



## Validated Instruments for Assessing Female Sexual Function

Cindy M. Meston & Leonard R. Derogatis

To cite this article: Cindy M. Meston & Leonard R. Derogatis (2002) Validated Instruments for Assessing Female Sexual Function, Journal of Sex & Marital Therapy, 28:S1, 155-164, DOI: [10.1080/00926230252851276](https://doi.org/10.1080/00926230252851276)

To link to this article: <http://dx.doi.org/10.1080/00926230252851276>



Published online: 19 Jan 2011.



[Submit your article to this journal](#)



Article views: 391



[View related articles](#)



Citing articles: 82 [View citing articles](#)

## Validated Instruments for Assessing Female Sexual Function

CINDY M. MESTON

*Department of Psychology, University of Texas at Austin, Austin, Texas, USA*

LEONARD R. DEROGATIS

*School of Nursing and Department of Epidemiology and Preventive Medicine,  
University of Maryland, Baltimore, Maryland, USA*

*In this article, we review five instruments for assessing female sexual dysfunction (FSD): the Brief Index of Sexual Functioning for Women (BISF-W; Taylor, Rosen, & Leiblum, 1994), the Changes in Sexual Functioning Questionnaire (CSFQ; Clayton, McGarvy, & Clavet, 1997), the Derogatis Interview for Sexual Functioning (DISF/DISF-SR; Derogatis, 1997), the Female Sexual Function Index (FSFI; Rosen et al., 2000), and the Golombok Rust Inventory of Sexual Satisfaction (GRISS; Rust & Golombok, 1986). The purpose of this article is to highlight the psychometric properties of these questionnaires in an effort to assist researchers in selecting effective measurement tools for FSD.*

Research into the diagnosis and treatment of female sexual dysfunction (FSD) has rapidly advanced over the past few years. This has created an increased need for psychometrically sound instruments for diagnosing FSD and for effectively monitoring treatment-induced changes. In May 2000, the Food and Drug Administration (FDA) Center for Drug Evaluation and Research drafted guidelines for the industry regarding the clinical development of drug products for FSD. Included in these guidelines was the following excerpt relating to the use of scales and questionnaires for diagnosing FSD:

New scales, questionnaires, and other instruments to diagnose FSD and/or particular components or to measure response to treatment over time should be developed, tested, and validated in women with FSD. The validation process for the scale should demonstrate its reliability, validity, responsiveness, and ability to measure minimal meaningful differences.

---

Address correspondence to Cindy M. Meston, University of Texas at Austin, Department of Psychology, Austin, Texas 78712, USA. E-mail: meston@psy.utexas.edu

Diagnostic scales and questionnaires should be able to distinguish patients with FSD from those without the disorder. If a scale is meant to detect a specific component of FSD, such as decreased sexual desire or decreased sexual arousal, the scale should also attempt to differentiate between women who have the component for which approval is sought and women having a different component of the disorder.

A number of inventories for assessing FSD have been published and several others currently are in the process of being validated. The five instruments discussed in this article were selected from Medline and Psychinfo searches through the literature dated from 1980 to current. In our literature searches, we discovered many questionnaires for assessing FSD that pertained to specific aspects of sexual function, such as "arousability" or sexual anxiety or aversions. Many were validated for use in the treatment of specific medically related sexual disorders such as cancer or diabetes. The instruments reviewed here include the Brief Index of Sexual Functioning for Women (BISF-W; Taylor, Rosen, Leiblum, 1994); the Changes in Sexual Functioning Questionnaire (CSFQ; Clayton, McGarvey, & Claver, 1997); the Derogatis Interview for Sexual Functioning (DISF/DISF-SR; Derogatis, 1997); the Female Sexual Function Index (FSFI; Rosen et al., 2000); and the Golombok Rust Inventory of Sexual Satisfaction (GRISS; Rust & Golombok, 1986). These five inventories currently are available for use and were chosen because they each assess various domains of sexual function in women and, in our opinion, they most clearly meet the psychometric requirements in diagnosing FSD as delineated by the FDA. The purpose of this article is to compare and highlight the psychometric properties of these questionnaires in order to assist researchers in selecting a measurement tool for FSD that best meets their individual research needs.

