Reproductive Health

REVIEW

Open Access

The psychometric properties of fear of childbirth instruments: a systematic review

Bibi Leila Hoseini[1], Abbas Ebadi[2], Ali Mashhadi[3], Mohammad Hassan Rakhshani[4] and Raheleh Babazadeh[5,6*]

Abstract

Background Anxiety disorders with a specific focus on fear of childbirth (FOC) are the most common mental health challenges in perinatal women. The accurate measurement of FOC is important for correctly identifying women with FOC, as well as for identifying the target population for treatment. We aimed to review FOC scales and evaluate their psychometric properties via the COSMIN methodology to identify the most suitable available instruments.

Methods We conducted this systematic review via a comprehensive search of databases, including PubMed, Web of Science, Scopus, Science Direct and ProQuest, to identify articles published from inception to May 2024 via combined keywords related to tools that assess FOC in women during pregnancy or postpartum period. The quality of the psychometric properties of the included studies was evaluated via the COSMIN checklist.

Results Of the 1160 records found initially, 47 articles were included in this review, 24 of which were related to the Wijma Delivery Expectancy/Experience Questionnaire (W-DEQ). According to the recommended categorization of the COSMIN methodology, among the 18 assessed scales, the Fear of Childbirth Questionnaire (FCQ) was categorized as A, and 11 scales, including the Fear-of-delivery Questionnaire (FDQ), W-DEQ-A & B, Delivery Fear Scale (DFS), Fear of Birth Scale (FOBS), Birth Anticipation Scale (BAS), Childbirth Fear Questionnaire (CFQ), Childbirth Fear Scale (CSF), Slade–Pais Expectations of Childbirth Scale (SPECS), and unnamed tools by Melender et al. (2005) and Eriksson et al. (2005) were categorized as B.

Conclusion According to the findings, the FCQ can be recommended for evaluating pregnant women with FOC. The measures categorized as B are potentially recommended for use, but further research is needed to evaluate the quality of this group.

Keywords Fear, Childbirth, Tokophobia, Scale, Psychometric, Pregnant women

*Correspondence:
Raheleh Babazadeh
babazadehr@mums.ac.ir; raheleh621@yahoo.com
[1] Department of Midwifery and Reproductive Health, Student Research Committee, School of Nursing & Midwifery, Mashhad University of Medical Sciences, Mashhad, Iran
[2] Nursing Care Research Center, Clinical Sciences Institute, Baqiyatallah University of Medical Sciences, Tehran, Iran
[3] Psychology & Clinical Neuropsychology, Department of Psychology, Ferdowsi University of Mashhad (FUM), Mashhad, Iran
[4] Department of Biostatistics & Epidemiology, School of Public Health, Sabzevar University of Medical Sciences, Sabzevar, Iran
[5] Nursing and Midwifery Care Research Center, Mashhad University of Medical Sciences, Mashhad, Iran
[6] Department of Midwifery, School of Nursing and Midwifery, Mashhad University of Medical Sciences, Mashhad, Iran

Background

For many women, perinatal transition is accompanied by hope and excitement, but it can concurrently be a frightening and negative experience. Anxiety-related disorders are the most common mental health problems experienced by women during the perinatal period. Childbirth fear can be a very specific focus of anxiety [1].

Fear of childbirth (FOC) was first described by Louis Victor Marcé, a French psychiatrist (1858) [2, 3]. It has been defined as a health subject for pregnant women related to an anxiety disorder or a phobic fear interfering with daily functioning and well-being, including physical complications, concentration disorders, nightmares, and

requests for cesarean section (CS). However, more moderate fear was defined as significant anxiety that had not interfered with the pregnant women's daily activities [3–5]. Another idiom is "tokophobia", which is characterized by unreasoning FOC, a particular and hurtful condition that includes "pathological dread" and "childbirth avoidance" [3, 6].

Worldwide, FOC rates vary from 6.3 to 14.8% in different countries [3]. In addition, the prevalence of FOC differs among various regions of a country, ranging from 20 to 61.2% in Africa [7] to 17.3–89.30% in different areas of Iran [8]. The reasons for these variances are unknown; however, various measurements of FOC, in addition to cultural differences, can play a role in different prevalence rates [2, 9–11].

The health and wellbeing of women are affected by FOC, which is a common problem in the perinatal period. In addition, FOC has several life-long consequences for women, including poor relationships with the baby [3, 11–13], the partner and the family [3, 11]. FOC can lead to the inability to cope with childbirth, which may lead to the avoidance of pregnancy, pregnancy termination, increased levels of perceived pain during labor and childbirth, and obstetric complications, including increased length of labor, increased likelihood of emergency cesarean delivery, decreased childbirth satisfaction, postpartum depression [12, 14], and posttraumatic stress disorder (PTSD) [9, 14]. Furthermore, it can often lead to elective cesarean section (CS) by women who are escaping from an exposed situation [3, 9].

However, there are opportunities to reduce FOC and the abovementioned negative outcomes, such as psychotherapy and educational interventions, including counseling provided by maternity care providers or educational programs on childbirth at prenatal care centers or hospitals [14]. The accurate measurement of FOC is very important for correctly identifying those experiencing FOC, as well as for identifying the target population for treatment [14]. There are several measures for FOC according to the literature, ranging from one-item tools such as those described by Laursen et al. [15] and Rouhe et al [16] to 53-item tools such as Melender et al.'s tool (2002) [17]. It seems that some of these instruments are invalid for use because they have not offered direct evidence of measurement properties [18]. On the other hand, it is not clear which of the other tools is more valid and reliable for use in clinical practice. Applying measures of poor or unknown quality is unethical and wastes resources. Selection of the best measurement for the outcome of interest requires first high-quality studies with the psychometric properties of the outcome measurement in the target population and, second, a high-quality systematic review of the studies with psychometric properties of the measurement, including all the required information in a transparent way [19].

It is important to assess the quality of studies' methodology with a specific tool so that we can trust the results. The COnsensus-based Standards for the Selection of Health Measurement INstruments (COSMIN) checklist is a tool that evaluates the quality of studies' methodology in terms of measurement properties [18]. Recently, the COSMIN checklist has been widely used in systematic reviews of outcome measurements [19].

According to the literature review, we found one systematic review on FOC instruments by Richens et al. in 2018 [20]. However, this research did not include several recent studies, which are among the most important measurements among FOC studies. Moreover, Richens et al. appraised the included studies via the tool developed by Hawker et al. (2002) [21], which is not a specialized tool for outcome measurement. Another systematic review in this field was conducted by Zhao et al. (2022) in Chinese, which assessed only five FOC Patient Reported Outcome Measures (PROMs) [22]. However, many more instruments are available for evaluating FOC, and we assess all FOC instruments that were developed for the first time. Moreover, its Chinese language makes it difficult to apply. Therefore, it is necessary to conduct a comprehensive and high-quality systematic review to identify FOC scales and evaluate their measurement properties via the COSMIN methodology to identify the most suitable available instruments.

Methods

This systematic review was reported according to the COSMIN Guideline for Systematic Reviews of Measurement Properties of PROMs [18], the COSMIN methodology for evaluating the content validity of PROMs [23], the COSMIN Methodology for Systematic Review of PROMs, the COSMIN Risk of Bias Checklist for Systematic Reviews of PROMs and the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 statement [24]. The protocol of the present systematic review was registered in the International Prospective Register for Systematic Review (PROSPERO) with the registration code CRD42024522599.

Eligibility criteria

Eligibility criteria were determined according to the key elements of the aim of the PROMs, including measuring the construct of interest, i.e., FOC, especially for the first time; the target population, including pregnant or postpartum women; and the research reporting psychometric characteristics or instrument revision or translated PROMs. We excluded studies that focused only on

PROM as an outcome measurement tool, such as clinical trials.

Information sources

A comprehensive search of global databases, including PubMed, Web of Science, Scopus, and Science Direct, was performed. Further searches were conducted in ProQuest Theses and Dissertations and in Google Scholar for gray literature. Finally, the references of the retrieved articles were also manually searched.

Search strategy

We combined keywords related to tools that assess FOC in pregnant or postpartum women. For the present systematic review, all the references were searched with search terms in MeSH, including #1 Construct search ((Fear*) OR (Phobia*)) AND ((Parturition*) OR (Birth*) OR (Childbirth*) OR (Obstetric Delivery) OR (Obstetric Deliveries)) OR Tocophobia AND #2 Population ((Pregnant wom*n) OR (pregnanc*) OR (postpartum wom*n) OR (Puerperium) OR (postpartum)) AND #3 Instrument search ((Questionnaire*) OR (Measure*) OR (measurement*) OR (Scale*) OR (Survey*)) AND #4 ((psychometric*) OR (reliability) OR (valid*) OR (Development)). The authors assessed the studies published from inception to March 2024. The search was updated for each source in May 2024. The full texts of the eligible studies were retrieved.

Selection process

We used Endnote X8 software to manage the identified studies. After duplicate articles were removed, two researchers independently screened the identified articles. The authors removed irrelevant studies. In case of any disagreements, they discussed and agreed by consensus with the other researchers.

Data collection process

Two reviewers independently extracted the required data from the eligible studies. They discussed and agreed about the collected data.

Data items

The extracted data consisted of the PROM name, author's name, year of publication, country, target population, item generation, mode of administration (self-report, interview), response options, number of items, dimensions of the scale, range of scores/scoring, available translations and measurement properties (Table 1).

