



Published in final edited form as:

*West J Nurs Res.* 2021 April ; 43(4): 364–373. doi:10.1177/0193945920942252.

## Development and Testing of the Dysmenorrhea Symptom Interference (DSI) Scale

Chen X. Chen<sup>1</sup>, Tabitha Murphy<sup>1</sup>, Susan Ofner<sup>2</sup>, Lilian Yahng<sup>3</sup>, Peter Krombach<sup>1</sup>, Michelle LaPradd<sup>2</sup>, Giorgos Bakoyannis<sup>2</sup>, Janet S. Carpenter<sup>1</sup>

<sup>1</sup>Indiana University School of Nursing, Indianapolis, IN, USA

<sup>2</sup>Indiana University School of Medicine Department of Biostatistics, Indianapolis, IN, USA

<sup>3</sup>Indiana University Center for Survey Research, Bloomington, IN, USA

### Abstract

Dysmenorrhea affects most reproductive-age women and increases the risk of future pain. To evaluate dysmenorrhea interventions, validated outcome measures are needed. In this two-phase study, we developed and tested the Dysmenorrhea Symptom Interference scale. During the scale-development phase (n=30), we created a 9-item scale based on qualitative data from cognitive interviews. During the scale-testing phase (n=686), we evaluated reliability, validity, and responsiveness to change. The scale measures how dysmenorrhea symptoms interfere with physical, mental, and social activities. Internal consistency was strong with Cronbach's  $\alpha > 0.9$ . Test-retest reliability was acceptable ( $r = 0.8$ ). The scale showed satisfactory content validity, construct validity (supported by confirmatory factor analysis), concurrent validity, and responsiveness to change. The minimally important difference was 0.3 points on a scale with a possible total score ranging from 1 to 5. This new psychometrically sound scale can be used in research and clinical practice to facilitate the measurement and management of dysmenorrhea.

### Keywords

Dysmenorrhea; Patient Reported Outcome Measures; Pelvic Pain; Psychometrics

Dysmenorrhea affects 45–95% of women of reproductive age (Iacovides et al., 2015). It is characterized by abdominal pain just before and/or during menstruation. Some women with dysmenorrhea also experience menstrual pain in other body locations (e.g., low back pain, headache) and gastrointestinal (GI) symptoms (e.g., bloating, nausea, and change in bowel movements; Chen et al., 2015). In addition to being a risk factor for developing other chronic conditions such as irritable bowel syndrome (Altman et al., 2006; Olafsdottir et al., 2012) and noncyclic pelvic pain (Westling et al., 2013), it can cause school and work absences (Iacovides et al., 2015). On a recurring basis, dysmenorrhea negatively affects women's physical activity, sleep, and quality of life (Iacovides et al., 2015).

**Correspondence:** Chen X. Chen, 600 Barnhill Drive NU E415, Indianapolis, IN, USA, Tel: (317) 274-7441, cxchen@iu.edu.

**Disclosure:** The authors declare no conflicts of interest.

To evaluate dysmenorrhea interventions, validated outcome measures are needed. Symptom interference with daily activities has been widely recognized as a core outcome in pain and symptom research (Dworkin et al., 2005). For example, pain interference with daily activities is a recommended core outcome in pain clinical trials by academic researchers, funding agencies, and federal regulatory agencies (Dworkin et al., 2005). Yet to our knowledge, a symptom interference scale has not been developed in the context of dysmenorrhea.

Outside the context of dysmenorrhea, generic pain interference scales exist, and they measure consequences of pain on relevant aspects of a person's life, such as social, cognitive, emotional, and physical activities (Amtmann et al., 2010; Cleeland & Ryan, 1994). However, they are not designed to measure cyclic pain (i.e., not dysmenorrhea-specific), and each scale applies only to one age group (i.e., either a pediatric or an adult population).

Therefore, the purposes of this study were to develop and test a dysmenorrhea symptom interference (DSI) scale, one that would capture a cyclic experience and apply to both adolescent girls and women. Our research team conducted a two-phase study guided by established methods for scale development and testing (American Educational Research Association et al., 2014; Nunnally & Bernstein, 1994; Turk et al., 2006). In Phase 1, we drafted the initial pool of items based on research literature and used cognitive interviewing methods to develop the scale further. In Phase 2, we quantitatively tested the scale's psychometric properties. Figure 1 shows the study design schema. In this paper, we first describe the methods and results for Phase 1 and then the methods and results for Phase 2.

## Methods: Phase 1 Scale Development

### Design

Using a cross-sectional design, we collected qualitative data using cognitive interviews (See Figure 1 for study schema).

### Sample

We used a purposive sampling strategy (Miles et al., 2013) to ensure the inclusion of individuals with a broad range of symptom severity, race/ethnicity, education, employment, and age. To assess understandability in low literacy and school-age populations, we had a recruitment goal of 25% with high school education and 10% younger than 18 years old. Participants were recruited from March to June 2018 through a study information page hosted at our home institution, study flyers posted locally, and Facebook and Instagram ads.

Eligibility criteria were: (1) female, (2) aged 14–42 (upper age limit reduced the likelihood of enrolling perimenopausal women; Avis et al., 2009), (3) able to read and converse in English, (4) currently living in the United States, and (5) menstrual pain (e.g., abdominal cramps, low back pain, headache) or menstrual GI symptoms (e.g., nausea, diarrhea, more bowel movements than usual, bloating) in the last 6 months.

There is no consensus on the sample size needed for instrument development. According to Willis (2015), a survey containing multiple distinct concepts may require significantly more interviews to reach saturation than a survey with a single well-defined concept. As we intended to measure a single defined concept (i.e., dysmenorrhea symptom interference), we believed a sample size of 30 would be appropriate (Willis, 2015).