## CRITERIA FOR MEASURES OF FEMALE SEXUAL DYSFUNCTION

The two primary characteristics of scientific measurement are reliability and validity. Reliability refers to the consistency or replicability of measurement and may be thought of as the reverse of measurement error; it is the proportion of measurement that is due to true variations in the construct of interest relative to total variation. A reliability coefficient of 1.00 represents measurement without error; a coefficient of 0.00 represents a complete absence of consistent variation. Three types of reliability coefficients are discussed with regard to FSD questionnaires. Test-retest reliability involves repeated administration of the questionnaire (usually at two to four week intervals) to the same population and serves as an indicator of the measure's stability over time. Internal consistency indicates the homogeneity of item within a domain, and inter-rater reliability indexes the consistency between raters for clinician-administered inventories.

Test validity refers to the degree to which the instrument measures what

it purports to measure. This includes issues such as: Are the test items appropriate and complete? Are the test scores related to some currently available external measure of the same domain (concurrent validity)? Can the test differentiate between clinical and nonclinical samples (discriminant validity)? How are the scores associated with those from a related but different domain (divergent validity)? Unlike test reliability, which is established through a series of highly prescribed studies, the validation of an instrument is programmatic, more of an ongoing process that involves numerous studies and trials that test and extend the generalizability of results.

In addition to its reliability and validity, an instrument's sensitivity should also be established for it to meet the optimal criteria for use in clinical trials of FSD. Sensitivity in this context has two major referents: the instrument's ability to differentiate between individuals with and without sexual dysfunction and the capacity to detect treatment-induced changes. In order of increasing sensitivity, this would mean being able to detect differences between drug versus placebo treatments, being able to discriminate between differential dose responses, and being able to discriminate between two distinct drug treatments. The degree to which the instrument is sensitive to differences between sexually functional and dysfunctional individuals determines its effectiveness as a diagnostic tool. The degree to which it is sensitive to detecting treatment-induced changes determines its effectiveness for measuring outcome efficacy.

A number of other factors may come into play when selecting a measurement tool for FSD aside from its ability to meet psychometric criteria. One major consideration is whether there are published norms available for both the sample of women with the sexual dysfunction as well as for a matched, sexually functional control group. Ideally, the norms would be based on women of an age range comparable to that being studied, and the normative sample would be derived from a population comparable to that being studied (for example, a nondepressed, community sample). Tests that offer interpretation at global, domain, and item levels offer considerably more information on outcome efficacy than those that provide only a "total score" of sexual functioning. The specific domains assessed (for example, desire, arousal, orgasm, satisfaction, anxiety) might also be a factor when selecting a specific measurement tool. Other considerations include test brevity, ease of administration and scoring, cost efficiency, computer compatibility, and availability of language translations.

## INSTRUMENTS FOR ASSESSING FEMALE SEXUAL FUNCTION

### The Brief Index of Sexual Functioning for Women (BISF-W)

The BISF-W (Taylor, et al., 1994) is a brief 22-item self-report inventory designed to assess current levels of female sexual functioning and satisfaction. The questionnaire takes 15–20 minutes to administer. Based on a sample of