Assessment the quality of measurement properties of the scales

Two reviewers independently assessed the quality of the included PROMs, with the checklist designated by the CONSensus-based Standards for the Selection of Health Measurement INstruments (COSMIN). The authors agreed by consensus in controversial cases through discussion with the senior researchers. This checklist contains 10 boxes that assess "the methodological quality of the included studies" and "the quality of their psychometric properties". To evaluate the methodological quality of the included studies, the authors used the user manual of the COSMIN methodology for assessing the content validity of the PROMs (to assess PROM development and content validity, i.e., boxes 1 and 2) [25]. Then, they assessed the other measurement properties (i.e., structural validity, internal consistency, measurement invariance/cross-cultural validity (boxes 3–5 refer to the internal structure of the PROM), reliability, measurement error, criterion validity, hypothesis testing and responsiveness (boxes 6–10)) on the basis of the COSMIN risk of bias checklist via the user manual of the COSMIN methodology for systematic reviews of PROMs [19, 26]. Each box is evaluated with a four-point Likert scale, including very good (V), adequate (A), doubtful (D) and inadequate (I). The final evaluation of each box was assessed on the basis of the 'worst score counts' principle [19].

Internal consistency was assessed by Cronbach's alpha for unidimensional scales or each subscale of multidimensional scales. According to the COSMIN checklist, it is preferable for the review team to formulate hypotheses to test hypotheses for construct validity. Several generic hypotheses are suggested in the manual. In this way, we are able to compare all related records of the included studies with the same hypotheses. Here are some generic hypotheses (H) in the manual that we decided to use, H1. Correlations with convergent instruments ≥ 0.50; H2. Correlations with divergent instruments $= 0.30–0.50$; and H3. Correlations with tools measuring unrelated constructs < 0.30) [18, 26].

Assessment the quality of measurement properties of the scales

The quality criteria for measuring the properties of the childbirth instruments were assessed in accordance with the updated criteria for good measurement properties. The rating of each property for a single study is as follows: insufficient (−), sufficient (+) or indeterminate (?) [18].

Table 1 Characteristics of the included PROMs

PROM	Authors (year)	Country	Target population	Item generation	Mode of Administration	Response options	N. of scales/ dimensions (N. of item)	Range of scores	Available translations	Measurement properties
Fear of child-birth question-naire	Areskog et al. [37]	Sweden	Pregnant women 19–38 y	–	Self-report	Yes/No	1 dimension (19 items)	–	Swidish– Eng-lish– Finnish	–
Fear–of–deliv-ery question-naire (FDQ) (A revised questionnaire of Areskog et al.)	Saisto et al. [34]	Finland	Pregnant women and their partners	–	Interview	Yes/No	1 dimension (10 items)	–		Cronbach's α = 0.76; Correlation: FDQ–PAS (r = 0.65, p < 0.001)
Wijma Delivery Expectancy/ Experience Questionnaire A (w–DEQ–A)	Wjma et al. [31]	Sweden	Pregnant women18–39 y	Qualitative research with Two authors' clinical experiences of women with FOC	Self-report	Six–point Likert: (0 = not at All to 5 = extremely"	1 dimension (33 items)	–165 (items positively formulated including 2, 3, 6, 7, 8, 11, 12, 15, 19, 20, 24, 25, 27, 31)	Sweden (Swed-ish), Iran (Farsi), Italy (Italian), Malawi (Chichew), Tur-key (Turkish), China (Chinese), Belgium (Flem-ish), Iceland (Icelandic), Denmark (Dan-ish), Estonia (Estonian), Rus-sian, Norway, (Norwegian), Hungary (Hungarian), Spain, (Spanish), Kenya, (Swahili), Slovakia, (Slo-vak), Japan (Japanese), Tan-zania (Kiswahili), Germany (Ger-man), Portugal (European Portuguese), Greece (Greek)	Nulliparous & Parous groups, respectively: Correlation of Wijma–A– SRI (r = 0.52; r = 0.65; P < 0.0001) Wijma–A—FQ (child-birth) (r = 0.43; r = 0.78; P < 0.0001); Wijma–A—FQ (agora, social, injury, gyneco-logical examination, elevator, darkness) (ranging from r = 0.07 in dark-ness to 0.34 in agora; r = 0.2 in elevator to r = 0.44 in social; P < 0.0001); Wijma–A—STAI (r = 0.54 & r = 0.55; P < 0.0001); Wijma –A–KSP (r = 0.43 & r = 0.47; P < 0.0001); Wijma–A –BDI Cronbach's α: α = 0.93; Split–half (α = 1.00)

Table 1 (continued)

PROM	Authors (year)	Country	Target population	Item generation	Mode of Administration	Response options	N. of scales/ dimensions (N. of item)	Range of scores	Available translations	Measurement properties
Wijma Delivery Expectancy/ Experience Questionnaire B (w–DEQ–B)	Wjma et al. [31]	Sweden	Pregnant women18–39 y	Qualitative research with Two authors' clinical experiences of women with FOC	Self-report	Six–point Likert: (0 = not at All to 5 = extremely"	1 dimension (33 items)	0–165 (items positively formulated including 2, 3, 6, 7, 8, 11, 12, 15, 19, 20, 24, 25, 27, 31)	Sweden (Swedish), Iran (Farsi), Italy, (Italian), India (Hindi), Turkey (Turkish), China (Chinese), Spain (Spanish), Japan (Japanese), Tanzania (Kiswahili)	Cronbach's α: (2 h after delivery: α = 0.93; 5 w after delivery: α = 0.94) Split–half (2 h after delivery: α = 0.95; 5 w after delivery: α = 0.96)

Table 1 (continued)

PROM	Authors (year)	Country	Target population	Item generation	Mode of Administration	Response options	N. of scales/ dimensions (N. of item)	Range of scores	Available translations	Measurement properties
Delivery Fear Scale (DFS)	Wjma', et al. [32]	Sweden	Pregnant women in labor ≥ 19 y	A list of 60 items written by 2 experts and then comment by 8 experienced midwives	Self–report	1– 10 (1 = 'do not agree at all', 10 = 'agree totally')	1 dimension (10 items)	1–100	English– Swedish– Turkish– Farsi	Cronbach's α = 0.88
	Sercekus et al. [35]	Turkey	Pregnant women in labor	–	Interview	–	1 dimension (10 items)	1–100	–	EFA (factor loading .63 to .83, total variance: 52.3%) CFA (n = 96, RMSEA = 0.000, P = 0.631); Cronbach α = 0.90; Split–half: α = 0.81; Correlation: DFS– STAI: (r = 0.80, p < .001)
	Shakarami et al. [36]	Iran	Pregnant women in maternity care	–	Interview	–	2 dimensions (10 items)	1–100	–	CFA (n = 200, RMSEA = 0.034, P < 0.001, CFI = 0.994, TLI = 0.992); Cronbach α = 0.77; Split–half: α = 0.83; Correlation: DFS–PRAQ: (r = 0.74, P <); DFS–CAQ: (r = 0.72, P < 0.001); DFS– STAI–Y1 (r = 0.71, P < 0.001); DFS– STAI–Y1 (r = 0.63, P < 0.001); DFS–short form of Lowe's childbirth self–efficacy inventory (r = –0.75 & –0.76, P < 0.001)

Table 1 (continued)

PROM	Authors (year)	Country	Target population	Item generation	Mode of Administration	Response options	N. of scales/ dimensions (N. of item)	Range of scores	Available translations	Measurement properties
Unnamed	Melender et al. [29]	Finland	–	Literature review and semi-structured interviews with 20 postpartum women 19–37 y	–	–	–	–	–	Content analysis
	Melender et al. [17]	Finland	Pregnant women 17–44 y	–	Self–report	4–point Likert: (1 = agree, 2 = agree to some extent, 3 = disagree to some extent, 4 = do not agree); and a dichotomous scale (1 = yes, 2 = no)	3 dimensions (Objects, Causes, and Manifestation of Fears) (53 items)	–		Factor analysis: (n = 329; total variance = 63.7) Cronbach's α = 0.91 (objects); α = 0.70 (causes); α = 0.78 (manifestations)
Unnamed	Eriksson et al. [28]	Sweden	The parents aged ≥ 20 y with an at least one infant	Literature review and Grounded theory with 3 women experienced intensive childbirth fear	Self–report	4–point Likert: totally disagree = 1, partly disagree = 2, partly agree = 3 and totally agree = 4	4 dimension including 'exposedness and inferiority', 'communicative difficulties', 'norms of harmony' and 'insecurity and danger' (29 items)	–		Factor analysis (N women = 328); Variance of four factors and total variance (women), repectively = 20– 15– 11– 6%, 52%) Cronbach's α: varied from 0.72 to 0.49 for women, for different factors
Unnamed	Waldenstro¨m et al. [40]	Sweden	Pregnant women of childbearing age	–	Self–report	very positive, fairly positive, mixed feelings, rather negative and very negative	1 question	–		–

Table 1 (continued)

PROM	Authors (year)	Country	Target population	Item generation	Mode of Administration	Response options	N. of scales/ dimensions (N. of item)	Range of scores	Available translations	Measurement properties
Unnamed	Laursen et al. [15]	Denmark	Pregnant women who were fluent in Danish	NR	Interview	Not at all', 'A little' or 'A lot'. Only the last response was considered to represent FOC	1 item	–		Odds ratio: Symptoms of anxiety: 4.80–fold in women with fear– Symptoms of depression: 2.70–fold in women with fear
Visual analog scale (VAS)	Rouhe et al. [16]	Finland	Pregnant women 16–47 y	–	Self–report	1– 10	1 item	0–10		N = 1400; Correlation: VAS– W–DEQ A (r = 0.7, P = 0.01); In VAS threshold was 6.0: Sensitivity = 89.2%; Specificity = 76.3%
Fear of Birth Scale (FOBS)	Haines et al. [71]	Sweden	Swedish and Australian pregnant women	–	Self–report	Two10–cm lines anchors (a) 'calm' and 'worried' and (b) 'no fear' and 'strong fear'	1 dimension (A two–item visual analog scale)	0– 100		Cronbach's α: 0.91
	Zhang et al. [72]	China	Chinese pregnant women	–	Self–report	Two 10–cm lines anchors (a) "no worry/strong worry"	1 dimension (A two–item visual analog scale)	0– 100	–	Cronbach's α: 0.897; Test–retest reliability (0.860); I–CVI of Q1 & Q2: 0.933 and 0.800; S–CVI: 0.867; Correlation: FOBS–CAQ; FOBS– Wijma; FOBS–EPDS; FOBS–GAD–7
Numeric rating scale (NRS)	Storksen et al. [39]	Norway	Pregnant women 18–45 y	–	Self–report	0 (not at all) to 10 ('extremely much')	1 dimension (1 item)	0–10		N = 1642; Correlation: NRS–the W–DEQ (R = 0.57); NRS– the SCL–anxiety scores and the EPDS scores (r = 0.29 for both)
Birth Anticipation Scale (BAS)	Elvander et al. [41]	USA– Pennsylvania	Nulliparous pregnant women aged 18–35	–	Self– report	extremely, quite a bit, moderately, a little, and not at all	1 dimension (6 items)	6–30		N = 3006; Cronbach's α = 0.82