## Procedures

The local institutional review board approved the study and granted a waiver of written (signed) informed consent. Potential study participants were directed to a short online screening survey with questions on eligibility, demographics, dysmenorrhea symptoms, and menstrual pain severity. An information sheet about the study was emailed to eligible and potentially interested participants. Those willing to participate, completed an online survey with an initial pool of 21 potential items developed by our team based on previous research on dysmenorrhea (Chen, Draucker, et al., 2018; Iacovides et al., 2015), pain interference (Amtmann et al., 2010; Cleeland & Ryan, 1994), and hot flashes (an episodic symptom; Carpenter, 2001; Carpenter et al., 2017). A phone interview was then scheduled, and verbal consent was obtained at the start of the phone interview.

The audio-taped semi-structured telephone interviews were conducted by a trained research assistant using a pre-developed interview guide protocol. The interviewer used cognitive interviewing techniques such as think-aloud and verbal probing to (1) investigate how well the initially drafted items captured the concept of interest (i.e., dysmenorrhea symptom interference with daily life) and what aspects of the concept were missing; (2) identify potentially problematic areas through feedback on clarity, comprehensibility, redundancy, and relevance of items (Willis, 2015); and (3) collect feedback on an appropriate recall period, the questionnaire format, and response options. Interviews lasted 30–45 minutes.

## Qualitative Analysis

The interviewer took detailed notes during each interview. Using the notes and audio recordings, three research team members analyzed participants' responses. All three team members were female and had training and experience with qualitative data analysis. A fourth team member was a female, who had nearly 30 years of experience with qualitative interviews and measurement development, also guided the data analysis. We created summary tables and graphs using the qualitative data to visualize how individuals interpreted and understood the questions we had asked (Miles et al., 2013). These items were binned (i.e., grouped items according to meaning to eliminate redundancy) and winnowed (i.e., removed items that had low applicability or were confusing; Willis, 2015). We did not have a target number of items to retain a priori. The process led to a preliminary questionnaire of nine items for the DSI scale. The nine items were reviewed by three research team members, with expertise in symptom science, women's health, and survey methodology, to verify understandability, non-ambiguity, comprehensiveness, and face validity (e.g., each item reflected the underlying construct of interest).

## Results: Phase 1 Scale Development

### Sample Characteristics (n=30)

Among the 106 who screened eligible for the study, 30 were purposively selected and enrolled (n=30) to ensure heterogeneity of demographic (age, race/ethnicity, education, and employment) and symptom characteristics (severity of menstrual pain). The mean age was 24 years (SD =6.3, range 14–42), and four participants (13.3%) were adolescents younger than 18 years. Table 1 shows other demographic characteristics. We reached our goal of 10% of individuals < 18 years old and 25% of individuals with lower levels of educational attainment (i.e., high school).

The mean “menstrual pain on the average” was 4.2 on a 0–10 scale (SD=1.5, range: 1–7), and the mean “worst menstrual pain” was 6.6 (SD=1.8, range: 3–9). Among these participants, 25 (83.3%) also had menstrual GI symptoms, including bloating (83.3%), more bowel movements than usual (60%), diarrhea/loose stool (50%), nausea (46.7%), constipation (3.3%), and vomiting (3.3%). Seven (23.3%) were on menses when the interviews were conducted.

### Findings from Qualitative Cognitive Interviews

We found that the initial drafted items largely captured the concept of interest. Participants described that dysmenorrhea interfered with physical, mental, and social activities. However, individuals were affected by dysmenorrhea symptoms differently based on their lifestyles and types of daily activities. We found that an established adult or pediatric pain interference scale would not be sufficient to accommodate girls and women of different ages, employment statuses, abilities, and lifestyles. For example, as some participants were not in school or not employed, items asking about performing schoolwork or work duties did not apply to them. Consequently, we decided to have one item related to work and to provide examples of work (i.e., work, housework, schoolwork, and homework). For another example, some participants stated they did not run or walk (due to lifestyle or physical disability not related to dysmenorrhea), so an umbrella term of “physical activity” was found to be more broadly applicable.

We identified problematic areas and addressed issues related to clarity, comprehensibility, relevance, and redundancy based on participants’ feedback. To improve clarity, we defined certain items (e.g., the item on sleep) to communicate the intent or meaning of the question. To enhance comprehensibility, we selected terms that participants found easy to understand and provided examples for certain items. To address item relevance, we removed items that participants perceived as irrelevant. Specifically, we found that activities that happen less frequently were less relevant to a cyclic/episodic pain condition. For example, “recreational activities” were perceived as irrelevant by several participants. They explained that when their menstruation hit during workdays, recreational activities were not applicable. To reduce item redundancy, we removed items that participants perceived as unnecessary and redundant (e.g., housework captures “household chores”).

Ultimately, nine items were retained for the DSI scale. These items measure how dysmenorrhea symptoms interfere with individuals’ physical activities, sleep, daily

activities, work, concentration, enjoyment of life, leisure activities, social activities, and mood (See Table 3).

We also received feedback on the recall period and response options. For the recall period, all participants were comfortable recalling how their menstrual symptoms interfered with their life in their current period (if they were on menses) or past menstrual period (if they were not on menses). Yet, some participants had difficulty recalling their experience over a longer time period (6 months). This was especially true for those whose symptoms and period length changed from cycle to cycle. As a result, we designed two DSI scale versions with two recall periods: an on-menses version asking participants to recall their experience in the last 24-hours, and an off-menses version asking participants to recall their experience with their last menstrual period. We chose a response scale that participants found intuitive and easy to use: 1 (not at all) to 5 (very much).

## Methods: Phase 2 Psychometric Testing

### Design

As shown in Figure 1, we used quantitative methods for psychometric testing. Both on-menses and off-menses versions were evaluated using anonymous surveys. Participants who were menstruating completed the on-menses version, while those not on menses completed the off-menses version. Those who were on days 1 to 3 of their cycle were invited to participate in a follow-up survey 24 hours after the initial survey. This follow-up survey allowed us to (1) test the DSI's test-retest reliability, (2) evaluate the DSI's responsiveness to detect change, and (3) estimate its minimally important difference (MID).

