Open Access Indonesian Journal of Medical Reviews
Volume 6, Issue 6, 2026 | Article ID: 1002 | Pages: 685–702 | e-ISSN 2807-6257

OAIJMR

ORIGINAL ARTICLE

Item Performance and Internal Consistency of the KAP–SADARI Breast Self-Examination Questionnaire among Indonesian Women: A Pilot Validation Study

Erlyn Tusara Putri Harjanti[1*], Widyanti Soewoto[2]

[1] Surgery Residency Program, Department of General Surgery, Faculty of Medicine, Universitas Sebelas Maret / Dr. Moewardi General Hospital, Surakarta, Indonesia

[2] Department of Oncology Surgery, Dr. Moewardi General Hospital, Surakarta, Indonesia

ARTICLE INFO

Article history

Received: September 5, 2026
Revised: September 20, 2026
Accepted: October 18, 2026
Available Online: October 28, 2026
Published: December 5, 2026

Keywords

Breast neoplasms; Breast self-examination; Health knowledge, attitudes, practice; Psychometrics; Surveys and questionnaires

***Corresponding author**

Erlyn Tusara Putri Harjanti
E-mail: erlyn.tusara@gmail.com

DOI

https://doi.org/10.37275/oaijmr.v6i6.1002

Access license

CC BY-NC-SA 4.0: readers may share and adapt the work for non-commercial purposes with attribution, indication of changes, and the same license.

Copyright

© 2026 The Authors. Published by HM Publisher in collaboration with CMHC Research Center.

ABSTRACT

Background: Breast cancer is the most common cancer in Indonesia, and most women present at an advanced stage, so programs that teach breast self-examination (BSE) need brief instruments with sound measurement properties to judge their effect. This study examined one such instrument.

Objective: To assess item performance and internal consistency of the 13-item knowledge, attitude and practice questionnaire on BSE (KAP–SADARI).

Methods: In this cross-sectional pilot study, 30 literate Indonesian women completed the knowledge, attitude and practice domains online in July–August 2026. Uncorrected item–total correlations and Cronbach's α were analyzed with Fisher-z and Feldt intervals. Corrected item–rest correlations and α-if-item-deleted were identified exactly for the attitude domain. For knowledge and practice, integer programming bounded them over every dataset that matches the observed statistics.

Results: All uncorrected correlations exceeded the critical r of 0.361 (0.422–0.919), yet two independent attitude items would meet this criterion with probability 1.000. Practice item P3 had a corrected correlation of 0.117–0.182 in every admissible dataset, and deleting it raised α in every possible dataset. Knowledge item K3 fell below 0.361 in 97% of randomly sampled admissible marginal sets. Cronbach's α was 0.733 (95% confidence interval [CI] 0.545–0.859) for knowledge, 0.725 (0.422–0.869) for attitude and 0.758 (0.594–0.871) for practice. None exceeded 0.70 significantly (p = 0.204–0.408).

Conclusion: KAP–SADARI shows acceptable but imprecise internal consistency. It is not yet ready for classifying women or evaluating programs. P3, and probably K3, need revision. The attitude domain needs more items before confirmatory validation in at least 319 women.

1. Introduction

Female breast cancer accounted for an estimated 11.8% of all new cancer cases and 7.1% of cancer deaths worldwide in 2024. Only lung cancer was more common.[1] In Indonesia, GLOBOCAN 2024 national estimates attribute 93,236 new cases (19.7% of all cancers) and 38,012 deaths to breast cancer. It is the most common cancer in the country and the second cause of cancer death after lung cancer.[1] Late presentation defines this burden. The national surgical-oncology registry covers 4,959 patients, and 20.1% of them had stage IV disease at diagnosis. Lower education and residence outside Java and Bali independently predicted late stage.[2] In Yogyakarta, 43.3% of women delayed presentation for three months or longer. Diagnostic delay was associated with stage III–IV disease (odds ratio [OR] 3.03), and failure to recognize the symptom was the most common reason for delay.[3] A 2026 Indonesian path analysis found painless symptoms to be the strongest direct predictor of delayed presentation.[4] Malaysia shows the same pattern. In Sarawak, 46.5% of patients presented at stage III–IV,[5] and in Kelantan 19.8% delayed care for more than 90 days although three in four had found the lump themselves.[6] Across 81 countries, the share of women with distant metastases at diagnosis varies several-fold with income and health-system capacity.[7]

Breast self-examination (BSE; *pemeriksaan payudara sendiri*, SADARI, in Indonesia) costs nothing, and every woman can perform it. It is no longer advocated as a stand-alone screening test for reducing mortality.[8] It remains central to breast awareness. That means knowing how one's breasts normally look and feel, recognizing a change and seeking care promptly. A systematic review of Indonesian studies estimated that 43.14% of women practice BSE. It named knowledge, attitude, family support, family history and exposure to information as determinants.[9] In the nationally representative Indonesian Family Life Survey, participation in community health programs was associated with BSE (adjusted OR 1.73).[10] In a nationwide sample, knowledge was associated with breast cancer screening (adjusted OR 1.75).[11] Knowledge does not guarantee practice, however. In Surabaya, 72.5% of women intended to perform BSE, but only 2.9% had done so monthly in the previous year.[12] In Saudi Arabia, 79% had heard of BSE and 42.6% knew the technique,[13] and only 6% of web-surveyed women practiced BSE well.[14] An umbrella review of 110,622 women in low- and middle-income countries estimated a pooled BSE prevalence of 32.15%. Knowledge (OR 3.95) and attitude (OR 2.73) were the strongest correlates.[15] Ethiopia shows a similar gap: 28.9% of university students had good knowledge, but 17.7% practiced BSE.[16–19] The same gap has

been described in India,[20] Thailand,[21] the Philippines[22] and Malaysia.[23,24]

Because education is intended to change BSE behavior, the measurement properties of the outcome instrument determine what intervention studies can show. A 2026 meta-analysis of 24 studies reported an OR of 8.30 for BSE practice after education.[25] A cluster randomized trial in Ethiopia raised BSE practice from 33.3% to 59.9%,[26] hands-on training outperformed theory-only teaching in a Turkish trial,[27] and augmented-reality teaching improved knowledge among Indonesian adolescent girls.[28] Most of these studies use locally assembled knowledge, attitude and practice (KAP) questionnaires, and their measurement properties are reported incompletely. Recently validated instruments vary widely in quality. A Chinese KAP scale reached excellent structural fit with McDonald's ω of 0.70–0.92,[29] the knowledge domain of a Taiwanese KAP questionnaire had a Cronbach's α of only 0.37,[30] and a review of the Breast Cancer Awareness Measure found a high pooled α but often insufficient evidence on structural validity and test–retest reliability.[31]

KAP–SADARI was developed for Indonesian women from three published instruments. Each has its own limitation. The questionnaire of Masawa and Mboineki has an attitude domain that lengthens it. Its knowledge domain relies mainly on dichotomous agree/disagree responses.[32] Sayi's instrument covers knowledge and perceptions of breast cancer beyond BSE. It includes many knowledge items that add to respondent burden, and its open-ended responses make scoring subjective.[33] The questionnaire of Mengie et al. is concise, but its practice domain does not assess whether BSE is performed with the recommended technique.[34] KAP–SADARI was therefore designed to be brief and to grade each response by its agreement with recommended practice instead of as simply right or wrong. It also includes six technique items.

Item analysis in many Indonesian validation studies relies on the Pearson correlation between each item and a total score that contains the item. The coefficient is judged against the critical *r* for the sample size.[35,36] Because it includes the item's correlation with itself, it is inflated, most severely in short scales. The corrected item–rest correlation removes the overlap. It is standard in recent instrument development, including Indonesian work.[37,38] Methodological studies have shown three things. Item–score correlations are biased indicators of item quality.[39,40] Corrected item–total correlations and α need explicit standard errors.[41] Reliability coefficients behave differently with few items, small samples and dichotomous responses.[42–44] With 30 respondents, point estimates without intervals can make an instrument look adequate. The data may be equally compatible with poor performance.

This study evaluated the item performance and internal consistency of the 13-item KAP–SADARI questionnaire in a pilot sample of Indonesian women recruited online by investigators at Universitas Sebelas Maret and Dr. Moewardi General Hospital, Surakarta. The specific aims were (1) to describe the respondents and compare their exposure to BSE information with published surveys; (2) to estimate item–total correlations and Cronbach's α with confidence intervals and to test α against conventional thresholds; (3) to remove the part–whole bias from item analysis by estimating corrected item–rest correlations and α-if-item-deleted, exactly where the data identify them and as attainable ranges where they do not; and (4) to derive the sample size and scale length that a confirmatory validation study requires.

2. Methods

2.1. Study design and setting

This was a cross-sectional pilot study of the measurement properties of a newly adapted questionnaire, covering item performance and internal consistency. It is reported following the COSMIN reporting guideline for studies on measurement properties, with the sampling frame described according to STROBE items for cross-sectional studies.

After ethical clearance had been granted, data were collected online with Google Forms from late July to August 2026. The study was run by investigators at Universitas Sebelas Maret and Dr. Moewardi General Hospital, Surakarta, and the questionnaire was administered in Bahasa Indonesia.

2.2. Participants and eligibility

The target population was women in Indonesia; the accessible population was literate women in Indonesia. Women who could read and write Bahasa Indonesia were eligible. Respondents who left any questionnaire item unanswered were excluded. Convenience sampling continued until 30 complete responses had been obtained, the conventional size for the pilot phase of instrument testing. Only complete responses were retained and the questionnaire did not route respondents by prior BSE performance, so every respondent answered every item, including the practice items.

2.3. Variables and data collection

The instrument was developed by adapting and modifying the instruments of Masawa and Mboineki,[32] Sayi[33] and Mengie et al.[34] Development comprised item identification and selection, simplification of language, adjustment to the Indonesian context and construction of a scoring system, as shown in Figure 1. The final instrument has three scored domains: knowledge (five items, K1–K5), attitude (two items, A1–A2) and practice (six items, P1–P6). Table 1 lists the items, keyed responses and scores. Each response was weighted by its agreement with recommended BSE practice, with 50 points for the most appropriate response, 10 for a partially appropriate response and 0 for an inappropriate response. Item K5 consists of four yes/no sub-items on signs of abnormality, each worth 12.5 points. Domain maxima are therefore 250 (knowledge), 100 (attitude) and 300 (practice). A domain score of at least 75% of the maximum was classified as good; this threshold is close to the 70% cut-off used in a published BSE KAP survey[14] and was not calibrated against an external criterion. The English wording in Table 1 is a translation of the Bahasa Indonesia questionnaire, which is provided in full in Supplementary File 1. An identity section recorded age band, marital status, education and occupation. Unscored questions recorded prior receipt and main source of BSE information and prior BSE performance.

Table 1. Domain structure, items and scoring key of the KAP–SADARI questionnaire.

Domain (maximum; 'good')	Item	Content	Keyed response (50 points)	Other scores
Knowledge (250; ≥187.5)	K1	Purpose of BSE	To help detect breast lumps early	Recognize breast shape and size 10; others 0
	K2	Age to start BSE	At first menstruation	Before marriage, in pregnancy, before breastfeeding, at menopause 0
	K3	Frequency of BSE	Every month	Every day or every week 10; every year 0
	K4	Timing of BSE	One week after menstruation	One week before or during menstruation 0
	K5	Signs to look for (four sub-items)	Lump; abnormal nipple discharge; skin change: 'yes' (12.5 each)	Nipple inverted since childhood: 'no' 12.5
Attitude (100; ≥75)	A1	Regularity of BSE	Monthly	Daily, weekly, yearly, no fixed schedule 0
	A2	Action on finding an abnormality	Visit a clinic for further diagnosis	Traditional healer, ignore it, do not know 0
Practice (300; ≥225)	P1	Body position	Standing	Sitting or lying down 10
	P2	Palpating hand	Opposite hand	Same-side hand or one hand for both 0
	P3	Palpating fingers	Three middle fingers	Thumb and index finger, whole palm 0
	P4	Direction of movement	Circular or vertical	Triangular 0
	P5	Nipple-discharge check	Squeeze nipple with thumb and index finger	Observation 10; pressing with paper 0
	P6	Area examined beyond the breast	Armpit	Neck 10; back 0

Abbreviations: BSE, breast self-examination; KAP, knowledge, attitude and practice; SADARI, pemeriksaan payudara sendiri (breast self-examination). Notes: 'Good' denotes ≥75% of the domain maximum. Because each attitude item scores 0 or 50, a good attitude (≥75 of 100) requires the keyed response to both items. The questionnaire was administered in Bahasa Indonesia; the English wording is a translation.

