

Psychometric validation of the Danish BREAST-Q reconstruction module

Cecilie Balslev Willert^{a,*}, Karl Bang Christensen^b, Pernille Envold Bidstrup^c,
Lene Mellemkjær^c, Anne-Marie Axø Gerdes^{d,e}, Lisbet Rosenkrantz Hölmich^{a,e}

^a Department of Plastic Surgery, Copenhagen University Hospital, Herlev and Gentofte, Denmark

^b Section of Biostatistics, Department of Public Health, University of Copenhagen, Denmark

^c Danish Cancer Institute, Copenhagen, Denmark

^d Department of Genetics, Copenhagen University Hospital, Rigshospitalet, Denmark

^e Department of Clinical Medicine, University of Copenhagen, Denmark

ARTICLE INFO

Keywords:

Psychometric validation
The BREAST-Q
Health-related quality of life
Rasch analysis
Breast reconstruction

ABSTRACT

Aim/background: Patient-reported outcome measurement instruments are important tools in understanding a breast reconstruction's impact on the patients' quality of life. A psychometric validation is essential before applying a patient-reported outcome measurement instrument in clinical practice and research. The BREAST-Q is a specific, validated questionnaire for breast surgery outcomes that has been translated from English to Danish. It consists of 167 items in 7 pre-operative scales and 15 post-operative scales. This validation study aims to validate the Danish BREAST-Q reconstruction module.

Material and methods: Eligible women were included from January 2019 to June 2020. Multiple-item scales with summated scores and more than 40 complete responses were eligible for psychometric validation, and psychometric analyses examined reliability and validity using Rasch Analyses and Classical Test Theory. Measurements included test for local response dependence, item fit, differential item functioning, and more. Clinical validity was assessed using known-groups hypotheses.

Results: We obtained 115 and 201 complete responses pre- and postoperatively, respectively. We validated 120 items in four preoperative and nine postoperative scales. The Rasch analyses disclosed evidence of local response dependence in eight scales. Chronbach's α ranged 0.81–0.95 after adjustment. Item fit was evaluated using item-rest correlations and showed good fit in 98 % of items. Differential Item Functioning was found in four items but had very little effect on the model. Clinical validity was supported by the know-group analyses.

Discussion/conclusion: The Danish BREAST-Q reconstruction module has good acceptability, feasibility and validity, and adequate reliability. The results support the use in a Danish population.

1. Introduction

Breast reconstruction following mastectomy due to invasive or in situ breast cancer or for risk-reduction is performed increasingly [1,2]. It is a major surgical procedure with inherent risks of complications and permanent side effects, such as impaired sensibility of the breast, pain, and altered breast shape and firmness, which can affect the patient's health-related quality of life (HRQOL). It is important to assess the alterations in HRQOL to provide evidence-based counseling to patients considering breast reconstruction, and patient-reported outcome measurement (PROM) is ideal for this purpose. Existing validated instruments in Danish, such as the SF-36, are not optimal because they are

generic and not specific for breast reconstruction.

The BREAST-Q reconstruction module is a specific questionnaire that measures HRQOL and satisfaction in women undergoing breast reconstruction. It is developed with adherence to international PROM guidelines, and is designed to measure change in the underlying construct because it consists of both a pre- and postoperative survey [3]. With validation for both paper and electronic versions using Rasch Analysis (RA) [4,5] and translations to more than 30 languages [6,7], the BREAST-Q reconstruction provides a valid, specific, and comprehensive patient-reported information on HRQOL following breast cancer surgery. We have previously performed a Danish linguistic validation of all BREAST-Q modules [8]. However, a psychometric scale validation

* Corresponding author. Department of Plastic Surgery, Copenhagen University Hospital, Herlev and Gentofte Hospital, Borgmester Ib Juuls Vej 5, 2730, Herlev, Denmark.

E-mail address: cecilie.balslev.willert@regionh.dk (C.B. Willert).

<https://doi.org/10.1016/j.breast.2025.103872>

Received 25 June 2024; Received in revised form 13 December 2024; Accepted 1 January 2025

Available online 2 January 2025

0960-9776/© 2025 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

should follow a translation to evaluate the measurement properties in the new linguistic and cultural context [9].

Psychometric validation entails examination of a PROM instrument's measurement properties, such as reliability and validity. There are two main psychometric theories: Classical Test Theory (CTT) and Modern Test Theory (MTT), and the latter includes RA. RA has several advantages and is increasingly used in developing and validating PROM instruments [10]. The current study aims to psychometrically validate the Danish BREAST-Q reconstruction module using CTT and RA for use in clinical practice and research.

2. Material and methods

2.1. Population

Eligible patients were women above 18 years old who were either facing or had undergone any kind of breast reconstruction (e.g., implant-based or autologous) and who understand and speak Danish. We distinguished between pre- and post-reconstruction responses and refer hereafter to them as pre- and postoperative. In the postoperative group, at least three months should have passed since the reconstruction of the breast mound. Women were invited to participate in the study at their outpatient clinic appointments at two Danish university hospitals (Herlev and Gentofte Hospital and Rigshospitalet).

2.2. Data collection

We collected data from January 2019 until June 2020 through electronic questionnaires via REDCap [11]. Non-responders were sent a reminder after three weeks and received a phone call if they remained non-responders after another three weeks. The surveys included the Danish BREAST-Q reconstruction module version 2.0, which consists of 167 items in seven preoperative and 15 postoperative scales. Of these, 20 are multiple-item scales, and two are single-item scales. Eighteen of the 20 multiple-item scales have summated scores of 0–100. The scale regarding sexual well-being was optional. Preoperative patients were invited to fill out both the pre- and postoperative BREAST-Q scales. Furthermore, we collected patient-reported information on sociodemographics (e.g., age and level of education) and, to enable clinical validation, we asked postoperative patients to state their overall satisfaction with the breast reconstruction and provide information about complications. Information on the date of the surgery, surgical indication (oncological or risk-reducing), uni- or bilateral surgery, timing of the breast reconstruction (immediate or delayed), and breast cancer treatment (radiation and chemotherapy) was collected from electronic patient files. The study was approved by the Regional Data Agency and run in accordance with the Declaration of Helsinki, including obtainment of informed consent [12].

