

Psychometric Evaluation of an Instrument to Measure Prospective Pregnancy Preferences

The Desire to Avoid Pregnancy Scale

Corinne H. Rocca, PhD,* Lauren J. Ralph, PhD,* Mark Wilson, PhD,† Heather Gould, MPH,* and Diana G. Foster, PhD*

Background: Existing approaches to measuring women's pregnancy intentions suffer important limitations, including retrospective assessment, overly simple categories, and a presumption that all women plan pregnancies. No psychometrically valid scales exist to prospectively measure the ranges of women's pregnancy preferences.

Materials and Methods: Using a rigorous construct modeling approach, we developed a scale to measure desire to avoid pregnancy. We developed 60 draft items from existing research, assessed comprehension through 25 cognitive interviews, and administered items in surveys with 594 nonpregnant women in 7 primary and reproductive health care facilities in Arizona, New Jersey, New Mexico, South Carolina, and Texas in 2016–2017. We used item response theory to reduce the item set and assess the scale's reliability, internal structure validity, and external validity. Items were included based on fit to a random effects multinomial logistic regression model (partial credit item response model), correspondence of item difficulty with participants' pregnancy preference levels, and

consistency of each item's response options with overall scale scores.

Results: The 14 final items covered 3 conceptual domains: cognitive preferences, affective feelings, and practical consequences. Items fit the unidimensional model, with a separation reliability of 0.90 (Cronbach α : 0.95). The scale met established criteria for internal validity, including correspondence between each item's response categories and overall scale scores. We found no important differential item functioning by participant characteristics.

Conclusions: A robust measure is available to prospectively measure desire to avoid pregnancy. The measure can aid in identifying women who could benefit from contraceptive care and research on less desired pregnancy.

Key Words: pregnancy intention, pregnancy preferences, latent trait theory, measurement, item response model, item response theory, United States

(*Med Care* 2019;57: 152–158)

From the *Department of Obstetrics, Gynecology and Reproductive Sciences, Bixby Center for Global Reproductive Health, Advancing New Standards in Reproductive Health (ANSIRH), University of California, San Francisco, School of Medicine, San Francisco; and †University of California, Graduate School of Education, Berkeley, CA.

Supported by the Eunice Kennedy Shriver National Institute of Child Health and Human Development, Office of Research on Women's Health, Building Interdisciplinary Research Careers in Women's Health grant #HD052163 and funding from an anonymous private foundation.

Preliminary results were presented at the North American Forum on Family Planning, November 2016, Denver, CO, and the Annual Meetings of the Population Association of America (PAA), April 2017, Chicago, IL.

The authors declare no conflict of interest.

Reprints: Corinne H. Rocca, PhD, Department of Obstetrics, Gynecology and Reproductive Sciences, Bixby Center for Global Reproductive Health, Advancing New Standards in Reproductive Health (ANSIRH), University of California, San Francisco, School of Medicine, 1330 Broadway Street, Suite 1100, Oakland, CA 94612. E-mail: corinne.rocca@ucsf.edu.

Supplemental Digital Content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's website, www.lww-medicalcare.com.

Copyright © 2018 The Author(s). Published by Wolters Kluwer Health, Inc. This is an open-access article distributed under the terms of the Creative Commons Attribution-Non Commercial-No Derivatives License 4.0 (CCBY-NC-ND), where it is permissible to download and share the work provided it is properly cited. The work cannot be changed in any way or used commercially without permission from the journal.

ISSN: 0025-7079/19/5702-0152

Unintended pregnancy is high in the United States, with 45% of pregnancies considered unintended.¹ Rates have declined only slightly over decades, despite efforts to understand and address the root causes.^{1,2} Absent from research, however, has been a robust approach to measuring the intendedness of pregnancies. Indeed, accurate and reliable measurement of pregnancy intentions is essential to identifying and understanding unintended pregnancies and distinguishing women in need of contraceptive care.^{3–6} However, no psychometrically validated measures have been developed to prospectively measure this complex latent construct. Reproductive health experts have expressed an urgent need for improved measures of pregnancy intention.^{3–9}

Current measurement approaches suffer important conceptual and practical limitations. Most commonly, pregnancies are categorized as intended, mistimed, or unintended based on questions asking women to report their intentions at the time of conception.^{1,10,11} These categories, however, do not capture the ranges of feelings many women have about a potential pregnancy.⁸ Research illustrates that while some women may strongly desire to become pregnant or to prevent pregnancy, others hold mixed feelings or are uncertain.^{12–15} For example, a woman may feel that a baby would make her

happy but also that she is not able to care for one.¹⁶ Continuous measures may better capture such nuances.

