

A Review and Critique of the Sexual Experiences Questionnaire (SEQ)

Barbara A. Gutek,^{1,3} Ryan O. Murphy,¹ and Bambi Douma²

This paper reviews and critiques the Sexual Experiences Questionnaire (SEQ), "... a self-report inventory representing the first attempt to assess the prevalence of sexual harassment in a manner that met traditional psychometric standards" (Fitzgerald, Gelfand, & Drasgow, 1995, p. 427). Widely used by its developers and others as a measure of sexual harassment, the SEQ is not a finished product, has a number of problems, and has weak psychometric properties. Because of inconsistencies (e.g., in time frame, number of items, wording of items), the SEQ lacks the advantages of standardized measures, such as the ability to assess changes over time. It defines sexual harassment very broadly, having the effect of distorting findings about sexual harassment. Most importantly, it is not clear what or whose definition of sexual harassment the SEQ assesses.

KEY WORDS: Sexual harassment; measurement; validity; sex discrimination; employment law.

The Sexual Experiences Questionnaire (SEQ) developed by Louise Fitzgerald and her colleagues and students was, according to its authors, developed to address the absence in the sexual harassment literature of the attempt to link data collection to a conceptual framework or to ascertain reliability and validity of the measures used (Fitzgerald et al., 1988). According to its developers, it is "... a self-report inventory representing the first attempt to assess the prevalence of sexual harassment in a manner that met traditional psychometric standards" (Fitzgerald, Gelfand, & Drasgow, 1995, p. 427). It has been described as being "... generally acknowledged as the most theoretically and psychometrically sophisticated instrument available" (Fitzgerald, Gelfand, et al., 1995, p. 428; see also Cortina, Swan, Fitzgerald, & Waldo, 1998) and as "... the most psychometrically sound measure of sexual harassment" (e.g., Fitzgerald, Magley, Drasgow, & Waldo, 1999, p. 245; see also Magley, Hulin, Fitzgerald, & DeNardo, 1999, p. 394; Schneider, Swan, & Fitzgerald, 1997, p. 404). The

¹Department of Management and Policy, University of Arizona, Tucson, Arizona.

²School of Business Administration, University of Montana, Missoula, Montana.

³To whom correspondence should be addressed at Department of Management and Policy, Eller College of Business and Public Administration, University of Arizona, PO Box 210108, Tucson, Arizona 85721; e-mail: bgutek@eller.arizona.edu.

SEQ has also been used to assess the level of sexual harassment in lawsuits brought against employers. An APA Monitor article (Murray, 1998) lauded the SEQ and quoted a forensic psychologist who described it as, "the gold standard in this area" and saying further that the SEQ "makes for solid science in the courtroom."

But does it? In a recent federal case, a judge issued a pretrial order dismissing results from the SEQ. In that instance the EEOC filed a federal civil suit against Dial Corporation (EEOC v. Dial Corporation, Nov. 17, 2002. Ill, No. 99 C 3356). The suit alleged that a group of nearly 100 female workers at Dial were routinely subjected to unwanted touching, verbal abuse, and sexual threats from male coworkers. As part of its civil action, the EEOC used evidence garnered through administration of the SEQ to Dial personnel presently or previously employed at an Illinois Dial soap production plant. The results from the SEQ were integrated into a report intended to serve as a part of an expert witness' testimony.

The federal judge excluded evidence resulting from the SEQ. Briefly, any expert testimony that relied on the SEQ was ruled inadmissible under federal evidence rule 702 because the "survey instrument presents inherent reliability problems" and "the SEQ portion of the survey lacks validity" (EEOC v. Dial Corporation, p. 13). The September 17, 2002 order reads in part, "The validity problems with the use of a scale purporting to measure survey respondents' offensive sex-related experiences at work weighted against receiving the portion of the expert's testimony that relied upon that scale into evidence" and "This lack of comparability of SEQ scores seems problematic to me, because it seems to render the SEQ scores devoid of any objective meaning" (p. 8). The judge concluded that the "survey materials [SEQ] are too flawed to be useful in assisting the fact finder in this case" (p. 14).

The use of the SEQ in particular legal proceedings as well as its prevalence in academic research (for example, in addition to those cited above, see: Barak, Fisher, & Houston, 1992; Brooks & Perot, 1991; Houston & Hwang, 1996; O'Connell & Korabik, 2000; O'Hare & O'Donohue, 1998) suggests that the scale's merits and appropriateness warrant some examination. Welsh (1999) recommended that someone other than the developers of the SEQ conduct such an evaluation.