### Sampling

Similar to phase 1, eligibility criteria included being: (1) female, (2) aged 14–42, (3) able to read and converse in English, (4) currently living in the United States, and (5) menstrual pain or menstrual gastrointestinal symptoms in the last 6 months. In addition, on menses participants were defined as being on days 1 to 3 of their menstrual cycle, whereas off menses participants were defined as not menstruating at the time of the study. Participants who were on day 4+ of their menses were excluded from the Phase 2 study because for the test-retest reliability, they were less likely to still experience symptoms after 24 hours. Participants from Phase 2 did not overlap with those who participated in Phase 1. Participants for Phase 2 were recruited from January to March 2019.

Different guidelines are available for judging the adequacy of sample sizes for factor analysis. Some methodologists recommend having at least 10–20 cases for each item in the scale being used (e.g.,  $20 \times 9$  items = 180; Everitt, 1975; Hair, 1998), while others suggest obtaining a sample size of at least 500 for factor analysis (Comrey & Lee, 2013). We opted for the larger and more conservative minimum sample size of at least 500, which our study exceeded.

## Procedures

The local institutional review board approved Phase 2 of the study. Participants were recruited from the opt-in survey panel registrants maintained by a web-based service (Qualtrics, Provo, UT). Eligibility criteria for Phase 2 were the same as Phase 1. The survey panel service provider sent email invitations to 65,625 women aged 14–42 years old. Potentially interested participants clicked a hyperlink to the survey that was embedded in the email invitation (n=3754). Potential participants were further screened for eligibility (n=1654). If eligible, a study information sheet explaining consent appeared. Survey completion implied informed consent. Of those eligible (n=1032), 836 responded to the survey, which after data cleaning, resulted in a final sample size of 686.

## Measures

**Initial Survey Measures**—The initial anonymous online survey included questions on participants' demographic and health information, the DSI scale, menstrual pain severity, perceived stress, and sleep disturbance.

**The DSI:** The only difference between the on- and off-menses versions was the recall period. The on-menses version asked participants to recall their last 24-hour experiences (i.e., the instructions read, “Over the last 24 hours, how much did your menstrual pain and menstrual gastrointestinal (GI) symptoms interfere with...”). The off-menses version asked participants to recall their experiences from the last menstrual period (i.e., the instructions read, “During your last menstrual period, how much did your menstrual pain and menstrual gastrointestinal (GI) symptoms interfere with...”). The online survey allowed for branching logic. Participants responded to different versions based on whether they were on menses or off menses. They also rated each item as essential, useful but not essential, or not necessary (Lawshe, 1975) based on their own, friends', or family members' experiences.

Both versions had the same nine items (see Table 3) with response options of 1 (not at all) to 5 (very much). Individual item scores were averaged to generate a total scale score (possible range 1–5) with higher scores indicating greater interference (more negative outcome).

**Menstrual Pain Severity:** We used the validated numerical rating scale to assess menstrual pain severity (Chen et al., 2015). For participants who were on menses, we asked what number best described their worst, least, and average pain in the last 24 hours and what number best described their current pain. Response options were from 0 (“no pain”) to 10 (“extremely severe”). For participants who were not on menses, we asked them to rate their worst, least, and average menstrual pain in their last menstrual period from 0 (“no pain”) to 10 (“extremely severe”).

**Perceived Stress Scale:** We measured perceived stress using the 10-item validated perceived stress scale (PSS; Cohen et al., 1983). Each question asks participants about their feelings and thoughts during the last month on a 0 “never” to 4 “very often” scale. Scale scores were calculated by reverse scoring four items and summing across all items.

**PROMIS® Sleep Disturbance Scale:** We used the 8-item PROMIS® sleep disturbance scale short form (8b), which has been validated with diverse samples (Yu et al., 2011). Each of the eight questions has five response options ranging from 1 to 5. Following the scoring manual, item scores were totaled to generate the raw scores. Raw scores were further converted to a T-score using a conversion table (Health Measures, 2019).

**Follow-up Survey Measures—**Only participants who were on days 1 to 3 of their menstrual cycle (i.e., on menses) were invited to complete a follow-up survey at 24 hours. The follow-up survey included the DSI on-menses version and an additional item asking participants to rate how their symptoms changed over the last 24 hours on a 7-point scale with the response options of much worse, moderately worse, a little worse, no change, a little better, moderately better, or much better. This global rating of change has been widely used to assess responsiveness to change and MID (Revicki et al., 2008). We did not invite participants who were off-menses to complete the follow-up survey because our goal was to assess test-retest reliability and responsiveness to change using current ratings rather than to assess consistency in recalled ratings.

### Data Quality Control Measures

To safeguard data quality, attention filters (i.e., “trap questions”) were also used. Responses from those who failed the attention filters were removed as were responses from “speeders” who were defined as having survey completion times less than one-third of the median survey duration. In total, we removed 150 problematic responses (from those who failed a trap question or who “sped”) before analysis. This left a sample size of 686 for psychometric analysis.

### Psychometric Data Analysis

**Reliability—**Internal consistency was evaluated using Cronbach’s alpha for both versions of the DSI scale. A threshold of 0.90 was used for internal consistency (Nunnally & Bernstein, 1994). For test-retest reliability, Pearson’s correlation coefficients were calculated for participants who were on menses and whose symptoms did not change over the 24 hours. As few standards exist for judging the minimum acceptable value for a test-retest estimate, we used 0.7 as the threshold (Crocker & Algina, 2008).

**Validity: Content Validity—**Content validity was evaluated by the content validity ratio (CVR) and the percentage of participants who rated the item as essential, useful but not essential, or not necessary (Lawshe, 1975). Item CVRs were calculated as  $CVR = (n_e - n/2)/(n/2)$ , where  $n_e$  indicates the number of participants indicating “essential” and  $n$  indicates the total number of participants. For a sample size of  $N=686$ , a CVR higher than 0.064 indicates excellent content validity for a given item (Lawshe, 1975; Wilson et al., 2012). This critical value of 0.064 was calculated based on Wilson et al.’s (2012) formula, with an alpha level of 0.05 and a sample size of 686. To complement CVR, we also calculated the percentage of participants who rated a given item as essential or useful. For a given item, if 90% of participants rated it as essential or useful (i.e., <10% rated as “not necessary”), we concluded that the item had reasonable content validity.