Relatedly, approaches to measuring pregnancy intention typically assume that women proactively plan their pregnancies.⁹ Consistent with intentionality-based behavior theory,^{17,18} these approaches assume that women hold and can articulate clear intentions, and that they engage in contraceptive and sexual behavior, to the extent possible, accordingly. Yet research reveals that women interpret the terms trying, planning, wanting, and intending a pregnancy differently, and many do not intuitively apply those categorizations to their own pregnancies.^{12,14} Some view pregnancies as only somewhat within their control, and for some, preferences about pregnancy can be vague, underspecified, and even unconscious.^{19,20} Furthermore, approaches tend to focus on the desire for a pregnancy and less on the desire to prevent pregnancies, which may be more important for the provision of contraceptive services.

Third, most measures, including those developed with robust methodologies, are retrospective, asking women about pregnancies they have experienced.^{21,22} While important for categorizing previous pregnancies and for population research in which prospective measurement is not possible, it remains unclear the degree to which individual women modify their perceived intentions in retrospect.^{23,24} Furthermore, measures categorizing pregnancies retrospectively are not useful for describing women's feelings when pregnancy has not occurred or identifying women in need of contraceptive services.

Finally, approaches to measuring prospective pregnancy intention have generally not been purposively developed or undergone rigorous psychometric evaluation. The development and evaluation of valid and reliable measurement instruments for latent variables are common in fields such as psychology and education, with norms and standards for psychometric analyses long-established in these disciplines. These approaches are only beginning to be applied in the reproductive health field.^{15,25}

We used construct modeling and item response theory-based methods to develop and evaluate a psychometrically sound measure of prospective pregnancy preferences, the desire to avoid pregnancy (DAP) measure. Our aim was to develop a scale that balanced precision with accuracy and construct validity; captured the multiple domains of pregnancy preferences; and that had a maximum of 15–20 items to improve usability.

MATERIALS AND METHODS

Theory and Key Conceptual Features

We grounded conceptualization of pregnancy preferences and the development of the DAP scale in preference constructionist theory from psychology and behavioral decision science,²⁶ as well as item response theory, or item response modeling (IRM) from measurement science.^{27–29} In contrast to rational choice models, preference construction theory posits that individuals often do not have clear preferences, particularly preferences that involve complex choices or are context-specific. When called upon to express a

preference, however, uncertain individuals will construct a preference (as opposed to drawing one from memory). The approach aligns with Bachrach and Morgan's cognitive-social model in which intentions are typically formed only when circumstances motivate one,³⁰ and Johnson-Hanks's arguments about the contingency of preferences given the uncertainty of the future.³¹ IRM, similarly, is based on the premise that individuals hold attitudes and attributes that are unobservable.²⁹ However, in responding to scale items, these latent traits become manifest through the individual's responses. When administering a psychometric measure with items assessing feelings and thoughts about a potential pregnancy or child, individuals will construct a preference, which is then manifest to observers in the items' responses.

These theories are reflected in key conceptual features of the DAP measure. First, we state deliberately that the instrument measures pregnancy "preferences" vs. "intentions," recognizing that uncertainty and ambiguity are legitimate stances when it comes to pregnancy. We treat preferences as latent and unobservable, allowing them to be either conscious or unconscious and multifaceted (eg, heart vs. mind, cognitive vs. affective). We include a middle response category (neither agree nor disagree), acknowledging that not having a preference regarding an item is a legitimate stance. Items cover considerations about both pregnancy and childbearing.^{15,32} Furthermore, instead of focusing on the desire *for* pregnancy, we focused the scale on desire to *avoid* pregnancy, which is more relevant to addressing "unintended" pregnancy and directing contraceptive services to those in need. Responses are coded so that a higher score indicates a stronger desire to avoid pregnancy. Finally, recognizing that preferences about pregnancy can change over time based on life circumstances, we used a relatively short time frame: 3 months for a possible pregnancy and 1 year for a new baby.^{33,34}

Item Development

We used a construct modeling approach to develop the items for the measure.²⁹ On the basis of an extensive literature review and input from experts, we defined the construct of desire to avoid pregnancy as a woman's underlying will to prevent herself from becoming pregnant, or her underlying predisposition against pregnancy. As a framework within which to ensure we included items covering all aspects of pregnancy preferences, we identified 3 conceptual domains: cognitive self-evaluation of preferences around pregnancy and childbearing; affective feelings about a potential pregnancy and child; and anticipated practical consequences if pregnancy and childbearing were to occur. We created a construct map for pregnancy preferences and situated them in a directed acyclic graph, delineating their relationship to social and structural factors (eg, sex norms, partner desires) and reproductive outcomes (eg, contraceptive use, pregnancy).

