International Journal of Advanced Community Medicine

Daprim S Ogaji
Department of Preventive and
Social Medicine, University of
Port Harcourt, Nigeria

Effect of questionnaire response scale on items response characteristics in primary care patients survey: Comparing unipolar versus bipolar response formats

Daprim S Ogaji

Abstract
Quantitative surveys on patient reported experiences of primary health care (PHC) are often conducted with questionnaires designed with scales or categorical responses. This cross-sectional analytic study involving 322 regular adult users of primary health centre in Nigeria compared the acceptability and score characteristics of 5-point bipolar and 5-point unipolar response formats in a multi-dimensional questionnaire. Outcome measured were the item score characteristics, effect sizes on item scores, response rate, difficulty rating. Significantly higher item scores were associated with the bipolar response format with an overall effect size of 0.12 (95%CI: 0.08 - 0.17). The Odds of responding to bipolar compared to the unipolar response format was 0.65 (95% CI: 0.53 – 0.79, p = 0.000). The response formats in scaled survey questionnaires could influence findings on patients' self-report. This have implications for the conduct of meta-analysis with data pooled from individual studies with different response formats.

Keywords: Response format, unipolar scale, bipolar scale, Likert scale, adjectival scale, primary health care, patient reported experiences measure

Introduction

Quantitative scales developed for patients' evaluation of primary care are commonly in the form of self-administered questionnaires designed to obtain structured feedback from patient who have recently had encounter with health care [1]. The ability to differentiate different levels and performance of health care delivery from the patient survey could depend on the response style used [2, 3]. The effects of response styles on patient-reported experience of health care may result in an inflation or deflation of respondent's score [2]. Besides the quantitative scores, the discriminative property of the questionnaire is also an important feature that may be affected by the response style. Thus, response formats may influence the ability of questionnaire to elicit enough variance among groups or entities being measured as well as influence the psychometric properties of the questionnaire [4].

There are tendencies for respondents to respond systematically to questions based on the response style and regardless of their content. Previous comparative studies on scale responses have shown how bias could result from the use of different response styles [2, 5-7]. Common biases with response styles include: acquiescence response style (ARS), disacquiescence response style (DARS), net acquiescence response style (NARS), extreme response style (ERS), response range (RR), midpoint responding (MPR) and non-contingent responding (NCR) [2, 6]. Extreme response style for example, is present where respondents preferentially select the endpoints of the scale when responding to the various items in the questionnaire and this can be explained by the floor and/or ceiling effects [8]. The group's culture is reported to be an important influence in both the extreme and acquiescent response styles [5].

The development of questionnaires for patient-reported experiences survey are often context and setting specific. Such development may either follow extensive processes involving patients or shorter processes that often depend more on subject experts [1, 8, 9]. Despite remarkable progress in the development and use of valid and reliable scales for patient-reported experiences on health care globally, only few of the available questionnaires for PHC were developed with reference to settings in sub Saharan Africa [10-12]. As progress n scale development for patient-reported experiences measures is till at a very early stage in

Correspondence
Daprim S Ogaji
Department of Preventive and
Social Medicine, University of
Port Harcourt, Nigeria

Nigeria, it is needful to investigate the relative performance of common response formats among the Nigeria PHC users. The objective of this research was to compare the item response, scores, floor, ceiling effects, difficulty grading by patients of the unipolar and bipolar response variants of a semantically identical questionnaire for patient-reported experience of primary health care in Nigeria.

Methods

Study setting
Study was conducted at the primary health centre, Aluu. This health centre which offers preventive, promotive and curative services for ambulatory and in-patients is supervised by the Rivers State Primary Health Care Board. The health centre manned by doctors, nurses, community health practitioners and other support staff runs 24 hours service.

Sample size
This study was designed to detect a difference of 10% in item response characteristics with a confidence of 95%. Sample size was calculated using formula for comparative design [13]

$$n\ per\ group = \frac{2(z_{\alpha/2} + z_\beta)^2 * \rho(1-\rho)}{(P_1 - P_2)^2}$$

Where

n = sample size in each group (assumed equal sized groups)

ρ = (P_1 + P_2)/2 is the mean of the two proportions. Previous pilot survey gave an item response rate of 93.3%. a difference of 10% would give the proportion as 93.7 and 83.7% respectively

$(P_1 - P_2)^2$ = Effect size (the difference in proportion)

$Z_{\alpha/2}$ = standard normal deviation corresponding to selected α level of 0.025 in each tail = 1.96

Z_β = desired power (typically 0.84 for 80% power)