269 sexually active women (age range 20–73 years), and using the original 3-factor scoring solution of the BISF-W (interest/desire, sexual activity, sexual satisfaction; Taylor et al., 1994), test-retest reliability the acceptable range, and concurrent validity was demonstrated by comparing the BISF-W with relevant subscales of the Derogatis Sexual Functioning Inventory (DSFI) (Derogatis, 1976). A new quantitative scoring algorithm, designed from a conceptual basis, recently was developed to facilitate the use of the BISF-W in clinical trials (Mazer, Leiblum, & Rosen, 2000). This scoring procedure provides an overall composite for sexual function, as well as seven dimension scores: thoughts/desire, arousal, frequency of sexual activity, receptivity/initiation, pleasure/orgasm, relationship satisfaction, and problems affecting sexual function. Internal consistencies for the desire and pleasure/orgasm domains were within the acceptable range (0.70 or greater; Nunnally, 1967); internal consistencies for the arousal, receptivity/initiation, relationship/satisfaction, and problems affecting sexual function were relatively low. Norms for the composite score and for each of the seven dimension scores are available on a sample of 225 healthy women (age range 20–55 years), 187 with regular sex partners and 38 without regular partners, and on 104 surgically menopausal, sexually active women with complaints of impaired sexual function (age range 20–55 years; Mazer et al., 2000). The BISF-W reliably discriminated between these samples of women with and without sexual complaints on each of the seven sexuality dimensions. In a placebo-controlled study of the effects of transdermal testosterone treatment on sexual function in 75 oophorectomized women, the BISF-W was sensitive to detecting differences between treatment groups on two of the seven sexuality dimensions and on the overall composite BISF-W score (Shiffrin et al., 2000).

### The Changes in Sexual Functioning Questionnaire (CSFQ)

The CSFQ (Clayton, McGarvey, & Clavet, 1997) is a 35-item (36 items for males) structured interview designed to measure illness- and medication-related changes in sexual functioning. A gender-specific self-report version of the inventory also is available (CSFQ-F for females and CSFQ-M for males). A total CSFQ score and five domain scores can be obtained. The five domains assessed are supported by factor analyses and include sexual desire frequency (two items), sexual desire/interest (three items), sexual pleasure (one item), sexual arousal (three items), and orgasm (three items). Additional questions ascertain the degree to which sexual functioning has changed over time, how extensive the change is, and the nature and probable cause of the change. Scores for each of the five domains, as well as for the total CSFQ can be computed. The CSFQ was standardized on a sample of 122 medical students (68 males, 54 females; age range 22–35 years) and 33 psychiatry residents (17 males, 16 females; age range 25–43 years). The CSFQ has been shown to discriminate reliably between a clinical sample of depressed patients and a nonclinical sample (separate by gender) on each of

the five sexuality domains and on the total CSFQ score (Clayton, McGarvey, Clavet, & Piazza, 1997). Norms, by gender, are available for depressed patients, medical students, psychiatry residents, and nonpsychiatric outpatients (Clayton, McGarvey, Clavet, & Piazza, 1997). There are no published norms available for women with FSD. Internal consistency and test-retest reliabilities for the CSFQ were within the acceptable range. Concurrent validity has been established between the CSFQ and DISF-SR scales. Correlations between comparable scales on these two measures ranged from .42 (sexual pleasure sexual behavior [DISF-SR]) to .76 (orgasm). The DSFQ takes approximately 20 min to administer.

### The Derogatis Interview for Sexual Functioning (DISF/DISF-SR)

The DISF (Derogatis, 1979) is a brief, 25-item semistructured interview designed to provide a multidimensional assessment of sexual functioning in males and females. A separate self-report version of the interview, the DISF-SR, also is available. The items from the DISF/DISF-SR are arranged into five domains that are supported by factor analyses: sexual cognition/fantasy (five items), sexual arousal (five items), sexual behavior/experience (five items), orgasm (six items), and sexual drive/relationship (four items). Interpretation of the DISF/DISF-SR scores is possible at three distinct levels: discrete item, domain, and global summary scores. A total score that summarizes level of sexual functioning across the five domains can also be computed. Both the DISF and the DISF-SR have been normed on community samples ( $n = 399$ ; age range 19–64). The community sample norms are available for males and females and are presented as standardized scores (area T-scores). Standardized area scores enable meaningful comparisons of strengths and weaknesses of an individual's profile to be made regardless of whether the raw scores are normally distributed. To date, there are no norms available on the DISF/DISF-SR for women with sexual dysfunction. Test-retest, internal consistency, and inter-rater reliabilities were well within the acceptable ranges. The DISF/DISF-SR takes approximately 15 minutes to administer and is available in eight foreign language translations (Danish, Dutch, French, German, Italian, Norwegian, Spanish).