We developed a library of 60 draft items, with each item derived directly from empirical qualitative work on how women conceptualize a potential pregnancy. Items were worded in both directions, so agreement with an item could mean either high or low desire to avoid pregnancy, depending on the item. We tested item comprehension through cognitive interviews with 25 women at 2 reproductive health facilities

in California. Participants were women aged 15 or older, who read and spoke English or Spanish and did not report current pregnancy. Participants were asked to think aloud as they responded to the items to assess whether they interpreted items as intended. We probed understanding of particular phrases, including “partner,” “end of the world,” and “stressed out.” We offered different response options, assessing which ones participants preferred and whether they felt a middle category was important. We tested to see whether women perceived similar items asking about a potential pregnancy versus child differently, and probed about the 3-month pregnancy and 1-year childbirth timeframes. We asked whether women felt upset by any items and, among Spanish speakers, about the interpretation of the translations. Finally, we asked about the ordering of items and whether items omitted anything. Items, response categories, and translations were honed and finalized based on this feedback.

Field Test

We recruited women from 7 reproductive and primary health facilities in Arizona, New Jersey, New Mexico, South Carolina, and Texas in 2016–2017, as a part of a study examining women’s pathways to suspicion and confirmation of new pregnancies. Our recruitment approach targeted socio-demographically and geographically diverse US women with all ranges of pregnancy preferences who were “at risk” of pregnancy. Scale development and evaluation can be validly conducted using nonprobability samples.³⁵

A trained research assistant approached all women in the waiting room with a study flier, and eligible women who were willing to participate provided verbal informed consent. To participate, women had to be aged 15–45, sexually active in the last year, not sterilized, and to speak and read English or Spanish. To align with the patchwork of minor consent laws across states, we only recruited minors in states or clinics where they could legally consent to receive facility services; we thus included minors aged 15–17 from all sites except 2 in South Carolina, where minors aged 15 were excluded. Participants completed a 30-minute anonymous survey on a tablet, providing information on sociodemographics and contraceptive use. Participants reporting they were not pregnant or did not know whether they were pregnant responded to the 60 candidate pregnancy preferences items. Participants received a \$20 gift card for completing the study.

Psychometric Analysis

We used IRM-based methods to reduce the item set and assess the performance of items as a scale, and to place individuals along the continuum of the underlying latent construct.^{27,28} IRM is considered a leading statistical paradigm for the development and evaluation of measurement scales, including for patient-reported measures.^{28,29,36–39} By fitting item responses to statistical models, IRM avoids key assumptions of classical approaches, including that the “distance” between response categories within each item and between items are equal, and that measurement precision is uniform across the range of scale scores. IRM also provides a richer description of each item’s performance, helping to ensure the best performing items are selected.

Analyses were conducted using ACER ConQuest version 4.5.2⁴⁰ and were consistent with guidelines for psychometric testing.^{29,41} An expansion of the Rasch model for polytomous data, the partial credit model (PCM), was used in the analyses.⁴² In this formulation, the location parameters can be decomposed into an item difficulty and step parameter, which corresponds to k-1 item response categories. We selected the more parsimonious PCM over the graded response model for several reasons, including that it allows researchers not familiar with IRM to use the raw summed scores to estimate trait scores, which is not sufficient with the graded response model. Also, because discrimination is invariant across all items with the PCM, we can directly and consistently interpret the relative values of person and item and step parameters—for items fitting the PCM—as a probability of a person endorsing a given response option. This feature will be important in future work to identify relevant cut-points.

We used an iterative process to select from the 60 candidate items those that would remain in the measure and assessed the psychometric properties of the final item set. We first assessed item acceptability and distribution of responses, removing items with >5% missing responses and for which >75% of responses fell in a single response category. After fitting item responses to the unidimensional PCM, we assessed item fit using a weighted mean-squared index, removing items falling outside of 0.67–1.33 as a general guideline.⁴³ We also considered removing items that covered the same conceptual territory as other, better performing items. We ensured that, for each item, women endorsing the highest response also had higher scores, on average, on the scale overall. We plotted item characteristic curves, which depict the expected probability of endorsement of each item’s response categories along all ranges of the scale; we removed items for which the ordering of responses was inconsistent with overall scale scores. We balanced construct validity with reliability in item selection, ensuring that the final measure included items that captured all conceptual domains. We aimed to find a solution with the fewest items possible while maintaining validity to create a scale that is practical to use in research.