3. Statistical analyses

We used CTT and RA, and analyses were based on complete cases. Descriptive statistics are reported for all 22 scales, while psychometric analyses are reported for multiple-item scales with summated scores 0–100 and more than 40 complete responses.

3.1. Overall concepts

Acceptability: was assessed by calculating the response rates, missing data, and the maximum endorsement frequency (MEF, the percentage of respondents who have chosen the response option that is the least (MEF, low) or the most (MEF, high) popular). As in previous validations, missing data >5 % and MEF >80 % were considered violations of data quality [4].

Reliability: refers to a scale's measurement precision [13] and was measured using Cronbach's coefficient alpha (see below).

Construct validity: refers to how well an instrument measures the construct it is supposed to measure [13] and was measured as the fit of the observed data to an MTT model, which is a requirement due to assumptions of unidimensionality and invariance for the MTT model.

Interpretability: is the degree to which the total score of a scale can be assigned qualitative meaning and is evaluated through a clinical validation [13]. We tested two known group hypotheses: 1) patients with complications would have a lower score than patients without complications, and 2) patients who were more satisfied with their breasts after the breast reconstruction would have a higher score than patients who were less satisfied with their breast after the breast reconstruction. Scores were compared across these groups using box plots and the Hodges-Lehman estimator of the median difference.

3.2. Classical Test Theory

Floor and ceiling effects: were defined as > 15 % of the respondents obtaining either the lowest or the highest scale score, as this makes it difficult to discriminate among respondents, thus compromising reliability [14].

Chronbach's alpha: is a reliability coefficient that measures the internal consistency between items [15]. Values range from 0 to 1, with a good correlation between 0.7 and 0.95 [14]. Chronbach's alpha should, however, be interpreted with care since it will rise with the number of items in a scale despite the average inter-item correlation being maintained [16]. If local dependency (LD, see below) is present, Chronbach's alpha increases and will thus falsely inflate reliability [17]. In this case, Chronbach's alpha calculated on testlets (see below) can be used to estimate reliability.

3.3. Rasch Analysis

Local dependency: Fundamental assumptions in RA include that items should be related through the latent variable alone, they should be conditionally independent of each other, and the response to one item should not depend directly on the response to another [17]. LD can occur if items are similar in content, i.e., items are asking the same in different ways. We examined the correlation between item residuals using $Q3^*$, defined as the difference between the largest and average residual correlation [18]. The $Q3^*$ differs from the more commonly used $Q3$ by taking into account how the number of items impacts residuals. A $Q3^*$ value of 0.2, which we used as a cut-off, is proven reasonably stable compared to the standard of 0.3 [18]. Evidence of LD led to an adapted scale where the item pair with the maximal residual correlation was collapsed into a testlet (i.e., a group of questions about the same topic), operationalized as the sum [19].

Differential Item Functioning (DIF): occurs if an item is answered differently by a subgroup of the study population due to an external factor such as for instance age. A fundamental assumption in RA is that DIF is not present, which is called invariance. We examined DIF for age (age below/above 60 years) and educational level (short, medium, and long) using ordinal logistic regression. A graphical evaluation assessed the magnitude of DIF (not shown).

Overall fit: The Rasch model's overall fit for each scale is an indicator of validity and unidimensionality. We evaluated the overall fit using conditional likelihood ratio tests (significant p-values indicate misfit) [20].

Item fit: evaluates how well the observed data fits the Rasch model expectations, and misfit implies that the scale's validity may be compromised. Evidence of misfit can also occur due to LD or DIF and should lead to a close examination of the relevant items. We used a fit statistic comparing observed and expected item rest-score associations (significant p-values indicate misfit) [21]. When item misfit was statistically significant after adjustment for multiple testing, we used a graphical evaluation of conditional item characteristics curves to inspect the magnitude [22]. The false discovery rate (FDR) was kept at 5 % using

the Benjamini-Hochberg procedure [23].

Unidimensionality: The Rasch model postulates that all items in a scale measure a single unidimensional latent variable responsible for the observed correlation between the observed item scores [24]. We evaluated unidimensionality by first evaluating the overall fit of the Rasch model for each scale. In case of any misfit, we evaluated item fit, LD, and DIF and, if possible, adapted the scale as outlined below.

Analysis of adapted scales: When we disclosed evidence of both LD and misfit, we took LD into account by testing DIF and model fit in the adapted testlet scales. Based on the results of these analyses, we synthesized a conclusion about the scale.

See Fig. 1 for an abbreviation overview.

We used SAS software version 9.4 and R version 4.2.1 [25] for statistical analysis. The R packages *eRm*, *iarm*, *RASCHplot*, and *sirt* were used for Rasch analyses [22,26–28]. A significance level of 0.05 was used.

4. Results

We included 140 preoperative and 257 postoperative patients (56 were both), and 115 and 201 provided complete responses, yielding response rates of 82 % and 78 %, respectively. Missing data was ≥ 5 % in 23 of 167 items (14 %), all of which can be found in the preoperative *Sexual well-being scale* and the postoperative *Satisfaction with breasts scale*, except for one item in the *Physical well-being scale*. MEF was >80 % in 25 of 167 items (15 %) (Supplementary Table 1).