Table 1 (continued)

PROM	Authors (year)	Country	Target population	Item generation	Mode of Administration	Response options	N. of scales/ dimensions (N. of item)	Range of scores	Available translations	Measurement properties
Slade–Pais Expectations of Childbirth Scale (SPECS)	Slade et al. [27]	UK	England pregnant women 17– 39 y	Semis-tructured interviews with 18 Pregnant women 10–38 w aged 17–39 y	Self– report	5–point Likert: 1 (strongly agree), 2 (agree), 3 (undecided), 4 (disagree), 5 (strongly disagree)	6 dimensions including coping and robustness to pain, staff and service responsive to needs, out of control and embarrassed, partner's coping, positive anticipation of birth. (50 items)	Reverse scoring of 34 items of the scale		PCA: (n = 148; Variance of 42.5%) Cronbach's α = 0.89 (for components ranged from 0.77– 0.86); The correlation between SPECS and STAI including state anxiety (r = 0.43, p < 0.001) and trait anxiety (*r* = .38, *p* < .001)
Fear of childbirth	Prelog et al. [38]	Slovenia	Nulliparas Pregnant women ≥ 18 y	–	Self– report	5–point Likert: 1 (not at all) to 5 (very much)	1 dimension (6 items)	–		Exploratory factor analysis (N = 325, all six items loaded substantially (0.37–0.79) on a single factor); Cronbach α = 0.82; The correlations amongFOC–anxiety (r = 0.37; p < 0.001) FOC–depression (r = 0.36; p < 0.001)

Table 1 (continued)

PROM	Authors (year)	Country	Target population	Item generation	Mode of Administration	Response options	N. of scales/ dimensions (N. of item)	Range of scores	Available translations	Measurement properties
Fear of Childbirth Questionnaire (FCQ)	Slade et al. [30]	UK	–	Semistructured interviews with pregnant women 25– 43 y (n = 10) and consultant midwives (n = 13) & published meta–synthesis	–	–	–	–		Thematic analysis
	Slade et al. [1]	UK	Pregnant women 25–43 y	–	Self– report	4–point Likert: Strongly disagree (0), Slightly Disagree (1), Slightly Agree (2), Strongly agree (3)	1 dimension (20 items)	0–60		–
	Sanjari et al. [43]	Iran	Pregnant women	–			4 dimensions including "uncertainty and injury", with 18.39%, the "unprofessional behavior of maternity staff" with 14.51%, "the unpredictable" with 14.44%, and "negative emotions" with 10.54% of the variance			EFA (N = 400; total variance = 57.87%; the variance of the first to fourth factor = 18.39%, 14.51%, 14.44, and 10.54%, respectively); CFA (N = 200; SRMR = 0.09, RMSEA = 0.09, CFI = 0.91); CVI = 0.83%; CVR for the 20 items = 81–100%; Cronbach α = 0.84 (for components ranged from 0.70 to 0.84); Test–Retest (r = 0.6, P < 0.01); Split–half > 0.7; Correlation: Slade scale– Wijma (r = 0.79, P > 0.01; Slade scale– CAQ = (r = 0.81;, P > 0.01)

Table 1 (continued)

PROM	Authors (year)	Country	Target population	Item generation	Mode of Administration	Response options	N. of scales/ dimensions (N. of item)	Range of scores	Available translations	Measurement properties
Childbirth Fear Questionnaire (CFQ)	Fairbrother et al. [14]	Canada, the United Kingdom, or the United States	English speaking, pregnant women ≥ 18 y	Literature review	Self– report	5–point Likert: 0 (not at all) to 4 (extremely)	9 dimensions including Fear of loss of sexual pleasure/ attractiveness; Fear of pain from a vaginal birth; Fear of medical interventions; Fear of embarrassment; Fear of harm to baby; Fear of cesarean birth; Fear of mum or baby dying; Fear or insufficient pain medication; Fear of body damage from a vaginal birth (40 items)	0–160	English Spanish	EFA (N = 643; CFI = 0.977, TLI = 0.974, RMSEA = 0.064 (90% CI: 0.062, 0.066), SRMR = 0.055); CFA model and invariance between parity (primiparous vs. multiparous) groups: (SRMR = 0.06, $\Delta X^2 = -$, $\Delta df = -$); Cronbach α = 0.94 (for components ranged from 0.71 to 0.94); The correlations CFQ–WDEQ–A (fear scales): r = 0.57 (p < 0.001); CFQ –EPDS: r = 0.35 (p < 0.001), CFQ– MQ: r = 0.28 (p < 0.001); CFQ–PDS–5: r = 0.24 (p = 0.001)
	González–de la Torre et al. [42]	Spain	Pregnant women ≥ 18	–	Self– report	5–point Likert: 0 (not at all) to 4 (extremely)	4 dimensions including fear of medical Interventions, fear of harm and dying, fear of pain, fears relating to sexual aspects, and embarrassment (37 items)	0–148		CVI–Total = 0.77; EFA (n = 279, RMSEA = 0.000); CFA (n = 278, RMSEA = 0.022); Cronbach α = 0.947 (for components ranged from 0.71 to 0.94); Total omega coefficient = 0.945

Table 1 (continued)

PROM	Authors (year)	Country	Target population	Item generation	Mode of Administration	Response options	N. of scales/ dimensions (N. of item)	Range of scores	Available translations	Measurement properties
Childbirth Fear Scale (CFS)	Nuraliyeva et al. [33]	Turkey	Women of childbearing age 18– 49 y including pregnant women	Literature review and the most frequently used Scales		5–point Likert: strongly agree and strongly disagree	3 dimensions including "Fear of Pregnancy, Childbirth, and Maternal Role"; "Fear of Inability to Meet Physical and Social Needs"; "Fear of Pregnancy and Childbirth problems" (20 items)	20–100 (The positive items include 4, 7, 9, 11, 13, 16, 18, and 19. The negative items coded reversely include: 1, 2, 3, 5, 6, 8, 10, 12, 14, 15, 17, and 20)	–	PCA (n = 500; KMO = 0.88, ×2 = 3673.824 in Bartlett's test of sphericity; The variance of the same 3 factors = from 8.02% to 29.54%, Total variance = 51.93%); Cronbach α = 0.86 (for the factors 0.88, 0.76, 0.75); Test–retest reliability (r = 0.88; *P* = .000); CFS– WCF–PPS (r = 0.53, P < 0.01

The S-R Inventory of Anxiousness (SRI); Fear of Childbirth (FOC); The Fear Questionnaire (FQ); The State-Trait Anxiety Inventory (STAI); The Karolinska Scales of Personality (KSP); Beck's Depression Inventory (BDI); Edinburgh Postnatal Depression Scale (EPDS); Mutilation Questionnaire (MQ); Pregnancy-related anxiety (PAS); Quality of Life (QOL); Total content validity (CVI-T); Content validity index per expert (CVI-E); Content validity of each item (CVI-I); posttraumatic stress disorder (the PDS-5); Childbirth Attitude Questionnaire (CAQ); Pregnancy-related anxiety questionnaire (PRAQ); Generalized Anxiety Disorder Scale (GAD-7); WCF-PPS, Women Childbirth Fear–Prior to Pregnancy Scale

Synthesis

First, the overall measurement property of the instrument (insufficient (−), sufficient (+), inconsistent (±) or indeterminate (?)), the results of all included studies on a measurement property were qualitatively summarized for comparison against the criteria for good measurement properties. In the overall rating step, we focused on the PROM, whereas the focus in the previous steps was on single studies.

The quality of the summarized evidence was subsequently rated with the Grading of Recommendations Assessment, Development and Evaluation (GRADE) approach, which indicates how trustworthy the summarized results are (high, moderate, low or very low) [18]. For content validity evaluation, three factors were applied: risk of bias, indirectness and inconsistency. To evaluate the other measurement properties, one more factor, imprecision (the total sample size of the included studies), was taken into account [25, 26].

Finally, the included PROMs were categorized into three recommended categories: (A) the most appropriate tools for sufficient content validity (at any level) AND at least low-level evidence for sufficient internal consistency; (B) the instruments not categorized into any A or C categories; and C) the instruments with high quality evidence for insufficient measurement properties [18].

All three abovementioned steps were independently assessed by two reviewers.

Results

Study selection

We obtained 1160 articles by searching several databases. After duplications were removed, two reviewers screened 934 remaining studies on the basis of title and abstract. Among them, 56 articles were potentially relevant to the aim of the current study. In the next step, their eligibility was assessed by retrieving the full texts. Finally, 47 eligible studies were included in the review, 24 of which were related to the Wijma Delivery Expectancy/Experience Questionnaire (W-DEQ) scale and its translated versions (Fig. 1).