### The Female Sexual Function Index (FSFI)

The FSFI (Rosen et al., 2000) is a brief, 19-item self-report measure of female sexual function that provides scores on five domains of sexual function as well as a total score. The domains assessed have been confirmed using factor analyses and include desire (two items), arousal (four items), lubrication (four items), orgasm (three items), satisfaction (three items), and pain (three items). The FSFI was developed on a female sample of 131 normal controls (age range 21–68) and 128 age-matched subjects (age range 21–69) who met *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV;

American Psychiatric Association, 1994) criteria for female sexual arousal disorder (FSAD). Norms are available for FSAD patients and controls at the item level, the domain level, and the full scale score. The FSFI has been shown to discriminate reliability between FSAD and control patients on each of the five domains of sexual function as well as on the full scale score. Internal consistency and test-retest reliabilities were within the acceptable range. Divergent validity has been established using the Locke-Wallace Marital Adjustment Test (1959). Correlations between the FSFI and Locke-Wallace were generally modest in magnitude (.53 for the control group and .22 for the FSAD group), with the strongest relation observed for the satisfaction domain of the FSFI. This provides further support for the construct validity of the FSFI because it demonstrates a statistical dissociation between FSFI scores and a theoretically different, albeit related, construct (that is, marital satisfaction). The FSFI takes approximately 15 minutes to administer. The FSFI may be accessed on the Web at [www.fsfi-questionnaire.com](http://www.fsfi-questionnaire.com).

### The Golombok-Rust Inventory of Sexual Satisfaction (GRISS)

The GRISS is a 56-item (28 items for males and 28 for females) self-report inventory designed to evaluate both the quality of a heterosexual relationship and each partner's level of sexual functioning within that relationship. The GRISS provides 12 domain scores (five female, five male, two common gender domains), confirmed using factor analyses, as well as a global score that summarizes the overall quality of relationship and sexual functioning in the couple. Domains pertaining to female sexuality include anorgasmia (four items), vaginismus (four items), avoidance (four items), nonsensuality (four items), and dissatisfaction (four items). The two domains common to both genders are frequency of sexual contact (four items) and noncommunication (four items). Subscale scores are transformed to standardized stanine scores and may be plotted on a common profile for the couple. The GRISS was standardized on a sample of 88 sex-therapy clients from clinics throughout the United Kingdom. Transformation keys for the main scale and subscales appear on the scoring sheet. The distributions of the transformed scales are approximately normed for the clinical sample but are skewed toward the lower end of the scale to facilitate measurement in nonclinical populations. Verification of the scales for nonclinical samples was conducted using data from a student sample and a random sample of 59 general practitioners. Internal consistencies, based on the 88-subject standardization sample were within the acceptable range (.70 or greater) for all female domains, with the exception of sexual communication (.61). Test-retest reliabilities were calculated from the pre- and post-treatment correlation of 41 clinical couples receiving either marital or sexual therapy. Test-retest reliabilities for the female domains ranged from .47 (dissatisfaction) to .82 (vaginismus). The GRISS has been shown to discriminate reliably between sexually functional ( $n = 30$ ) and dysfunctional ( $n = 42$ ) women on the five female domains, and on

**TABLE 1.** Measurement Characteristics of Five Brief Inventories for Assessing Female Sexual Functioning

Inventory name	# Items	Standardization sample	Administration time/modality	Domains measured
BISF-W	22	225 healthy women (20–55 years); 104 surgically menopausal women with impaired sexual function (21–55 years)	15–20 min self-report female only	Thoughts/desire, arousal, frequency of sexual activity, receptivity, pleasure/orgasm, relationship satisfaction, problems affecting sexuality
CSFQ	35	122 male and female medical students (22–35 yrs) 35 psychiatry residents (25–43 yrs)	15–20 min interview male & female	Sexual pleasure, sexual desire/frequency, sexual desire/interest, arousal, orgasm
DISF/ DISF-SR	25	Community sample of 122 (DISF) and 277 (DISF-SR) males and females (19–69 years)	10–15 min interview & self report male & female	Sexual cognition & fantasy arousal, sexual behavior & experience, orgasm, sexual drive and relationship
FSFI	19	131 normal women (21–68 years); 128 women with FSAD (21–69 years)	10–15 min self-report female only	Desire, arousal, lubrication orgasm, satisfaction, pain
GRISS	28	88 sex therapy clients (males and females)	15–20 min self-report male & female	Anorgasmia, vaginismus, sexual frequency, sexual communication, satisfaction, nonsensuality, sexual avoidance