Mean ages were 49.8 and 51.4 years, respectively. Most patients were recruited from Herlev and Gentofte Hospital (Table 1). Of the 18 multiple-item scales with summated scores, 13 had more than 40 complete responses and were eligible for validation. Hence, 120 items were validated using RA (Table 2).

4.1. Classical Test Theory

Floor or ceiling effects were found in 11/18 scales with summated scores, and Chronbachs alfa calculated on adapted testlet scales ranged from 0.81 to 0.95 (Supplementary Table 2 and Table 2).

Abbreviations

DIF = Differential item functioning

FDR= False discovery rate

HRQOL= Health-related quality of life

LD= Local dependency

MEF= Maximum endorsement frequency

MTT= Modern Test Theory

RA= Rasch Analyses

Table 1

Patient characteristics; values are no. (%) unless otherwise indicated.

Characteristics	Pre-operative responses (N = 115) ^a	Post-operative responses (N = 201) ^a	
Age, mean (range) in years	49.9 (29–75)	51.4 (20–81)	
Recruitment site			
- Herlev and Gentofte Hospital	100 (87.0)	182 (90.6)	
- Rigshospitalet	15 (13.0)	19 (9.5)	
Highest level of education^b			
- Primary School	3 (2.6)	7 (3.5)	
- High School	7 (6.1)	12 (6.0)	
- Short-cycle higher education	19 (16.5)	34 (16.9)	
- Medium-cycle higher education	53 (46.1)	76 (37.8)	
- Long-cycle higher education	31 (27.0)	65 (32.3)	
- Other	2 (1.7)	7 (3.5)	
Household income (n) (%)			
- < 500,000 DKK	29 (25.2)	45 (22.4)	
- 500–800,000 DKK	37 (32.2)	62 (30.9)	
- >800,000 DKK	49 (42.6)	91 (45.3)	
- Unknown	0	3 (1.5)	
Marital status			
- Married/cohabiting	85 (73.9)	145 (72.1)	
- Partner/significant other	8 (7.0)	12 (6)	
- Widowed or single	22 (19.2)	43 (21.4)	
- Unknown	0	1 (0.5)	
Employment status			
- Employed	80 (69.6)	158 (78.6)	
- Unemployed	3 (2.6)	4 (2.0)	
- Retired ^c	8 (7.0)	13 (6.5)	
- Sick leave	19 (16.5)	13 (6.5)	
- Other	5 (4.4)	13 (6.5)	
Radiation therapy			
- Yes	45 (39.1)	30 (14.9)	
- No	70 (60.9) ^d	171 (85.1) ^d	
Chemotherapy			
- Yes	82 (71.3)	123 (61.2)	
- No	33 (28.7) ^d	78 (38.8) ^d	
Months since surgery, median (range)	–	10.1 (3.0–345)	
Type of surgery		Unilateral surgery (N=123)	Bilateral surgery (N = 78)
Surgical indication			
- Oncological	–	122 (99.2)	14 (18.0)
- Risk-reducing		1 (0.8)	15 (19.2)
- Both (for bilateral surgeries)		–	49 (62.8)
Timing of the breast reconstruction			
- Immediate	–	56 (45.5)	46 (59.0)
- Delayed		67 (54.5)	8 (10.3)
- Both (for bilateral surgeries)		–	24 (30.8)
Type of breast reconstruction			
- Expander-implant		28 (23.0)	18 (23.1)
- Direct-to-implant		33 (27.1)	32 (41.0)
- Autologous with implants		14 (11.5)	4 (5.1)
- Autologous		47 (38.5)	3 (3.9)
- A combination of the above (for bilateral surgeries)		–	21 (26.9)

^a 35 patients have contributed with both pre- and postoperative responses.

^b Classified according to DISCED-15 as described by Statistics Denmark <https://www.dst.dk/en/Statistik/dokumentation/nomenklaturer/disced15-audd>.

^c Both incapacity beneficiaries and pensioners.

^d Either the patient did not receive radiation or chemotherapy, or the data was not applicable.

Fig. 1. Abbreviation box in alphabetic order.

Table 2

Overview of the psychometric validation results of the Danish BREAST-Q reconstruction module.

Scales	Number of items	Sample size	Rasch Analyses					Classical Test Theory	
			Local independence (no LD)	No differential item functioning (invariance)		Overall Rasch model fit	Item fit ^a	No floor or ceiling effects	α^b
				Age groups	Education groups				
Preoperative scales									
Psychosocial wellbeing	10	114	÷	✓	✓	✓	✓	✓	0.91
Sexual wellbeing	6	77	÷	✓	÷	✓	✓	✓	0.81
Satisfaction with breasts	4	112	✓	✓	✓	✓	✓	✓	0.85
Physical wellbeing: chest	10	109	÷	✓	✓	✓	✓	÷	0.86
Physical wellbeing: abdomen	4	33	–	–	–	–	–	÷	–
Satisfaction with abdomen	1	33	–	–	–	–	–	–	–
Lattisimus dorsi module: Physical well-being: back and shoulder	11	20	–	–	–	–	–	÷	–
Postoperative scales									
Psychosocial wellbeing	10	196	÷	✓	✓	✓	✓	÷	0.93
Sexual wellbeing	6	158	÷	✓	✓	✓	✓	✓	0.81
Satisfaction with breasts	15	174	✓	÷	✓	✓	✓	✓	0.92
Satisfaction with implants	2	146	–	–	–	–	–	–	–
Physical wellbeing: chest	11	191	÷	✓	✓	✓	✓	÷	0.86
Physical wellbeing: abdomen	7	45	✓	✓	✓	✓	✓	÷	0.85
Satisfaction with abdomen	3	45	–	–	–	–	–	–	–
Satisfaction with nipple reconstruction	1	79	–	–	–	–	–	–	–
Lattisimus dorsi module: Physical well-being: back and shoulder	11	27	–	–	–	–	–	✓	–
Lattisimus dorsi module: satisfaction with back	8	28	–	–	–	–	–	÷	–
Adverse effects of radiation	6	31	–	–	–	–	–	÷	–
Satisfaction with information	15	185	÷	✓	✓	✓	÷	✓	0.93
Satisfaction with surgeon	12	193	÷	÷	✓	✓	÷	÷	0.93
Satisfaction with medical team	7	195	✓	✓	÷	✓	✓	÷	0.94
Satisfaction with office staff	7	192	✓	✓	✓	✓	✓	÷	0.95

✓ = Present.