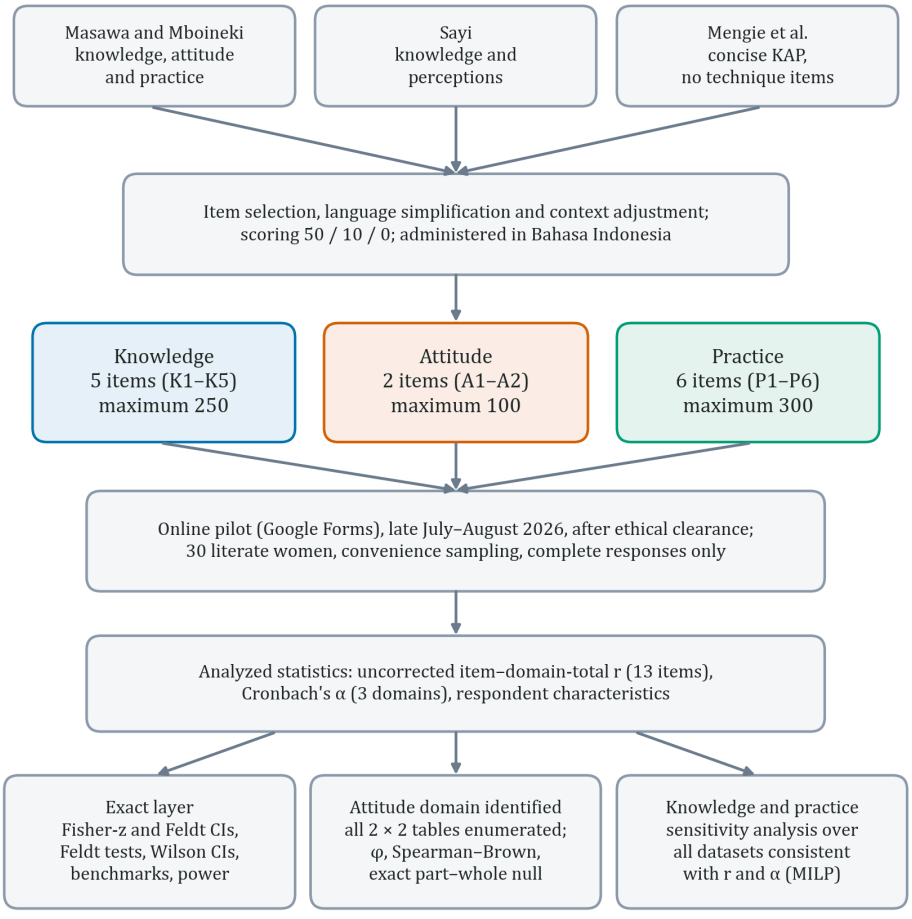


Figure 1. Development of the KAP–SADARI questionnaire and analysis workflow. Items were adapted from three published instruments into three scored domains, administered online to 30 women, and evaluated in three analytic layers: exact computations from the observed statistics, exact identification of the two-item attitude domain, and a sensitivity analysis over every dataset consistent with the observed item–total correlations and Cronbach's α.

2.4. Statistical analysis

With 30 respondents, the Wilson 95% confidence interval (CI) for a proportion of 50% has a half-width of ±16.8 percentage points, the Fisher-*z* interval for a correlation of 0.50 extends from 0.17 to 0.73, and the Feldt interval for an α of 0.73 in a five-item scale extends from 0.54 to 0.86. The study was therefore designed to estimate item and scale statistics with explicit imprecision and to inform the sample size of a confirmatory study, not to establish that any coefficient exceeds a threshold.

Respondent characteristics are reported as frequencies and percentages with Wilson score 95% CIs, where x is the count, n the sample size and $z = 1.96$:

$$\mathrm{CI} = \frac{x + \frac{z^2}{2} \pm z\sqrt{\frac{x(n - x)}{n} + \frac{z^2}{4}}}{n + z^2}$$

Exposure to BSE information was compared with the proportions of women who had heard of, or were aware of, BSE in five published surveys conducted online, in the community, in antenatal clinics and among university students[13,14,45–47] by two-sided exact binomial tests with Holm adjustment, and differences were expressed as Cohen's h. Because 'had received information' and 'had heard of BSE' are related but not identical constructs, these comparisons are descriptive:

$$h = 2 \arcsin\sqrt{p_1} - 2 \arcsin\sqrt{p_2}$$

Item validity was assessed in SPSS version 31 as the Pearson correlation between each item score and its domain total (r_{iT}), with the item included in the total, against the critical r of 0.361 for $n = 30$ at a two-sided α of 0.05 (a one-sided test would use 0.306). Each coefficient was tested exactly with a t test on 28 degrees of freedom, and its 95% CI was obtained by the Fisher z transformation:

$$z = \frac{1}{2}\ln\left(\frac{1 + r}{1 - r}\right), \quad 95\%\ \mathrm{CI} = \tanh\left(z \pm \frac{1.96}{\sqrt{n - 3}}\right)$$

Because the item contributes to the total, r_{iT} is positive even when the items of a domain are mutually independent. Under independence its expected value is $\sigma_i/\sqrt{(\Sigma\sigma_i^2)}$, where σ_i is the standard deviation of item i. For the attitude domain, this expectation and the probability that each item would exceed 0.361 under independence came exactly from the hypergeometric distribution. For knowledge and practice, each item column was permuted independently (1,000 permutations per dataset). The corrected item–rest correlation removes the part–whole overlap. It is the correlation between an item and the sum of the other items in its domain, where σ_T is the standard deviation of the domain total:

$$r_{i,T\text{-}i} = \frac{r_{iT}\sigma_T - \sigma_i}{\sqrt{\sigma_T{}^2 + \sigma_i{}^2 - 2r_{iT}\sigma_i\sigma_T}}$$

Corrected correlations were judged against the conventional threshold of 0.30, which was the decision criterion for item revision, and are also shown against the critical value of 0.361 for comparability with the uncorrected analysis. Cronbach's α for each domain of k items was computed as

$$\alpha = \frac{k}{k - 1}\left(1 - \frac{\sum_{i=1}^{k}\sigma_i{}^2}{\sigma_T{}^2}\right)$$

with α-if-item-deleted computed from the remaining $k - 1$ items. Feldt's 95% CI was used, with F quantiles on $n - 1$ and $(n - 1)(k - 1)$ degrees of freedom, together with one-sided Feldt tests of H_1: $\alpha > 0.70$ and H_1: $\alpha > 0.60$; because these procedures assume essentially τ-equivalent, approximately normal items, their coverage with coarse items and 30 respondents is approximate:

$$95\%\ \mathrm{CI} = [1 - (1 - \hat{\alpha})F_{0.975},\ 1 - (1 - \hat{\alpha})F_{0.025}],$$
$$F \sim F\big(n - 1, (n - 1)(k - 1)\big)$$

Because α for two items depends on their variance ratio, the attitude domain was also summarized by the inter-item φ coefficient and the Spearman–Brown coefficient, and the reliability expected with m parallel items was projected with the Spearman–Brown formula:

$$\rho_{\mathrm{SB}} = \frac{2r}{1 + r}, \quad \rho_m = \frac{m\rho}{1 + (m - 1)\rho}$$

For knowledge and practice, McDonald's ω for the unweighted sum score was estimated from a one-factor model fitted by unweighted least squares to the Pearson inter-item correlation matrix, the same basis as α; ordinal (polychoric) estimates would require the response matrix. The mean inter-item correlation was also reported.

The analyzed data comprised the 13 uncorrected item–total correlations, the three α coefficients and the respondent frequencies. Item-level responses were not part of the analyzed data, nor was their cross-classification with respondent characteristics. Statistics that depend on item-level responses were therefore estimated in a sensitivity analysis. This analysis covers the space of datasets that match the observed statistics, and no single dataset is treated as observed. An admissible dataset was any matrix of 30 respondents × k item scores, restricted to the scoring key in Table 1, that reproduced every observed r_{iT} and α of its domain to within ±0.0005. This is the rounding range of three-decimal reporting. Two identities link the observed statistics to the item standard deviations:

$$\sum_{i=1}^{k} r_{iT}\sigma_i = \sigma_T, \quad \sigma_T{}^2 = \frac{\sum_{i=1}^{k}\sigma_i{}^2}{1 - \alpha(k - 1)/k}$$

Given the item variances, the observed statistics fix the variance of the domain total and the covariance of each item with the total. Through the item–rest formula, they also fix the corrected item–rest correlation and α-if-item-deleted to within rounding. For the two binary attitude items, every possible 2 × 2 table was enumerated. A table and its 0↔50 mirror image can both reproduce the statistics. They share every coefficient and differ only in endorsement rates, so the table consistent with the information exposure of the sample was reported. Knowledge has 180 possible response patterns and practice 144. Each dataset was represented by the number of respondents with each pattern. Once the item marginal distributions are fixed, every observed statistic becomes a linear constraint on these counts. Mixed-integer linear programming with the HiGHS solver decided admissibility. Item marginal distributions were sampled uniformly over count compositions. They were screened by the two identities and by the integer lattice of attainable cross-products. Targeted searches toward the lowest and highest variance of each item supplemented the sample. In total, 1.16×10^9 knowledge and 2.02×10^{10} practice marginal sets were sampled. Of these, 3.60×10^6 and 84,640 passed the screens. About 196,000 and 84,640 integer programs were solved, and 223 and 116 distinct admissible marginal sets were found. Within each admissible set, integer programming gave the minimum and maximum of the proportion in the good category, the proportion at the domain maximum and the mean inter-item correlation. ω was computed in up to five admissible datasets per set. A search cannot guarantee that

every admissible marginal set is found. Ranges are therefore reported as the most extreme values found, that is, inner bounds of the attainable range. Percentages of admissible sets are computed over the sets found by random sampling only, because targeted searches over-represent extremes. They describe the random search and are not probabilities that a conclusion is true. Outer bounds came from a continuous relaxation. In it, the ratio $w_i = \sigma_i/\sigma_T$ was maximized and minimized subject to the two identities, the rounding ranges and a positive semidefinite covariance matrix. The corrected item–rest correlation and α-if-item-deleted depend on the observed statistics only through w_i. These limits were not exceeded in any of 40 random starts per bound; as results of a local solver they are numerical bounds, not certified global bounds. Sequential quadratic programming solved the relaxation from 40 random starts per bound, and every range found by integer programming lay inside its outer bound. Doubling the rounding range to ±0.001 moved no outer bound by more than 0.007. A conclusion was regarded as robust when it held in every admissible dataset found. It was regarded as further supported when it also held across the outer bound. When a range spanned a decision threshold, the quantity was reported as not identified by the data.

The sample size for a confirmatory study was derived from three quantities. The first was the exact power of the one-sided Feldt test of $\alpha > 0.70$ and $\alpha > 0.60$ at 80% power, assuming the population α equals the observed value. The second was the sample size at which the lower 95% limit would reach 0.70. The third was the sample size giving a CI half-width of 0.05. The same calculations were repeated for a planning value of $\alpha = 0.80$, the target for the revised domains (with the attitude domain lengthened to three items). The Spearman–Brown prophecy formula gave the number of parallel items needed for α of 0.80 and 0.90. Other analyses used Python 3.14 with NumPy 2.4.4 and SciPy 1.18.1. The HiGHS interface of SciPy handled integer programming, and its SLSQP routine produced the outer bounds. Permutation analyses used fixed random seeds, and the code is provided in Supplementary File 2. All tests were two-sided at $\alpha = 0.05$ unless stated otherwise. The *p* values are reported to three decimal places.

2.5. Ethical approval

This study was approved by the Health Research Ethics Committee of Dr. Moewardi General Hospital, Surakarta, Indonesia (ethical clearance no. 1.312/VII/HREC/2026, issued on 29 July 2026). All respondents provided informed consent on the first page of the online questionnaire before completing it, and the confidentiality of respondents' data was maintained by the investigators. The study was conducted in accordance with the Declaration of Helsinki.

3. Results

3.1. Respondents and exposure to information

The characteristics of the 30 respondents are detailed in Table 2 and shown in Figure 2. The most frequent age band was 21–30 years (30.0%; 95% CI 16.7–47.9%), followed by 31–40 years (26.7%), and the median respondent was in the 31–40-year band. Half were married (50.0%) and 43.3% had never married. Every respondent had completed at least senior high school, and 53.3% held a bachelor's degree. Private-sector employees (50.0%) and students (36.7%) predominated. Twenty-nine respondents (96.7%; 95% CI 83.3–99.4%) had received information about BSE; the main source was a doctor, nurse or midwife for 20 respondents (66.7% of all respondents and 69.0% of those informed), the internet for five (16.7%), and a family member or friend or a book for two each (6.7%).

Table 2. Characteristics of the 30 respondents and exposure to information on breast self-examination.

Characteristic	n	%	Wilson 95% CI (%)	Half-width (pp)
Age (years)				
≤20	4	13.3	5.3–29.7	12.2
21–30	9	30.0	16.7–47.9	15.6
31–40	8	26.7	14.2–44.4	15.1
41–50	5	16.7	7.3–33.6	13.1
> 50	4	13.3	5.3–29.7	12.2
Marital status				
Never married	13	43.3	27.4–60.8	16.7
Married	15	50.0	33.2–66.8	16.8
Divorced	1	3.3	0.6–16.7	8.0
Widowed	1	3.3	0.6–16.7	8.0
Education				
Senior high school	14	46.7	30.2–63.9	16.8
Bachelor's degree	16	53.3	36.1–69.8	16.8
Occupation				
Private-sector employee	15	50.0	33.2–66.8	16.8
Student	11	36.7	21.9–54.5	16.3
Civil servant/teacher/lecturer	3	10.0	3.5–25.6	11.1
Housewife	1	3.3	0.6–16.7	8.0
Ever received BSE information				
Yes	29	96.7	83.3–99.4	8.0
No	1	3.3	0.6–16.7	8.0
Source of BSE information				
Doctor/nurse/midwife	20	66.7	48.8–80.8	16.0
Internet	5	16.7	7.3–33.6	13.1

Characteristic	n	%	Wilson 95% CI (%)	Half-width (pp)
Family member/friend	2	6.7	1.8–21.3	9.7
Book	2	6.7	1.8–21.3	9.7
None	1	3.3	0.6–16.7	8.0

Abbreviations: BSE, breast self-examination; CI, confidence interval; pp, percentage points. Notes: Wilson score intervals. With n = 30 the half-width at a proportion of 50% is ±16.8 pp; 93 and 381 respondents are needed for ±10 and ±5 pp. 'Source of BSE information' gives the single main source; 'None' is the respondent who had not received information. Of the 29 informed respondents, 20 (69.0%) named a doctor, nurse or midwife.