Once we decided on a final set of items, we reanalyzed the data to check item and scale psychometric properties, including item fit and mean locations. We assessed internal consistency reliability with the separation reliability coefficient. To assess internal structure validity, we plotted women’s overall scores on a scale next to item threshold difficulty levels (ie, Wright Maps). We examined the plot to ensure that items’ thresholds spanned the range of participants’ pregnancy preferences and served to differentiate women along the construct. We also fit the data to a multidimensional PCM and assessed correlation of scores across domains, using an a priori >0.90 correlation between domains as a threshold of important differences.

Because no psychometric instruments to measure pregnancy intention exist, we compared DAP scores among women reporting using contraception in the past 30 days, not using contraception, and those who had not had sex within 30 days, to garner evidence of validity based on external variables. We used explanatory IRM, integrating contraceptive use terms

into the PCM.²⁷ We hypothesized women not using contraception would have lower DAP scores than those contracepting or not currently sexually active.

Finally, we assessed differential item functioning (DIF) to investigate whether any items performed differentially between women based on their age, race/ethnicity, main partnership, or whether they had children. DIF examines group (eg, sociodemographic) differences in the probability of a response to an item, conditional on the underlying trait (pregnancy preferences)—or group differences in item parameters. The presence of DIF can sometimes indicate that an item is biased (ie, if young women are more likely to agree to an item than older women who otherwise have similar pregnancy preferences) and can also provide insight into how different groups form pregnancy preferences. We assessed DIF by fitting a new partial credit DIF model, which expanded on the PCM by incorporating item-by-group interactions, so that item responses were predicted by item difficulty, step, group, and item-by-group effects.⁴⁴ We considered items for removal based on a priori effect sizes rather than log-likelihood ratio tests,^{45,46} using an item-by-group parameter effect size of ≥ 0.6 logits as evidence of DIF.^{47,48} We confirmed results by calibrating the underlying trait separately in each subgroup and comparing the resulting item parameter estimates, again considering a 0.6 logit difference as evidence of DIF.⁴⁹

To translate scale properties into a more traditional scale evaluation framework, we evaluated the final scale using a classical approach. We summed raw scores across items and evaluated the distribution, calculated item-total correlations, and assessed internal consistency reliability with Cronbach’s α .

RESULTS

A total of 810 women enrolled into the pregnancy suspicion and confirmation study. Among the 810, 207 reported they were pregnant and 10 were missing a pregnancy status. Among the 602 reporting they were not pregnant or did not know whether they were pregnant, 8 were called into their clinical appointment before completing the DAP items, leaving 594 in the analysis sample.

The 594 participants were, on average, 27 years old, and 49% identified as Latina, 27% as black, 16% white, and 8% multiracial or other (Table 1). About 26% were married, and the large majority (83%) had a partner they considered to be main or serious. Over 60% had ≥ 1 children. Almost half (47%) received some sort of public assistance, and 47% had experienced food insecurity within a year. Most participants present to a clinic for reproductive health care, including contraception (49%), a pelvic examination or Pap smear (24%), pregnancy testing (10%), or sexually transmitted infection (9%) testing.

The final DAP measure included 14 items (Supplement 1, Supplemental Digital Content 1, <http://links.lww.com/MLR/B654>). Participants expressed the full range of emotions and desires about pregnancy in response to the DAP items. For instance, 23% agreed or strongly agreed that it would be “the end of the world” for them if they were to have a baby in the next year, while 63% disagreed or strongly disagreed. At the same time, when asked to think about becoming pregnant, over half (54%) disagreed or strongly

TABLE 1. Respondent Characteristics (N = 594)