$n/group = 158$

This was adjusted slightly upwards to 161

Study instrument
The research questionnaire was obtained at a phase during the iterative development of the patient evaluation scale [14]. The developmental phases of this patient evaluation scale included the generation of items from a review of literature and content analysis of qualitative interviews with PHC patients, face and content validation by local PHC experts, cognitive testing of the items and instructions in the questionnaire with PHC patients and quantitative pre-test of questionnaire's acceptable across patients' groups drawn from 4 PHC centres in the north and south of Nigeria. The two response formats of the questionnaire prepared by the research team were semantically identical in content and instructions but phrased in response to the answering scale. All questions were closed ended and worded in similar manner and direction. The core questionnaire contained 27 items. Form A was structured in statement format with bipolar 5-point Likert response options ('strongly disagree', 'disagree', 'neither agree nor disagree', 'agree', and 'strongly agree') while Form B was phrased as questions with unipolar 5-point adjectival response options ('poor', 'fair', 'good', 'very good' and 'excellent').

Study procedure
Participants were regular adult outpatients to the centre that gave consent to participate. Matching was done by gender and health problems and consecutively recruited patients were sequentially assigned into either of the two response groups and offered the questionnaire to complete. Completed questionnaires were handed over to trained research assistants who were always present at the facility during the period of data collection.

Study variables
The variables measured were:

Measures of acceptability [8]
- Questionnaire response rate: Proportion of questionnaires returned after appropriate completion.
- Items response rate: Proportions of individual item responded to by the respondents.
- Endorsement frequencies: Distribution of responses across the various item response options.

Score distribution
Distribution characteristics of item score showing the floor effect (proportion of responses clustered at lowest respond end of the scale), ceiling effect (proportion of responses clustered at highest respond end of the scale), means, standard deviations, skewness and ranges of item scores.

Ease of completion
- Time of completion of the questionnaire by respondents
- Difference in patients' difficulty rating of the questionnaire on a scale of 1 – 7.

Statistical analysis
Depending on the measurement scale, the differences between the two questionnaires were analysed using the Odd's ratio or the standardised difference between means. Determination of standardized difference between proportions comparing difference in proportion was achieved with the Odds-Ratio. The Odds-Ratio is calculated based on a 2 by 2 contingency table as shown below [15]:

	Item responded to	
	Yes	No
Questionnaire format A	a	b
Questionnaire format B	c	d

$$\overline{ES} = \frac{ad}{bc}$$

The calculated Odds Ratio is the Odds of responding to items with bipolar response format in relation to the Odds of responding to same item with the unipolar response format. The effect statistics ($\overline{ES}$) were computed in terms of the calculated Odds ratio and its 95% confidence intervals for each PHC attribute (items).

Determination of standardized difference between means
This was computed for the differences in the means of the various item scores, using the Cohen's effect size. The Forest plot of this finding on all items and overall effect size was then presented.

$$d = \frac{\mu_a - \mu_b}{S_{pooled}}$$

[16]

Cohen effect size
Where
d = Cohen's effect size
- μ_a = Mean (scores of items in form A)
- μ_b = Mean (scores of items in form B)
- S = Pooled standard deviation
- The pooled standard deviation from both measure was similarly calculated using the formula [16]:

$$S_{pooled} = \sqrt{\frac{[(n_a - 1)S_a^2 + (n_b - 1)S_b^2]}{n_a + n_b}}$$

[16]

Pooled standard deviation
Where
s = standard deviation
n = number of subjects

Results
Most of the respondents in this survey were female (93%), aged 20-29 years (51%) and married (81%). There was no significant difference in the socio-demographic characteristics of the respondents in both group except for marital status where higher proportion of patients that received questionnaire with unipolar response format were single (Table 1).

Figure 1, presents findings on the floor and ceiling effects of the unipolar response format of the questionnaire. The floor effects range from 0.0– 21.1 and the ceiling effects from 11.4-62.9 across items. Items with high ceiling effects were related to experience on consultation (62.9), staff availability (60.9) and likelihood of return to the centre for care should the need arise in the future (60.7).

Similar presentation of the floor and ceiling effects with the use of the bipolar-type response format is shown in Figure 2. This showed a mean floor effect of 2.6 with an observed range from 0.0 – 15.3% and a mean ceiling effect of 33.4 with an observed range of 17.3 – 45.8. Although both formats generally showed low floor effects, the observed ceiling effects were more pronounced in the unipolar response format.