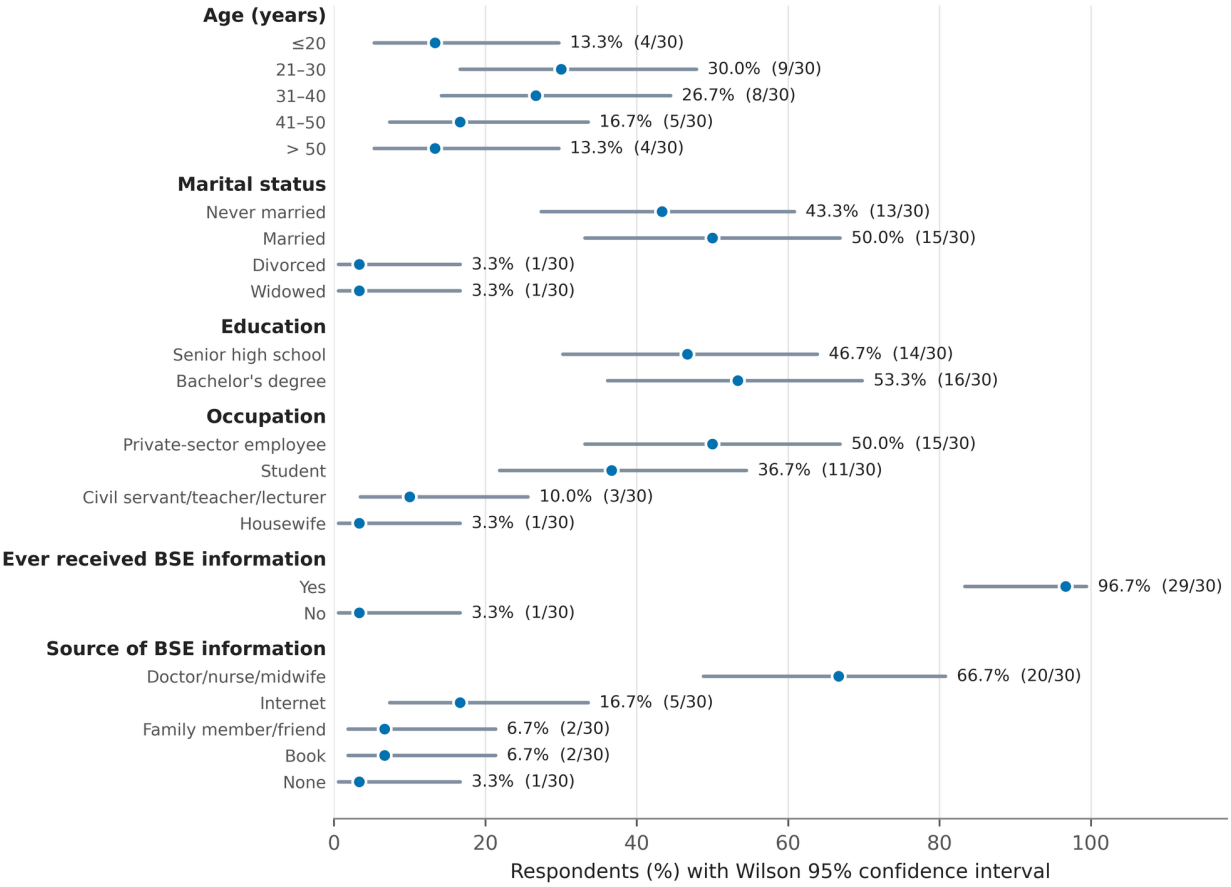


Figure 2. Characteristics of the 30 respondents. Points show proportions and horizontal bars Wilson 95% confidence intervals; labels give the percentage and count.

The proportion of respondents who had received BSE information was higher than the proportion of women who had heard of BSE in four of five published surveys after Holm adjustment (Table 3 and Figure 3). The differences were medium to large. Cohen's *h* was 0.57 against a Saudi web-based survey (79.6%; Holm-adjusted $p = 0.041$) and 0.58 against a Saudi online survey (79.0%; $p = 0.037$). It was 0.73 against university students in three Arab countries (72.7%; $p = 0.006$) and 0.99 against Bangladeshi students (60.5%; $p < 0.001$). The difference from Ghanaian antenatal attendees (84.0%) was not significant ($p = 0.076$; $h = 0.46$).

Table 3. Exposure to breast self-examination information compared with published surveys.

Comparator population	n	Reference proportion (%)	Exact binomial p	Holm-adjusted p	Cohen's h
Saudi Arabia, online survey[13]	930	79.0	0.012	0.037	0.58
Saudi Arabia, web-based community survey[14]	397	79.6	0.020	0.041	0.57
Egypt, Jordan and Syria, university students[45]	1,361	72.7	0.002	0.006	0.73
Ghana, antenatal clinics[46]	536	84.0	0.076	0.076	0.46
Bangladesh, university students[47]	400	60.5	<0.001	<0.001	0.99

Abbreviations: BSE, breast self-examination; CI, confidence interval. Notes: Present study: 29 of 30 respondents (96.7%; Wilson 95% CI 83.3–99.4%) had received information on BSE. Each row tests the observed proportion against the published proportion of women who had heard of, or were aware of, BSE (two-sided exact binomial test); Holm adjustment across the five comparisons. Cohen's h of 0.2, 0.5 and 0.8 indicates small, medium and large differences.

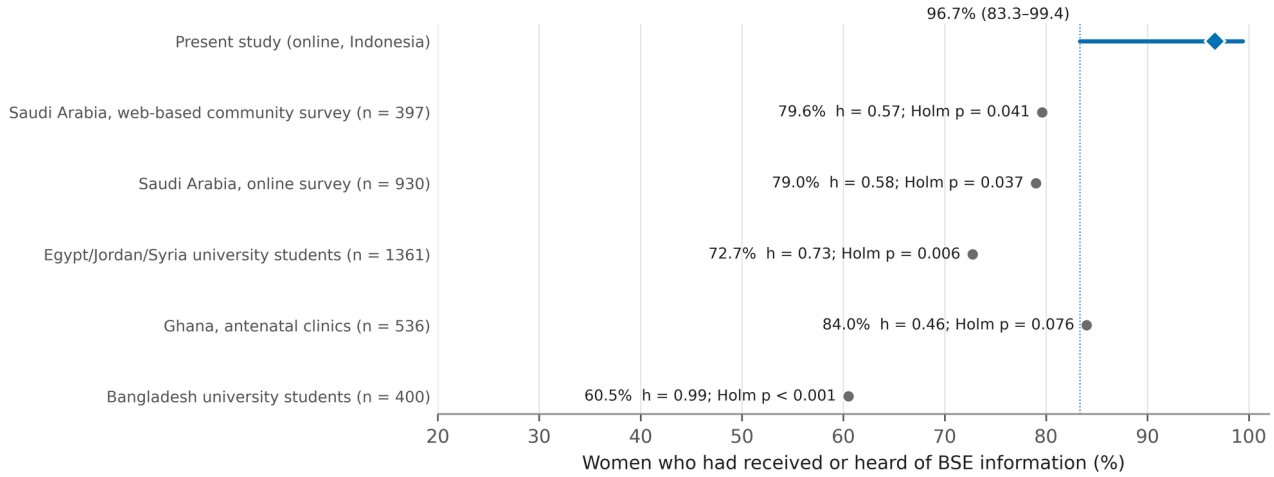


Figure 3. Exposure to breast self-examination information compared with published surveys. The diamond and bar show the proportion of respondents in this study who had received BSE information with its Wilson 95% CI; gray points show the proportions of women who had heard of or were aware of BSE in five published surveys, labeled with Cohen's h and the Holm-adjusted exact binomial p value.

3.2. Uncorrected item–total correlations

All 13 uncorrected item–total correlations exceeded the critical value of 0.361 (exact $p \leq 0.020$; Table 4 and Figure 4). Coefficients ranged from 0.530 (K3) to 0.917 (K5) in the knowledge domain. They were 0.919 and 0.860 for the attitude items and ranged from 0.422 (P3) to 0.823 (P6) in the practice domain. The lower confidence limits of two items fell below the critical value: K3 (95% CI 0.210–0.748) and P3 (0.073–0.679). The criterion itself had little power to separate items. Suppose the items of a domain were mutually independent. Across admissible datasets the expected uncorrected correlation would then range from 0.103 to 0.595 for knowledge items and from 0.225 to 0.536 for practice items. Individual items would exceed 0.361 by chance alone with probabilities of 0.08–0.97 (knowledge) and 0.20–0.91 (practice). All items of a domain would do so together with probability 0.03–0.16 and 0.02–0.04, respectively.

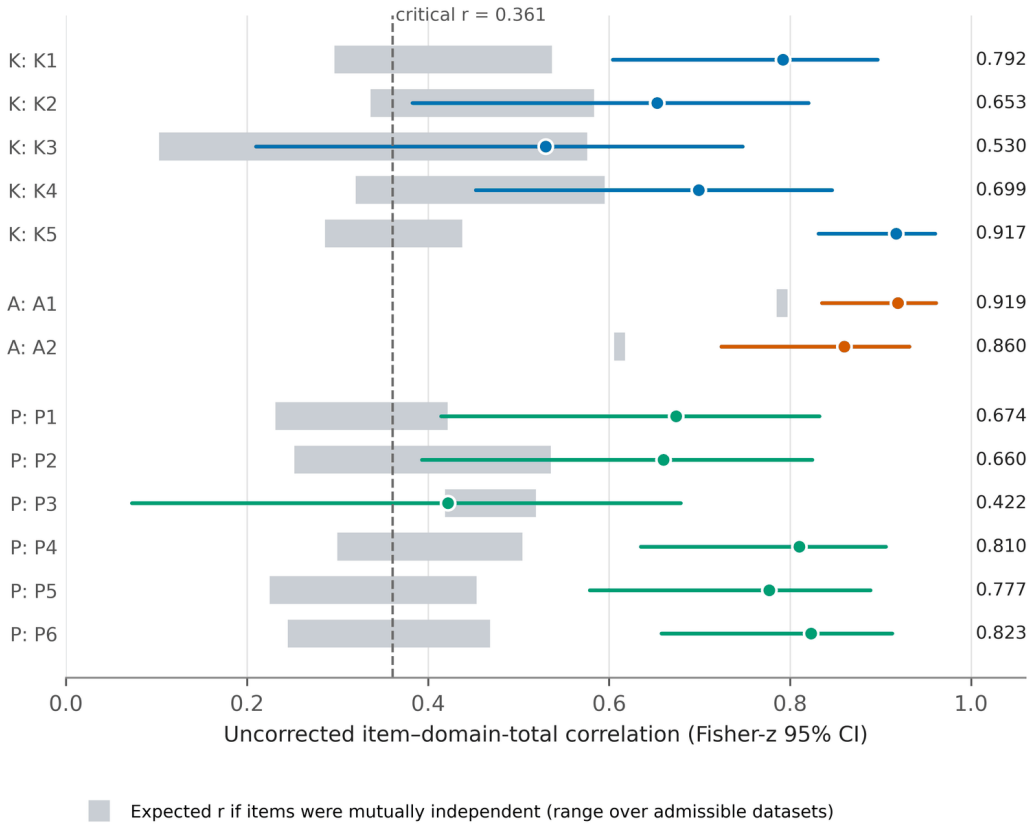


Figure 4. Uncorrected item–domain-total correlations with Fisher-z 95% confidence intervals. The dashed line marks the critical r of 0.361 used for the uncorrected validity criterion. Grey bands show the correlation expected for each item if all items of its domain were mutually independent (range over admissible datasets; exact value for attitude), which overlaps or exceeds the critical value for every item.

Table 4. Uncorrected item–domain-total correlations, their confidence intervals and their expected values under mutual independence of the items.

Domain	Item	r (uncorrected)	95% CI (Fisher z)	t (df = 28)	p	Expected r if items independent	P(r > 0.361 | independence)
Knowledge	K1	0.792	0.604–0.896	6.86	<0.001	0.297–0.537	0.34–0.92
	K2	0.653	0.383–0.820	4.56	<0.001	0.337–0.583	0.46–0.96
	K3	0.530	0.210–0.748	3.31	0.003	0.103–0.576	0.08–0.95
	K4	0.699	0.453–0.846	5.17	<0.001	0.320–0.595	0.39–0.97
	K5	0.917	0.831–0.960	12.16	<0.001	0.286–0.438	0.34–0.73
Attitude	A1	0.919	0.835–0.961	12.33	<0.001	0.791	1.000[a]
	A2	0.860	0.724–0.932	8.92	<0.001	0.612	1.000[a]
Practice	P1	0.674	0.414–0.832	4.83	<0.001	0.231–0.422	0.25–0.68
	P2	0.660	0.393–0.824	4.65	<0.001	0.252–0.536	0.27–0.91
	P3	0.422	0.073–0.679	2.46	0.020	0.419–0.519	0.68–0.89
	P4	0.810	0.635–0.906	7.31	<0.001	0.300–0.504	0.35–0.85
	P5	0.777	0.579–0.889	6.53	<0.001	0.225–0.454	0.20–0.77
	P6	0.823	0.658–0.913	7.67	<0.001	0.245–0.469	0.25–0.80

Abbreviations: CI, confidence interval. Notes: p values are exact two-sided t tests on 28 degrees of freedom. The expected uncorrected r under independence equals $\sigma_i/\sqrt{(\Sigma\sigma_j^2)}$; for knowledge and practice it is shown as the range over all admissible datasets, and the last column gives the permutation probability (1,000 permutations per dataset) that the item exceeds the critical r when every item is permuted independently. With equal item variances the expectation is $1/\sqrt{k}$: 0.71, 0.45 and 0.41 for k = 2, 5 and 6. [a] Exact hypergeometric probability that both attitude items exceed 0.361 when they are independent. Probability that all items of a domain exceed 0.361 under independence: knowledge 0.03–0.16, practice 0.02–0.04, attitude 1.000.