÷ = Not present.

^a Item fit was not present if a p-value remained significant after adjustment for multiple testing.^b Cronbachs alpha was calculated on adapted testlet scales.

4.2. Rasch Analyses

Evidence of LD, indicated by Q3,* values above 0.2, was found in 8/13 validated scales. Evidence of DIF, calculated on adapted testlet scales taking LD into account, was found for age in the postoperative *Satisfaction with breasts* and *Satisfaction with surgeon* scale. DIF for education was found in the preoperative *Sexual well-being* scale and in the postoperative *Satisfaction with medical team* scale. However, none of these were significant after adjustment for multiple testing.

The initial analysis of overall fit showed a very poor fit of the data to the Rasch model for nearly all scales (results not shown). However, testlet analyses considering LD showed adequate overall Rasch model fit for all scales. Item fit was evaluated on 72 items and 23 testlets across 13 scales. Of these 95 items or testlets, 16 item fit statistics (17 %) were significant at the 5 % FDR level (Supplementary Table 3). However, after adjustment for multiple testing, only two (2 %) remained significant: item o *What the scars would look like in the Satisfaction with information scale* and item f *Made you feel comfortable?* In the *Satisfaction with surgeon*

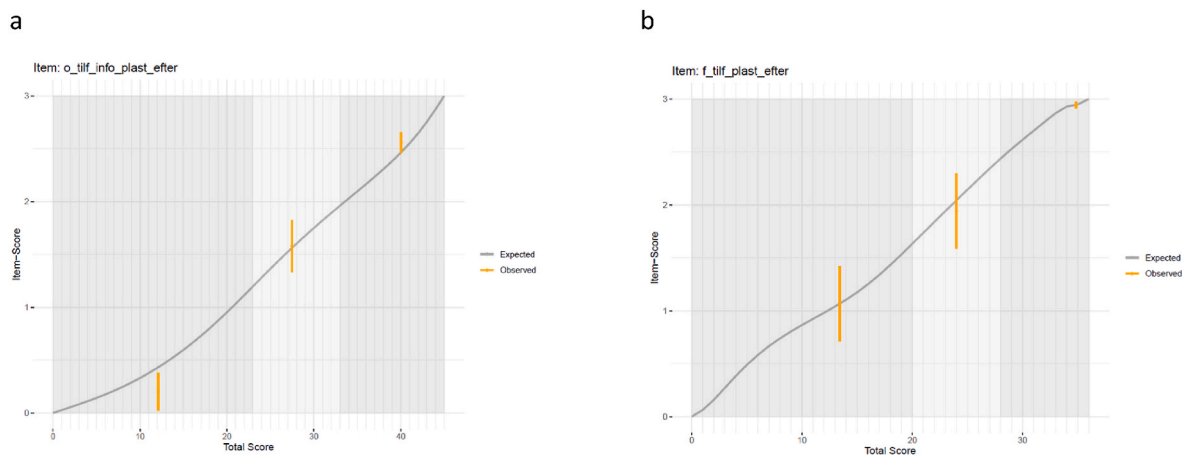


Fig. 2. Conditional item characteristic curves of item o: What the scars would look like in the Satisfaction with information scale (a) and item f: Made you feel comfortable? in the Satisfaction with surgeon scale (b).

scale. Only item o's conditional item characteristics curve deviates from expected means, and the magnitude of misfit is small (Fig. 2).

The known-group analyses of clinical validity comparing patients with *more* or *less* postoperative self-reported satisfaction with breasts and patients *with* or *without* surgical complications, respectively, reached higher scores in all scales, except for the *Physical well-being: abdomen* scale (Table 3). In the *Satisfaction with breasts* scale, the differences in means were 9 (95 % CI 3–14) and 4 (95 % CI 0–9) points in the two known-groups, respectively (Fig. 3).

5. Discussion

The psychometric validation of the Danish BREAST-Q reconstruction module using both CTT and RA of 115 preoperative and 201 postoperative responses showed a good overall fit in all scales, a good item fit in most scales, and good clinical validity. LD was found in 8/13 scales. However, reliability measured using Chronbach's α was acceptable when taking LD into account.

5.1. Acceptability

Acceptability was good, with response rates above proposed standards [29] and missing data ≤ 5 % and MEF < 80 % for most items. The missing data was concentrated in two scales: the preoperative *Sexual well-being* scale (range 11.3–13.0 %), and the postoperative *Satisfaction with breasts* scale (range 5–6.5 %). Similarly, the German and American validations of the BREAST-Q Breast Conserving Therapy module found missing data to be highest in the *Sexual well-being* scale [30,31]. This, along with the fact that 22/115 and 43/201 patients, pre-and post-operatively, respectively, preferred not to answer this optional scale, signifies the delicacy of the content, similar to what the Japanese and American BREAST-Q validation studies found [31,32]. Patients also expressed this in the Danish linguistic validation [8].