Figure 3 shows the sizes of the effects of response formats on item scores. Approximately, 56 (95% CI: 54-57) of items in the unipolar response format had mean scores that were significantly lower than corresponding items in the bipolar response format. The higher standardized mean differences above 0.5 were observed with items 11 to 14 while the overall effect of response format on scale's score was 0.12 (95% CI: 0.08 – 0.17; p = 0.02) in favour of the bipolar response format from the forest plot.

Figure 4 shows similar comparison between the unipolar and bipolar scale with respect to item response. Patients who received the unipolar response format had significantly higher item response rate than the responders to the bipolar response format. The effect size of 0.65 (95% CI: 0.53 – 0.79) in the forest plot indicates that patients who responded to the unipolar questionnaires were at least 50% (OR = 1.54, 95% CI: 1.27 – 1.89, *p*<0.001) more likely to respond to the items in the questionnaire than those who responded to the bipolar response scale.

Discussion
Summary of the findings
This study shows that difference could exist between questionnaire with bipolar and unipolar response formats. These differences are related to the items scores, items responses and the ceiling effect. There was no significant difference in the floor effect, time of completion of the questionnaire and patients' grad of the ease of completion of the variants of the questionnaire.

Comparison with other studies
We found no report on similar comparison of scores, items' response and other indices for determining the acceptability or feasibility between unipolar and bipolar scales or more specifically, between unipolar and bipolar response formats. Items with bipolar response formats are believed to perform better than other response scales during psychometric validation of questionnaire [4, 8]. This type of scale is also more commonly used in published scales for patient-reported experiences of primary health care [17]. The use of the unipolar response format had also been adopted in fewer questionnaires that were developed for patient-reported experiences of primary care [17].

Multipoint ordinal data in this study were analysed using frequency of endorsement of various categories and rating scores applied to single items or the summation of items. The conduct of parametric analysis on multipoint ordinal data and item summation are still subjects of intense debates [18]. However, treating ordinal scale as interval data to allow application of more robust parametric tests is reported to aid interpretation of the findings without causing great loss in accuracy even where sample sizes and response options are not large enough for scores to attain normal distribution pattern [19-21]. Furthermore, summing the score of items is advantageous because single items have considerable random measurement error; are less reliable, lack precision and rarely possess sufficient information for the estimation of validity, accuracy and reliability [22, 23].

The observed floor effects for both variants were generally low but the bipolar response variants provided a significantly lower ceiling effects from this study. While it is appropriate to keep both indices below 10% or maintain an even endorsement of all response options [8]. It is also of great advantage to reduce the floor and ceiling effects in questionnaire surveys as doing so will improve the efficiency of the study as well as reduce the sample size required to elicit a difference in the entities being measured [24]. Questionnaires for patient-reported experiences are often designed as *evaluative* and/or *discriminative* scale. The evidence from this research, if confirmed using stronger research design, could form the evidence base for the choice of response format in the Nigerian cultural and practice setting. From the findings of this study, the preference for the unipolar response format over the bipolar, could be premised on the need to pay more attention to issues of validity (associated with higher response rate) in a trade-off for higher scale scores.

Implications of the findings
Findings show that response formats could have significant effects on item response rates and also item scores. There was significantly lower item non-response with the unipolar adjectival response format and also, significantly higher score was reported with the Likert-type bipolar response

format. This has important implications where data from different survey research are compared or pooled together during the conduct of meta-analysis. The procedure in the latter, often involve the transformation of ratings/scores from individual studies that used different response formats to allow comparison across studies, use of more robust parametric tests [19], and eventual pooling for meta-analysis [25-27]. The findings from this study suggest the possibility of bias with the pooling of data from surveys that used various response formats.

Limitations of this study

This research was restricted to five-point unipolar and bipolar scales and each respondent did not serve as his/her own control as in a cross-over design. Questionnaires with 5-7-point response scales are reported to be associated with more impressive response performance than when less or

more response steps are applied [8].

Although the clinical and socio-demographic characteristics of patients in both study groups in this research are similar; they differ markedly from the demographic structure of the general population. This could affect the generalisability of the findings of this research. However, no report on the socio-demographic characteristics of PHC users in Nigeria was available to compare with this study population.

The cross-sectional design used in this study could limit the use of causal interpretation. Further research using cross-over design may be needed to confirm or refute these hypotheses.

There is additional need to explore the effects of semantics and culture in questionnaire response and also take into cognizance the fact that observed response rates from questionnaire surveys are also indicative of the extent respondents are willing and able to complete the surveys.