**Validity: Construct Validity**—To assess construct validity, we performed confirmatory factor analysis as opposed to exploratory factor analysis, because items were expected to measure a unidimensional construct of dysmenorrhea interference. Acceptable model fit was noted as a Comparative Fit Index (CFI) > 0.90, Goodness of Fit Index (GFI) > 0.9, Bentler-Bonett Normed Fit Index > 0.9, Root Mean Square Error of Approximation (RMSEA) < 0.10, and Standardized Root Mean Square Error (SRMR) < 0.08 (Bartholomew et al., 2008; Hooper et al., 2008).

**Validity: Concurrent Validity**—Concurrent validity was assessed separately for on-menses and off-menses groups using Pearson's correlations of symptom interference with (1) menstrual pain severity, (2) stress, and (3) sleep disturbance. We expected correlations to be weak to moderate ( $r=0.1-0.6$ ) and positive (i.e., greater interference with greater menstrual pain severity, perceived stress, and sleep disturbance) as the latter are conceptually different from dysmenorrhea symptom interference (Iacovides et al., 2015; Ju et al., 2014). We expected stronger correlations with menstrual pain severity than with perceived stress and sleep disturbance as the construct of menstrual pain severity is conceptually closer to the construct of dysmenorrhea symptom interference than perceived stress and sleep disturbance.

**Minimally Important Difference (MID)**—A MID is “the smallest difference in score in the domain of interest that patients perceived as important, either beneficial or harmful, and that would lead the clinician to consider a change in the patient's management” (p. 377, Guyatt et al., 2002). We used both distribution- and anchor-based approaches to estimate MID. Distribution-based approaches are based on the statistical distribution of the measure scores, while anchor-based approaches are based on external criteria (also referred to as anchors; Revicki et al., 2008). For distribution-based approaches, we analyzed only those on menses women who responded to the question about symptom change. We calculated the standard deviation (SD) and the standard error of measurement (SEM). Because 0.2 SD approximates a small effect size and 0.5 SD approximates a medium effect size, a score difference between those boundaries (e.g., 0.35 SD) was used as a reasonable MID estimate (Chen, Kroenke, et al., 2018; Eton et al., 2004). The anchor-based analysis consisted of calculating the mean DSI change from baseline to 24-hour follow-up for one category shift in the symptom change score (e.g., between “no change” and “a little worse”; Amtmann et al., 2010).

**Responsiveness**—Responsiveness to change was estimated by calculating the standardized response mean (SRM; Revicki et al., 2008), which is the DSI mean at baseline minus the DSI mean at follow up divided by the standard deviation of the DSI change score. An absolute SRM value of 0.2 to 0.5 is considered a small change, 0.5 to 0.8 is moderate, and 0.8 is large (Revicki et al., 2008). Some researchers suggest that an absolute SRM value 0.3 indicates responsiveness (Askew et al., 2016). In addition, we compared the amount of DSI change across menstrual pain improved, unchanged, and worsened groups based on the retrospective rating of change. Omnibus analysis of variance was used to compare mean change across improved, unchanged, and worsening groups. The Tukey-



Kramer adjustment was used to control the type I error for pairwise comparisons of unchanged versus worse and unchanged versus better.

## Results: Phase 2 Quantitative Psychometric Testing (n=686)

### Sample Characteristics

Phase 2 participants were diverse in race/ethnicity, educational level, and employment status (See Table 1). Among 686 participants, 260 (37.9%) were on menses and responded to the on-menses version of the DSI scale.

For the on-menses subset, the mean age was 28.6 years (SD=6.9). The mean menstrual pain at its worst was 6.4 (SD=2.4), menstrual pain at its least was 3.0 (SD=2.3), menstrual pain on average was 5.0 (SD=2.3), and menstrual right now was 4.4 (SD=2.8). For participants who were on menses during the initial survey, 100 (38.5%) completed the second survey. Among participants who were invited to participate in the second survey, those who completed and those who did not complete the survey were not statistically different in demographic characteristics (age, race, ethnicity) and menstrual pain level.

The remaining 62.1% (n=426) were off menses and responded to the off-menses version of the DSI. For the off-menses subset, the mean age was 27.6 years (SD=8.1). The mean menstrual pain at its worst was 6.4 (SD=2.0), menstrual pain at its least was 2.6 (SD=2.2), and menstrual pain on average was 4.9 (SD=1.9).

**Reliability: Internal Consistency and Test-retest Reliability**—For the on-menses version (i.e., 24-hour recall), Cronbach's alpha was 0.93 at Time 1 and 0.95 at Time 2, respectively. For participants whose symptoms did not change in the last 24 hours (n=32), the test-retest reliability was satisfactory ( $r=0.79$ ,  $p<.0001$ ).

For the off-menses version (i.e., recalling the last menstrual period), the scale was internally consistent with a Cronbach's alpha of 0.91.

**Validity: Content Validity**—The content validity ratios were satisfactory for most items. For two items, leisure activities and social activities, the CVR was slightly below the critical value of 0.06 (See Table 3). The negative CVR for these two items indicated less than 50% of participants indicated the items as essential. Because only 6% of participants rated the leisure activities and social activities items as "not necessary," we retained these two items for comprehensiveness.

**Validity: Construct Validity**—Confirmatory factor analysis supported unidimensionality. As shown in Table 4, fit indices suggested a good fit of the one-factor model for both on-menses and off-menses versions. In addition, all items for both on-menses and off-menses versions had large factor loadings (See Table 3).

**Validity: Concurrent Validity**—As shown in Table 2, the DSI scale was significantly correlated in the expected directions with pre-specified measures of menstrual pain severity, perceived stress, and sleep disturbance.

**MID**—As shown in Table 2, the distribution-based MID estimates were between 0.27 to 0.36 for both on-menses and off-menses version. The anchor-based estimate was 0.28 for minimally important improvement and 0.18 for minimally important worsening. Taken together, on a 5-point scale, the MID estimate for DSI was in the vicinity of 0.3 points.