Our intent here is to review the development of the SEQ and evaluate it. We will discuss the characteristics of the SEQ (stability of the instrument; wording of the questions; test-retest and internal consistency reliability; and content, construct, and criterion validity). We then discuss the erroneous substantive conclusions that have been drawn by virtue of using the SEQ to measure sexual harassment, and finish with a discussion of appropriate uses for the versions of the SEQ.

It is important to note that our review of the SEQ should not be taken as a critique of the whole literature on sexual harassment. Other such critiques already exist (e.g., Arvey & Cavanaugh, 1995; Lengnick-Hall, 1995). Despite some criticisms about the body of literature, social scientists have learned a substantial amount about sexual harassment through research (see for example, Guttek & Done, 2001; O'Donohue, 1997; Stockdale, 1996). Nor is this critique a personal attack or dismissal of the research done by the developers of the SEQ who have, in general, contributed substantially to our knowledge of sexual harassment.

DEFINING SEXUAL HARASSMENT

Sexual harassment has both a legal connotation and a broader “lay” meaning. The legal definition rests on the Equal Employment Opportunity Commission’s (EEOC) statement on sexual harassment and has evolved with changing case law. The EEOC guidelines (1980, 1993, 1997) define sexual harassment as “unwelcome sexual advances, requests for sexual favors, and other verbal or physical conduct of a sexual nature” that are either a condition of employment (quid pro quo harassment) or that create an intimidating or hostile work environment. In quid pro quo harassment, getting or keeping a job or other conditions of employment are contingent on granting sexual favors or other sexual activity. In the case of a hostile work environment, these unwelcome behaviors must be severe and/or pervasive enough to meet a reasonable person’s and the complainant’s standard of a hostile work environment.

But sexual harassment also has a broader lay meaning including unwelcome sexual overtures or other social-sexual behavior at work that clearly would not satisfy the legal requirements. An instrument said to measure sexual harassment may focus on only the legal definition or some broader concept of social-sexual behavior that is not necessarily illegal. The SEQ is said to measure “psychological” sexual harassment (Fitzgerald, Swan, & Magley, 1997). More specifically, the developers of the SEQ say that the “. . . advantage of our model is that it articulates the relation between the legal and psychological constructs without in any sense equating them” (Fitzgerald, Gelfand, et al., 1995, p. 438). In addition, Fitzgerald, Gelfand, et al. propose that their operationalization of sexual harassment (the items of the SEQ) links that psychological definition to the legal concept. They conclude that their “. . . model provides a parsimonious yet comprehensive answer to the question, ‘What is sexual harassment?’” (Fitzgerald, Gelfand, et al., p. 430).

DEVELOPMENT OF THE SEQ IN HISTORICAL CONTEXT

The first peer-reviewed publication of the SEQ appeared in 1988 (Fitzgerald et al.). We rely heavily on this article and a second article by Fitzgerald, Gelfand, and Drasgow published in 1995 because these are frequently referenced as sources of information on the development of the SEQ. However, neither article contains an actual SEQ instrument.

The early development of the SEQ is described in Fitzgerald et al. (1988). According to the authors, “the original intent had been to design an inventory of 50 items” (p. 158) about students’ experiences in higher education, but it was too difficult to come up with that many questions without some redundancy. So the pilot version contained 30 self-report items based on Till’s five categories of sexual harassment (Till, 1980). It was administered to 468 undergraduate and graduate students of both sexes. The final item of the SEQ, “I have been sexually harassed,” was described as the “criterion item.” Response options for the items were *never*, *once*, and *more than once*. Based on the feedback of this pilot sample, two items were dropped, resulting in a 28-item measure that was used for two samples in the first study (903 graduate and undergraduate women and 491 undergraduate men from

one university; 843 graduate and undergraduate women and 362 undergraduate men from another university).

Cronbach's internal reliability coefficient for the 28-item measure for the first sample was .92. The second sample's reliability was reported as "comparable." Test-retest reliability on a small graduate student sample ($n = 46$) "... yielded a stability coefficient of .86 over a 2-week period" (Fitzgerald et al., 1988, pp. 158–159). Presumably the sample on which the test-retest reliability analysis was conducted included only females (consistent with the samples used in the first study), but no information is provided in the article.