5.2. Classical Test Theory

Floor or ceiling effects were found in 11/18 multiple-item scales with summated scores, which makes it difficult to discriminate among patients and thus affects the reliability of the particular scales. This should be kept in mind when interpreting results. With the exception of the *Satisfaction with surgeon/medical team/office staff*, the scales with the higher floor or ceiling effects (> 25 %) were scales with < 40 complete responses: the preoperative *Physical well-being: abdomen* scale ($n = 33$), the postoperative *Lattissimus dorsi: satisfaction with back* scale ($n = 27$) and the postoperative *Adverse effects of radiation* ($n = 31$), thus the estimated floor effects should be interpreted with caution. Similarly, the Japanese validation of the BREAST-Q: Mastectomy module found floor effects in the *Sexual and Physical well-being* scale and ceiling effects in the

Satisfaction with care scales [32]. The latter was also found in validating the BREAST-Q: Breast Conserving Therapy Module and the BODY-Q, a PROM instrument for massive weight loss patients. The authors suggest that it was because the study sample came from a sizeable specialized cancer center or simply reflected the patient's satisfaction with treatment [31–33]. In our case, the study sample also primarily came from a single specialized unit. Reliability was backed by Chronbach's α , ranging from 0.81 to 0.95 in adapted scales with testlets, which is within the acceptable range, as in the original validation [5].

5.3. Rasch Analyses

The analyses disclosed evidence of LD for 8/13 scales. Examples of items with local dependence are item a *Was competent* and item g *Was thorough* from the *Satisfaction with surgeon* scale, and item e *Confident sexually about how your breast(s) look when unclothed*, and item f *Sexually attractive when unclothed* from the *Sexual well-being* scale. As local independence is one of the fundamental assumptions in RA, LD may lead to item misfit, and it is important to consider why it occurs and what the consequence should be. One option could be to exclude items because they are too similar. However, since we found the content of the items to be distinct but related and continue to find all items clinically relevant, we recommend leaving the scales unchanged. In our study, we handled the disclosed LD by generating testlets, so the LD did not influence the remaining analyses.

The value of the Q3 test is influenced by factors like the number of items; therefore, the cut-off value should differ across studies. We chose to use Q3, * because it takes this into account.

LD and DIF were not evaluated in the validation of the original BREAST-Q reconstruction module [5]. It may be because LD was not commonly addressed at the time of the study [17]. Given our findings, the original reliability measures of Chronbach's α and Person Separation Index could be higher than they would have been had LD been considered. Both LD and DIF are incorporated in the validation of later added scales [34].

The results from the fit analyses on adapted testlet scales indicate good validity. Hence, the scales measure what they are intended to measure and support the assumption of unidimensionality.

The clinical validity tested using known-group hypotheses demonstrates, except for the *Physical well-being: abdomen* scale, possibly due to a small sample size, that the scores of the BREAST-Q are in line with both the patient's perception and with our expectation that surgical complications negatively affect the result of a breast reconstruction, supporting its use in clinical practice. Our results are akin to those from the original validation [5].

As samples between 100 and 500 [35,36] are recommended for Rasch analyses, we believe that our overall sample size of 109–196 is sufficient. Still, a low number of respondents prevented us from

Table 3

Clinical validity for the multiple-item post-operative scales with summated scores. Scale scores are reported as means with 95 % confidence intervals and compared across groups using differences between means with 95 % confidence intervals.

Scale	Satisfaction with breasts after surgery compared with before ^a			Patient-reported complications ^b		
	Less	More	Difference	Yes	No	Difference
Psychosocial wellbeing	65 (58–67)	71 (68–75)	9 (3–14)	68 (65–71)	74 (70–79)	6 (1–12)
Sexual wellbeing	44 (39–50)	56 (51–60)	11 (4–18)	52 (47–56)	57 (52–62)	5 (–1–12)
Satisfaction with breasts	52 (48–55)	68 (64–71)	16 (11–21)	61 (58–65)	66 (62–70)	4 (0–9)
Physical wellbeing: chest	70 (65–75)	78 (74–83)	9 (2–15)	73 (69–76)	81 (77–86)	8 (3–14)
Physical wellbeing: abdomen	73 (55–91)	71 (65–78)	2 (–17–21)	68 (61–74)	78 (69–87)	10 (0–21)
Satisfaction with information	61 (56–67)	69 (65–73)	8 (1–14)	64 (61–68)	74 (69–79)	10 (4–16)
Satisfaction with surgeon	82 (76–87)	87 (84–91)	6 (–1–12)	84 (81–88)	92 (89–96)	8 (3–13)
Satisfaction with medical team	88 (84–93)	94 (91–97)	6 (0–11)	92 (89–94)	95 (91–99)	3 (–2–7)
Satisfaction with office staff	91 (87–95)	94 (91–97)	3 (2–8)	92 (89–94)	95 (91–99)	3 (–2–8)

^a Patients were asked to state their overall satisfaction with their breast after reconstruction compared to before reconstruction as *less*, *the same* or *more*. If both breasts were reconstructed, they should answer with regard to the breast they were the least satisfied with. Results for patients with *the same* satisfaction are not shown.

^b Complications were patient-reported and not validated.