Characteristics	Mean (SD)
Age (range:15–45)	27.3 (7.5)
	n (%)
Age group	
15–24	243 (40.9)
24–45	351 (59.1)
Race/ethnicity (N = 583)	
Latina	284 (48.7)
Non-Latina black	160 (27.4)
Non-Latina white	95 (16.3)
Multiracial/other	44 (7.6)
Married (N = 589)	152 (25.8)
Has a main/serious partner (N = 586)	487 (83.1)
Children (N = 576)	
0	226 (39.2)
1	118 (20.5)
2	112 (19.4)
3+	120 (20.8)
Education (N = 565)	
Less than high school	87 (15.4)
High school or general education diploma	344 (60.9)
Some community or technical college or degree	76 (13.5)
College degree or more	58 (10.3)
Receives public assistance (N = 573)	268 (46.8)
Experiences food insecurity (N = 584)	272 (46.6)
Current contraceptive use (N = 591)	
Yes	401 (67.9)
No	123 (20.8)
No sex with male in last 30 d	67 (11.3)
Reason for clinic visit (N = 535)	
Contraceptive care	260 (48.6)
Pelvic examination/Pap smear	126 (23.6)
Pregnancy test	52 (9.7)
Sexually transmitted infection test	49 (9.2)
Injury, illness, chronic condition	48 (9.0)

disagreed that it makes them feel excited, whereas 25% agreed or strongly agreed. Although 57% agreed or strongly agreed that they would have difficulty managing to raise the child, 30% disagreed or strongly disagreed.

Fit to the item response model, participant DAP levels fell in a rough bell-shaped curve, with a slight left skew or tendency toward higher preference to avoid pregnancy (Supplement 2, Supplemental Digital Content 2, <http://links.lww.com/MLR/B655>). The Wright Map illustrated that items and category thresholds spanned the range of participant levels, indicating the categories served to differentiate participants along the scale. Corresponding averaged raw scores fell from 0 to 4 (mean, 2.2; SD, 1.1).

The final 14 items fit the unidimensional PCM well (Table 2). The item with the lowest location was “It would be a good thing for me if I became pregnant in the next 3 months” (location: -0.92 logits), indicating that participants did not need to desire to avoid pregnancy very much to disagree with the item. The item “It would be the end of the world for me to have a baby in the next year” had the highest location, indicating that participants had to have a strong desire to avoid pregnancy to agree with the item. Notably, when fit to a multidimensional model, responses on the 3 domains were highly correlated (0.97, 0.94, and 0.93), indicating all items tapped into a single construct and supporting the use of a unidimensional model.

TABLE 2. Scale Reliability and Properties, IRM, and Classical Test Theory

	IRM (Separation Reliability: 0.90)		Classical Test Theory (Cronbach α : 0.95)
	Item Fit	Item Location	Item-Total Correlation
Pregnant: would not mind (+)	0.92	-0.60	0.79
Pregnant: good thing for me (+)	0.97	-0.92	0.75
Pregnant: unhappy (-)	0.87	0.28	0.81
Pregnant: excited (+)	0.82	-0.36	0.83
Pregnant: closer to partner (+)	1.80	0	0.53
Baby: want (+)	0.84	-0.61	0.80
Baby: bad for life (-)	0.80	0.69	0.82
Baby: positive addition to life (+)	0.86	0	0.80
Baby: end of the world for me (-)	1.04	0.93	0.75
Baby: makes me smile (+)	0.92	0.24	0.80
Baby: stressed out (-)	1.04	-0.16	0.77
Baby: loss of freedom (-)	1.19	0.42	0.73
Baby: hard for me to manage (-)	1.03	0.33	0.77
Baby: harder to achieve other things (-)	1.09	-0.23	0.77

Item fit and location are derived from a unidimensional partial credit item response model for polytomous items. Item fit is the weighted mean-squared fit *t* statistic. Item location is the difficulty parameter in logits.

(+): Item coded so that “strongly agree” = 0 and “strongly disagree” = 4. (-): Item coded so that “strongly agree” = 4 and “strongly disagree” = 0.

IRM indicates item response modeling.

The 14-item DAP scale (IRM-derived scores) had a separation reliability of 0.90. Using raw scores, the DAP had a Cronbach’s α of 0.95. The item “Becoming pregnant in the next 3 months would bring me closer to my (main) partner” fell outside of the prespecified bounds for good fit and

reduced the scale’s reliability; however, the item was retained to ensure that all key areas of the construct of pregnancy preferences were covered.

DAP scale items met prespecified criteria for internal structure validity, including having category responses that corresponded to participant scores on the scale overall. As hypothesized, women not using contraception scored significantly higher on the DAP than those using a method (0.75 logits higher) and those who had not had sex in 30 days (0.74 logits higher; $P < 0.001$).

We did not find important DIF by race/ethnicity or main partnership (Table 3). There was some evidence of DIF for the item “Becoming pregnant in the next 3 months would bring me closer to my (main) partner,” with women aged 15–24 and those without children less likely to disagree that becoming pregnant would bring them closer to their partner than women aged 24–45 and those with children, respectively. These results were confirmed comparing item parameter estimates on models calibrated separately by subgroup.