Study characteristics

The features of the included studies are described in Table 1. The ages of the included women were mostly above 18 years, except for those in the Rouhe et al. and Slade et al. (2016) studies [16, 27]. Among the included PROMs, only eight reported the origin of the item generation. Five out of eight PROMs performed qualitative research to generate items [27–31], three of which involved semi-structured interviews [27, 29, 30], one used grounded theory [28], and one used an unknown method involving interviews with two authors about the clinical experiences of women who feared childbirth [31]. Wijma et al. (2002) composed a 60-item list of negative and positive items expressing fear-related appraisals and their contrasts [32]. Two out of eight PROMs were identified on the basis of a literature review [14, 33]. All instruments were self-reported except the Fear-of-Delivery Questionnaire (FDQ), which is a revised version of Areskog's instrument; the Deliver Fear Scale (DFS), which was developed by Serçekuş et al. (2017) and Shakarami et al. (2021); and the Unnamed Tool, which was developed by Laursen et al. (2008) [15, 34–36]. The response options for all PROMs were on a Likert scale except for two PROMs (Areskog's questionnaisre and FDQ), which were dichotomous (yes/no) [34, 37]. The number of instrument dimensions and items varied from 1–9 to 1–53, respectively.

Methodological quality of the studies

The methodological quality for PROM development and content validity of the included PROMs are shown in Table 2. The basis of all the scales was classical test theory (CTT). No PROM development or content validity was reported in six studies [15, 16, 37–40]. However, most of the other studies reported PROM development and content validity unclearly or incompletely. In relation to the design of the PROM, the W-DEQ questionnaire was developed on the basis of qualitative research on two authors' clinical experiences with women who feared childbirth [31]. The DFS originated from a list of 60 items written by two experts and then commented upon by 8 experienced midwives, not the target population [31]. The Childbirth Fear Questionnaire (CFQ) and the Childbirth Fear Scale (CFS) were developed according to a literature review [14, 33]. Therefore, the quality of the PROM design at these scales was inadequate. Only four PROMs, the Fear of Childbirth Questionnaire (FCQ), the Slade–Pais Expectations of Childbirth Scale (SPECS), and unnamed questionnaires designed by Melender et al. (1999) and Eriksson et al. (2005), were developed on the basis of qualitative research on the target populations [27–30]. However, these studies had some ambiguity in conducting or reporting qualitative research, such as the presence of skilled group interviewers, interviews or group meetings based on an appropriate topic or interview guide. This lack of data leads to a doubtful rating (D) design. The rest of the studies did not report how or where the scale originated. Saisto (2001), Melender (1999), Elvander (2013), Salde (2016), Slade (2019) and González-de la Torre performed pilot studies or cognitive interviews [27, 29, 30, 34, 41, 42]. However, there was some ambiguity regarding the comprehensibility and comprehensiveness of the questionnaire from the

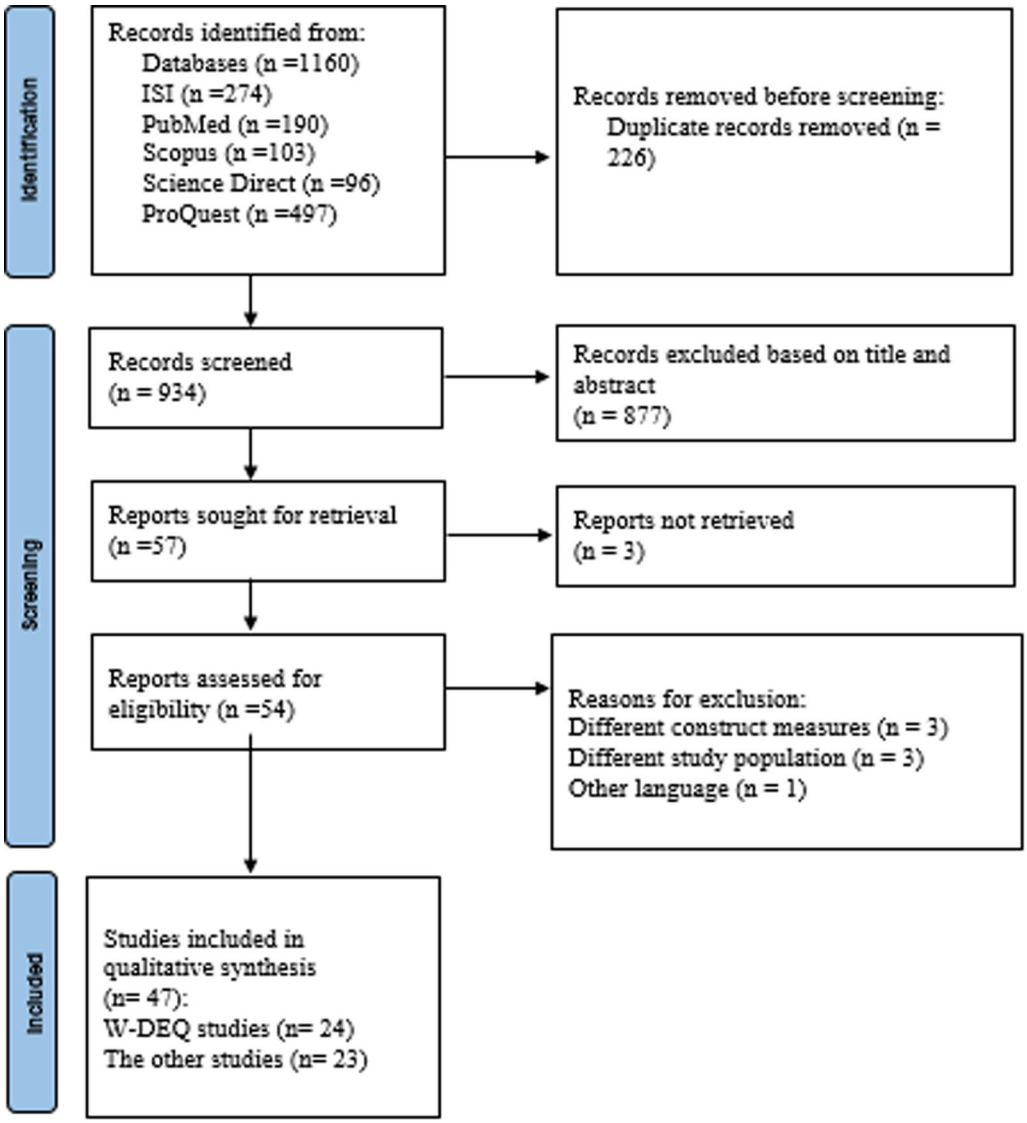

Fig. 1 PRISMA flow diagram of study selection

target population. Content validity has been reported in studies by [1, 14, 27, 28, 33, 36, 42, 43, 72].

The other measurement properties (i.e., boxes 3–10) of the included studies were rated for methodological quality with the COSMIN Risk of Bias Checklist (Table 3). Ten studies reported structural validity. Internal consistency was reported in all studies except for six, which included the Areskog, Slade 2021, and 1-item scales, i.e., the Waldenstro¨m, Laursen, Rouhe, Storksen et al. [1, 15, 16, 37, 39, 40]. None of the studies reported cross-cultural validity by measuring multiple-group CFA except Fairbrother et al. (2022) for parity and nationality. Multiple-group CFA was performed to compare national and parity changes. They showed measurement invariance across parity groups (i.e., the measurement model of the CFQ was generalizable across parity groups) [14]. Among the included studies, six reported reliability via split-half methods [35, 36, 43] or test–retest methods [33, 43, 72]. Hypothesis testing for construct validity was reported in 12 studies that measured correlations with similar/dissimilar or unrelated instruments [14–16, 27, 33–36, 38, 39, 43, 72]. None of the studies measured or reported measurement errors, criterion validity or responsiveness. Furthermore, none of the cross-cultural studies performed multiple-group CFA to validate the translated versions.

To avoid very long tables for faster scanning, we have shown the methodological quality rating in addition

Table 2 Quality of the PROM development and content analysis

PROM	Authors (year)	PROM development			Content validity				
		Design	Cognitive interview/Pilot study	Total PROM development	Asking patients			Asking experts	
					Relevance	Comprehensiveness	Comprehensibility	Relevance	Comprehensiveness
Fear of childbirth questionnaire	Areskog et al. [37]	NR	NR	NR	NR	NR	NR	NR	NR
FDQ (revised version of Areskog's tool)	Saisto et al. [34]	NR	D	D	NR	NR	NR	NR	NR
W–DEQ–A	K. Wjma', et al. [31]	I	D	I	NR	NR	NR	NR	NR
W–DEQ_B	K. Wjma', et al. [31]	I	D	I	NR	NR	NR	NR	NR
Unnamed	Melender et al. [29]	D	D	D	–	–	–	–	–
	Melender et al. [17]	–	–	–	NR	NR	NR	NR	NR
DFS	K. Wjma', et al. [32]	I	I	I	D	D	D	I	I
	SERÇEKUŞ [35]	–	–	NR	NR	NR	NR	NR	NR
	Shakarami et al. (2021)	–	I	I	D	I	D	D	I
Unnamed	Eriksson [28]	I	I	I	D	NR	D	D	D
Unnamed	Waldenstro¨m et al. [40]	NR	NR	NR	NR	NR	NR	NR	NR
Unnamed	Laursen [15]	NR	NR	NR	NR	NR	NR	NR	NR
VAS	Rouhe [16]	NR	NR	NR	NR	NR	NR	NR	NR
FOBS	Haines et al. [71]	NR	NR	NR	NR	NR	NR	NR	NR
	Zhang et al. [72]	–	NR	NR	NR	NR	NR	V	NR
NRS	Storksen et al. [39]	NR	NR	NR	NR	NR	NR	NR	NR
BAS	Elvander et al. [41]	NR	D	D	NR	NR	NR	NR	NR
SPECS	Slade et al. [27]	D	D	D	D	D	D	D	D

Table 2 (continued)