**Responsiveness**—The DSI scale on-menses version was very responsive to detect menstrual pain improvement (SRM=0.72, large effect size) but was not as responsive to detect menstrual pain worsening (SRM =−0.06, small effect size). The DSI discriminated the pain improved, unchanged, and worsened groups ( $p < .01$  for the omnibus test). Pairwise comparisons showed the DSI successfully detected differences among pain improved and unchanged groups ( $p=0.046$ ).

## Discussion

We developed the DSI scale from the perspectives of adolescent girls and women aged 14 to 42. When tested in a diverse large sample in the United States, the DSI was shown to be reliable, valid, and responsive to detect menstrual pain change. The rigorous scale development process and strong psychometric properties make the tool useful for research and clinical practice.

The DSI is advantageous over other existing pain interference measures because the DSI is specific to cyclic menstrual pain. In addition, it can be used for diverse age, race/ethnicity, education level, employment status, and menstrual pain severity.

The tool captures a concept not previously considered in the field (i.e., dysmenorrhea symptom interference). Other measures of dysmenorrhea pain have not fully captured symptom interference. For example, one measure assessed “working ability” as the only dysmenorrhea impact (Teheran et al., 2018), while another only assessed impact on “things the person usually does” without asking what specific aspects of life are affected (Wyrwich et al., 2018). Pain interference with daily activities has been acknowledged as a core outcome in pain research, especially in clinical trials (Dworkin et al., 2005). As a valid outcome measure, the DSI can be used to further develop and test interventions for dysmenorrhea.

The DSI can be administered flexibly at different phases of the menstrual cycle, given that the two DSI versions (on-menses and off-menses) with different recall periods were shown to be both reliable and valid. Compared to the off-menses version, the on-menses version is likely less subject to recall bias due to the shorter recall period. However, when daily measurement during menstruation is not feasible, recalling dysmenorrhea symptom interference with the most recent menstrual period can be appropriate. Using longitudinal designs, future research can evaluate the concordance between the DSI on-menses version (with daily recall) and the off-menses version (recalling the most recent menstrual period).

The DSI can be used for women with and without other gynecological conditions (e.g., endometriosis, uterine fibroids). Conceptually, the measure was intended to address dysmenorrhea pain regardless of clinical diagnosis. Other research also supports women with and without comorbid gynecological conditions (e.g., endometriosis, uterine fibroids)

had little differences in dysmenorrhea experiences (Nguyen et al., 2015). Psychometrics from this mixed sample of women were strong, suggesting the measure is appropriate for broad clinical use.

The DSI was very responsive to detect menstrual pain improvement, as shown by a large effect size estimate. However, the scale was not responsive to pain worsening. This may be because participants in this study had restricted lower bounds for worsening. For the on-menses version, we collected baseline data when participants were on days 1 to 3 of their menstrual cycle and collected follow-up data 24 hours after. Most women experience their highest menstrual pain on days 1 to 2, which may have contributed to a small magnitude of change in the pain worsening groups.

We also estimated the MID for the scale to see how much of a change in the DSI score was clinically meaningful. A change of 0.3 points in this 5-point scale indicated a clinically meaningful change had occurred. This MID estimate can help clinicians interpret scores about whether a meaningful change in dysmenorrhea has occurred, which can then guide treatment decision-making. The MID estimate also can be used to inform power calculations for clinical studies.

There are several strengths to this study. First, we developed and tested the scale using rigorous methods. Second, we developed and tested the scale using diverse samples in terms of age, race/ethnicity, education level, lifestyle, and menstrual pain severity. Third, both on-menses and off-menses versions were developed and evaluated. The availability of two versions gives clinicians and researchers a choice in regards to the timing of DSI assessment.

We acknowledge several study limitations. First, two DSI items (i.e., the leisure activities and social activities items) had low CVRs. This may be because dysmenorrhea symptoms affected individual women differently. Some women might not have seen these two items as essential. However, as only a very small percentage of participants rated the two items as “not necessary,” we retained them for comprehensiveness. These items need to be further evaluated and possibly dropped in the future. Second, we did not assess test-retest reliability of the off-menses version, as we were less interested in assessing consistency in recall based on a longer recall period. Third, for the MID anchor-based analysis, the sample size was small in a few anchor categories. Estimating MID based on fewer observations may result in unstable estimates (Yost et al., 2011). Given that the sample size of the “somewhat worse” and “much worse” categories were both below 10, we only performed MID anchor-based analysis on other categories. Fourth, the samples from both phases of the study were self-selected rather than randomly selected. We acknowledge coverage bias and self-selection bias. Fifth, the clinical data were self-reported. Sixth, because this was a descriptive study, the DSI scale’s responsiveness to intervention effects needs to be further evaluated in clinical trials.

In conclusion, we developed and tested the DSI scale in this two-phase study. In phase 1, we developed this 9-item scale to measure how dysmenorrhea symptoms interfere with physical, mental, and social activities based on qualitative data from cognitive interviews. The DSI has two versions (i.e., on-menses and off-menses versions) with different recall

periods. In phase 2, we evaluated the psychometric properties of both versions of DSI. The DSI was shown to be reliable, valid, and responsive to detect menstrual pain improvement. It can be adopted in research and clinical practice to facilitate the measurement and management of dysmenorrhea.