Each DAP item has response options that range from 0 to 4. It is recommended that researchers using IRM fit item responses to a PCM. Researchers using a classical approach should sum raw item scores and divide by 14 to obtain an average pregnancy preferences score (final range: 0–4). Higher scores reflect a higher desire to avoid pregnancy. The DAP is intended to be used as a continuous measure; rounding is not recommended (Supplement 1, Supplemental Digital Content 1, <http://links.lww.com/MLR/B654>).

CONCLUSIONS

The DAP instrument measures the latent construct of preference to avoid pregnancy, capturing a range of pregnancy and childbearing preferences including cognitive evaluations of preferences, feelings, and anticipated practical consequences. The scale showed good reliability and met criteria for internal structure and external validity. All items fit a unidimensional

TABLE 3. DIF Parameter Estimates for Item-by-Group Interactions for Desire to Avoid Pregnancy Scale Items

Items	Age Group	Race/Ethnicity		Has Main Partner	Has Children
	> 24 (vs. 15–24)	Black (vs. White)	Latina (vs. White)	Partner (vs. Not)	Children (vs. Not)
Pregnant: would not mind	0.03	0.32	0.08	-0.04	0.14
Pregnant: good thing for me	0.02	-0.12	-0.20	0.12	0.19
Pregnant: unhappy	0.05	0.18	0.06	0.04	0.22
Pregnant: excited	0.18	0.13	0.04	0.02	0.15
Pregnant: closer to partner	0.70*	0.12	-0.18	-0.30	0.74*
Baby: want	0.14	0.08	0.02	-0.05	0.19
Baby: bad for life	0.25	-0.15	-0.28	0.12	-0.24
Baby: positive addition to life	0.12	0.14	-0.02	-0.20	0.05
Baby: end of the world for me	0.23	0.11	0.44	0.11	-0.13
Baby: makes me smile	0.24	0.26	-0.14	-0.13	0.16
Baby: stressed out	-0.38	-0.44	-0.27	0.06	-0.48
Baby: loss of freedom	-0.24	-0.18	0.02	0.14	-0.46
Baby: hard for me to manage	-0.18	-0.30	0.12	-0.07	-0.28
Baby: harder to achieve other things	-0.18	-0.12	0.32	0.16	-0.26

Parameters presented are item-by-group estimates from a partial credit model that predicted item response by item, step, sociodemographic group, and item-by-group interactions. They represent the difference in location on the logit scale between women in each sociodemographic group.

*Estimate surpasses the ± 0.6 logit threshold for DIF.

DIF indicates differential item functioning.

model, indicating that the items, which capture multiple aspects of pregnancy preferences, all tap into a single pregnancy preferences construct. The DAP scale is the first prospective measure of pregnancy preferences to be developed and evaluated using rigorous psychometric methods.

The item “Becoming pregnant in the next 3 months would bring me closer to my (main) partner” had poorer than desired fit to the PCM and exhibited some evidence of DIF by age and whether the woman had children. We elected to maintain the item because partnerships are a critical element of pregnancy preferences.⁵⁰ In addition, we interpret the DIF we detected to be less an issue of bias and more a reflection of how women in these groups view the prospect of pregnancy. Perhaps older women and women with children have learned from experience that childbearing would not necessarily result in becoming closer to one’s partner. Research into the performance of this item and alternative ways to express it are needed, as are examinations of DIF in other samples.

One disadvantage of the DAP scale is the inability to differentiate respondents who feel indifference or uncertainty about pregnancy (eg, respond “neither” on many items) from those who feel ambivalence (eg, respond “agree” to both negative and positive items), as these women would have similar overall scores. Researchers wanting to distinguish these groups may choose to examine responses to positively worded items separately from negatively worded ones.

A robust prospective measure of pregnancy preferences will contribute greatly to research on less intended pregnancy. Almost half of US pregnancies are considered “unintended” based on a retrospective categorical item, with stark differences by age, race/ethnicity, socioeconomic status, and other groupings.¹ Use of the DAP measure can lead to more precise and nuanced estimates, enabling us to better understand differences in and causes of less intended pregnancy, and to identify the actual consequences of pregnancies that were strongly undesired by women. The measure can be used to examine inconsistencies between reported intention and contraceptive use and pregnancy on the individual level. It can also help to elucidate pathways to pregnancy and the extent to which preferences are an independent risk factor versus a mediator of underlying risk factors.