PROM	Authors (year)	PROM development			Content validity				
		Design	Cognitive interview/Pilot study	Total PROM development	Asking patients			Asking experts	
					Relevance	Comprehensiveness	Comprehensibility	Relevance	Comprehensiveness
FCQ	Slade et al. (2021)	D	–	–	–	–	–	–	–
	Sanjari et al. [43]	–	D	D	D	I	D	D	I
	Prelog et al. [38]	–	NR	NR	V	V	V	V	I
Fear of childbirth	Slade et al. [27]	NR	NR	NR	NR	NR	NR	NR	NR
CFQ	Fairbrother et al. [14]	I	I	I	I	I	I	I	I
	González–de la Torre et al. [42]	–	D	D	I	I	I	V	D
CFS	Nuraliyeva et al. [33]	I	I	I	I	I	D	D	D

Quality of the PROM development: V (Very good), A (Adequate), D (Doubtful), I (Inadequate), NR (Not Reported)

Table 3 Methodological quality of studies on measurement properties

PROM	Authors (year)	Measurement properties				
		Structural Validity	Internal consistency	Cross-cultural validity	Reliability	Construct validity
Fear of childbirth questionnaire	Areskog et al. [37]	NR	NR	NR	NR	NR
FDQ	Saisto et al. [34]	NR	V	NR	NR	V (in line with H 2)
DFS	K. Wjma', et al. [32]	NR	V	NR	NR	NR
	SERÇEKUŞ [35]	V	V	NR	NR	V (in line with H2)
	Shakarami et al. [36]	V	I	NR	V	V (in line with H1 & H2 & H3)
Unnamed	Melender et al. [17]	A	V	NR	NR	NR
Unnamed	Eriksson	A	V	NR	NR	NR
Unnamed	Waldenstro¨m et al	NR	NR	NR	NR	NR
Unnamed	Laursen et al. [15]	NR	NR	NR	NR	I (in line with H2)
VAS	Rouhe [16]	NR	NR	NR	NR	V (in line with H 1)
FOBS	Haines et al. [71]	NR	V	NR	NR	NR
	Zhang et al. [72]	NR	V	NR	V	V (in line with H 1 & H2)
NRS	Storksen et al. [39]	NR	NR	NR	NR	V (in line with H 1 & 2)
BAS	Elvander et al. [41]	NR	V	NR	NR	NR
SPECS	Slade et al. [27]	A	V	NR	NR	V (in line with H 2)
Slade FCQ	Slade et al. (2021)	NR	NR	NR	NR	NR
	Sanjari et al. [43]	V	V	NR	V	V (in line with H 1)
Fear of childbirth	Prelog et al. [38]	A	V	NR	NR	V (in line with H 2)
CFQ	Fairbrother et al. [14]	A	V	V	NR	V (in line with H1, 2)
	González-de la Torre et al. [42]	V	V	NR	NR	NR
CFS	Nuraliyeva et al. [33]	A	V	NR	V	V (in line with H1)

Scores for methodological quality using COSMIN Risk of Bias Checklist: V (Very good), A (Adequate), D (Doubtful), I (Inadequate), NR (Not Reported); Generic hypothesis for Construct validity: H1 (hypothesis 1), H2 (hypothesis 2), H3 (hypothesis 3)

to the quality of measurement properties of the studies relating to W-DEQ scales in a separate table (Table 4). We found 24 studies for the W-DEQ (i.e., W-DEQ versions A & B), 14 and five of which assessed the psychometric properties of the W-DEQ-A [44–57] and W-DEQ-B [58–62], respectively. Five studies assessed the psychometric properties of both the W-DEQ-A and B [31, 63–66]. All the translated studies that assessed the psychometric properties of the W-DEQ-A and B reported structural validity in addition to internal consistency [44, 46–67]. Structural validity was very good in 11 and six studies assessing the W-DEQ-A [46, 47, 49, 51, 52, 54–57, 63, 66] and W-DEQ-B [60, 62–66], respectively, and it was adequate in seven and three studies assessing the W-DEQ-A [44, 48, 50, 53, 64, 65, 67] and W-DEQ-B [58, 59, 61], respectively. The internal consistency was very good in all studies assessing the W-DEQ-B and the W-DEQ-A, except for two studies [47, 51]. Wijam et al. (1998) did not assess structural validity, but they reported internal consistency [31]. Reliability was very good in one and two studies assessing W-DEQ-A [31] and W-DEQ-B [31, 58], respectively, and adequate in three and two studies assessing W-DEQ-A [53, 57, 64] and W-DEQ-B [62, 64], respectively. Construct validity was reported in all studies except four and three studies assessing the W-DEQ-A [48, 52, 57, 65] and W-DEQ-B [31, 62, 65], respectively. None of the studies measured or reported measurement errors, criterion validity, or responsiveness. None of the cross-cultural studies performed multiple-group CFA to validate the translated versions.

Quality of measurement properties of the scales

The quality of the measurement properties of the included scales was rated on the basis of updated criteria for good measurement properties in each study. In general, we found 18 PROMs for FOC assessment. The number of studies per PROM were as follows: W-DEQ (24 studies), DFS (3 studies), unnamed Melender questionnaire (2 studies), FOBS (2 studies), FCQ (3 studies), and CFQ (2 studies). The remaining PROMs were reported in one study. The results of all the psychometric properties assessed in the abovementioned studies for each PROM, such as different versions of the W-DEQ scale, were qualitatively summarized in Table 5. In the case of

Table 4 Methodological quality of studies on the measurement properties of the W-DEQ-PROM

PROM	Authors, year, country	Measurement properties			
		Structural Validity	Internal consistency	Reliability	Construct validity
W-DEQ-A	K. Wjma', et al. [31], Sweden	NR	V (?)	V (+)	V (in line with H1,H2 & H3)
	Korukcu et al., [47], Turkey	V (–)	I (?)	NR	A (+) (in line with hypothesis 2) A (–) (in line with hypothesis 3) A (+) (in line with hypothesis 3)
	Takegata et al., [53], Japan	A (?)	V (?)	A (+)	V (+) (in line with hypothesis 1) A (+) (in line with hypothesis 2) V (+) (in line with hypothesis 3)
	Fenaroli & Saita, [63], Italy	V (+)	V (+)	NR	V (+) (in line with hypothesis 2) V (–) (in line with hypothesis 2)
	Lukasse et al., [48]	A (?)	V (?)	NR	NR
	Abedi et al., [44], Iran	A (?)	V (-)	NR	A (+) (in line with hypothesis 2) A (–) (in line with hypothesis 3)
	Mortazavi, [66], Iran	V (+)	V (+)	NR	V (+) (in line with hypothesis 1) V (+) (in line with hypothesis 2)
	Moghaddam Hosseini et al., [49], Hungary	V (+)	V (+)	NR	A (–) (in line with hypothesis 2 A (+) (in line with hypothesis 2) A (+) (in line with hypothesis 3)
	Andaroon et al., [67], Iran	A (?)	V (–)	NR	A (+) (in line with hypothesis 2) A (+) (in line with hypothesis 3)
	Khwepeya et al., [55], Malawi	V (–)	V (–)	NR	V (+) (in line with hypothesis 3) A (+) (in line with hypothesis 3)
	Pitel et al., [50], Slovakia	A (?)	V (–)	NR	I (+) (in line with hypothesis 2) I (+) (in line with hypothesis 2) I (–) (in line with hypothesis 3) I (+) (in line with hypothesis 3) I (+) (in line with hypothesis 3)
	Ortega-Cejas et al., [57], Spain	V (–)	V (?)	A (+)	NR
	Onchonga et al., [56], Kenya	V (–)	V (?)	NR	A (–) (in line with hypothesis 2) V (–) (in line with hypothesis 3)
	Lai et al., [64], China	A (?)	V (?)	A (+)	A (–) (in line with hypothesis 2) V (–) (in line with hypothesis 3)
	Han et al., [46], China	V (–)	V (?)	NR	V (+) (in line with H1)
	Massae et al., [65], Tanzania	A (?)	V (-)	NR	NR
	Roosevelt et al. [51]	V (?)	I (?)	NR	V (+) (in line with H1) V (+) (in line with H2)
	Varela et al. [54], Greece	V (+)	V (+)	NR	V (+) (in line with H2)
	Souto et al., [52], Portugal	V (+)	V (+)	NR	NR
W-DEQ-B	K. Wjma', et al. [31], Sweden	NR	V (?)	V (+)	NR
	Fenaroli & Saita, [63], Italy	V (+)	V (+)	NR	V (–) (in line with hypothesis 2) V (+) (in line with hypothesis 2)
	Korukcu et al., [60], Turkey	V (-)	V (?)	NR	A (–) (in line with hypothesis 2)
	Takegata et al., [61], Japan	A (?)	V (?)	NR	A (+) (in line with hypothesis 2)
	Mortazavi, [66], Iran	V (+)	V (+)	NR	V (+) (in line with hypothesis 2 and for factors 1, 2, 3 & 5) V (-) (in line with hypothesis 2 and for factors 4, 6, 7)
	Jha et al., [59], India	A (?)	V (-)	NR	I (–) (in line with hypothesis 3)
	Abbaspoor et al., [58], Iran	A (?)	V (-)	V (+)	V (-)(in line with hypothesis 2)
	Roldán-Merino et al., [62], Spain	V (-)	V (?)	A (+)	NR
	Lai et al., [64], China	V (-)	V (?)	A (+)	A (+) (in line with hypothesis 2) V (–) (in line with hypothesis 3)
	Massae et al., [65], Tanzania	V (+)	V (+)	NR	NR

Scores for methodological quality using COSMIN Risk of Bias Checklist: V (Very good), A (Adequate), D (Doubtful), I (Inadequate), NR (Not Reported); Generic hypothesis for Construct validity: H1 (hypothesis 1), H2 (hypothesis 2), H3 (hypothesis 3)

Table 5 Summary of Findings (Quality of the evidence for measurement properties results of the PROMS)