## Acknowledgments

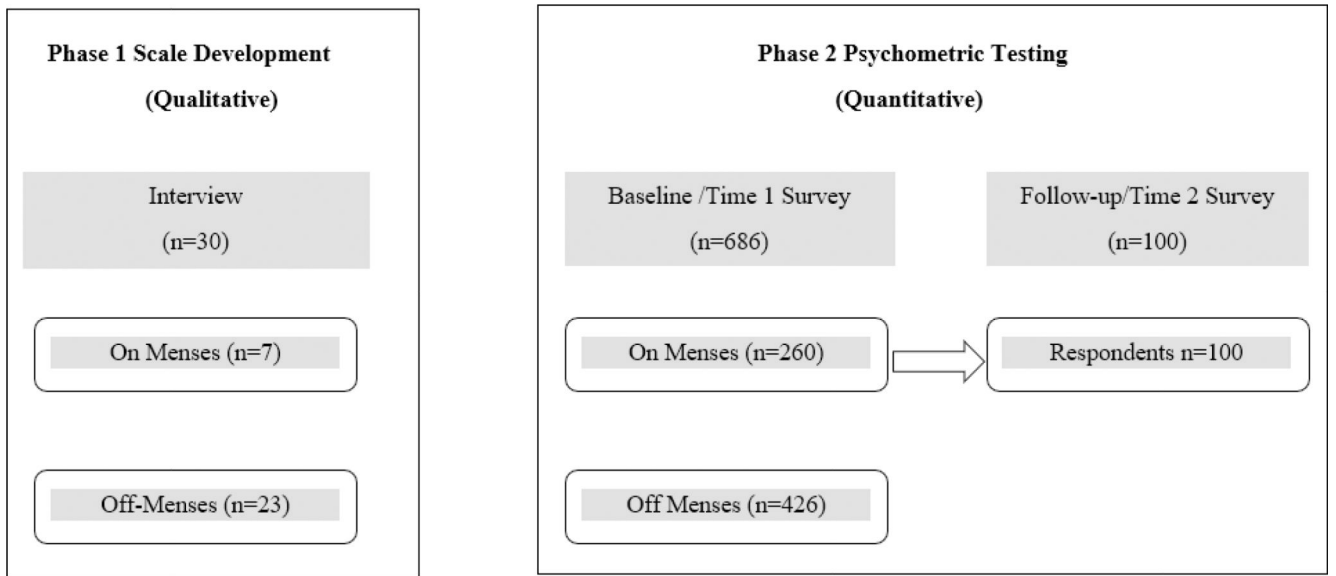
Dr. Chen is supported by the Grant Numbers KL2 TR002530 and UL1 TR002529 (A. Shekhar, PI) from the National Institutes of Health, National Center for Advancing Translational Sciences, Clinical and Translational Sciences Award and the EMPOWER Grant from the Indiana University–Purdue University Indianapolis (C. Chen, PI). The content is solely the responsibility of the authors and does not represent the official views of the National Institutes of Health.

## References

- Altman G, Cain KC, Motzer S, Jarrett M, Burr R, & Heitkemper M (2006). Increased symptoms in female IBS patients with dysmenorrhea and PMS. *Gastroenterology Nursing*, 29(1), 4–11. 10.1097/00001610-200601000-00002 [PubMed: 16552294]
- American Educational Research Association, American Psychological Association, National Council on Measurement in Education, Joint Committee on Standards for Educational, & Psychological Testing (US). (2014). *Standards for educational and psychological testing*. American Educational Research Association.
- Amtmann D, Cook KF, Jensen MP, Chen WH, Choi S, Revicki D, . . . Lai JS (2010). Development of a PROMIS item bank to measure pain interference. *Pain*, 150(1), 173–182. 10.1016/j.pain.2010.04.025 [PubMed: 20554116]
- Askew RL, Cook KF, Revicki DA, Cella D, & Amtmann D (2016). Evidence from diverse clinical populations supports clinical validity of PROMIS pain interference and pain behavior. *Journal of Clinical Epidemiology*, 73, 103–111. 10.1016/j.jclinepi.2015.08.035 [PubMed: 26931296]
- Avis NE, Colvin A, Bromberger JT, Hess R, Matthews KA, Ory M, & Schocken M (2009). Change in health-related quality of life over the menopausal transition in a multiethnic cohort of middle-aged women: Study of Women's Health Across the Nation (SWAN). *Menopause*, 16(5), 860. 10.1097/gme.0b013e3181a3cdaf [PubMed: 19436224]
- Bartholomew DJ, Steele F, & Moustaki I (2008). *Analysis of multivariate social science data*. Chapman and Hall/CRC.
- Carpenter JS (2001). The Hot Flash Related Daily Interference Scale: A tool for assessing the impact of hot flashes on quality of life following breast cancer. *Journal of Pain and Symptom Management*, 22(6), 979–989. 10.1016/s0885-3924(01)00353-0 [PubMed: 11738160]
- Carpenter JS, Bakoyannis G, Otte JL, Chen CX, Rand KL, Woods N, . . . Guthrie KA (2017). Validity, cut-points, and minimally important differences for two hot flash-related daily interference scales. *Menopause*, 24(8), 877–885. 10.1097/GME.0000000000000871 [PubMed: 28509814]
- Chen CX, Draucker CB, & Carpenter JS (2018). What women say about their dysmenorrhea: A qualitative thematic analysis. *BMC Womens Health*, 18(1), 47. 10.1186/s12905-018-0538-8 [PubMed: 29499683]
- Chen CX, Kroenke K, Stump TE, Kean J, Carpenter JS, Krebs EE, . . . Monahan PO (2018). Estimating minimally important differences for the PROMIS pain interference scales: results from 3 randomized clinical trials. *Pain*, 159(4), 775–782. 10.1097/j.pain.0000000000001121 [PubMed: 29200181]
- Chen CX, Kwekkeboom KL, & Ward SE (2015). Self-report pain and symptom measures for primary dysmenorrhoea: A critical review. *European Journal of Pain*, 19(3), 377–391. 10.1002/ejp.556 [PubMed: 25059384]
- Cleeland CS, & Ryan KM (1994). Pain assessment: Global use of the brief pain inventory. *Annals Academy of Medicine Singapore*, 23(2), 129–138. 10.1002/ejp.556
- Cohen S, Kamarck T, & Mermelstein R (1983). A global measure of perceived stress. *Journal of Health and Social Behavior*, 24(4), 385–396. 10.2307/2136404 [PubMed: 6668417]