3.3. Attitude domain

Of the 5,456 possible cross-classifications of the two binary attitude items (all 2 × 2 tables of 30 respondents), only one table and its mirror image reproduced r_{iT} = 0.919, r_{iT} = 0.860 and α = 0.725 to three decimals (Table 5). In the table in which most respondents chose a clinic visit, 19 respondents (63.3%; 95% CI 45.5–78.1%) gave the keyed response to both items. They therefore had a good attitude. Six chose a clinic visit without reporting regular monthly BSE, and five gave neither keyed response. None reported monthly BSE without choosing a clinic visit. The mirror-image table would imply that only 16.7% of respondents would visit a clinic after finding a breast abnormality, although almost all had received BSE information. The keyed response was given by 63.3% for A1 and 83.3% for A2. In a two-item domain, the inter-item φ equals the corrected item–rest correlation of each item. It was 0.588 (95% CI 0.289–0.782; Fisher exact p = 0.003), and the Spearman–Brown coefficient was 0.740. Had the two items been independent with the same endorsement rates, both would still have exceeded the critical r in every possible table (probability 1.000). An α of 0.70 or more would have occurred with probability 0.003 (Figure 5).

Table 5. Attitude domain: item cross-classification identified from the observed item–total correlations and α, with derived coefficients.

Quantity	Value	95% CI / test
A1 keyed (monthly BSE) and A2 keyed (visit a clinic)	19 (63.3%)	
A1 not keyed, A2 keyed	6 (20.0%)	
A1 keyed, A2 not keyed	0 (0.0%)	
Neither keyed	5 (16.7%)	
A1 keyed response, %	63.3	45.5–78.1
A2 keyed response, %	83.3	66.4–92.7
Good attitude (both keyed), %	63.3	45.5–78.1
Attitude score, mean ± SD (of 100)	73.3 ± 38.8	
Inter-item φ (= corrected item–rest r of each item)	0.588	0.289–0.782
Fisher exact test of association	OR = ∞	p = 0.003
Cronbach's α (= KR-20)	0.725	0.42–0.87
Spearman–Brown coefficient	0.740	
Parallel items needed for α ≥ 0.80 / ≥ 0.90	3 / 7	

Abbreviations: BSE, breast self-examination; CI, confidence interval; OR, odds ratio; SD, standard deviation; KR-20, Kuder–Richardson formula 20. Notes: Of the 5,456 possible 2 × 2 tables for two 0/50 items in 30 respondents, only this table and its mirror image (0↔50 on both items: 5, 6, 0 and 19 respondents; A1 36.7%, A2 16.7%, good attitude 16.7%) reproduce r = 0.919, r = 0.860 and α = 0.725 to three decimals; all coefficients in the lower part of the table are identical for the two. The empty cell (no respondent endorsed regular BSE while choosing an option other than a clinic visit) makes the odds ratio infinite. φ CI by Fisher z.

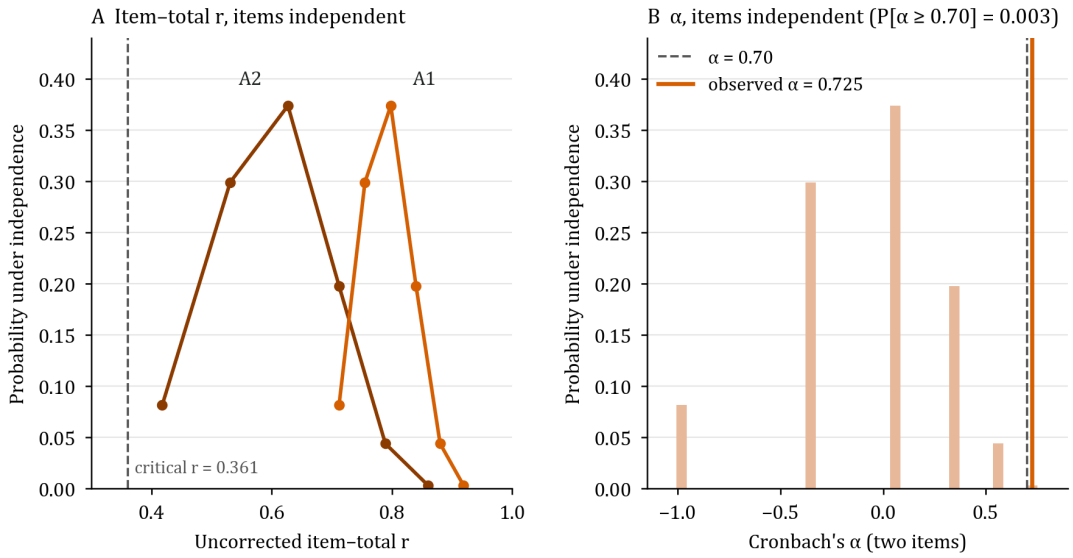


Figure 5. Exact behavior of the two-item attitude domain when the items are independent. Panel A gives the hypergeometric probability of each uncorrected item–total correlation for items with the identified endorsement rates (63.3% and 83.3%); both correlations exceed the critical value 0.361 in every possible table. Panel B gives the corresponding distribution of Cronbach's α; the observed α of 0.725 lies in the upper tail.

3.4. Corrected item statistics in the knowledge and practice domains

Admissible datasets were found for 223 sets of knowledge item marginals (180 from random sampling) and 116 sets of practice item marginals (73 from random sampling). Table 6 and Figures 6 and 7 show the corrected item statistics across these datasets. In the knowledge domain, the corrected item–rest correlation of K3 (frequency of BSE) ranged from 0.185–0.480 (outer bound 0.167–0.522). It was below 0.361 in 97% and below 0.30 in 82% of randomly sampled admissible sets. Without K3, α ranged from 0.729–0.838, above the domain α of 0.733 in 97% of sets. The observed statistics therefore do not identify whether K3 meets the corrected criterion, although most admissible datasets place it below. K2 (age to start BSE) had a corrected correlation of 0.344–0.499, below 0.361 in 2% of sets, and deleting it raised α in 40% of sets. K1 (0.590–0.702), K4 (0.404–0.567) and K5 (0.847–0.878) exceeded both thresholds in every admissible dataset found. Deleting K4 raised α in 3 of 223 admissible sets (1% of the randomly sampled sets).

For the practice domain, the corrected item–rest correlation of P3 (palpating fingers) ranged from 0.117–0.182. It was below 0.30 in every admissible dataset found and could not exceed 0.422 in any dataset (outer bound). Deleting P3 raised α from 0.758 to 0.803–0.846 in every admissible dataset found. The outer bound of 0.765–0.856 shows that α without P3 exceeds the domain α in every possible dataset. P1 (body position; 0.492–0.585) and P2 (palpating hand; 0.408–0.560) performed moderately. P4 (direction of movement; 0.652–0.730), P5 (nipple-discharge check; 0.623–0.714) and P6 (axillary examination; 0.688–0.765) were the strongest items. Deleting P2 raised α in 1 of 116 admissible sets (none of the randomly sampled sets), and deleting any other practice item except P3 lowered α in every admissible dataset found.

Table 6. Corrected item statistics for the knowledge and practice domains: ranges over admissible datasets and outer bounds over all possible datasets.

Item	Item–rest r: range found (median)	Item–rest r: outer bound	Below 0.30 / 0.361 (% of sets)	α if deleted: range found	α if deleted: outer bound	α rises if deleted (% of sets)	Keyed response (%): range
Knowledge (α = 0.733; 223 admissible marginal sets, 180 random)							
K1	0.590–0.702 (0.638)	0.465–0.739	0 / 0	0.628–0.648	0.628–0.796	0	6.7–76.7
K2	0.344–0.499 (0.396)	0.286–0.615	0 / 2	0.693–0.775	0.692–0.840	40	10.0–86.7
K3	0.185–0.480 (0.273)	0.167–0.522	82 / 97	0.729–0.838	0.728–0.857	97	0.0–80.0
K4	0.404–0.567 (0.462)	0.339–0.653	0 / 0	0.675–0.750	0.675–0.829	1	6.7–86.7
K5	0.847–0.878 (0.871)	0.715–0.880	0 / 0	0.541–0.606	0.524–0.700	0	0.0–56.7
Practice (α = 0.758; 116 admissible marginal sets, 73 random)							
P1	0.492–0.585 (0.520)	0.347–0.647	0 / 0	0.717–0.725	0.716–0.832	0	10.0–86.7
P2	0.408–0.560 (0.446)	0.331–0.635	0 / 0	0.721–0.764	0.720–0.834	0	6.7–76.7
P3	0.117–0.182 (0.148)	0.105–0.422	100 / 100	0.803–0.846	0.765–0.856	100	20.0–76.7
P4	0.652–0.730 (0.674)	0.535–0.771	0 / 0	0.668–0.683	0.667–0.794	0	10.0–80.0
P5	0.623–0.714 (0.672)	0.483–0.739	0 / 0	0.682–0.709	0.682–0.806	0	6.7–60.0
P6	0.688–0.765 (0.738)	0.556–0.785	0 / 0	0.662–0.693	0.661–0.789	0	6.7–83.3

Abbreviations: α, Cronbach's alpha; K1–K5, knowledge items; P1–P6, practice items; r, correlation coefficient. Notes: An admissible dataset is a 30 × k matrix of item scores, restricted to the scoring key, that reproduces every observed uncorrected item–total r and α to three decimals. Corrected item–rest r and α-if-item-deleted depend on the observed statistics only through the item's share of the total-score standard deviation, so one dataset per admissible set of item marginals suffices. 'Range found' gives the most extreme values over all admissible sets found by random and targeted integer-programming searches (inner bounds); percentages are over the sets found by random sampling. 'Outer bound' comes from a continuous relaxation (positive semidefinite covariance, rounding ranges) and was not exceeded in any of 40 random starts of a local solver (a numerical bound, not a certified one); for P3, α if deleted exceeds the domain α across the whole outer bound.

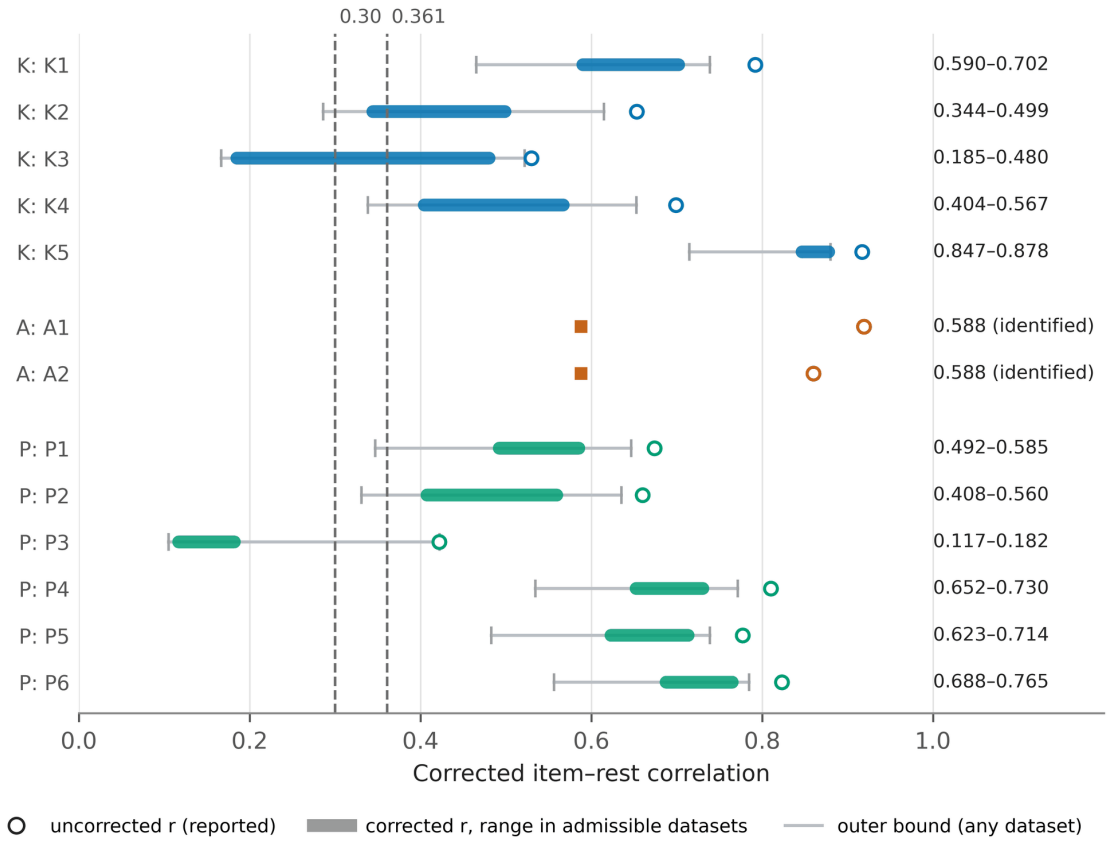


Figure 6. Corrected item–rest correlations. Colored bars show the range of the corrected correlation over every admissible dataset found for knowledge and practice items, and thin gray lines its outer bound over all possible datasets; squares show the identified value for the attitude items (equal to the inter-item φ); open circles show the uncorrected correlations. Dashed lines mark the conventional threshold of 0.30 and the critical value of 0.361.