PROM	Content validity		Structural validity		Internal consistency		Cross-cultural validity		Reliability		Construct validity	
	Overall rating[1]	QoE[2]	Overall rating	QoE	Overall rating	QoE	Overall rating	QoE	Overall rating	QoE	Overall rating	QoE
Fear of childbirth questionnaire (Areskog)	±	Very low[a,b]	?	NA	?	NA	?	NA	?	NA	?	NA
FDQ	±	Very low[a,b]	?	NA	?	High	?	NA	?	NA	–	High
W-DEQ-A	±	Very low[a,b]	–	Moderate[b]	–	Moderate[b]	?	NA	+	High	+ (in line with H1, H2 & H3)	Moderate [b]
W-DEQ-B	±	Very low[a,b]	–	Moderate[b]	–	Moderate[b]	?	NA	+	High	± (in line with H2); - (in line with H3)	Moderate [b]
DFS	±	Low[a,b]	+	High	+	High	?	NA	+	High	+ (in line with H1); - (in line with H2; + (in line with H3)	Moderate [b]
Unnamed (Melender et al.)	+	Low[a]	?	Moderate[a]	?	High	?	NA	?	NA	?	NA
Unnamed (Eriksson)	+	Moderate [a]	?	Moderate[a]	?	Moderate[b]	?	NA	?	NA	?	NA
Unnamed (Laursen et al.) [15]	±	Very low[a,b]	?	NA	?	NA	?	NA	?	NA	?	Very low [a]
Unnamed (Waldenstro¨m et al. [40])	±	Very low[a,b]	?	NA	?	NA	?	NA	?	NA	?	NA
VAS	±	Very low[a,b]	?	NA	?	NA	?	NA	?	NA	+	High
FOBS	±	Very low[a,b]	?	NA	?	High	?	NA	+	High	?	High
NRS	±	Very low[a,b]	?	NA	?	NA	?	NA	?	NA	+ (in line with H1 &2)	High
BAS	±	Very low[a,b]	?	NA	?	High	?	NA	?	NA	?	NA
SPECS	+	Moderate[a]	?	Moderate[a]	?	High	?	NA	?	NA	+	High
Slade FCQ	+	High	+	High	+	High	?	NA	+	High	+ (in line with H1)	High
Fear of childbirth (Prelog et al.)	?	Very low[a]	?	Moderate[a]	?	High	?	NA	?	NA	+	High
CFQ	?	Very low[a]	+	High	+	High	+	High	?	NA	+ (in line with H1) - (in line with H2)	High
CFS	?	Moderate[a]	?	Moderate[a]	?	High	?	NA	+	High	+ (in line with H1)	High

[1] Overall rating: + = Sufficient rating, – = Insufficient rating, ± = inconsistent rating, ? = indeterminate rating

[2] QoE (Quality of the evidence): using a modified GRADE approach (Rating: high, moderate, low, very low evidence); a: downgraded for Risk of Bias; b: downgraded for inconsistency

inconsistent results, we concluded on the basis of the majority of consistent results and then downgraded for inconsistency [26].

PROM selection

According to the recommended categorization of the COSMIN methodology, the Slade FCQ was categorized as A. The results of this PROM can be trusted and recommended for use. The PROMs, including FDQ, W-DEQ-A & B, DFS, FOBS, BAS, CFQ, CSF, SPECS, and unnamed tools by Melender et al. (2005) and Eriksson et al. (2005), were categorized as B. These PROMs are potentially recommended for use, but further research is needed to evaluate the quality of this group. We did not categorize PROMs with no PROM development or content validity information, including PROMs invented by Arekog, Waldenstrom, Laursen, Rouhe, Storksen and Prelog et al.

Discussion

The present systematic review evaluated the measurement properties of FOC in women during pregnancy and postpartum period to identify the best available tools. This study showed that the Slade FCQ was categorized as A. Since this scale was developed on the basis of interviews with a target population, pilot tests were performed, and content validity, i.e., the relevancy, comprehensibility and comprehensiveness of the scales, was assessed. Therefore, it had sufficient content validity. It also had sufficient internal consistency with a high quality of evidence. According to the COSMIN guideline, content validity is an appropriate reflection of the assessed construct. Therefore, it is the most important measurement property [18]. Content validity can affect other measurement properties, such as internal consistency and structural validity [23]. According to the quality of the measurement properties, internal consistency is evaluated on the basis of Cronbach's alpha for unidimensional or each subscale of multidimensional instruments. Although nearly all the included studies in this systematic review had evidence for internal consistency, the necessity of a sufficient rating for structural validity as a criterion for sufficient internal consistency suggests that the results of internal consistency may be interpreted with caution [26]. Therefore, we did not assess PROMs with no content validity. Some of these scales are one- or two-item tools, such as the unnamed tools by Waldenstrom et al. (2006) and Laursen et al. [15] and the FOBS, VAS and NRS. It seems that one- or two-item scales are not sufficient to estimate a stable evaluation of childbirth fear. They also cannot encompass fear possibilities experienced by pregnant or parturient women (e.g., fear of pain or harm, fear of medical interventions) [14]. The instrumentation experts agreed that the content of the measurements should be derived directly from the target population and respondents for both the subjects covered by the measurement and the wording of the items. By deriving the items from the respondents, appropriate issues can be included. Measures that contain questions written by experts or authors or derived from the literature review reflect professional views of the construct rather than the experience of potential respondents [68]. However, most of the present FOC scales were of unknown origin or were derived from a literature review. Only four out of 18 PROMs in the present study were derived from the instrument respondents [27–30]. To date, the most commonly used scale for evaluating FOC is the W-DEQ, which has been translated into several languages worldwide. They have also assessed its psychometric properties. As mentioned before, the W-DEQ was developed on the basis of qualitative research with two authors' clinical experiences of women who feared childbirth, not pregnant or postpartum women. The W-DEQ assesses a wide range of emotions during labor and delivery (e.g., during labor and delivery, do you think you will feel lonely, confident, strong, weak, afraid, deserted, safe, independent, desolate, tense, or happy?) In factor analytic studies of the W-DEQ, fear has been found to be one of three to nine reported factors [44, 46–67], in contrast to the unidimensional W-DEQ developed by Wijma et al. (1998) [31], suggesting that the W-DEQ is not only a measure of fear. Furthermore, some dimensions of FOC are not addressed in this scale (e.g., cognitive aspects of fear, such as fear of pain, social embarrassment, and mothers' safety; physiological reactions to fear, such as tachycardia, hyperventilation, and sweating; and behavioral responses, such as fight and escape responses). The lack of these items may be due to not deriving the content of the questionnaire from the target population, i.e., pregnant or postpartum women. The psychometric properties of the translated versions of the W-DEQ were assessed in a systematic review by [69]. Although the results of this study are nearly consistent with our research with respect to the W-DEQ scale, only 18 studies were included, while 24 related studies were included. Moreover, the structured review conducted by Richen et al. (2018) on the current measurement tools did not include several FOC instruments, including the unnamed tools by Eriksson et al. (2005) and Laursen et al. [15], the DFS, BAS, SPECS, fear of childbirth by prelog, the Slade FCQ, the CFQ and the CFS. Furthermore, the quality of the tools included in this review was evaluated using the tool developed by Hawker et al. (2002). This tool contains nine criteria, including title, abstract, introduction, aims, methods and material, sampling, ethics issues, bias, data analysis, results, transferability, generalizability, and implications ranging from "very poor" to "good" [20]. It

seems that this tool is a critical appraisal checklist of an article rather than a checklist for evaluating the measurement properties of an instrument.

Another systematic review by Zhao et al. (2022) assessed only five measurement tools for FOC in comparison with the 18 PROMs used in the present study. According to this research, the W-DEQ is recommended for use in measuring FOC [22]. This finding is consistent with the systematic review conducted by Mudgal et al. (2024), who assessed existing tools for assessing childbirth fear. They analyzed the psychometric properties of each tool [70]. However, this finding is inconsistent with the findings of our study and Valera et al.'s study (2024) in terms of structural validity, internal consistency and construct validity. There were inconsistent ratings of the methodological properties of the W-DEQ, including structural validity, internal consistency, and several hypotheses of construct validity. As mentioned in the results, we concluded on the basis of the majority of consistent results and then downgraded for inconsistency. According to the COSMIN methodology, another strategy is not to summarize inconsistent results and not to grade the evidence [26]. Since Zhao et al.'s study (2022) is in the Chinese language, in addition to its full text not being accessible, it is not clear how they rated the measurement properties of the included PROMs, which strategy they applied and which instruments needed to be investigated more according to them. On the other hand, Mudgal et al. (2024) did not apply the COSMIN guideline or other specific quality assessment scales or criteria for evaluating the measurement properties of the tools.

The reliability of the included PROMs was estimated using split-half or test–retest methods in the present study. It seems that the split-half method is preferable to the test–retest method. The transitional nature of pregnancy, delivery and postpartum and experience is an ongoing psychological process, and women's fears are supposed to change as pregnancy advances or after delivery [31].

As mentioned in the results, the psychometric properties of the included studies, including criterion validity, measurement error and responsiveness, were not measured or reported. This finding is in line with the studies of [22, 70].

Strengths, limitations and recommendations for future research

One of the strengths of this study was the application of the COSMIN checklist as a standard guideline for PROM assessment. To the best of our knowledge, the present study is the only one that has assessed all available FOC scales using the COSMIN checklist. The COSMN checklist is highly specialized, detailed and time-consuming. Therefore, to perform this project, the authors worked as a team and were supervised by an experienced specialist of instrumentation (AE) in the research team who guided the data analysis. Although the measurement properties of the included scales were rated carefully with the COSMIN guideline, this approach is somewhat subjective, especially for content validity. Therefore, to increase the trustworthiness of the findings, two reviewers independently extracted the required data and rated the methodological quality of the included studies and the quality of the measurement properties. The third and fourth authors were consulted in case of any ambiguity. Furthermore, conducting a comprehensive search without date and place restrictions helped the researchers dismiss the relevant studies. However, there were several limitations in this study. First, the authors had problems obtaining the COSMIN checklist manuals due to internet restrictions. Fortunately, Dr. Lidwine B Mokkink (PhD, VU University Medical Center, Department of Epidemiology and Biostatistics), a member of the research team of the COSMIN guideline, sent the data to the first author (BLH) via email and guided the authors in applying them. Second, there was a lack of content validity and some necessary information for the PROM assessment of some of the included studies, which resulted in either an inability to categorize them or a low score. Third, despite the researchers' attempts, there was some problem in reaching the original scales, which caused their content validity to not be assessed. On the basis of the COSMIN methodology, reviewers should assess the items of the scale. Finally, since only studies published in English were reviewed, studies published in other languages were not included.