- Comrey AL, & Lee HB (2013). *A first course in factor analysis*: Psychology Press.
- Crocker LM, & Algina J (2008). *Introduction to classical and modern test theory*. Cengage Learning.
- Dworkin RH, Turk DC, Farrar JT, Haythornthwaite JA, Jensen MP, Katz NP, . . . Witter J (2005). Core outcome measures for chronic pain clinical trials: IMMPACT recommendations. *Pain*, 113(1–2), 9–19. 10.1016/j.pain.2004.09.012 [PubMed: 15621359]
- Eton DT, Cella D, Yost KJ, Yount SE, Peterman AH, Neuberg DS, . . . Wood WC (2004). A combination of distribution- and anchor-based approaches determined minimally important differences (MIDs) for four endpoints in a breast cancer scale. *Journal of Clinical Epidemiology*, 57(9), 898–910. 10.1016/j.jclinepi.2004.01.012 [PubMed: 15504633]
- Everitt BS (1975). *Multivariate analysis: the need for data, and other problems*. *British Journal of Psychiatry*, 126, 237–240. 10.1192/bjp.126.3.237
- Guyatt G, Osoba D, Wu AW, Wyrwich KW, & Norman GR (2002). Methods to explain the clinical significance of health status measures. *Mayo Clinic Proceedings*, 77(4), 371–383. 10.4065/77.4.371 [PubMed: 11936935]
- Hair JF (1998). *Multivariate data analysis*. Prentice-Hall International.
- HealthMeasures (2019). PROMIS® scoring manuals. <http://www.healthmeasures.net/promis-scoring-manuals>
- Hooper D, Coughlan J, & Mullen MR (2008). Structural equation modelling: Guidelines for determining model fit. *Electronic Journal of Business Research Methods*, 6(1), 53–60. 10.21427/D7CF7R
- Iacovides S, Avidon I, & Baker FC (2015). What we know about primary dysmenorrhea today: a critical review. *Human Reproduction Update*, 21(6), 762–778. 10.1093/humupd/dmv039 [PubMed: 26346058]
- Ju H, Jones M, & Mishra G (2014). The prevalence and risk factors of dysmenorrhea. *Epidemiologic Reviews*, 36, 104–113. 10.1016/j.jpag.2019.09.004 [PubMed: 24284871]
- Lawshe CH (1975). A quantitative approach to content validity. *Personnel Psychology*, 28(4), 563–575. 10.1111/j.1744-6570.1975.tb01393.x
- Miles MB, Huberman AM, & Saldaña J (2013). *Qualitative data analysis: A methods sourcebook* (3rd ed.). SAGE Publications.
- Nunnally JC, & Bernstein IH (1994). *Psychometric theory* (3rd ed.). Tata McGraw-Hill Education.
- Nguyen AM, Humphrey L, Kitchen H, Rehman T, & Norquist JM (2015). A qualitative study to develop a patient-reported outcome for dysmenorrhea. *Quality of Life Research*, 24(1), 181–191. 10.1007/s11136-014-0755-z [PubMed: 25048731]
- Olafsdottir LB, Gudjonsson H, Jonsdottir HH, Bjornsson E, & Thjodleifsson B (2012). Natural history of irritable bowel syndrome in women and dysmenorrhea: A 10-year follow-up study. *Gastroenterology Research and Practice*, 2012, 534204. 10.1155/2012/534204 [PubMed: 22474441]
- Revicki D, Hays RD, Cella D, & Sloan J (2008). Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. *Journal of Clinical Epidemiology*, 61(2), 102–109. 10.1016/j.jclinepi.2007.03.012 [PubMed: 18177782]
- Teheran AA, Pineros LG, Pulido F, & Mejia Guatibonza MC (2018). WaLIDD score, a new tool to diagnose dysmenorrhea and predict medical leave in university students. *International Journal of Women's Health*, 10, 35–45. 10.2147/IJWH.S143510
- Turk DC, Dworkin RH, Burke LB, Gershon R, Rothman M, Scott J, . . . Wyrwich KW (2006). Developing patient-reported outcome measures for pain clinical trials: IMMPACT recommendations. *Pain*, 125(3), 208–215. 10.1016/j.pain.2006.09.028 [PubMed: 17069973]
- Westling AM, Tu FF, Griffith JW, & Hellman KM (2013). The association of dysmenorrhea with noncyclic pelvic pain accounting for psychological factors. *American Journal of Obstetrics & Gynecology*, 209(5), 422.e421–422.e410. 10.1016/j.ajog.2013.08.020 [PubMed: 23973396]
- Willis GB (2015). *Analysis of the cognitive interview in questionnaire design*. Oxford University Press.
- Wilson FR, Pan W, & Schumsky DA (2012). Recalculation of the critical values for Lawshe's content validity ratio. *Measurement and Evaluation in Counseling and Development*, 45(3), 197–210. 10.1177/0748175612440286

- Wyrwich KW, O'Brien CF, Soliman AM, & Chwalisz K (2018). Development and validation of the endometriosis daily pain impact diary items to assess dysmenorrhea and nonmenstrual pelvic pain. *Reproductive Sciences*, 25(11), 1567–1576. 10.1177/1933719118789509 [PubMed: 30033855]
- Yost KJ, Eton DT, Garcia SF, & Cella D (2011). Minimally important differences were estimated for six patient-reported outcomes measurement information system-cancer scales in advanced-stage cancer patients. *Journal of Clinical Epidemiology*, 64(5), 507–516. 10.1016/j.jclinepi.2010.11.018 [PubMed: 21447427]
- Yu L, Buysse DJ, Germain A, Moul DE, Stover A, Dodds NE, . . . Pilkonis PA (2011). Development of short forms from the PROMIS sleep disturbance and sleep-related impairment item banks. *Behavioral Sleep Medicine*, 10(1), 6–24. 10.1080/15402002.2012.636266 [PubMed: 22250775]



**Figure 1. Study Schema**

Note: Two phases included independent sets of participants.