Table 1: Patients' socio-demographic characteristic

Characteristics	Distribution		*Chi-square	P-value
	Bipolar (n = 153)	Unipolar (n = 152)		
Age – n (%)				
Less than 20	6 (4.1)	6 (4.0)		
20 - 29	84 (56.8)	77 (51.0)	0.63	0.43
30 – 39	40 (27.0)	48 (31.8)		
40 - 60	18 (12.2)	20 (13.2)		
Gender – n (%)				
Male	11 (7.3)	9 (7.0)	0.009	0.92
Female	139 (92.7)	119 (93.0)		
Marital status – n (%)				
Single	4 (2.7)	29 (19.2)	23.72	0.00
Married	138 (94.5)	122 (80.9)		
Widowed/divorced	4 (2.7)	0 (0.0)		
Level of schooling – n (%)				
None	0 (0.0)	0 (0.0)		
Primary	0 (0.0)	0 (0.0)		
Secondary	78 (52.7)	80 (53.0)	0.94	0.33
Intermediate	15 (10.1)	29 (19.2)		
First degree	47 (31.8)	36 (23.8)		
Higher degree	8 (5.4)	6 (4.0)		
Employment – n (%)				
Not working	22 (14.9)	16 (10.9)		
House wife	28 (18.9)	7 (4.8)		
Self-employed	43 (29.1)	73 (49.7)		
Private employed	11 (7.4)	15 (10.2)	0.82	0.36
Public employed	32 (21.6)	19 (12.9)		
Student	9 (6.1)	17 (11.6)		
Other	3 (2.0)	0 (0.0)		
Religion – n (%)				
Christian	139 (92.7)	138 (91.4)		
Moslem	2 (1.3)	4 (2.6)		
Traditional	5 (3.3)	5 (3.3)	0.22	0.64
Atheist	0 (0.0)	4 (2.6)		
Prefer not to disclose	4 (2.7)	0 (0.0)		
Consult at health Centre – n (%)				
Doctor	43 (31.9)	49 (32.7)		
Nurse	49 (36.3)	65 (43.3)	1.49	0.22
Community Health Worker	39 (28.9)	36 (24.0)		
Others	4 (3.0)	0 (0.0)		
Perceived health status (%, n)				
Poor	0 (0.0)	0 (0.0)		
Fair	13 (8.6)	9 (6.1)		
Good	41 (27.0)	47 (32.0)	0.08	0.78
Very good	48 (31.6)	47 (32.0)		
Excellent	50 (32.9)	44 (29.9)		
Duration using facility – n (%)				
Less than 1 month	17 (11.3)	36 (23.8)		
1 - < months	29 (19.2)	23 (15.2)		
3 months - < 1 year	43 (28.5)	33 (21.9)	1.58	0.21
1 year to < 3 years	26 (17.2)	19 (12.6)		
More than 3 years	36 (23.8)	40 (26.5)		
Payment modality – n (%)				
Free of charge	43 (30.7)	50 (34.0)		
Insurance/prepaid	9 (6.4)	0 (0.0)	0.54	0.46
Money from own pocket	66 (47.1)	59 (40.1)		
Money from other sources	22 (15.7)	38 (25.9		