Therefore, additional rigorous psychometric studies of the scales, especially the scales in group B, are needed. Furthermore, the present research calls for FOC scale development for some groups, such as adolescents. The present scales, specifically the Slade FCQ, which is in group A, are designed for adults.

Conclusions

The clinical reasoning of health care providers depends on the scale they use. In other words, healthcare professionals in midwifery practice can support women effectively in terms of FOC if they can evaluate and measure it properly. According to these findings, the FCQ can be recommended for use in clinical practice for evaluating pregnant women with FOC. The PROMs categorized as B, including FDQ, W-DEQ-A & B, DFS, FOBS, BAS, CFQ, CSF, SPECS, and unnamed tools by Melender et al. (2005) and Eriksson et al. (2005), are potentially recommended for use, but further research

is needed to evaluate the quality of this group through rigorous psychometric assessments.

Statement of significance

Women's health and well-being are affected by fear of childbirth (FOC) as a common problem in the perinatal period. Although there are many scales for FOC assessment and several approaches to manage this problem and its negative outcomes, accurate measurement of FOC is crucial for correctly identifying women experiencing FOC, as well as target population for treatment. According to this systematic review, no comprehensive research has determined the best FOC scale to use in clinical midwifery. This paper introduces the recommended scale to assess FOC appraised with the COSMIN checklist.

Abbreviations

FOC	Fear of childbirth
PROMs	Patient-reported outcome measures
COSMIN	The COnsensus-based Standards for the Selection of Health Measurement INstruments
GRADE	The grading of recommendations assessment, development and evaluation
W-DEQ	Wijma delivery expectancy/experience questionnaire
FCQ	Fear of childbirth questionnaire
FDQ	Fear-of-delivery questionnaire
DFS	Delivery fear scale
VAS	Visual analog scale
FOBS	Fear of birth scale
NRS	Numeric rating scale
BAS	Birth anticipation scale
SPECS	Slade–pais expectations of childbirth scale
CFQ	Childbirth fear questionnaire
CFS	Childbirth fear scale

Acknowledgements

The authors would like to thank Dr. Lidwine B Mokkink for providing us with the COSMIN checklist manuals and guiding how to use them.

Author contributions

BLH, AE, AM, MHR and RB contributed to the conception and design of the study; BLH and MHR contributed to acquisition of data; BLH, AE, AM, MHR and RB contributed to analysis and interpretation of data; BLH wrote the main manuscript and; BLH and MHR prepared the tables 1- 5; BLH, AE, AM, MHR and RB revised the manuscript critically for important intellectual content, and BLH, AE, AM, MHR and RB approved final version of the manuscript to be submitted.

Funding

This study was supported by the Vice Presidency of Mashhad University of Medical Sciences with research ID 4012017.

Availability of data and materials

No datasets were generated or analysed during the current study.

Declarations

Ethics approval and consent to participate

This study was approved by the Ethical Committee of Mashhad University of Medical Sciencess with ethical code number IR.MUMS.NURSE.REC.1402.016.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 12 June 2024 Accepted: 29 October 2024
Published online: 25 November 2024

References

1. Slade P, Balling K, Houghton G, Sheen K. A new scale for fear of childbirth: the fear of childbirth questionnaire (FCQ). J Reprod Infant Psychol. 2021. https://doi.org/10.1080/02646838.2021.1928615.
2. Clark M. Understanding women's experiences of fear of childbirth. Canterbury: Canterbury Christ Church University; 2020.
3. Nilsson C, Hessman E, Sjöblom H, Dencker A, Jangsten E, Mollberg M, et al. Definitions, measurements and prevalence of fear of childbirth: a systematic review. BMC Pregnancy Childbirth. 2018;18(1):28.
4. Kananikandeh S, Amin Shokravi F, Mirghafourvand M, Jahanfar S. Factors of the childbirth fear among nulliparous women in Iran. BMC Pregnancy Childbirth. 2022;22(1):547.
5. Ölmez F, Oğlak SC. The prevalence and associated factors of fear of childbirth among pregnant adolescents in Turkey: a cross-sectional study. Birth. 2021;14:15.
6. Shabakhti L, Rohani C, Matbouei M, Jafari N. Tokophobia; pathological fear of childbirth: a systematic review. Hayat J. 2020;26(1):84–104.
7. Gelaw T, Ketema TG, Beyene K, Gurara MK, Ukke GG. Fear of childbirth among pregnant women attending antenatal care in Arba Minch town, southern Ethiopia: a cross-sectional study. BMC Pregnancy Childbirth. 2020;20(1):672.
8. Moradi M, Nazi A, Mazloumi E. Prevalence and causes related to fear of vaginal delivery in Iran: a systematic review. Iran J Nurs Res. 2022;17(1):43–53.
9. Dencker A, Nilsson C, Begley C, Jangsten E, Mollberg M, Patel H, et al. Causes and outcomes in studies of fear of childbirth: a systematic review. Women Birth. 2019;32(2):99–111.
10. Hildingsson I, Rubertsson C, Karlström A, Haines H. Exploring the fear of birth scale in a mixed population of women of childbearing age-A Swedish pilot study. Women Birth. 2018;31(5):407–13.
11. O'Connell MA, Leahy-Warren P, Kenny LC, O'Neill SM, Khashan AS. The prevalence and risk factors of fear of childbirth among pregnant women: a cross-sectional study in Ireland. Acta Obstet Gynecol Scand. 2019;98(8):1014–23.
12. Çıtak Bilgin N, Coşkun H, Coşkuner Potur D, İbar Aydın E, Uca E. Psychosocial predictors of the fear of childbirth in Turkish pregnant women. J Psychosom Obstet Gynecol. 2021;42(2):123–31.
13. Tomazin KJDCG, Miot HA, Stoll K, Gonçalves IR, Spiri WC, Felipe TRL, et al. Cross-cultural adaptation and validation of the childbirth fear prior to pregnancy scale in Brazil. Open Nurs J. 2021. https://doi.org/10.2174/1874434602115010179.
14. Fairbrother N, Collardeau F, Albert A, Stoll K. Screening for perinatal anxiety using the childbirth fear questionnaire: a new measure of fear of childbirth. Int J Environ Res Public Health. 2022. https://doi.org/10.3390/ijerph19042223.
15. Laursen M, Hedegaard M, Johansen C. Fear of childbirth: predictors and temporal changes among nulliparous women in the Danish national birth cohort. BJOG: An Int J Obstet Gynaecol. 2008;115(3):354–60.
16. Rouhe H, Salmela-Aro K, Halmesmäki E, Saisto T. Fear of childbirth according to parity, gestational age, and obstetric history. BJOG: An Int J Obstet Gynaecol. 2009;116(1):67–73.
17. Melender H-L. Experiences of fears associated with pregnancy and childbirth: a study of 329 pregnant women. Birth. 2002. https://doi.org/10.1046/j.1523-536X.2002.00170.x.
18. Prinsen CAC, Mokkink LB, Bouter LM, Alonso J, Patrick DL, de Vet HCW, et al. COSMIN guideline for systematic reviews of patient-reported outcome measures. Qual Life Res. 2018;27(5):1147–57.
19. Mokkink L, de Vet HCW, Prinsen CAC, Patrick DL, Alonso J, Bouter LM, et al. COSMIN risk of bias checklist for systematic reviews of patient reported outcome measures. Qual Life Res. 2018;27:1171–9.