**Table 1.**

## Demographic Characteristics of the Phase 1 and Phase 2 Samples

	Phase 1	Phase 2	
	Qualitative	Quantitative (n=686)	
	(n=30) n (%)	On Menses (n=260) n (%)	Off Menses (n=426) n (%)
Race			
White	20 (66.6)	159 (61.2)	305 (71.6)
Black or African American	2 (6.7)	50 (19.2)	41 (9.6)
American Indian or Alaska Native	1 (3.3)	5 (1.9)	6 (1.4)
Asian	4 (13.3)	29 (11.2)	25 (5.9)
Native Hawaiian or Pacific Islander	0 (0)	1 (0.4)	2 (0.5)
Other	1 (3.3)	5 (1.9)	21 (4.9)
More than One Race	2 (6.7)	6 (2.3)	19 (4.5)
Prefer not to answer	0 (0)	5 (1.9)	7 (1.6)
Ethnicity			
Hispanic, Spanish, or Latino	2 (6.7)	22 (8.5)	61 (14.3)
Not Hispanic, Spanish, or Latino	28 (93.3)	237 (91.2)	362 (85.0)
Prefer not to answer	0 (0)	1 (0.4)	3 (0.7)
Education Level Completed			
8th grade or less	0 (0)	3 (1.2)	8 (1.9)
Some high school	2 (6.7)	16 (6.2)	68 (16.0)
High school degree or GED	5 (23.3)	49 (18.8)	83 (19.5)
Some college	10 (33.3)	83 (31.9)	107 (25.5)
2-year college degree	1 (3.3)	26 (10.0)	44 (10.3)
4-year college degree	8 (26.7)	63 (24.2)	87 (20.4)
Postgraduate degree	4 (13.3)	20 (7.7)	24 (5.6)
Prefer not to answer	0 (0)	0 (0.0)	5 (1.2)
Employment <sup>1</sup>			
Full-time	8 (26.7)	108 (41.5)	126 (29.6)
Self employed	1 (3.3)	13 (5.0)	21 (4.9)
Part-time	2 (6.7)	31 (11.9)	64 (15.0)
Student	18 (60.0)	49 (18.8)	112 (26.3)
Homemaker	1 (3.3)	37 (14.2)	63 (14.8)
Unemployed	0 (0)	26 (10.0)	60 (14.1)
No answer	0 (0)	2 (0.8)	6 (1.4)
Other Gynecologic Conditions <sup>1</sup>			
Endometriosis	1 (3.3)	14 (5.4)	19 (4.5)
Uterine fibroids	1 (3.3)	12 (4.6)	9 (2.1)

	Phase 1	Phase 2	
	<u>Qualitative</u>	<u>Quantitative (n=686)</u>	
	(n=30)	On Menses (n=260)	Off Menses (n=426)
	n (%)	n (%)	n (%)
Pelvic Inflammatory disease	1 (3.3)	9 (3.5)	10 (2.3)
Polycystic ovary syndrome	1 (3.3)	18 (6.9)	15 (3.5)

<sup>1</sup>Participants could select more than one category.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

**Table 2.**

Psychometrics for the DSI Scale (N=686)

Version	Sample Size	Mean (SD)	Cronbach's Alpha	Concurrent Validity (Correlation Coefficients)						MID	
				MP at its least	MP at its worst	MP on average	MP right now	Perceived Stress	Sleep Disturbance	0.35 SD	1 SEM
On Menses <sup>I</sup>	260	2.9 (1.3)	0.93	0.53**	0.66**	0.58**	0.63**	0.14*	0.18*	0.36	0.27
Off Menses <sup>I</sup>	426	2.8 (1.2)	0.91	0.43**	0.52**	0.53**	--	0.31**	0.12*	0.32	0.27

<sup>I</sup>Time 1 measure was used.

\* <.05

\*\* <.001.

MP: Menstrual Pain; MID: minimally important difference; SD: Standard Deviation; SEM: Standard Error of Measurement

**Table 3.**

## DSI Scale Item Level Psychometrics

Item	Content Validity Index <sup>1</sup> (n=686)	% rated item "not necessary" <sup>2</sup> (n=686)	Item total correlation		Factor Loading	
			On-Menses <sup>3</sup> (n=260)	Off-Menses <sup>3</sup> (n=426)	On-Menses <sup>3</sup> (n=260)	Off-Menses <sup>3</sup> (n=426)
<b>Physical Activities</b> (e.g. walk, run, swim, yoga, & other exercises)	0.15	4.7%	0.73	0.67	.76 ***	.71 ***
<b>Sleep</b> (falling or staying asleep)	0.48	4.1%	0.64	0.57	.66 ***	.60 ***
<b>Daily Activities</b>	0.40	3.3%	0.84	0.77	.87 ***	.81 ***
<b>Work</b> (including work, housework, schoolwork, and homework)	0.39	7.1%	0.75	0.67	.79 ***	.71 ***
<b>Concentration</b>	0.16	5.1%	0.73	0.70	.76 ***	.74 ***
<b>Enjoyment of Life</b>	0.30	6.0%	0.81	0.74	.84 ***	.78 ***
<b>Leisure Activities</b> (time spent relaxing, having fun, doing hobbies, etc.)	-0.03	6.0%	0.76	0.77	.80 ***	.82 ***
<b>Social activities</b> (time spent with family, friends, etc.)	-0.04	6.3%	0.76	0.73	.79 ***	.77 ***
<b>Mood</b> (irritable, anxious frustrated, depressed, etc.)	0.60	3.5%	0.65	0.58	.67 ***	.61 ***

<sup>1</sup> Above critical value 0.064 indicates excellent content validity for a given item.

<sup>2</sup> Below 10% indicates acceptable content validity for a given item.

<sup>3</sup> Time 1 measure was used.

\*\*\* p<.0001.

**Table 4.**

## Confirmatory Factor Analysis Fit Indices

	Comparative Fit Index (CFI)	Goodness of Fit Index (GFI)	Bentler-Bonett Normed Fit Index	Root Mean Square Error of Approximation (RMSEA, 90% CI)	Standardized Root Mean Square Error (SRMR)
Goal <sup>1</sup>	> 0.9	>0.9	>0.8	<0.1	< 0.08
On Menses <sup>2</sup> (n=260)	0.95	0.92	0.94	0.1 (0.09–0.11)	0.03
Off Menses <sup>2</sup> (n=426)	0.96	0.95	0.94	0.08 (0.07–0.09)	0.03

<sup>1</sup>(Bartholomew et al., 2008; Hooper et al., 2008).

<sup>2</sup>Time 1 measure was used.