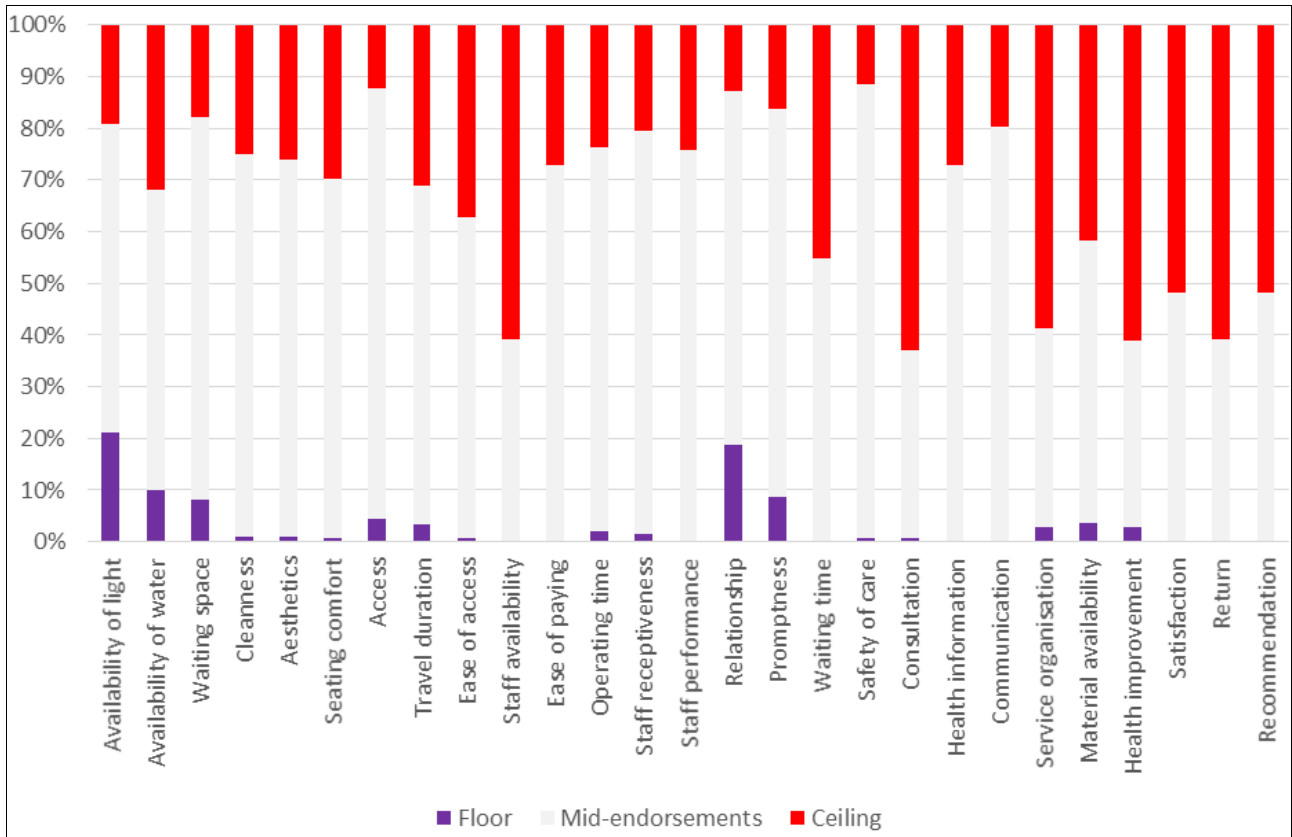*Linear by linear Chi-Square test

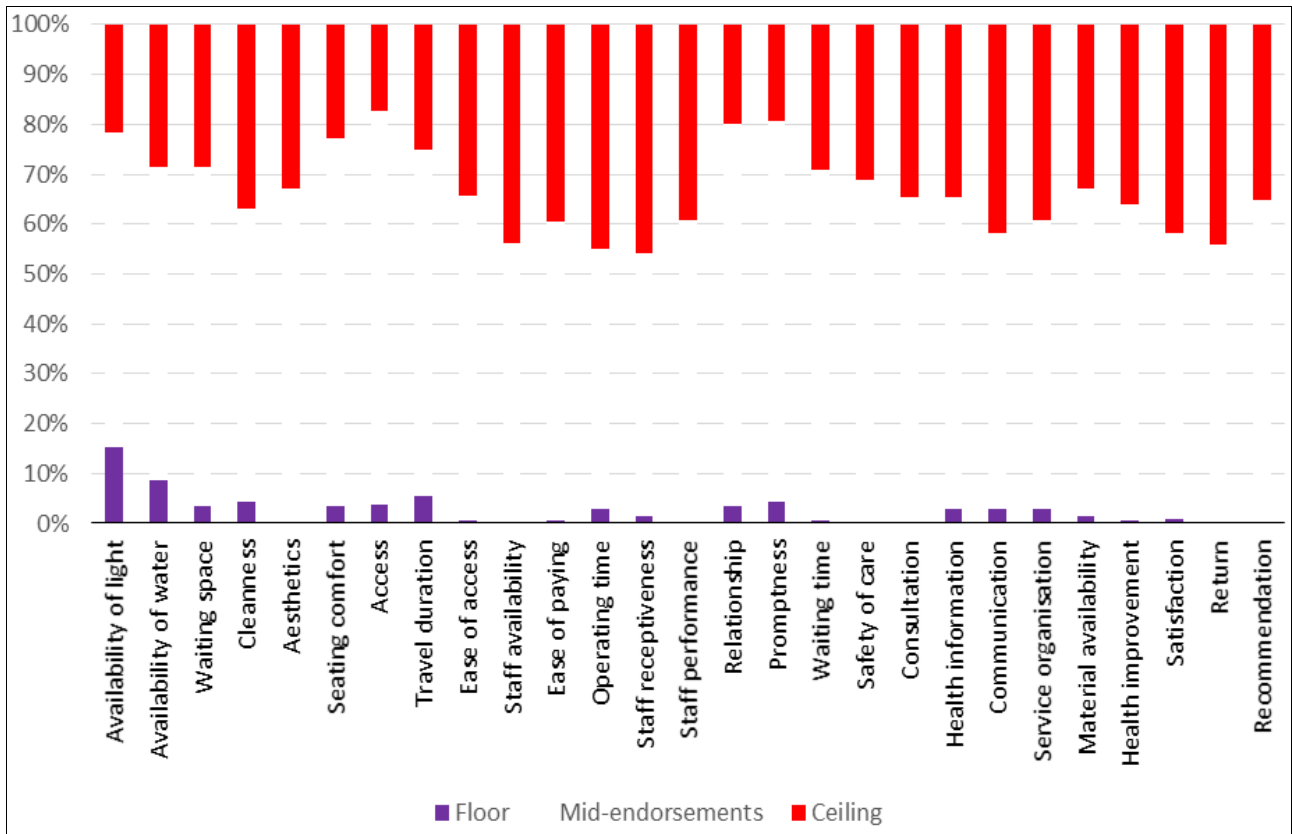**Fig 1:** Floor and ceiling effects of PES with unipolar response format