20. Richens Y, Smith DM, Lavender DT. Fear of birth in clinical practice: a structured review of current measurement tools. Sex Reprod Healthcare. 2018;16:98–112.
21. Hawker S, Payne S, Kerr C, Hardey M, Powell J. Appraising the evidence: reviewing disparate data systematically. Qual Health Res. 2002;12(9):1284–99.
22. Ya-nan ZLY, Jing Z, Jia-wen H, Xi-mei M, Si-min Z, Guang-hong X, Yan-hong WG. The instruments of measuring fear of childbirth in pregnancy based on COSMIN: a systematic review of psychometric properties. Chin J Nurs Educ. 2022;19(4):378–84.
23. Terwee CB, Prinsen CAC, Chiarotto A, Westerman MJ, Patrick DL, Alonso J, et al. COSMIN methodology for evaluating the content validity of patient-reported outcome measures: a Delphi study. Qual Life Res. 2018;27(5):1159–70.
24. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ. 2021;372:n71.
25. Terwee CB, Prinsen C, Chiarotto A, De Vet H, Bouter LM, Alonso J, et al. COSMIN methodology for assessing the content validity of PROMs–user manual. Amsterdam: VU University Medical Center; 2018.
26. Mokkink LB, Prinsen C, Patrick DL, Alonso J, Bouter L, De Vet H, et al. COSMIN methodology for systematic reviews of patient-reported outcome measures (PROMs). User Manual. 2018;78(1):6–3.
27. Slade P, Pais T, Fairlie F, Simpson A, Sheen K. The development of the slade-pais expectations of childbirth scale (SPECS). J Reprod Infant Psychol. 2016;34(5):495–510.
28. Eriksson C, Westman G, Hamberg K. Experiential factors associated with childbirth-related fear in Swedish women and men: a population based study. J Psychosom Obstet Gynecol. 2005;26(1):63–72.
29. Melender H-L, Lauri S. Fears associated with pregnancy and childbirth—experiences of women who have recently given birth. Midwifery. 1999;15(3):177–82.
30. Slade P, Balling K, Sheen K, Houghton G. Establishing a valid construct of fear of childbirth: findings from in-depth interviews with women and midwives. BMC Pregnancy Childbirth. 2019;19(1):96.
31. Wijma K, Wijma B, Zar M. Psychometric aspects of the W-DEQ; a new questionnaire for the measurement of fear of childbirth. J Psychosom Obstet Gynecol. 1998;19(2):84–97.
32. Wijma K, Alehagen S, Wijma B. Development of the delivery fear scale. J Psychosom Obstet Gynecol. 2002;23(2):97–107.
33. Nuraliyeva Z, Nurten K. Development of the childbirth fear scale. Arch Health Sci Res. 2022;9(3):161–9.
34. Saisto T. Obstetric, psychosocial, and pain-related background, and treatment of fear of childbirth. 2001.
35. Serçekuş P, İsbir GG, İnci FH. Reliability and validity of the delivery fear scale. Dokuz Eylül Üniversitesi Hemşirelik Fakültesi Elektronik Dergisi. 2017;10(4):179–85.
36. Shakarami A, Iravani M, Mirghafourvand M, Jafarabadi MA. Psychometric properties of the Persian version of delivery fear scale (DFS) in Iran. BMC Pregnancy Childbirth. 2021;21(1):147.
37. Areskog B, Kjessler B, Uddenberg N. Identification of women with significant fear of childbirth during late pregnancy. Gynecol Obstet Invest. 1982;13(2):98–107.
38. Prelog PR, Makovec MR, Šimic MV, Sršen TP, Perat M. Individual and contextual factors of nulliparas' levels of depression, anxiety and fear of childbirth in the last trimester of pregnancy: intimate partner attachment a key factor? Slovenian J Public Health. 2019;58(3):112–9.
39. Storksen HT, Eberhard-Gran M, Garthus-Niegel S, Eskild A. Fear of childbirth; the relation to anxiety and depression. Acta Obstet Gynecol Scand. 2012;91(2):237–42.
40. Waldenström U, Hildingsson I, Ryding E. Antenatal fear of childbirth and its association with subsequent caesarean section and experience of childbirth. BJOG: An Int J Obstet Gynaecol. 2006;113(6):638–46.
41. Elvander C, Cnattingius S, Kjerulff KH. Birth experience in women with low, intermediate or high levels of fear: findings from the first baby study. Birth. 2013;40(4):289–96.
42. González-de la Torre H, Domínguez-Gil A, Padrón-Brito C, Rosillo-Otero C, Berenguer-Pérez M, Verdú-Soriano J. Validation and psychometric properties of the spanish version of the fear of childbirth questionnaire (CFQ-e). J Clin Med. 2022;11(7):1843.
43. Sanjari S, Fakhraei AA, Soleimani MRM, Alidousti K. Validation of the slade fear of childbirth scale for pregnancy in a sample of Iranian women: a crosssectional study. Crescent J Med Biol Sci. 2022;9(3):138–46.
44. Abedi P, Hazeghi N, Afshari P, Fakhri A. The validity and reliability of Persian version of Wijma delivery expectancy/experience questionnaire (version a) among Iranian nulliparous women. Glob J Health Sci. 2016;9(2):269.
45. Andaroon N, Kordi M, Kimiaee SA, Esmaeili H. The effect of individual counseling on attitudes and decisional conflict in the choice of delivery among nulliparous women. J Educ Health Promot. 2020;9:35.
46. Han L, Wu J, Wu H, Liu J, Liu Y, Zou Z, et al. Validating the use of the Wijma delivery expectancy/experience questionnaire in mainland China: a descriptive, cross-sectional study. BMC Pregnancy Childbirth. 2022;22(1):931.
47. Korukcu O, Kukulu K, Firat MZ. The reliability and validity of the Turkish version of the Wijma delivery expectancy/experience questionnaire (W-DEQ) with pregnant women. J Psychiatr Ment Health Nurs. 2012;19(3):193–202.
48. Lukasse M, Schei B, Ryding EL. Prevalence and associated factors of fear of childbirth in six European countries. Sex Reprod Healthcare. 2014;5(3):99–106.
49. MoghaddamHosseini V, Makai A, Dweik D, Varnagy A. Factor analysis study of the Hungarian translation of Wijma delivery expectancy/experience questionnaire (version A). Curr Psychol. 2020;39(4):1098–105.
50. Pitel L, Psenkova P, Jones CCG, Zahumensky J. Validation of the Slovak version of the Wijma delivery expectancy/experience questionnaire (W-DEQ), version A. Bratislava Med J. 2020;121(6):415–21.
51. Roosevelt L, Zielinski R, Seng J, Low LK. Measuring fear of childbirth among a diverse population in the United States: a revised Wijma delivery expectancy/experience scale (WDEQ-10). J Midwif Womens Health. 2023;68(5):581–7.
52. Souto S, Prata AP, Albuquerque RS, Almeida S. Psychometric properties of the European Portuguese version of the Wijma delivery expectancy/experience questionnaire in pregnant women. Health Care Women Int. 2024;45(6):603–20.
53. Takegata M, Haruna M, Matsuzaki M, Shiraishi M, Murayama R, Okano T, et al. Translation and validation of the Japanese version of the Wijma delivery expectancy/experience questionnaire version A. Nurs Health Sci. 2013;15(3):326–32.
54. Varela P, Zervas I, Vivilaki V, Lykeridou A, Deltsidou A. Validity and reliability of the Greek version of Wijma delivery expectancy/experience questionnaire (Version A) among low-risk pregnant women. BMC Psychol. 2024;12(1):165.
55. Khwepeya M, Huang H-C, Lee GT, Kuo S-Y. Validation of the Wijma delivery expectancy/experience questionnaire for pregnant women in Malawi: a descriptive, cross-sectional study. BMC Pregnancy Childbirth. 2020;20(1):455.
56. Onchonga D, Várnagy Á, Amer F, Viktoria P, Wainaina P. Translation and validation of the Swahili version of the Wijma delivery expectancy/experience questionnaire version A (W-DEQ-A). Sex Reprod Healthc. 2021;29:100626.
57. Ortega-Cejas CM, Roldán-Merino J, Lluch-Canut T, Castrillo-Pérez MI, Vicente-Hernández MM, Jimenez-Barragan M, et al. Reliability and validity study of the Spanish adaptation of the "Wijma delivery expectancy/experience questionnaire" (W-DEQ-A). PLoS ONE. 2021;16(3):e0248595.
58. Abbaspoor Z, Haghighizadeh MH, Abedi P. Psychometric properties of the Iranian version of Wijma delivery expectancy/experience questionnaire in women who experience fear of childbirth: version B. J OBGYN. 2021;8(1):39–45.
59. Jha P, Larsson M, Christensson K, Svanberg AS. Fear of childbirth and depressive symptoms among postnatal women: a cross-sectional survey from Chhattisgarh, India. Women Birth. 2018;31(2):E122–33.
60. Korukcu O, Bulut O, Kukulu K. Psychometric evaluation of the Wijma delivery expectancy/experience questionnaire version B. Health Care Women Int. 2016;37(5):550–67.
61. Takegata M, Haruna M, Matsuzaki M, Shiraishi M, Okano T, Severinsson E. Psychometric evaluation of the Japanese Wijma delivery expectancy/experience questionnaire version B. Open J Nurs. 2017;7(1):15–27.
62. Roldán-Merino J, Ortega-Cejas CM, Lluch-Canut T, Farres-Tarafa M, Biurrun-Garrido A, Casas I, et al. Validity and reliability of the Spanish version of the "Wijma delivery expectancy/experience questionnaire" (W-DEQ-B). PLoS ONE. 2021;16(4):e0249942.

63. Fenaroli V, Saita E. Fear of childbirth: a contribution to the validation of the Italian version of the Wijma delivery expectancy/experience questionnaire (WDEQ). TPM Test Psychom Methodol Appl Psychol. 2013;20(2):131–54.
64. Lai THT, Kwok ST, Wang W, Seto MTY, Cheung KW. Fear of childbirth: validation study of the Chinese version of Wijma delivery expectancy/experience questionnaire version A. Midwifery. 2022;104:103188.
65. Massae AF, Larsson M, Leshabari S, Mbekenga C, Pembe AB, Svanberg AS. Fear of childbirth: validation of the Kiswahili version of Wijma delivery expectancy/experience questionnaire versions A and B in Tanzania. BMC Pregnancy Childbirth. 2022;22(1):882.
66. Mortazavi F. Validity and reliability of the Farsi version of Wijma delivery expectancy questionnaire: an exploratory and confirmatory factor analysis. Electron Physician. 2017;9(6):4606–15.
67. Andaroon N, Kordi M, Ghasemi M, Mazlom R. The validity and reliability of the Wijma delivery expectancy/experience questionnaire (Version A) in Primiparous Women in Mashhad, Iran. Iran J Med Sci. 2020;45(2):110–7.
68. Doward LC, Meads DM, Thorsen H. Requirements for quality of life instruments in clinical research. Value Health. 2004;7(s1):S13–6.
69. Varela P, Zervas I, Lykeridou A, Deltsidou A. A systematic review of the psychometric properties of the translated versions of the Wijma delivery expectancy/experience questionnaire. Eur Rev Appl Psychol. 2024;74(1):100911.
70. Mudgal S, Shafqat N. Tools for assessing childbirth fear: a comprehensive review and psychometric evaluation. WHO South-East Asia J Public Health. 2024;13(1):16–23.
71. Haines H, Pallant JF, Karlström A, Hildingsson I. Cross-cultural comparison of levels of childbirth-related fear in an Australian and Swedish sample. Midwifery. 2011;27(4):560–7.
72. Zhang Q, McAra-Couper J, Lou Y, Guo S, Qiu P. Validation of the Chinese version of the Fear of Birth Scale among pregnant women. Midwifery. 2024;133:103986.

Publisher's Note