ACKNOWLEDGMENTS

The authors are very grateful for the contributions of Rana Barar for project direction and Brenly Rowland, Clara Finley Baba, and Jasmine Powell for data collection and study administration. The authors thank the staff of the recruitment sites for hosting their research team.

REFERENCES

- Finer LB, Zolna MR. Declines in unintended pregnancy in the United States, 2008–2011. *N Engl J Med*. 2016;374:843–852.
- US Department of Health and Human Services. Health people 2020, family planning objectives. 2011. Available at: HealthyPeople.gov. Accessed August 10, 2015.
- Gipson JD, Koenig MA, Hindin MJ. The effects of unintended pregnancy on infant, child, and parental health: a review of the literature. *Stud Fam Plann*. 2008;39:18–38.
- Mumford SL, Sapra KJ, King RB, et al. Pregnancy intentions—a complex construct and call for new measures. *Fertil Steril*. 2016;106:1453–1462.
- Santelli J, Rochat R, Hatfield-Timajchy K, et al. The measurement and meaning of unintended pregnancy. *Perspect Sex Reprod Health*. 2003;35:94–101.
- Petersen R, Moos MK. Defining and measuring unintended pregnancy: issues and concerns. *Womens Health Issues*. 1997;7:234–240.
- Santelli JS, Lindberg LD, Orr MG, et al. Toward a multidimensional measure of pregnancy intentions: evidence from the United States. *Stud Fam Plann*. 2009;40:87–100.
- Bachrach CA, Newcomer S. Intended pregnancies and unintended pregnancies: distinct categories or opposite ends of a continuum? *Fam Plann Perspect*. 1999;31:251–252.
- Aiken ARA, Borrero S, Callegari LS, et al. Rethinking the pregnancy planning paradigm: unintended conceptions or unrepresentative concepts? *Perspect Sex Reprod Health*. 2016;48:147–151.
- Casterline JB, El-Zeini LO. The estimation of unwanted fertility. *Demography*. 2007;44:729–745.
- National Center for Health Statistics. *National Survey of Family Growth 2015*. Atlanta, GA: Centers for Disease Control and Prevention; 2015.
- Barrett G, Wellings K. What is a “planned” pregnancy? Empirical data from a British study. *Soc Sci Med*. 2002;55:545–557.
- Borrero S, Nikolajski C, Steinberg JR, et al. “It just happens”: a qualitative study exploring low-income women’s perspectives on pregnancy intention and planning. *Contraception*. 2015;91:150–156.
- Kendall C, Afbale-Munsuz A, Speizer I, et al. Understanding pregnancy in a population of inner-city women in New Orleans: results of qualitative research. *Soc Sci Med*. 2005;60:297–311.
- Rocca CH, Harper CC, Raine-Bennett TR. Young women’s perceptions of the benefits of childbearing: associations with contraceptive use and pregnancy. *Perspect Sex Reprod Health*. 2013;45:23–32.
- Aiken ARA, Dillaway C, Mevs-Korff N. A blessing I can’t afford: factors underlying the paradox of happiness about unintended pregnancy. *Soc Sci Med*. 2015;132:149–155.
- Ajzen I. From intentions to actions: a theory of planned behavior. In: Kuhl J, Beckmann J, eds. *Action Control: From Cognition to Behavior*. New York, NY: Springer-Verlag; 1985:11–39.
- Fishbein M, Ajzen I. *Belief, Attitude, Intention and Behavior: An Introduction to Theory and Research*. Reading, MA: Addison-Wesley; 1975.
- Gribaldo A, Judd MD, Kertzer DI. An imperfect contraceptive society: fertility and contraception in Italy. *Popul Dev Rev*. 2009;35:551–584.
- Johnson-Hanks J. Demographic transitions and modernity. *Ann Rev Anthropol*. 2008;37:301–315.
- Barrett G, Smith SC, Wellings K. Conceptualisation, development, and evaluation of a measure of unplanned pregnancy. *J Epidemiol Community Health*. 2004;58:426–433.
- Morin P, Payette H, Moos M-K, et al. Measuring the intensity of pregnancy planning effort. *Paediatr Perinat Epidemiol*. 2003;17:97–105.
- Joyce T, Kaestner R, Korenman S. On the validity of retrospective assessments of pregnancy intention. *Demography*. 2002;39:199–213.
- Guzzo KB, Hayford SR. Revisiting retrospective reporting of first-birth intendedness. *Matern Child Health J*. 2014;18:2141–2147.
- Rocca CH, Krishnan S, Barrett G, et al. Measuring pregnancy planning: an assessment of the London Measure of Unplanned Pregnancy among urban, south Indian women. *Demogr Res*. 2010;23:293–334.
- Bhrolchain MN, Beaujouan E. *How Real are Reproductive Goals? Uncertainty and the Construction of Fertility Preferences*. Southampton, UK: ESRC Centre for Population Change; 2015.
- De Boeck P, Wilson M. *Explanatory Item Response Models: a Generalized Linear and Nonlinear Approach*. New York, NY: Springer-Verlag; 2004.
- Embretson SE, Reise SP. *Item Response Theory for Psychologists*. Mahwah, NJ: MEA; 2000.
- Wilson M. *Constructing Measures: An Item Response Modeling Approach*. Mahwah, NJ: Lawrence Erlbaum Associates; 2005.
- Bachrach CA, Morgan SP. A cognitive-social model of fertility intentions. *Popul Dev Rev*. 2013;39:459–485.
- Johnson-Hanks J. When the future decides: uncertainty and intentional action in contemporary Cameroon. *Curr Anthropol*. 2005;46:363–385.
- Hayford SR, Guzzo KB, Kusunoki Y, et al. Perceived costs and benefits of early childbearing: new dimensions and predictive power. *Perspect Sex Reprod Health*. 2016;48:83–91.