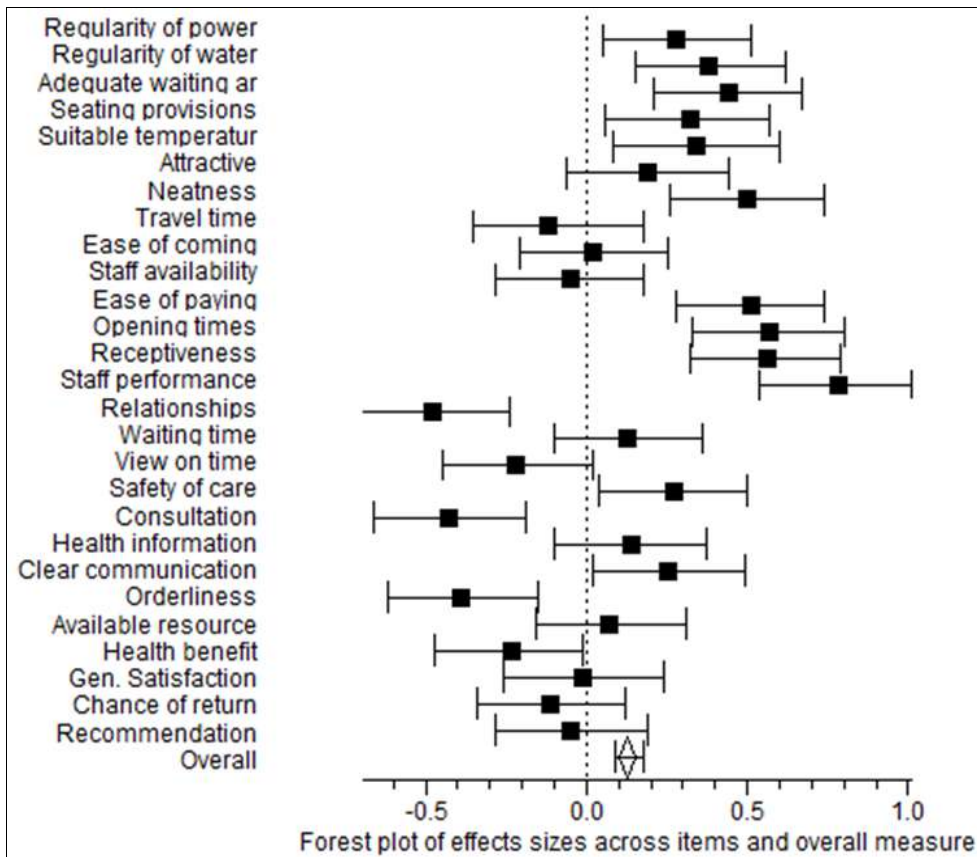**Fig 2:** Floor and ceiling effects of PES with bipolar response format