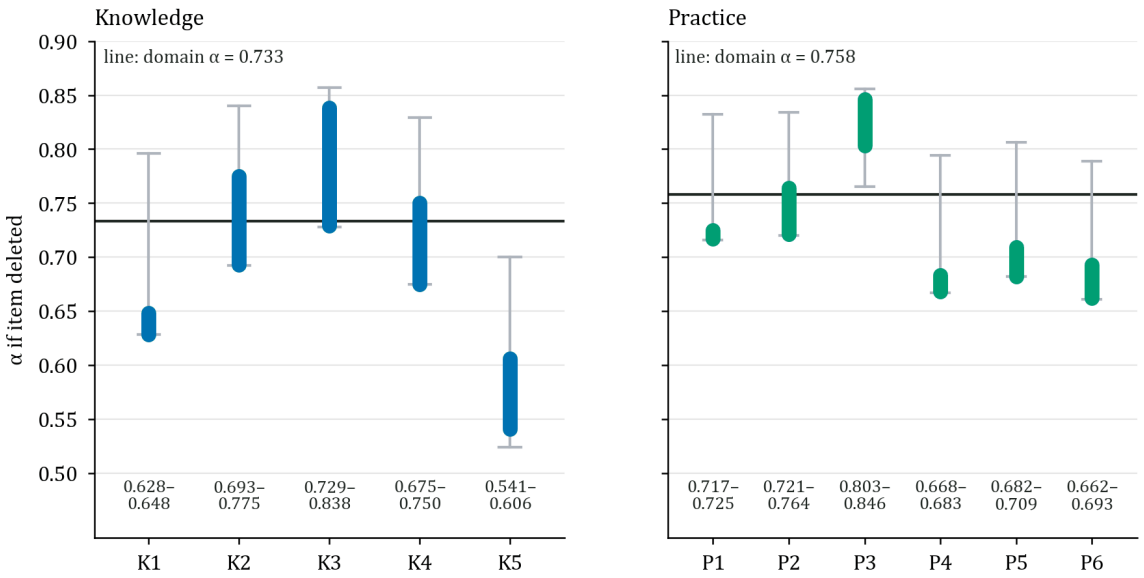


Figure 7. Cronbach's α if each item is deleted, as ranges over admissible datasets (colored bars) and outer bounds over all possible datasets (thin gray lines), relative to the domain α (horizontal line). An item whose range lies wholly above the line lowers internal consistency in every admissible dataset; for P3 this holds across the whole outer bound.

3.5. Internal consistency

Cronbach's α was 0.733 (95% CI 0.545–0.859) for knowledge, 0.725 (0.422–0.869) for attitude and 0.758 (0.594–0.871) for practice (Table 7 and Figure 8). No domain had an α significantly greater than 0.70 (Feldt one-sided $p = 0.323$, $p = 0.408$ and $p = 0.204$, respectively). Only practice had an α significantly greater than 0.60 ($p = 0.029$; knowledge $p = 0.069$; attitude $p = 0.159$). At the observed values, a sample of 30 gave only 11.1%, 7.8% and 19.8% power to show α > 0.70. McDonald's ω ranged from 0.743 to 0.782 for knowledge and from 0.761 to 0.794 for practice across admissible datasets, close to α. The mean inter-item correlation was narrowly bounded at 0.395–0.423 for knowledge and 0.379–0.390 for practice.

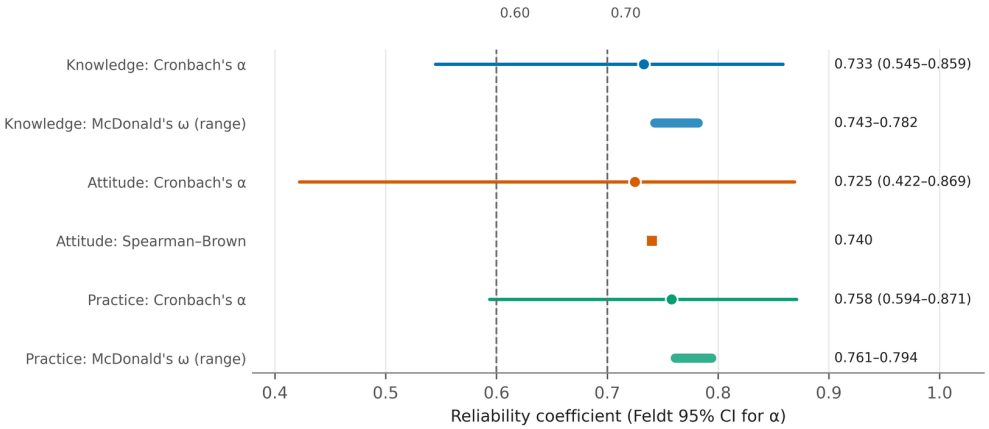


Figure 8. Internal consistency of the three domains. Circles and thin bars give Cronbach's α with Feldt 95% confidence intervals; thick bars give the range of McDonald's ω over admissible datasets (knowledge, practice); the square gives the Spearman–Brown coefficient of the two-item attitude domain. Dashed lines mark 0.60 and 0.70.

Table 7. Internal consistency of the three domains with interval estimates and one-sided tests against conventional thresholds.

Domain	k	Cronbach's α (Feldt 95% CI)	Test α > 0.70: F; p	Test α > 0.60: F; p	Power vs 0.70 at n = 30	McDonald's ω: range	Mean inter-item r: range
Knowledge	5	0.733 (0.545–0.859)	1.124; 0.323	1.498; 0.069	11.1%	0.743–0.782	0.395–0.423
Attitude	2	0.725 (0.422–0.869)	1.091; 0.408	1.455; 0.159	7.8%	SB 0.740	0.588 (identified)
Practice	6	0.758 (0.594–0.871)	1.240; 0.204	1.653; 0.029	19.8%	0.761–0.794	0.379–0.390

Abbreviations: CI, confidence interval; SB, Spearman–Brown coefficient for the two-item attitude domain. Notes: Feldt $F = (1 - \alpha_0)/(1 - \hat{\alpha})$ with df = 29 and 29(k − 1); one-sided p for H_1: $\alpha > \alpha_0$. Power is for the one-sided Feldt test at α = 0.05 assuming the population α equals the observed value. ω (one-factor model, unweighted sum score) and the mean inter-item r depend on the joint item responses and are shown as ranges over admissible datasets.

3.6. Score distributions

Table 8 shows which quantities the observed statistics could not identify for knowledge and practice. These are the quantities that depend on how responses are distributed instead of on how items covary. Across admissible datasets the mean knowledge score ranged from 67.1 to 162.5 of 250 points. The mean practice score ranged from 84.0 to 197.3 of 300 points. The proportion with good knowledge ranged from 3.3% to 53.3%, and the proportion with good practice from 6.7% to 56.7%. For the attitude domain, the identified table gives a mean score of 73.3 ± 38.8 of 100 points. The good category coincides with the maximum score.

Table 8. Score-distribution quantities and their identification from the observed statistics.

Quantity	Knowledge	Practice	Attitude	Identification
Domain score, mean (points)	67.1–162.5	84.0–197.3	73.3	Fixed by item marginals; marginals not identified
Domain score, SD (points)	61.6–81.7	76.3–90.1	38.8	Fixed by item marginals and α
Respondents in the 'good' category (%)	3.3–53.3	6.7–56.7	63.3	Not identified for knowledge and practice
Respondents at the domain maximum (%)	0.0–26.7	0.0–16.7	63.3	Not identified for knowledge and practice
Mean inter-item r	0.395–0.423	0.379–0.390	0.588	Narrowly bounded

Abbreviations: SD, standard deviation. Notes: Ranges are over admissible datasets (for % good, % at maximum and mean inter-item r: integer-programming minimum and maximum within each admissible set of item marginals, over the sets examined). For the attitude domain the good category coincides with the maximum score (both items keyed).

3.7. Requirements for a confirmatory study

Table 9 and Figure 9 give the sample sizes required for a confirmatory study. Showing with 80% power that α exceeds 0.70 at the observed values would require 319 respondents for practice, 1132 for knowledge and 3269 for attitude. Showing α > 0.60 would require 59, 94 and 178. Suppose the revised domains reach α = 0.80. Then 80% power would require only 90 (practice), 94 (knowledge) and 113 respondents (a three-item attitude domain), and a CI half-width of 0.05 would require 150, 156 and 189. At the observed values, a CI half-width of 0.05 would require 218, 277 and 473 respondents. The Spearman–Brown projections in Figure 10 show that, at the observed inter-item φ, the attitude domain would need 3 parallel items to reach α ≥ 0.80 (projected 0.811) and 7 to reach 0.90.

Table 9. Sample size and scale length required by a confirmatory validation study.

Domain	Observed α	n: 80% power, α > 0.70	n: 80% power, α > 0.60	n: lower 95% limit ≥ 0.70	n: CI half-width ≤ 0.05	If revised α = 0.80: n for 80% / 90% power; ±0.05	Parallel items for 0.80 / 0.90
Knowledge	0.733	1132	94	684	277	94 / 131; 156	6 / 13
Attitude	0.725	3269	178	2032	473	113 / 158; 189 (k = 3)	3 / 7
Practice	0.758	319	59	187	218	90 / 126; 150	7 / 15

Abbreviations: α, Cronbach's alpha; CI, confidence interval; k, number of items. Notes: Power from the exact Feldt F distribution (one-sided, α = 0.05) assuming the population α equals the observed value; precision from the Feldt interval at the observed α. The planning scenario assumes the revised domains reach α = 0.80 (attitude with one added item, k = 3). Items needed: Spearman–Brown prophecy for parallel items at the observed inter-item correlation (attitude φ; midpoint of the admissible mean inter-item r for knowledge and practice).

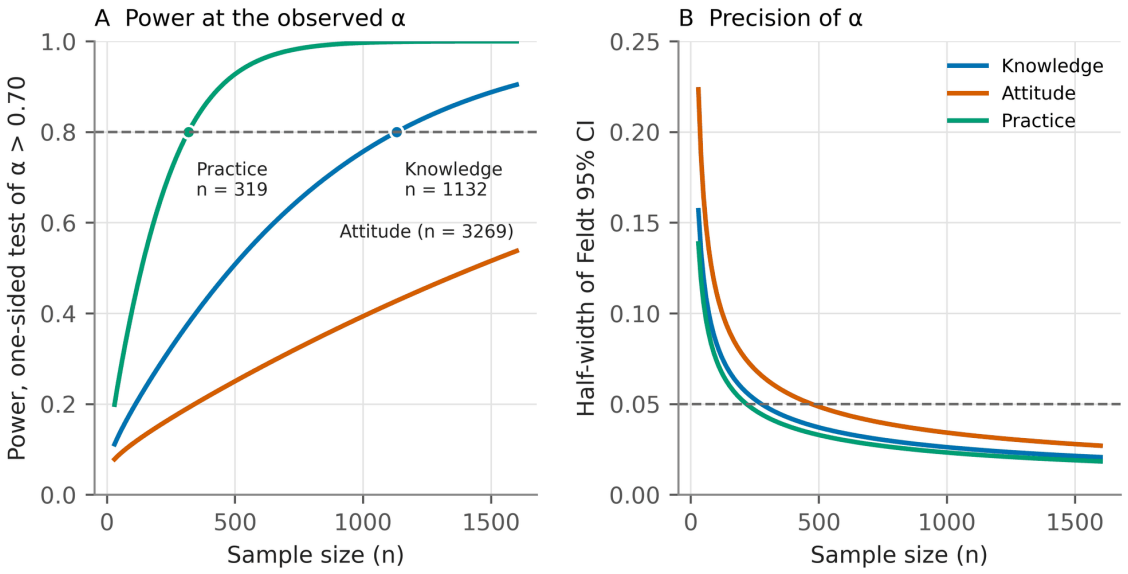


Figure 9. Sample size for a confirmatory validation study. Panel A shows the power of the one-sided Feldt test of α > 0.70 as a function of sample size, assuming the population α equals the observed value; points mark 80% power. Panel B shows the half-width of the Feldt 95% confidence interval; the dashed line marks ±0.05.

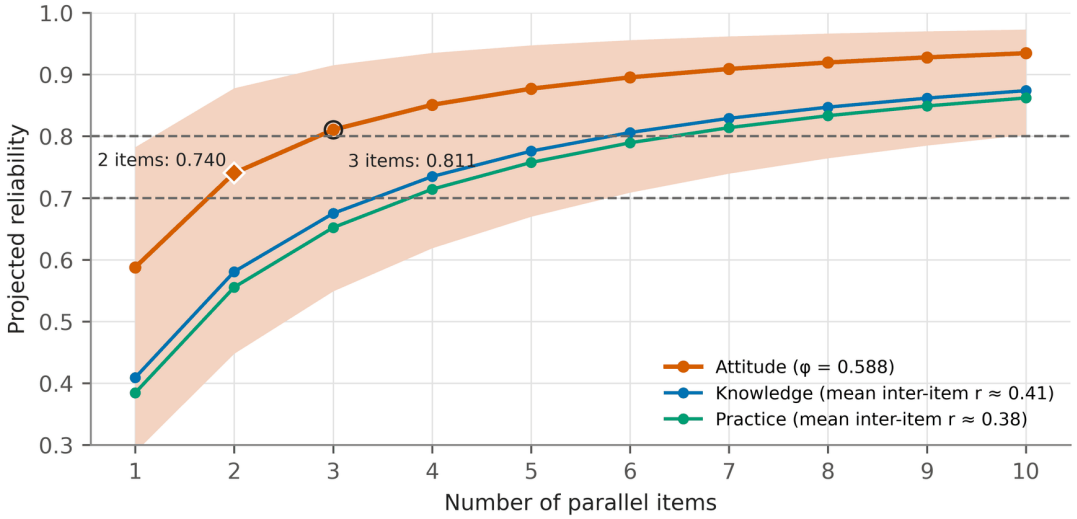


Figure 10. Projected reliability by number of parallel items (Spearman–Brown prophecy). The attitude curve uses the identified inter-item φ with its 95% CI (shaded); knowledge and practice curves use the midpoint of the admissible mean inter-item correlation. Dashed lines mark 0.70 and 0.80.