33. Jones RK. Change and consistency in US women's pregnancy attitudes and associations with contraceptive use. *Contraception*. 2017;95:485–490.
34. Rocca CH, Hubbard AE, Johnson-Hanks J, et al. Predictive ability and stability of adolescents' pregnancy intentions in a predominantly Latino community. *Stud Fam Plann*. 2010;41:179–192.
35. Kline P. *A Handbook of Test Construction: Introduction to Psychometric Design*. London, UK: Methuen; 1986.
36. Embretson SE. The new rules of measurement. *Psychol Assess*. 1996;8:341–349.
37. Hays RD, Morales LS, Reise SP. Item response theory and health outcomes measurement in the 21st century. *Med Care*. 2000;38:II28–II42.
38. Nguyen TH, Han H-R, Kim MT, et al. An introduction to item response theory for patient-reported outcome measurement. *Pat Cen Out Res*. 2014;7:23–35.
39. Wilson M, Allen DD, Li JC. Improving measurement in health education and health behavior research using item response modeling: comparison with the classical test theory approach. *Health Educ Res*. 2006;21:19–32.
40. Adams RJ, Wu ML, Macaskill G, et al. *ACER ConQuest version 4.5.0: Generalized Item Response Modeling Software*. Camberwell, Australia: Australian Council for Educational Research and University of California, Berkeley; 2016.
41. American Educational Research Association, American Psychological Association, National Council for Measurement in Psychology. *Standards for Educational and Psychological Testing*. Washington, DC: American Psychological Association; 2014.
42. Masters GN. Rasch model for partial credit scoring. *Psychometrika*. 1982;47:149–174.
43. Wright BD, Masters GN. *Rating Scale Analysis (Rasch Measurement Series)*. Chicago, IL: MESA Press; 1982.
44. Meulders M, Xie Y. Person-by-item predictors. In: De Boeck P, Wilson M, eds. *Explanatory Item Response Models: A Generalized Linear and Nonlinear approach*. New York, NY: Springer-Verlag; 2004:213–226.
45. Steinberg L, Thissen D. Using effect sizes for research reporting: examples using item response theory to analyze differential item functioning. *Psychol Methods*. 2006;11:402–415.
46. Uebelacker LA, Strong D, Weinstock LM, et al. Use of item response theory to understand differential functioning of DSM-IV major depression symptoms by race, ethnicity and gender. *Psychol Med*. 2009;39:591–601.
47. Longford NT, Holland PW, Thayer DT. Stability of the MH D-DIF statistics across populations. In: Holland PW, Thayer DT, eds. *Differential Item Functioning*. Hillsdale, NJ: Lawrence Erlbaum Associates; 1993:67–113.
48. Paek I. *Investigations of Differential Item Functioning: Comparisons Among Approaches, and Extension to a Multidimensional Context [unpublished doctoral dissertation]*. Berkeley: University of California; 2002.
49. Wilson M. *Validity Constructing Measures: An Item Response Modeling Approach*. Mahwah, NJ: Lawrence Erlbaum Associates; 2005:165–169.
50. Zabin LS, Huggins GR, Emerson MR, et al. Partner effects on a woman's intention to conceive: "not with this partner". *Fam Plann Perspect*. 2000;32:39–45.