** 591–600.
27. Kamper SJ, Ostelo RW, Knol DL *et al*. Global Perceived Effect scales provided reliable assessments of health transition in people with musculoskeletal disorders, but ratings are strongly influenced by current status. *J Clin Epidemiol* 2010; **63** (7)**:** 760–6.e1.
28. Fung CSC, Yu EYT, Guo VY *et al*. Development of a Health Empowerment Programme to improve the health of working poor families: protocol for a prospective cohort study in Hong Kong. *BMJ Open* 2016; **6** (2)**:** 5–7.
29. Fung CS, Wan EY, Yu CL, Wong CK. Validity and reliability of the 19-item Audit of Diabetes-Dependent Quality of Life (ADDQoL-19) questionnaire in Chinese patients with type 2 diabetes mellitus in primary care. *Qual Life Res* 2016; **25** (9)**:** 2373–8.
30. Lam TH, Kleevans JW, Wong CM. Doctor-consultation in Hong Kong: a comparison between findings of a telephone interview with the general household survey. *Community Med* 1988; **10** (3)**:** 175–9.
31. Yu X, Tam WW, Wong PT, Lam TH, Stewart SM. The Patient Health Questionnaire-9 for measuring depressive symptoms among the general population in Hong Kong. *Compr Psychiatry* 2012; **53** (1)**:** 95–102.
32. McHorney CA, Tarlov AR. Individual-patient monitoring in clinical practice: are available health status surveys adequate? *Qual Life Res* 1995; **4** (4)**:** 293–307.
33. MacCallum RC, Browne MW, Sugawara HM. Power analysis and determination of sample size for covariance structure modeling. *Psychol Methods* 1996; **1** (2)**:** 130–49.
34. Bentler PM. Comparative fit indexes in structural models. *Psychol Bull* 1990; **107** (2)**:** 238–46.
35. Forza C, Filippini R. TQM impact on quality conformance and customer satisfaction: a causal model. *Int J Prod Econ* 1998; **55** (1)**:** 1–20.
36. Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Struct Equ Model* 1999; **6** (1)**:** 1–55.
37. Nunnally JC, Bernstein IH. *Psychometric Theory*. New York, NY: McGraw-Hill, 1994.
38. Gwet KL. Computing inter-rater reliability and its variance in the presence of high agreement. *Br J Math Stat Psychol* 2008; **61** (Pt 1)**:** 29–48.
39. Norman GR, Sloan JA, Wyrwich KW. Interpretation of changes in health-related quality of life: the remarkable universality of half a standard deviation. *Med Care* 2003; **41** (5)**:** 582–92.
40. Ni MY, Yao XI, Leung KSM *et al*. Depression and post-traumatic stress during major social unrest in Hong Kong: a 10-year prospective cohort study. *Lancet* 2020; **395** (10220)**:** 273–84.