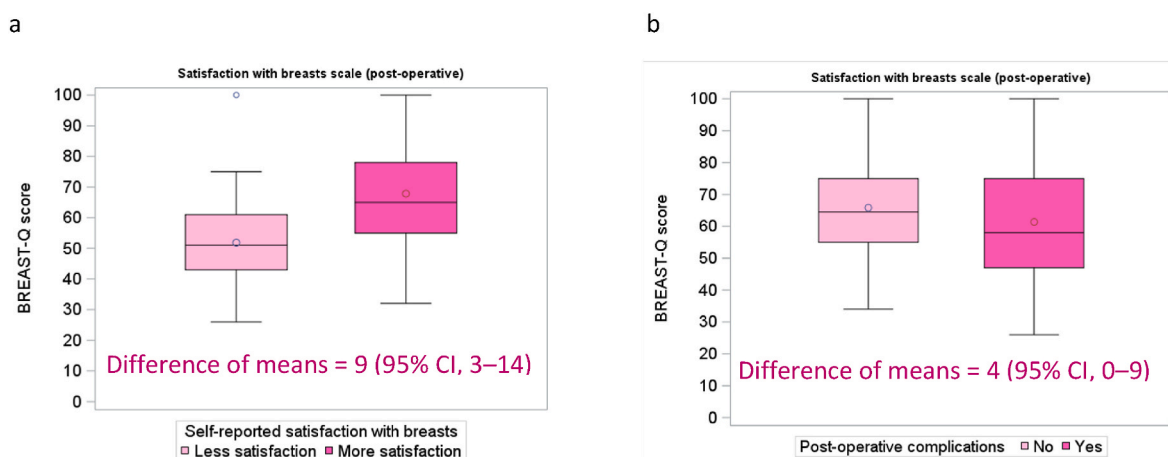


Fig. 3. Box plots of the known-group analyses show the BREAST-Q-scores for the groups with more ($n = 90$) and less ($n = 61$) postoperative self-reported satisfaction compared with before surgery (a) and the groups with ($n = 126$) and without ($n = 66$) one or more self-reported surgical complications (b) for the Satisfaction with breasts scale. Results from the group that reported the same post-operative self-reported satisfaction with breasts compared with before surgery ($n = 41$) in (a) are not shown.

conducting RA in five scales, and in the preoperative *Sexual well-being* scale ($n = 77$) and the postoperative *Physical well-being: abdomen* scale ($n = 45$), the results should be interpreted with caution. Another limitation is that we primarily recruited patients from one institution, and the study population may, therefore, not be representative of Denmark as a whole. However, Denmark only has 6.0 million inhabitants [37], and even though cultural differences exist across the country, we expect that the impact is negligible. Additionally, previous BREAST-Q validations have examined response category threshold order [5]. Our sample size is substantially smaller, making it difficult to conclude from such analyses. Also, we have not accounted for the type of breast reconstruction in our known-group analyses that examine the interpretability of the BREAST-Q. Autologous reconstruction has been associated with higher satisfaction with breasts but also a higher number of complications [38]. Lastly, we could not evaluate responsiveness (the instrument's ability to measure clinically important change over time in the measured construct) since only a small subgroup ($n = 35$) completed both surveys, and we did not examine test-retest reliability.

The study's strengths include a comprehensive RA comprising analyses of LD and DIF and taking disclosures of such into account in the remaining reliability and validity analyses. RA has several advantages, such as independence from the test sample [10]. In addition, we included both pre-and postoperative patients and validated scales that are identical pre-and postoperatively separately, which has not been done before.

A recent update on the content validity of the breast cancer modules in BREAST-Q found that the questionnaire is still very relevant for these patient groups [7]. However, to ensure comprehensiveness, new scales for upper extremity lymphedema, breast sensation, fatigue, cancer worry, and work impact have been developed [7,34,39,40]. Danish translation and validation of these additional scales would also be of great value. International use of the BREAST-Q may provide great insight into the well-being and satisfaction of women going through breast surgery and allow for comparison across cultures. We welcome other psychometric validations using RA, particularly including an assessment of LD and DIF, and a similar validation of the remaining translated Danish BREAST-Q modules is warranted. An evaluation of the responsiveness could add to the clinical interpretation of the scores. In addition, establishing normative BREAST-Q scores in a Danish population, as has been done in other countries, would add value to the clinical use since they vary across cultures [41–46].

6. Conclusion

The Danish BREAST-Q reconstruction module has now been both linguistically and psychometrically validated. It has demonstrated good acceptability, feasibility, validity, and adequate reliability and can be used in clinical practice and research. However, the evidence of LD and floor or ceiling effects in several scales reduces the measurement precision.

CRediT authorship contribution statement

Cecilie Balslev Willert: Writing – review & editing, Writing – original draft, Resources, Project administration, Methodology, Funding acquisition, Formal analysis, Conceptualization. **Karl Bang Christensen:** Writing – review & editing, Writing – original draft, Validation, Supervision, Methodology, Formal analysis. **Pernille Envold Bidstrup:** Writing – review & editing, Supervision. **Lene Mellemkjær:** Writing – review & editing, Supervision. **Anne-Marie Axø Gerdes:** Writing – review & editing, Supervision. **Lisbet Rosenkrantz Hölmich:** Writing – review & editing, Writing – original draft, Supervision, Conceptualization.

Ethical approval

Approval from the Danish Health Research Ethics Committee was not necessary.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the authors used Grammarly Premium in order to improve readability and language. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Funding

This work was supported by Trygfonden (grant number 123060). Trygfonden had no involvement in the study.

Conflict of interests

The authors have no conflict of interests.

Acknowledgements

Thank you to Professor Emeritus Niels Kroman at Clinic of Breast Surgery, Copenhagen University Hospital, Herlev and Gentofte, Denmark for supervision. Thank you to the outpatient clinics at Department of Plastic Surgery, Copenhagen University Hospital, Herlev and Gentofte Hospital and at Department of Plastic Surgery, Copenhagen University Hospital, Rigshospitalet for assisting with the inclusion of the participants.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.breast.2025.103872>.