Fig 3: Effect of response format on item scores

Note: Black square boxes and tails show the estimate (standardized mean difference) and the 95% CIs of these estimates. The combined effect from all individual item mean difference gave an overall standardized mean difference between bipolar and unipolar response formats is 0.12 (95% CI: 0.08 – 0.17). This mild effect was statistically significant (p = 0.02).
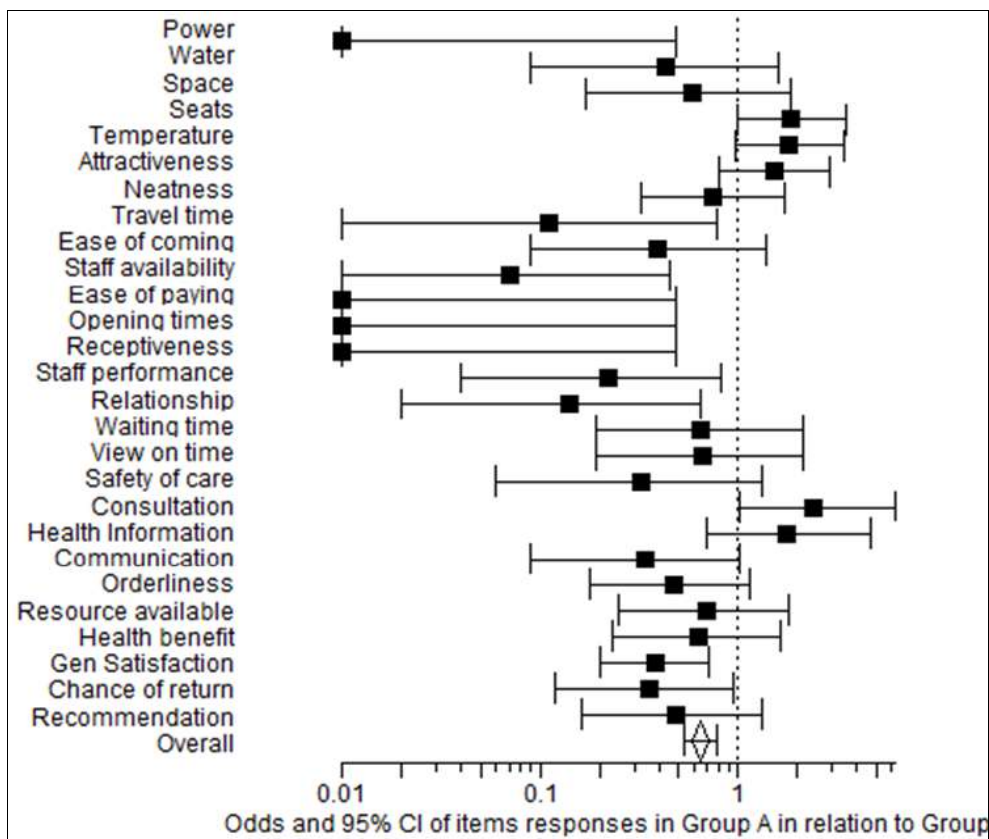

Fig 4: Effect of response format on item response

Note: Black square boxes and tails show the effect statistics (estimate) and the 95% CIs of these estimates of the Odds of item response in with bipolar compared to unipolar formats.

Table 2: Comparing response, score distribution and acceptability indices between response formats

Index	Response format		Standardized difference (95% CI)
	Bipolar	**Unipolar**	
Measure response rate (%)	95.0	94.4	[a]1.13 (0.43, 3.01)
Item response rate – % (range)	91.0 (75.2 – 94.1)	93.7 (75.7 – 100.0)	[a]0.65 (0.53, 0.79)***
Floor endorsement – %Mean (item range)	4.5 (0.0 - 28.1)	4.2 (0.0 – 46.1)	[a]1.06 (0.86, 1.32)
Ceiling endorsement – %Mean (item range)	28.2 (0.0 – 38.2)	30.6 (3.3 – 60.5)	[a]0.89 (0.81, 0.98) *
Skewness – Mean (item range)	0.9 (-2.0 - -0.06)	0.5 (-2.05 - -0.06)	-
Scale mean (SD)	4.04 (0.5)	3.91 (0.6)	-
Time of completion - Mean (SD), n	12.8 (5.5), 140	12.7 (5.5), 150	[d]0.01 (-0.22, 0.24)
Patients' grade for measure - Mean, (SD), n	5.6 (1.6), 133	5.5 (1.4), 132	[d]0.10 (-0.14, 0.35)
Scale score - Mean, (SD), n	4.04(0.9), 153	3.91(1.0), 152	[d]0.12 (0.08, 0.17) *

Odds ratio, d – Standardised mean difference, P values (*<0.05; **<0.01; ***<0.001)

Conclusion

Patients in this study were more likely to respond to items with unipolar response scale than with bipolar response scale. By contrast, mean item score was statistically higher in the bipolar variant. The high ceiling effect observed with both response formats is expected in evaluative instrument like this scale. However, validity and generalizability are enhanced with lower items or questionnaire non-response rates for discriminative and/or evaluative scale. The decision to trade higher scale score for validity may require stronger evidence from cross-over design.