4. Discussion

4.1. Principal findings

In this pilot evaluation of the 13-item KAP–SADARI questionnaire among women recruited online by investigators at Universitas Sebelas Maret and Dr. Moewardi General Hospital, Surakarta, all domains had Cronbach's α in the acceptable range (0.725–0.758), and every item exceeded the conventional uncorrected validity criterion. Three findings qualify this picture. First, the uncorrected criterion carried little information. Two independent attitude items would satisfy it with certainty, and individual knowledge and practice items would satisfy it by chance alone with probabilities of up to 0.97. Second, once the part–whole overlap was removed, P3 (palpating fingers) had a corrected item–rest correlation below 0.30 in every admissible dataset found, and deleting it raised the practice α in every possible dataset. K3 (frequency of BSE) fell below the critical value in most, but not all, admissible datasets. Third, the confidence intervals of all three α coefficients extended well below 0.70, so a sample of 30 could not show that any domain reaches the conventional threshold. The instrument is therefore promising but not yet validated. The pilot data identify which items to revise and how large the confirmatory study must be.

4.2. Uncorrected and corrected item statistics

The uncorrected item–total correlation remains widespread in Indonesian validation work. A validation of the Beliefs about Medicine Questionnaire in South Sulawesi judged uncorrected item–total correlations against an *r*-table threshold.[35] An Indonesian validation of the five-item Brief Symptom Rating Scale reported item validity coefficients of 0.634–0.781 alongside α of 0.770.[36] The SADAR-PrEP instrument, developed in six Indonesian regions, took a different route. It reported corrected item–total correlations of 0.538–0.676 together with known-groups effect sizes,[37] and a Malaysian breast cancer screening health-literacy questionnaire reported corrected item–total correlations of at least 0.71.[38] The difference is not cosmetic. When items are unrelated, an item's uncorrected correlation with a total that

contains it approximates its share of the total variance. That is about $1/\sqrt{2} = 0.71$ for two items and 0.41–0.45 for five or six items of similar variance. All of these values exceed 0.361. Metsämuuronen has shown that item–score correlations distort item discrimination and reliability estimates,[39,40] and van der Ark derived standard errors for corrected item–total correlations because their sampling variability is large.[41] For short domains such as those of KAP–SADARI, corrected coefficients with intervals are the only item statistics that carry real information.

4.3. The two weakest items

The content and scoring of the two weakest items suggest likely explanations. For K3, whose statistical status is not identified, the content argument carries the decision to revise. Item K3 credits monthly BSE with 50 points. It awards 10 points for daily or weekly examination, so respondents with an inaccurate belief about frequency still earn partial credit. The ordering of its responses may not follow the knowledge continuum that the other items measure. Frequency may also be the BSE fact that varies most across lay sources. That can dilute its association with knowledge of purpose, timing and warning signs. Item P3 asks which fingers are used for palpation. Many women do not know this technical detail even when they examine their breasts. In Saudi Arabia, 79% had heard of BSE but only 42.6% knew the correct technique,[13] and in three Arab countries 41.92% of students did not know how to perform BSE.[45] The practice items ask what the respondent does. Women who have never performed BSE therefore answer them hypothetically. The practice domain then partly measures knowledge of technique, which can lower the association of the most technical item with the rest. The unscored question on prior BSE performance should route non-performers past the practice items. Alternatively, the domain should be reframed explicitly as knowledge of technique.

4.4. Reliability relative to published instruments

Against published instruments, the reliability of KAP–SADARI is in the expected range for scales of its length. An Indonesian diabetes knowledge test reported α of 0.73 and ω of 0.72,[48] the Iranian Breast Cancer Perception Scale reported α of 0.68 and ω of 0.70,[49] and the Malay Champion Health Belief Model Scale reported subscale α of 0.59–0.87.[50] Longer instruments reach higher values. The Turkish Breast Cancer Awareness Scale had a total α of 0.907,[51] a 36-item Indian KAP scale 0.87,[52] and the Breast Cancer Awareness Measure a pooled α of 0.89 across adaptations.[31] Precision differs more than level does. In a validation among 372 Iranian women, the 95% CI of ω spanned only 0.66–0.74;[49] with 30 respondents, the lower 95% limits of the KAP–SADARI α coefficients lie between 0.42 and 0.59. Simulation studies find that ω and α give similar estimates in unidimensional data,[42,43] as they did here, and that interval estimation, whether frequentist or Bayesian, is essential at small samples.[53,54]

Recent validations of breast cancer screening-belief, fear and awareness measures also show which property is most often missing. An Armenian health-belief survey administered by telephone reported test–retest correlations of 0.72–0.97 with α of 0.75–0.94.[55] The Greek Breast Cancer Screening Beliefs Questionnaire reached α of 0.814,[56] and the Chinese MARA awareness questionnaire combined α above 0.75 with retest correlations above 0.65.[57] A Bangla breast cancer fear scale administered online reached α of 0.939 but a test–retest coefficient of only 0.53.[58] High internal consistency in an online sample does not guarantee temporal stability. Short subscales can perform poorly within otherwise reliable instruments. A Turkish perceived-barriers scale had a total α of 0.87 but subscale α as low as 0.34.[59] For the Champion Breast Cancer Fear Scale, ω (0.877) and α (0.876) were almost identical.[60] For the Breast Cancer Screening Beliefs Questionnaire, an exploratory structural equation model fitted better than a confirmatory factor model.[61] BSE-specific transtheoretical scales recruited online were evaluated with factor analysis, item–total correlations and test–retest reliability together.[62] The coefficients of KAP–SADARI sit within this range, but its evidence base lacks the structural and temporal properties that these studies report.

4.5. The attitude and knowledge domains

The attitude domain illustrates the specific problem of two-item scales. Its items were clearly associated, with a φ of 0.588 and a Spearman–Brown coefficient of 0.740, so the domain is not noise. Two-item measures can be reliable when the items are near-parallel: the two-item Connor–Davidson Resilience Scale exceeded 0.72 in a national sample of 5,023 adults,[63] and a two-item empathy measure reached α of 0.919 with an inter-item correlation of 0.852.[64] The KAP–SADARI attitude items, however, differ in difficulty. Every respondent who reported regular monthly BSE also chose a clinic visit, but six who chose a clinic visit did not report regular BSE, a cumulative pattern in which the clinic item is the easier of the two. This structure caps the attainable correlation and makes α sensitive to endorsement rates. A third and fourth item from the constructs most associated with BSE, such as perceived self-efficacy or barriers,[65,66] would raise projected reliability above 0.80 and broaden the content of the domain beyond two behavioral intentions.

The knowledge domain combines distinct facts. Its internal consistency should be read with that in mind. Knowing the purpose of BSE, its timing and the warning signs of breast cancer need not be strongly correlated. The four sub-items of K5 act as a small testlet, which contributes four correlated responses to one item. Models designed for dichotomous and ordinal knowledge items serve such content better than α does. Model-based reliability coefficients outperformed α for dichotomous items in simulation.[44] The Japanese Cancer Knowledge Scale used a two-parameter item response model to select eight items with discrimination parameters of 1.13–1.86.[67] A confirmatory study should therefore fit an item response model to the knowledge domain and treat the sub-items of K5 explicitly. Relying on α alone would not be enough.

4.6. Scoring and reporting of domains

The three domains should also be scored and reported separately. Summing knowledge, attitude and practice into a single KAP score would create a multidimensional total for which α and ω-total overstate the reliability of any general factor, as shown in simulations of bifactor structures,[68] and for which hierarchical coefficients would be needed to justify a total score.[69] The different levels and correlates of knowledge, attitude and practice in the BSE literature[14,15] argue for domain scores as the primary outcomes.

4.7. Sample and recruitment

The respondents were unusually well informed. Exposure to BSE information (96.7%) exceeded that reported in Saudi online and community samples, among Arab and Bangladeshi university students, and was comparable only to antenatal attendees in Ghana, who are routinely counseled.[13,14,45–47] All respondents had completed senior high school. In the Indonesian national registry, lower education predicted late-stage presentation,[2] so the women at greatest risk may have been the least represented. Restricted variability in a

homogeneous, well-informed sample attenuates item correlations and α, so the coefficients observed here may underestimate rather than overestimate those in a more heterogeneous population; conversely, items answered correctly by almost all respondents here may discriminate quite differently among less informed women. Ceiling effects on knowledge items cannot be excluded because their distributions are not identified by the observed statistics. Online recruitment adds two further concerns: survey mode can shift reported levels of health-related constructs even when measurement is invariant,[70] and web surveys recruited through social media are vulnerable to fraudulent responses, which removed most submissions in one international study.[71] Future administrations should include verification questions, completion-time screening and recruitment through health facilities and community programs.

4.8. Information sources and implementation

Health workers were the main source of information for 66.7% of respondents, which matters for implementation. In a multinational survey of 5,435 women in the Middle East and North Africa, information from health professionals was associated with greater breast cancer awareness. Reliance on mass-media channels was associated with lower awareness.[72] Among Bangladeshi adolescents, information from health workers was associated with BSE practice (OR 9.62).[73] Social media was the leading source in Saudi Arabia (51.4%)[14] and among Arab students (38%).[45] Yet only 8% of YouTube videos on BSE were rated excellent for quality.[74] In Indonesia, participation in community health programs was associated with BSE practice.[10] Primary-care and community programs are therefore a plausible setting for both BSE education and the use of a brief evaluation instrument.

In the regional context, a brief Indonesian instrument has a clear role. The pooled Indonesian BSE prevalence of 43.14%[9] and the 20.1% of registry patients presenting with stage IV disease[2] call for education programs whose effects can be measured reliably. Recent Indonesian validations of instruments for dengue prevention,[75,76] financial toxicity in breast cancer,[77] cancer quality of life,[78] screen exposure in children[79] and paternal breastfeeding self-efficacy[80] show that confirmatory factor analysis, ω, test–retest intraclass correlation and known-groups validity are attainable standards in Indonesian samples of 171 to 462 respondents. In Malaysia, a randomized trial of mHealth education and navigation achieved clinical breast examination uptake of 45.8%, compared with 3.5% in the control arm,[81] and in Yogyakarta augmented-reality teaching produced high BSE knowledge in 71.4% of girls compared with 28.6% after a lecture.[28] Trials of this kind need outcome instruments that discriminate between women and detect change. The revised KAP–SADARI should be built to demonstrate that property.

Published education studies show what responsiveness requires. In a Nigerian pre–post study, mean knowledge, attitude and practice scores rose from 1.58 to 4.31, 2.37 to 4.80 and 1.97 to 5.81, respectively;[82] a comparison of podcast and lecture teaching used a knowledge tool with α of 0.76 together with observed practice scores;[83] a simulation-based trial among Afghan refugee women measured BSE skill as well as awareness;[84] and in an Iranian trial comparing face-to-face and WhatsApp education, knowledge correlated with practice only at $r = 0.30$ one month after face-to-face teaching.[85] The modest knowledge–practice correlation supports separate domains, and the use of observed technique as an outcome in these trials suggests an external criterion against which the KAP–SADARI practice domain could be validated.

4.9. Classification rule and content validity

The classification rule needs the same scrutiny. The good category is defined as 75% of the domain maximum, and it could not be evaluated. The proportions of respondents classified as having good knowledge or practice varied from a few percent to more than half across datasets that match the observed statistics. A cut-off chosen by convention can therefore yield almost any prevalence of 'good' knowledge in a given sample. The instrument should not be used to estimate the prevalence of good knowledge, attitude or practice, or to classify individual women, until the cut-off has been anchored by standard-setting methods or by known-groups and criterion data such as observed BSE technique.

Content validity rests on adaptation from published instruments and has not been confirmed by an expert panel. An expert panel should now review the scoring key. Guidance documents differ on the age at which to start BSE, the positions to use during palpation and whether to squeeze the nipple, and their advice has changed over time. Three keyed responses depend on these choices (K2, P1 and P5). A panel of breast surgeons, oncology nurses, midwives and health educators could document that the keyed responses reflect current Indonesian recommendations. It would report item-level and scale-level content validity indices, as was done for the Malay Champion Health Belief Model Scale (item content validity index ≥0.83)[50] and for a Ukrainian screening questionnaire.[86] Cognitive interviews would show whether women interpret the response options as intended.

4.10. Research agenda

The findings point to a clear research agenda. The next phase should combine an expert-panel review of all 13 items with cognitive interviews. It should revise K3 by removing partial credit for over-frequent examination, revise or reframe P3, and add at least one attitude item. A field test of at least 319 women should follow. That number is needed to show with 80% power that the practice α exceeds 0.70, and 473 would be preferable to estimate every α within ±0.05. The sample should be stratified by education, residence and age and recruited partly through primary-care facilities. The study should test the three-domain structure by confirmatory factor analysis and fit item response models to the knowledge items. It should estimate test–retest reliability at two weeks in a subsample large enough for a precise intraclass correlation.[87] It should also examine known-groups validity by comparing women who do and do not perform BSE and women with and without health-worker information. Responsiveness should then be assessed within a randomized comparison of hands-on and lecture-based education, a design already used to evaluate BSE teaching.[27,88] If the revisions bring every domain to an α of about 0.80, a sample of 189 would estimate each coefficient within ±0.05. A field test of at least 319 women therefore covers both the pessimistic and the planning scenario. Depositing the item-level responses with the confirmatory study would allow every bound reported here to be replaced by an estimate.