the common domain of sexual frequency, but not on the common domain of communication. Validation of the inventory's sensitivity for detecting therapy-induced changes in sexual functioning was assessed in 30 couples who had received 5 sex-therapy sessions. The correlation between GRISS scores and blind therapist ratings was .43 for females. The GRISS takes approximately 15 min to administer. A Dutch translated version of the GRISS is available (Kuileter, Vroege, & Van Lankveld, 1993).

## CONCLUSIONS

We have provided a brief psychometric review of five instruments applicable for use in assessing FSD: BISF-W, CSFQ, DISF/DISF-SR, FSFI, and GRISS. Each of these instruments assesses a number of specific domains of sexual functioning. For each inventory, interpretation is possible at the item, domain, and global levels. All five instruments have self-report versions available; the CSFQ and DISF also are available in interview formats. All five

**TABLE 2.** Reliability Coefficients and Validity Correlations for the BISF-W, CSFQ, DISF/DISF-SR, FSFI, and GRISS

Inventory	Test-retest reliability	Internal consistency reliability	Inter-rater reliability	Concurrent	Divergent
BISF-W	.68–.78 <sup>a</sup>	.39–.83 <sup>a</sup> –.08–.72 <sup>a</sup>	N/A	.46–.69 <sup>a</sup>	N/A
CSFQ	.45–1.00 <sup>c</sup>	.64–.80 <sup>d</sup>	N/A	.42–.76	N/A
DISF/DISF-SR	.80–.90 <sup>e</sup>	.74–.80 <sup>e</sup>	.84–.92 <sup>f</sup>	N/A	N/A
FSFI <sup>g</sup>	.79–.86	.89–.96	N/A	N/A	.19–.57
GRISS	.47–.82 <sup>h</sup>	.61–.83 <sup>i</sup>	N/A	N/A	N/A

*Note:* Coefficient ranges refer to the variability between specific female domains within scales.

<sup>a</sup>Based on the original 3-factor solution (Taylor et al., 1994).

<sup>b</sup>Based on the 7 dimension scoring algorithm (Mazer et al., 2000).

<sup>c</sup>Based on individual test items for females only.

<sup>d</sup>Females only.

<sup>e</sup>DISF-SR.

<sup>f</sup>DISF.

<sup>g</sup>Coefficients based on combined FSAD and control patients.

<sup>h</sup>Based on 41 clinical couples.

<sup>i</sup>Based on 88-subject standardization sample. Concurrent validity demonstrated using the DSFI and DISF-SR for the BISF-W and CSFQ, respectively.

questionnaires are brief and take a maximum of 20 minutes to administer. The CSFQ, DISF, and GRISS may be administered to males and females (the BISF-W and FSFI were developed for females only), and the DISF and GRISS have been translated into foreign languages. Internal consistency and test-retest reliabilities have been established and fall within the acceptable range for each of these tests, with the exception of low alpha coefficients for certain of the BISF-W dimensions. Acceptable inter-rater reliabilities statistics have been reported for the DISF, and acceptable divergent validity statistics have been published for the FSFI.