References

- [1] Jonczyk MM, Jean J, Graham R, Chatterjee A. Surgical trends in breast cancer: a rise in novel operative treatment options over a 12 year analysis. *Breast Cancer Res Treat* 2019;173:267–74. <https://doi.org/10.1007/s10549-018-5018-1>.
- [2] Alborno CR, Cordeiro PG, Pusic AL, McCarthy CM, Mehrara BJ, Disa JJ, et al. Diminishing relative contraindications for immediate breast reconstruction: a multicenter study. *J Am Coll Surg* 2014;219:788–95. <https://doi.org/10.1016/J.JAMCOLLSURG.2014.05.012>.
- [3] Pusic AL, Klassen AF, Scott AM, Klok JA, Cordeiro PG, Cano SJ. Development of a new patient-reported outcome measure for breast surgery: the BREAST-Q. *Plast Reconstr Surg* 2009;124:345–53. <https://doi.org/10.1097/PRS.0b013e3181ae807>.
- [4] Fuzesi S, Cano SJ, Klassen AF, Atisha D, Pusic AL. Validation of the electronic version of the BREAST-Q in the army of women study. *Breast* 2017;33:44–9. <https://doi.org/10.1016/j.breast.2017.02.015>.
- [5] Cano SJ, Klassen AF, Scott AM, Cordeiro PG, Pusic AL. The BREAST-Q: further validation in independent clinical samples. *Plast Reconstr Surg* 2012;129:293–302. <https://doi.org/10.1097/PRS.0b013e31823a2ec6b>.
- [6] Cohen WA, Mundy LR, Ballard TNS, Klassen A, Cano SJ, Browne J, et al. The BREAST-Q in surgical research: a review of the literature 2009–2015. *J Plast Reconstr Aesthet Surg* 2016;69:149–62. <https://doi.org/10.1016/j.bjps.2015.11.013>.
- [7] Kaur MN, Chan S, Bordeleau L, Zhong T, Tsangaris E, Pusic AL, et al. RESEARCH Open Access Re-examining content validity of the BREAST-Q more than a decade later to determine relevance and comprehensiveness. *J Patient Rep Outcomes* 2023;7:37. <https://doi.org/10.1186/s41687-023-00558-y>.
- [8] Willert CB, Gjørup CA, Hølmich LR. Danish translation and linguistic validation of the BREAST-Q. *Dan Med J* 2020;67.
- [9] Beaton DE, Bombardier C, Guillemin F, Ferraz MB. Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine* 2000;25:3186–91.
- [10] Cano SJ, Hobart JC. The problem with health measurement. *Patient Prefer Adherence* 2011;5:279–90. <https://doi.org/10.2147/PPA.S14399>.
- [11] Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)—a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform* 2009;42:377–81. <https://doi.org/10.1016/j.jbi.2008.08.010>.
- [12] Association WM. World medical association declaration of Helsinki: ethical principles for medical research involving human subjects. *JAMA* 2013;310:2191–4. <https://doi.org/10.1001/JAMA.2013.281053>.
- [13] Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *J Clin Epidemiol* n.d.;63. <https://doi.org/10.1016/j.jclinepi.2010.02.006>.
- [14] Terwee CB, Bot SDM, de Boer MR, van der Windt DAWM, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaire. *J Clin Epidemiol* 2007;60:34–42.
- [15] Cronbach LJ. Coefficient alpha and the internal structure of tests. *Psychometrika* 1951;16:297–334. <https://doi.org/10.1007/BF02310555>.
- [16] Streiner DL, Norman GR, Cairney J. *Health Measurement Scales: a practical guide to their development and use*. Oxford University Press; 2015. <https://doi.org/10.1093/MED/9780199685219.001.0001>.
- [17] Marais I. Local dependence. In: Christensen K, Kreiner S, Mesbah M, editors. *Rasch models in health*. Hoboken, NJ: ISTE Ltd/John Wiley and Sons Inc; 2013. p. 111–30.
- [18] Christensen KB, Makransky G, Horton M. Critical values for yen's Q3: identification of local dependence in the Rasch model using residual correlations. *Appl Psychol Meas* 2017;41:178–94. <https://doi.org/10.1177/0146621616677520>.
- [19] Wainer H, Kiely GL. Item clusters and computerized adaptive testing: a case for testlets. *J Educ Meas* 1987;24:185–201. <https://doi.org/10.1111/j.1745-3984.1987.tb00274.x>.
- [20] Kreiner S, Christensen KB. Overall tests of the Rasch model. In: Christensen K, Kreiner S, Mesbah M, editors. *Rasch models in health*. Hoboken, NJ USA: John Wiley & Sons, Inc; 2013. p. 105–10.
- [21] Christensen KB, Ks. Item fit statistics. In: Christensen K, Kreiner S, Mesbah M, editors. *Rasch models in health*. Hoboken, NJ USA: ISTE Ltd/John Wiley and Sons Inc; 2013. p. 83–104.
- [22] Buchardt A-S, Christensen KB, Jensen SN. Visualizing Rasch item fit using conditional item characteristic curves in R. *Psychol Test Assess Model* 2023;65:206–19.
- [23] Hochberg Y, Benjamini Y. More powerful procedures for multiple significance testing. *Stat Med* 1990;9:811–8. <https://doi.org/10.1002/sim.4780090710>.
- [24] Horton M, Marais I, Christensen KB. Dimensionality. In: Christensen K, Kreiner S, Mesbah M, editors. *Rasch models in health*. Hoboken, NJ USA: ISTE Ltd/John Wiley and Sons Inc; 2013. p. 137–58.
- [25] Team RC. R Development Core Team. R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2014. Google Scholar 2018.
- [26] Mueller M. Package: “iarm”: item analysis in Rasch models. 2022.