4.11. Strengths and limitations

This study has several strengths. It reports every coefficient with an interval and tests reliability against explicit thresholds rather than relying on point estimates. It removes the part–whole bias from item analysis and shows, by exact calculation, how little the conventional uncorrected criterion can reveal in short domains. Without distributional assumptions, it separates the quantities that the observed statistics identify (corrected item statistics, the attitude cross-classification and the mean inter-item correlation) from those

they do not (score distributions), so that the weakest item is flagged by a conclusion that holds across every dataset consistent with the observations and the second by the proportion of admissible datasets in which it fails. Finally, it converts the pilot results into the sample size and scale length that a confirmatory study needs.

The study also has limitations. The sample of 30 women recruited online by convenience sampling was small, highly educated and almost uniformly informed about BSE, which limits generalizability and may attenuate correlations. Because item-level responses were not part of the analyzed data, corrected item statistics for knowledge and practice are reported as ranges over admissible datasets found by search; these are inner bounds, and the attainable range may be wider, up to the outer bounds, which were themselves obtained by numerical optimization. Feldt intervals and Pearson-based ω are approximate for coarse, categorical items. Score distributions, the proportions in the good category, between-domain correlations and known-groups comparisons could not be estimated. Only item performance and internal consistency were examined; structural validity, test–retest reliability, criterion validity and responsiveness remain untested, and content validity rests on adaptation rather than a formal expert-panel index. Practice items were answered by respondents regardless of whether they had ever performed BSE. These results should therefore be regarded as a pilot evaluation that guides revision of the instrument.

5. Conclusion

The 13-item KAP–SADARI questionnaire showed acceptable but imprecise internal consistency in a pilot of 30 Indonesian women. Cronbach's α was 0.733, 0.725 and 0.758 for knowledge, attitude and practice, and none could be shown to exceed 0.70. The conventional uncorrected item–total criterion told us little in these short domains. After correction, the palpating-fingers item P3 performed poorly in every dataset that reproduces the observed statistics. The frequency item K3 probably does as well. The two attitude items are clearly associated ($\varphi = 0.588$) but too few for a reliable domain. P3 and K3 should be revised, at least one attitude item added and the keyed responses reviewed by an expert panel before the instrument is used to classify women or to evaluate BSE education in Indonesian primary care and community programs. A confirmatory study of at least 319 women, stratified by education and residence, should then establish structural validity, test–retest reliability, known-groups validity and responsiveness. It should use corrected item statistics and interval estimates throughout.

Declarations

Ethics Approval and Consent to Participate

This study was approved by the Health Research Ethics Committee of Dr. Moewardi General Hospital, Surakarta, Indonesia (ethical clearance no. 1.312/VII/HREC/2026). All respondents provided informed consent on the first page of the online questionnaire. The study was conducted in accordance with the Declaration of Helsinki.

Consent for Publication

Not applicable; no individual data are reported.

Availability of Data and Materials

The data generated and analyzed in this study are available from the corresponding author upon reasonable academic request.

Competing Interests

The authors declare that they have no competing interests.

Funding

None. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Authors' Contributions

E.T.P.H.: Conceptualization, Methodology (instrument development), Investigation (data collection), Formal analysis, Writing – original draft. W.S.: Conceptualization, Supervision, Writing – review & editing. All authors read and approved the final manuscript.

Acknowledgments

None.

Generative Artificial Intelligence Disclosure

No generative artificial intelligence tools were used in designing the study, collecting or analyzing the data, interpreting the findings, or preparing this manuscript.

Preprint and Related Manuscripts

This manuscript has not been posted as a preprint.

6. References

1. Sung H, Filho AM, Laversanne M, et al. Global cancer statistics 2024: GLOBOCAN estimates of incidence and mortality worldwide for 34 cancers in 186 countries. CA Cancer J Clin. 2026;76(4):e70090. doi:10.3322/caac.70090
2. Setyawan IB, Kurnia D, Setiaji K, et al. Sociodemographic disparities associated with advanced stages and distant metastatic breast cancers at diagnosis in Indonesia: a cross-sectional study. Ann Med Surg (Lond). 2023;85(9):4211–4217. doi:10.1097/MS9.0000000000001030
3. Hutajulu SH, Prabandari YS, Bintoro BS, et al. Delays in the presentation and diagnosis of women with breast cancer in Yogyakarta, Indonesia: A retrospective observational study. PLoS One. 2022;17(1):e0262468. doi:10.1371/journal.pone.0262468
4. Wiranata JA, Hutajulu SH, Astari YK, et al. Direct and Indirect Associations of Sociodemographic Factors and Patient-Perceived Barriers With Delayed Breast Cancer Presentation: A Cross-Sectional Path Analysis. World J Oncol. 2026;17(4):454–462. doi:10.14740/wjon2754
5. Lim MSH, Tan SSN, Wan Maharuddin IB, et al. Breast cancer survival rates in Sarawak, Malaysia: a central referral centre study. BMC Cancer. 2026;26(1):712. doi:10.1186/s12885-026-16065-4
6. Soh JY, Yahya MM, Bachok N, et al. Factors associated with delay in seeking care for breast symptoms. BMC Womens Health. 2022;22(1):316. doi:10.1186/s12905-022-01898-5
7. Benitez Fuentes JD, Morgan E, de Luna Aguilar A, et al. Global Stage Distribution of Breast Cancer at Diagnosis: A Systematic Review and Meta-Analysis. JAMA Oncol. 2024;10(1):71–78. doi:10.1001/jamaoncol.2023.4837
8. Kösters JP, Gøtzsche PC. Regular self-examination or clinical examination for early detection of breast cancer. Cochrane Database Syst Rev. 2003;(2):CD003373. doi:10.1002/14651858.CD003373
9. Azhar Y, Hanafi RV, Lestari BW, et al. Breast Self-Examination Practice and Its Determinants among Women in Indonesia: A Systematic Review, Meta-Analysis, and Meta-Regression. Diagnostics (Basel). 2023;13(15):2577. doi:10.3390/diagnostics13152577
10. Sujarwoto, Maharani A. Participation in community-based healthcare interventions and non-communicable diseases early detection of general population in Indonesia. SSM Popul Health. 2022;19:101236. doi:10.1016/j.ssmph.2022.101236
11. Solikhah S, Lianawati L, Matahari R, et al. Determinants of Breast Cancer Screening Practice among Women in Indonesia: A

Nationwide Study. Asian Pac J Cancer Prev. 2021;22(5):1435–1441. doi:10.31557/APJCP.2021.22.5.1435

12. Dewi TK, Ruiter RAC, Diering M, et al. Breast self-examination as a route to early detection in a lower-middle-income country: assessing psychosocial determinants among women in Surabaya, Indonesia. BMC Womens Health. 2022;22(1):179. doi:10.1186/s12905-022-01748-4
13. Khan S, Alzkari F, Alabdulkarim A, et al. Factors influencing breast self-exam practice and awareness among females in Saudi Arabia. Sci Rep. 2026;16(1):18114. doi:10.1038/s41598-026-48261-4
14. Kandasamy G, Almaghaslah D, Almanasef M, et al. Knowledge, attitude, and practice towards breast self-examination among women: a web based community study. Front Public Health. 2024;12:1450082. doi:10.3389/fpubh.2024.1450082
15. Tilahun BD, Ayele M, Abate BB, et al. Breast self-examination prevalence and determinants in low- and middle-income countries: an umbrella review of systematic reviews and meta-analyses. Front Glob Womens Health. 2026;7:1439187. doi:10.3389/fgwh.2026.1439187
16. Asmare K, Birhanu Y, Wako Z. Knowledge, attitude, practice towards breast self-examination and associated factors among women in Gondar town, Northwest Ethiopia, 2021: a community-based study. BMC Womens Health. 2022;22(1):174. doi:10.1186/s12905-022-01764-4
17. Mohammed AS, Tafese T, Menberu MS, et al. Practice of breast self-examination and its associated factors among women of reproductive age in seven public health facilities in Addis Ababa, Ethiopia: A multi-centre cross-sectional study. Womens Health (Lond). 2026;22:17455057261467912. doi:10.1177/17455057261467912
18. Hussein D, Oyato BT, Gashaw K, et al. Practice of breast self-examination and associated factors among women of reproductive age in the North Shoa Zone, Oromia, Ethiopia, 2022: a convergent mixed-methods study. BMJ Open. 2024;14(5):e073951. doi:10.1136/bmjopen-2023-073951
19. Hosh M, Abera L, Hailu M, et al. Knowledge and practice of breast self-examination among female undergraduate students in Jigjiga City, Ethiopia. Sci Rep. 2025;15(1):38020. doi:10.1038/s41598-025-21825-6
20. Jadhav BN, Abdul Azeez EP, Mathew M, et al. Knowledge, attitude, and practice of breast self-examination is associated with general self-care and cultural factors: a study from Tamil Nadu, India. BMC Womens Health. 2024;24(1):151. doi:10.1186/s12905-024-02981-9
21. Taneepanichskul S, Chuemchit M, Wongsasuluk P, et al. Practice, confidence and continuity of breast self-examination among women in Thailand during COVID-19 pandemic: a cross-sectional study. BMJ Open. 2023;13(8):e071306. doi:10.1136/bmjopen-2022-071306
22. Tiongco RE, Castro EJ, Baco AD, et al. Knowledge, practices, and perceived barriers on breast self-examination of female college students: Experience from a local higher education institution. J Educ Health Promot. 2024;13(1):413. doi:10.4103/jehp.jehp_1879_23
23. Mohan D, Su TT, Donnelly M, et al. Breast Cancer Screening in Semi-Rural Malaysia: Utilisation and Barriers. Int J Environ Res Public Health. 2021;18(23):12293. doi:10.3390/ijerph182312293
24. Noman S, Elarusy NME, Rahman HA, et al. Knowledge and beliefs on breast cancer screening and uptake among Yemeni female school teachers in Malaysia. Sci Rep. 2024;14(1):31878. doi:10.1038/s41598-024-83415-2
25. Mahar B, Osman MB, Fauzi FA, et al. The impact of educational interventions on breast self-examination practice, knowledge, and beliefs among women: A systematic review and meta-analysis. J Public Health Res. 2026;15(1):22799036261423725. doi:10.1177/22799036261423725
26. Agide FD, Garmaroudi G, Sadeghi R, et al. Designing and evaluation of the effect of community-based intervention on breast self-examination among reproductive-aged women in Ethiopia: A Cluster Randomized Controlled Trial. PLoS One. 2025;20(8):e0329839. doi:10.1371/journal.pone.0329839
27. Ağadayı E, Arby TE, Balta AT, et al. Effects of hands-on breast self-examination training on knowledge, health beliefs, and BSE practice among female university students: a randomized controlled trial. BMC Womens Health. 2026;26(1):501. doi:10.1186/s12905-026-04656-z
28. Khasanah N, Liliana A, Wahyuningsih M, et al. Augmented Reality-Based Breast Self-Examination Education and Its Effects on Knowledge and Perceived Behavioral Control Among Indonesian Adolescent Girls: A Quasi-Experimental Study. Asian Pac J Cancer Prev. 2026;27(7):2653–2662. doi:10.31557/APJCP.2026.27.7.2653
29. Liao Y, Hairon SM, Yaacob NM, et al. Psychometric validation of a knowledge, attitudes, and practices (KAP) scale for breast cancer screening among Chinese women. BMC Public Health. 2025;25(1):3478. doi:10.1186/s12889-025-24775-w
30. Lin JY, Hung CI, Li TC, et al. Development and validation of a Chinese version of a questionnaire to evaluate the knowledge, attitude, and practice about breast cancer screening among financial female workers in Taiwan. Front Oncol. 2025;15:1559622. doi:10.3389/fonc.2025.1559622
31. Fejer A, Atbaei MA, Zand A, et al. Psychometric Properties of the Breast Cancer Awareness Measure (Breast-CAM): A Systematic Review and Meta-Analysis. Cancers (Basel). 2026;18(6):956. doi:10.3390/cancers18060956
32. Masawa G, Mboineki JF. Assessing breast self-examination knowledge, attitude and practice as a secondary prevention of breast cancer among female undergraduates at the University of Dodoma: a protocol of analytical cross-sectional study. Front Epidemiol. 2024;4:1227856. doi:10.3389/fepid.2024.1227856
33. Sayi E. Knowledge, attitude and practice of breast self-examination among female students in institutions of higher learning: a case of Mount Kenya University [master's thesis]. Nairobi: University of Nairobi; 2018. Available from: http://hdl.handle.net/11295/104955
34. Mengie T, Yifru A, Fikre A, et al. Assessment of Knowledge, Attitude and Practice Regarding Breast Self-Examination Among Female Students in Debre Birhan University, North Shewa Ethiopia. Am J Clin Exp Med. 2020;8(3):42–48. doi:10.11648/j.ajcem.20200803.13
35. Halid Z, Perwitasari DA, Sulistyani N, et al. Beliefs about Medicine Questionnaire: Translation, Validation, and Adaptation for Post-Acute Coronary Syndrome Patients in the South Sulawesi Population, Indonesia. Explor Res Clin Soc Pharm. 2026;23:100794. doi:10.1016/j.rcsop.2026.100794
36. Sanjaya G, Kekalih A, Pakasi TA, et al. Validity and reliability test of the brief symptom rating scale 5 (BSRS-5) questionnaire in Indonesian version as an assessment tool for psychological disorders in Indonesia. Asian J Psychiatr. 2023;86:103676. doi:10.1016/j.ajp.2023.103676
37. Arifin B, Vinegia A, Usmar U, et al. Development and psychometric validation of SADAR-PrEP, an Indonesian PrEP knowledge and attitudes instrument. Sci Rep. 2026;16(1):28466. doi:10.1038/s41598-026-47715-z
38. Johari Chan M, Nik Farid ND, Thangiah N, et al. Developing and Validating a Breast Cancer Screening Health Literacy Questionnaire Guided by Sørensen's Integrated Model for Malaysian Adults. Malays J Med Sci. 2026;33(1):129–144. doi:10.21315/mjms-10-2025-719
39. Metsämuuronen J. Typology of Deflation-Corrected Estimators of Reliability. Front Psychol. 2022;13:891959. doi:10.3389/fpsyg.2022.891959
40. Metsämuuronen J. Attenuation-Corrected Estimators of Reliability. Appl Psychol Meas. 2022;46(8):720–737. doi:10.1177/01466216221108131
41. van der Ark LA. Standard Errors for Reliability Coefficients. Psychometrika. 2025;90(5):1679–1704. doi:10.1017/psy.2025.10050
42. Cho E. Beyond alpha and omega: The accuracy of single-test reliability estimators in unidimensional continuous data. Behav Res Methods. 2024;56(6):6299–6317. doi:10.3758/s13428-024-02361-z
43. Edwards AA, Joyner KJ, Schatschneider C. A Simulation Study on the Performance of Different Reliability Estimation Methods.