FDA guidelines suggest that questionnaires should be developed, tested, and validated in women with FSD. Toward this end, the FSFI has been validated and normed on a sample of women with diagnosed FSAD. The BISF-W has recently been normed on a sample of surgically menopausal women with self-reported sexual dysfunction. With regard to the questionnaires' sensitivity in detecting differences in sexual functioning between clinical and nonclinical samples, the FSFI has been shown to reliably discriminate between women with and without FSAD on each of the five domains of sexual functioning assessed (desire, arousal, lubrication, orgasm, satisfaction, pain). The BISF-W and the GRISS reliably discriminated between women with and without nonspecific sexual disorders. Studies assessing the ability of these inventories to detect treatment-induced changes in female sexual functioning are lacking. The GRISS has been shown to detect therapy-induced changes in global levels of sexual function among couples, and the

BISF-W has been shown to detect drug-induced changes in female sexual functioning beyond placebo.

As noted earlier, the validation of a test instrument is an enduring process that requires numerous studies and trials. The instruments reviewed here provide promise for use in research on the treatment of FSD. Further studies are needed to provide more comprehensive and representative norms and to assess the discriminant validity of these questionnaires for women with specific subtypes of FSD including, hypoactive sexual desire disorder and inhibited orgasm. Equally important, studies are needed to determine the ability of these instruments to register reliable treatment-induced changes, including specific changes in sexual function that may occur following therapeutic intervention, and differences that may exist in the effectiveness of varying types and intensities of treatment.

## REFERENCES

- American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.
- Clayton, A. H., McGarvey, E. L., & Clavet, G. J. (1997). The Changes in Sexual Functioning Questionnaire (CSFQ): Development, reliability and validity. *Psychopharmacology Bulletin*, *33*, 731–745.
- Clayton, A. H., McGarvey, E. L., Clavet, G. J., & Piazza, L. (1997). Comparison on sexual functioning in clinical and nonclinical populations using the Changes in Sexual Functioning Questionnaire (CSFQ). *Psychopharmacology Bulletin*, *33*, 747–753.
- Derogatis, L. R. (1976). The DSFI: A multidimensional measure of sexual functioning. *Journal of Sex & Marital Therapy*, *5*, 244–281.
- Derogatis, L. R. (1997). The Derogatis Interview for Sexual Functioning (DISF/DISF-SR): An introductory report. *Journal of Sex & Marital Therapy*, *23*, 291–304.
- Kuileter, M. M., Vroege, J. A., & van Lankveld, J. J. D. M. (1993). *The Golombok Rust Inventory of Sexual Satisfaction, Nederlandse vertaling en aanpassingen*. Leiden: Netherlands University Medical Center, [Dutch translation and adaptation].
- Locke, H. J., & Wallace, K. M. (1959). Short marital-adjustment and prediction tests: Reliability and validity. *Marriage & Family Living*, *21*, 251–255.
- Mazer, N. A., Leiblum, S. R., & Rosen, R. C. (2000). The Brief Index of Sexual Functioning for women (BISF-W): A new scoring algorithm and comparison of normative and surgically menopausal populations. *Menopause*, *7*, 350–363.
- Nunnally, J. (1967). *Psychometric theory*. New York: McGraw Hill.
- Rosen, R., Brown, C., Heiman, J., Leiblum, S., Meston, C. M., Shabsigh, R., Ferguson, D., & D'Agostino, R., Jr., (2000). The Female Sexual Function Index (FSFI): A multidimensional self-report instrument for the assessment of female sexual function. *Journal of Sex & Marital Therapy*, *26*, 191–208.
- Rust, J., & Golombok, S., (1986). *The Golombok-Rust Inventory of Sexual Satisfaction*. Odessa, FL: Psychological Assessment Resources.
- Shiffrin, J. L., Braunstein, G. D., Simon, J. A., Casson, P. R., Buster, J. E., Redmond, G.

- P., Burki, R. E., Ginsburg, E. S., Rosen, R. C., Leiblum, S. R., Caramelli, K. E., & Mazer, N. A. (2000). Transdermal testosterone treatment in women with impaired sexual function after oophorectomy. *New England Journal of Medicine*, *343*, 682–688.
- Taylor, J. F., Rosen, R. C., & Leiblum, S. R. (1994). Self-report assessment of female sexual function: Psychometric evaluation of the Brief Index of Sexual Functioning for Women. *Archives of Sexual Behavior*, *23*, 627–643.