References

1. Wensing M, Elwyn G. Research on patients' views in the evaluation and improvement of quality of care. BMJ Quality & Safety. 2002; 11(2):153-7.
2. Baumgartner H, Steenkamp JB. Response styles in marketing research: A cross-national investigation. Journal of marketing research. 2001; 38(2):143-56.
3. Donabedian A. Evaluating the quality of medical care. The Milbank Quarterly. 2005; 83(4):691-729.
4. Hinkin TR. A brief tutorial on the development of measures for use in survey questionnaires. Organizational research methods. 1998; 1(1):104-21.
5. Greenleaf EA. Improving rating scale measures by detecting and correcting bias components in some response styles. Journal of Marketing Research. 1992; 29(2):176.
6. Harzing AW. Response styles in cross-national survey research: A 26-country study. International Journal of Cross Cultural Management. 2006; 6(2):243-66.
7. Johnson T, Kulesa P, Cho YI, Shavitt S. The relation between culture and response styles: Evidence from 19 countries. Journal of Cross-cultural psychology. 2005; 36(2):264-77.
8. Streiner DL, Norman GR, Cairney J. Health measurement scales: a practical guide to their development and use. Oxford University Press, USA, 2015.
9. Fitzpatrick R, Davey C, Buxton MJ, Jones DR. Evaluating patient-based outcome measures for use in clinical trials. 2, 1-74.
10. Baltussen RM, Yé Y, Haddad S, Sauerborn RS. Perceived quality of care of primary health care services in Burkina Faso. Health policy and planning. 2002; 17(1):42-8.
11. Haddad S, Fournier P, Potvin L. Measuring lay people's perceptions of the quality of primary health care services in developing countries. Validation of a 20-item scale. International Journal for Quality in Health Care. 1998; 10(2):93-104.
12. Webster TR, Mantopoulos J, Jackson E, Cole-Lewis H, Kidane L, Kebede S *et al*. A brief questionnaire for assessing patient healthcare experiences in low-income settings. International Journal for Quality in Health Care. 2011; 23(3):258-68.
13. Naing L, Winn T, Rusli BN. Practical issues in calculating the sample size for prevalence studies. Archives of orofacial Sciences. 2006; 1:9-14.
14. Ogaji DS, Giles S, Daker-White G, Bower P. Development and validation of the patient evaluation scale (PES) for primary health care in Nigeria. Primary health care research & development. 2017; 18(2):161-82.
15. Lipsey MW, Wilson DB. Practical meta-analysis. Sage Publications, Inc, 2001.
16. Thalheimer W, Cook S. How to calculate effect sizes from published research: A simplified methodology. Work-Learning Research. 2002, 1-9.
17. Ogaji DS. Questionnaires for Patient Evaluation of Primary Health Care: A Systematic Review and Implications for the Nigerian Practice Setting. Journal of Community Medicine and Primary Health Care. 2017; 29(1):18-34.
18. Kuzon Jr WM, Urbanchek MG, McCabe S. The seven deadly sins of statistical analysis. Annals of plastic surgery. 1996; 37(3):265-72.
19. Norman G. Likert scales, levels of measurement and the laws of statistics. Advances in health sciences education. 2010; 15(5):625-32.
20. Traylor M. Ordinal and interval scaling. Journal of the Market Research Society. 1983; 25(4):297-303.
21. Acock AC, Martin JD. The Under measurement Controversy: Should Ordinal Data be Treated as Interval?. Sociology and Social Research. 1974; 58(4):427-33.
22. Spector PE. Summated rating scale construction: An introduction. Sage, 1992.
23. Nunnally JC, Bernstein IH. Psychological theory. New York, NY: MacGraw-Hill. 1994.
24. Bruce B, Fries J, Lingala B, Hussain YN, Krishnan E. Development and assessment of floor and ceiling items for the PROMIS physical function item bank. Arthritis research & therapy. 2013; 15(5):R144.
25. Hall JA, Dornan MC. Meta-analysis of satisfaction with medical care: description of research domain and analysis of overall satisfaction levels. Social science & medicine. 1988; 27(6):637-44.
26. Hall JA, Dornan MC. What patients like about their medical care and how often they are asked: a meta-

analysis of the satisfaction literature. Social science & medicine. 1988; 27(9):935-9.

27. Ogaji DS, Giles S, Daker-White G, Bower P. Systematic review of patients' views on the quality of primary health care in sub-Saharan Africa. SAGE open medicine. 2015, 30, 3:2050312115608338.