Educ Psychol Meas. 2021;81(6):1089–1117. doi:10.1177/0013164421994184

44. Béland S, Falk CF. A Comparison of Modern and Popular Approaches to Calculating Reliability for Dichotomously Scored Items. Appl Psychol Meas. 2022;46(4):321–337. doi:10.1177/01466216221084210
45. Gamal A, Awwad SA, Abdallfatah A, et al. Knowledge, attitudes, and awareness of breast self-examination among female university students in three Arab countries: a multi-center cross-sectional study. BMC Public Health. 2025;25(1):2961. doi:10.1186/s12889-025-23932-5
46. Dwumfour CK, Acheampong SM, Dankwah H, et al. Awareness and Practice of Breast Self-Examination Among Pregnant Women Attending Antenatal Clinics in Ghana: A Cross-Sectional Study. SAGE Open Nurs. 2026;12:23779608261477846. doi:10.1177/23779608261477846
47. Ishtiak ASM, Ahmed N, Gaffar F, et al. Knowledge, practice and associated factors of breast self-examination among female university students of Bangladesh. Heliyon. 2022;8(11):e11780. doi:10.1016/j.heliyon.2022.e11780
48. Cahyaningsih I, Rokhman MR, Maziyyah N, et al. Translation and Validation of the Diabetes Knowledge Questionnaire in Indonesian Patients With Type 2 Diabetes. Sci Diabetes Self Manag Care. 2024;50(6):484–496. doi:10.1177/26350106241287445
49. Mashayekh-Amiri S, Jafarabadi MA, Hosseinzadeh M, et al. Measurement properties of the Iranian version of the breast cancer perception scale (BCPS) according to the COSMIN checklist. BMC Cancer. 2024;24(1):743. doi:10.1186/s12885-024-12493-2
50. Htay MNN, Schliemann D, Dahlui M, et al. Validation of the Champion Health Belief Model Scale for an Investigation of Breast Cancer Screening Behaviour in Malaysia. Int J Environ Res Public Health. 2021;18(17):9311. doi:10.3390/ijerph18179311
51. Uzun SU, Mut ANU, Çöl M. Breast Cancer Awareness Scale: Turkish Validity and Reliability Study Among Adult Women. Int J Nurs Pract. 2025;31(5):e70046. doi:10.1111/ijn.70046
52. Sinha N. Cross-sectional validation of a breast cancer KAP scale for urban Indian women using social media. Indian J Cancer. 2026;63(2):191–199. doi:10.4103/ijc.ijc_359_26
53. Pfadt JM, van den Bergh D, Sijtsma K, et al. Bayesian Estimation of Single-Test Reliability Coefficients. Multivariate Behav Res. 2022;57(4):620–641. doi:10.1080/00273171.2021.1891855
54. DelosReyes JMV, Padilla MA. Obtaining a Bayesian Estimate of Coefficient Alpha Using a Posterior Normal Distribution. Educ Psychol Meas. 2025;85(4):829–852. doi:10.1177/00131644241311877
55. Tupper H, Ghukasyan R, Bayburtyan A, et al. Validation of an Eastern Armenian breast cancer health belief survey. PLOS Glob Public Health. 2023;3(5):e0001849. doi:10.1371/journal.pgph.0001849
56. Toska A, Georga D, Christakis A, et al. Translation and Validation of the Greek Version of the Breast Cancer Screening Beliefs Questionnaire. Asian Pac J Cancer Prev. 2025;26(2):663–670. doi:10.31557/APJCP.2025.26.2.663
57. Li J, Zhang X, Zhang F, et al. Adaptation and analysis of the psychometric properties of the Chinese version of the MARA questionnaire to assess breast cancer awareness. Cancer Causes Control. 2025;36(12):1837–1849. doi:10.1007/s10552-025-02050-y
58. Rahman MA, Hossen MM, Chowdhury MEH, et al. Psychometric Validation of the Bangla Version of the Breast Cancer Fear Scale Among Female University Students in Bangladesh. Int J Breast Cancer. 2025;2025(1):6811105. doi:10.1155/ijbc/6811105
59. Aker MN, Yılmaz Sezer N, Öner Cengiz H. Psychometric properties of the Turkish version of the perceived barriers of mammography scale. J Obstet Gynaecol Res. 2024;50(11):2113–2120. doi:10.1111/jog.16090
60. Teye-Kwadjo E. Understanding Fear of Breast Cancer Among Ghanaian Women via the Champion Breast Cancer Fear Scale. Breast Cancer (Auckl). 2025;19:11782234251353270. doi:10.1177/11782234251353270
61. Teye-Kwadjo E. Validity and Reliability of the Breast Cancer Screening Beliefs Questionnaire Among Ghanaian Women. Int J Breast Cancer. 2026;2026(1):6682781. doi:10.1155/ijbc/6682781
62. Sapci E, Gungormus Z. Breast self-examination behavior change process and early detection of breast cancer: the role of the transtheoretical model. J Health Psychol. 2026;31(8):3178–3195. doi:10.1177/13591053251412165
63. Waddimba AC, Baker BM, Pogue JR, et al. Psychometric validity and reliability of the 10- and 2-item Connor–Davidson resilience scales among a national sample of Americans responding to the Covid-19 pandemic: an item response theory analysis. Qual Life Res. 2022;31(9):2819–2836. doi:10.1007/s11136-022-03125-y
64. Takahashi N, Matsuhisa T, Nishikawa H, et al. Validity and reliability of the 2-item version compared with the 10-item version of the Japanese consultation and relational empathy (CARE) measure: primary analysis of a cross-sectional survey. BMC Prim Care. 2025;26(1):343. doi:10.1186/s12875-025-03018-2
65. Ghafri TA, Mahrouqi MA, Zadjali AA, et al. Determinants of performing breast self-examination in Omani women using health belief model: Confirmatory factor analysis with hierarchical logistic regression. J Cancer Policy. 2026;49:100772. doi:10.1016/j.jcpo.2026.100772
66. Noman S, Shahar HK, Rahman HA, et al. Factor structure and internal reliability of breast cancer screening Champion's Health Belief Model Scale in Yemeni women in Malaysia: a cross-sectional study. BMC Womens Health. 2021;21(1):437. doi:10.1186/s12905-021-01543-7
67. Kimata A, Kumagai K, Kondo N, et al. Development and validation of the Cancer Knowledge Scale for the general population: An item response theory approach. Patient Educ Couns. 2022;105(1):151–157. doi:10.1016/j.pec.2021.05.007
68. Trizano-Hermosilla I, Gálvez-Nieto JL, Alvarado JM, et al. Reliability Estimation in Multidimensional Scales: Comparing the Bias of Six Estimators in Measures With a Bifactor Structure. Front Psychol. 2021;12:508287. doi:10.3389/fpsyg.2021.508287
69. Reise SP, Haviland MG. Understanding Alpha and Beta and Sources of Common Variance: Theoretical Underpinnings and a Practical Example. J Pers Assess. 2025;107(3):267–282. doi:10.1080/00223891.2024.2420175
70. Zager Kocjan G, Lavtar D, Sočan G. The effects of survey mode on self-reported psychological functioning: Measurement invariance and latent mean comparison across face-to-face and web modes. Behav Res Methods. 2023;55(3):1226–1243. doi:10.3758/s13428-022-01867-8
71. McDuff K, Su TT, Brown DA, et al. Battling the Bots and Defending Against Fraudulent Responses in an International Community-Engaged Web-Based Survey With People Living With Long COVID: Methodological Study. J Med Internet Res. 2026;28:e88838. doi:10.2196/88838
72. Qutob IA, Soliman A, Aufy H, et al. Breast cancer awareness and breast self-examination among females in the Middle East and North Africa: a multinational cross-sectional study. BMJ Oncol. 2025;4(1):e000869. doi:10.1136/bmjonc-2025-000869
73. Alam R, Akter S, Mohammad T, et al. Knowledge of breast cancer and practice of breast self-examination among adolescent girls in semi-urban area – Savar, Dhaka, Bangladesh: A school-based cross-sectional study. PLoS One. 2026;21(6):e0334375. doi:10.1371/journal.pone.0334375
74. Bu Bshait MS, Almaqhawi A. Quality Assessment of YouTube Videos As Information Source for Breast Self-Examination. Cureus. 2024;16(9):e70227. doi:10.7759/cureus.70227
75. Hertanti NS, Chuang YH. Cross-cultural adaptation and psychometric evaluation of the dengue preventive attitudes questionnaire for Indonesian public health nurses. BMC Nurs. 2026;25(1):560. doi:10.1186/s12912-026-04928-x
76. Hertanti NS, Nguyen NTT, Chi DTK, et al. Psychometric evaluation of the Indonesian version of the confidence scale among primary healthcare nurses in dengue-related prevention

practice. BMC Med Educ. 2026;26(1):1062. doi:10.1186/s12909-026-09471-6

77. Pangestu S, Purba FD, Setyowibowo H, et al. Validity, test-retest reliability, and responsiveness of the Indonesian version of FACIT-COST measure for subjective financial toxicity. Health Qual Life Outcomes. 2024;22(1):89. doi:10.1186/s12955-024-02303-5
78. Rimayanti UP, Mizuno M, Kadar K, et al. Ensuring reliability and cultural validity of the Indonesian version of the Quality Of Life Index for patients with cancer. Res Nurs Health. 2023;46(5):527–537. doi:10.1002/nur.22334
79. Khatimah NFK, Jusriani J, Wulandari D, et al. Validity and reliability of Indonesian version of the digital screen exposure questionnaire (DSEQ) for young children. PLoS One. 2026;21(3):e0346133. doi:10.1371/journal.pone.0346133
80. Huda MH, Rustina Y, Waluyanti FT, et al. Psychometric evaluation of the Indonesian version of paternal breastfeeding self-efficacy scale- short form: A confirmatory factor analysis. Midwifery. 2024;139:104182. doi:10.1016/j.midw.2024.104182
81. Schliemann D, Jamil ASA, Mohan D, et al. The development and evaluation of a mHealth, community education and navigation intervention to improve clinical breast examination uptake in Segamat Malaysia: A randomised controlled trial. PLoS One. 2023;18(10):e0288437. doi:10.1371/journal.pone.0288437
82. Uruntie RO, Oputa CH, Peters E, et al. Effect of educational intervention on the knowledge, attitude and practice of breast self-examination among female students at a private university in Southern Nigeria. BMC Cancer. 2024;24(1):355. doi:10.1186/s12885-024-12116-w
83. Ogunmodede EO, Ajala DE, Aluko JO, et al. Promoting breast health among female adolescents: a comparative analysis of the effects of two didactic tools on knowledge and practice regarding breast self-examination in Southwest Nigeria. BMC Cancer. 2024;24(1):1183. doi:10.1186/s12885-024-12949-5
84. Kuru Alici N, Arikan Dönmez A, Aktaş O, et al. The effect of simulation-based breast health education on breast cancer awareness and breast self-examination skills of Afghan refugee women: a randomized controlled trial. BMC Public Health. 2025;25(1):2048. doi:10.1186/s12889-025-23313-y
85. Cheraghalizadeh H, Adib-Rad H, Pasha H, et al. Effectiveness of face to face and virtual education to promote breast self-examination based on the theory of planned behavior: a randomized controlled trial study. BMC Cancer. 2025;25(1):548. doi:10.1186/s12885-025-13935-1
86. Ponzel N, Kolesnik P. Validation of Ukrainian breast and cervical cancer screening questionnaire. Wiad Lek. 2026;79(7):1581–1590. doi:10.36740/WLek/222250
87. Williams B, FitzGibbon L, Brady D, et al. Sample size matters when estimating test–retest reliability of behaviour. Behav Res Methods. 2025;57(4):123. doi:10.3758/s13428-025-02599-1
88. Mahar B, Osman MB, Fauzi FA. From awareness to action: Health Belief Model-based educational intervention to improve breast self-examination practice among college teachers in Pakistan (CRCT). PLoS One. 2026;21(1):e0341804. doi:10.1371/journal.pone.0341804