- [27] Mair P, Rusch T, Hatzinger R, Maier MJ, Debelak R. Maintainer. Package “eRm”: Extended Rasch modeling. 2024.
- [28] Robitzsch A. Package “sirt”: Supplementary item response theory models. 2024.
- [29] Draugalis JLR, Coons SJ, Plaza CM. Best practices for survey research Reports: a synopsis for authors and reviewers. *Am J Pharm Educ* 2008;72. <https://doi.org/10.5688/AJ720111>.
- [30] Stolpner I, Heil J, Feißt M, Karsten MM, Weber WP, Blohmer J-U, et al. Clinical validation of the BREAST-Q breast-conserving Therapy module. *Ann Surg Oncol* 2019;26:2759–67. <https://doi.org/10.1245/s10434-019-07456-y>.
- [31] Klassen AF, Dominici L, Fuzesi S, Cano SJ, Atisha D, Locklear T, et al. Development and validation of the BREAST-Q breast-conserving Therapy module. *Ann Surg Oncol* 2020;27:2238–47. <https://doi.org/10.1245/s10434-019-08195-w>.
- [32] Saiga M, Taira N, Kimata Y, Watanabe S, Mukai Y, Shimozuma K, et al. Development of a Japanese version of the BREAST-Q and the traditional psychometric test of the mastectomy module for the assessment of HRQOL and patient satisfaction following breast surgery. *Breast Cancer* 2017;24:288–98. <https://doi.org/10.1007/s12282-016-0703-6>.
- [33] Poulsen L, Klassen A, Rose M, Roessler KK, Juhl CB, Støving RK, et al. Psychometric validation of the BODY-Q in Danish patients undergoing weight loss and body contouring surgery. *Plast Reconstr Surg Glob Open* 2017;5. <https://doi.org/10.1097/GOX.0000000000001529>.
- [34] Klassen AF, Kaur MN, Tsangaris E, de Vries CEE, Bordeleau L, Zhong T, et al. Development and psychometric validation of BREAST-Q scales measuring cancer worry, fatigue, and impact on work. *Ann Surg Oncol* 2021;28:7410–20. <https://doi.org/10.1245/S10434-021-10090-2/FIGURES/2>.
- [35] Hagell P, Westergren A. Sample size and statistical conclusions from tests of fit to the Rasch model according to the Rasch unidimensional measurement model (rumm) program in health outcome measurement. *J Appl Meas* 2016;17:416–31.
- [36] Stone M. The effect of sample size for estimating Rasch/IRT parameters with dichotomous items. *J Appl Meas* 2004;5:48–61.
- [37] Befolkningstal - Danmarks Statistik n.d. <https://www.dst.dk/da/Statistik/emner/borgere/befolkning/befolkningstal> (accessed June 20, 2024).
- [38] Sadok N, Krabbe-Timmerman IS, Buisman NH, van Aalst VC, de Bock GH, Werker PMN. Short-term quality of life after autologous compared with alloplastic breast reconstruction: a prospective study. *Plast Reconstr Surg* 2023;152:55S–68S. <https://doi.org/10.1097/PRS.00000000000010496>. 1963.
- [39] Klassen AF, Tsangaris E, Kaur MN, Poulsen L, Beelen LM, Jacobsen AL, et al. Development and psychometric validation of a patient-reported outcome measure for arm lymphedema: the LYMPH-Q upper extremity module. *Ann Surg Oncol* 2021;28:5166–82. <https://doi.org/10.1245/s10434-021-09887-y>.
- [40] Tsangaris E, Klassen AF, Kaur MN, Voineskos S, Bordeleau L, Zhong T, et al. Development and psychometric validation of the BREAST-Q sensation module for women undergoing post-mastectomy breast reconstruction. *Ann Surg Oncol* 2021;28:7842–53. <https://doi.org/10.1245/s10434-021-10094-y>.
- [41] Nelson JA, Chu JJ, McCarthy CM, Stern CS, Shamsunder MG, Pusic AL, et al. BREAST-Q react: clinical reference values for the BREAST-Q in post-mastectomy breast reconstruction patients. *Ann Surg Oncol* 2022;29:5280–93. <https://doi.org/10.1245/s10434-022-11521-4>.
- [42] Sadok N, Jansen L, de Zoete MD, van der Lei B, Werker PMN, de Bock GH. A Dutch cross-sectional population survey to explore satisfaction of women with their breasts. *Plast Reconstr Surg Glob Open* 2021;9. <https://doi.org/10.1097/GOX.0000000000004002>. 4002–e4002.
- [43] Crittenden TA, Smallman A, Dean NR. Normative data for the BREAST-Q Reconstruction module in an Australian population and comparison with US norms and breast reconstruction patient outcomes. *J Plast Reconstr Aesthetic Surg* 2022;75:2219–28. <https://doi.org/10.1016/j.bjps.2022.01.033>.
- [44] Klifto KM, Aravind P, Major M, Payne RM, Shen W, Rossion GD, et al. Establishing institution-specific normative data for the BREAST-Q reconstruction module: a prospective study. *Aesthet Surg J* 2020;40:NP348–N355. <https://doi.org/10.1093/asj/sjz296>.
- [45] Jepsen C, Paganini A, Hansson E. Normative BREAST-Q reconstruction scores for satisfaction and well-being of the breasts and potential donor sites: what are Swedish women of the general population satisfied/dissatisfied with? *J Plast Surg Hand Surg* 2023;58:124–31. <https://doi.org/10.2340/jphs.v58.15301>.
- [46] Mundy LR, Homa K, Klassen AF, Pusic AL, Kerrigan CL. Breast cancer and reconstruction: normative data for interpreting the BREAST-Q. *Plast Reconstr Surg* 2017;139:1046E–55E. <https://doi.org/10.1097/PRS.0000000000003241>. 1963.