The same 1988 article describes a second study designed to measure sexual harassment at work, along with a revision to the measure used in the first study. The instrument was changed to address sexual harassment from males only, as well as adding five items (based on a focus group's input) and modifying the wording of some of the original items. This sample consisted of 642 female faculty members, administrators, and staff at a university. The response options remained the same and Cronbach's alpha for this 33-item work-related version of the SEQ (called SEQ2) equaled .86 (see Appendix A for a summary of versions and samples).

The second article describing the development of the SEQ was published in 1995 (Fitzgerald, Gelfand, et al.). It reviewed some of the material covered in the earlier review (Fitzgerald et al., 1988), explained some shortcomings of that version (pp. 428–429; although it is not clear if the SEQ or SEQ2 is being discussed), and described the development of a revised version. That article (and a later book chapter [Fitzgerald, Swan, et al., 1997]) also described a "tripartite model" of sexual harassment. The authors contended that sexual harassment consists of three components: sexual coercion (analogous to quid pro quo harassment), gender harassment, and unwanted sexual attention (both of which are related to a hostile work environment). Presumably to be consistent with the tripartite model, new items, 54 in all (including revisions of previous items), were generated and a 5-point scale, replacing the three options used in the earlier version, was adopted. (See Appendix B for a synopsis of these types of changes.) The response options for the 5-point scale, described as a Likert scale, are not provided in that article. The ordering of items was also changed to reflect the intensity of the type of harassing behavior, with the mildest behavioral items appearing first. The 54-item version was piloted on 150 female graduate students. Based on analyses, 20 items were then selected for this revised version of the SEQ, referred to as SEQ-W. The authors mentioned in a footnote that a "... parallel form, designed for students, is designated the SEQ-E" (p. 433). We assume that the changes identified above were also applied to that student version. In this article, the authors also changed the wording of the "criterion" question to "Have you ever been sexually harassed?" even when they were describing the 1988 article, which clearly states that the item used at that time was "I have been sexually harassed."

This new version was given to employees from randomly selected work groups in a west coast public utility company along with other materials. Of the 1,156 respondents, 448 were women. It is not clear if the analyses reported in the article are based on responses from both sexes or only from women's responses. While this new version of the SEQ was described as a 20-item scale, the global "criterion" item ("Have you ever been sexually harassed?") and an item from the unwanted sexual attention

subscale were examined separately, so in reality, the measure was 18 items. Apparently one item from these 18 (in the sexual coercion subscale) was dropped from the measure because of “extremely restricted variance,” leaving 17 items (Fitzgerald, Gelfand, et al., 1995, see Appendix B, p. 435). The 17 items formed three factors consistent with the tripartite model; they were labeled gender harassment (five items, $\alpha = .82$), unwanted sexual attention (seven items, $\alpha = .85$), and sexual coercion (five items, $\alpha = .42$). With the exception of the sexual coercion subscale, these reliabilities are somewhat similar to those found for the pilot sample (150 respondents, 54 items; gender harassment $\alpha = .86$, unwanted sexual attention $\alpha = .75$, and sexual coercion $\alpha = .87$).

Subsequent reports did not make it clear that the version of the SEQ described in Fitzgerald et al. (1988; SEQ or SEQ2) may be quite different from the version described in Fitzgerald, Gelfand, et al. (1995; SEQ-E or SEQ-W). For example, in a 1997 article (Fitzgerald, Drasgow, Hulin, Gelfand, & Magley, 1997), when discussing their use of another version of the SEQ (labeled SEQ-R for revised), the authors cited the 1988 article when they provided information about the reliability of the SEQ, but they cite Gelfand, Fitzgerald, and Drasgow (1995) when they provide information about the SEQ consisting of three factors. Presumably, the items had changed quite a bit since the first version (see Appendices A and C).

Modifications to the SEQ have continued and versions have proliferated (see Appendices A, B, and C). The various versions of the SEQ all consist of a set of questions (varying in number, but typically 18 or more) that measure specific kinds of behavior that the respondent experienced over some time frame (which also varies). Items of the SEQ are summed (e.g., Glomb, Munson, Hulin, Bergman, & Drasgow, 1999; Magley et al., 1999), averaged (e.g., Fitzgerald, Drasgow, et al., 1997; Glomb et al., 1997; O’Connell & Korabik, 2000), or presented solely as a frequency of endorsement (e.g., Fitzgerald et al., 1988, 1999; Schneider et al., 1997) and all are weighted equally when aggregate scores are given. Anyone who has experienced at least one of the behaviors at least once over the specified time frame is generally classified as having been sexually harassed (Fitzgerald et al., 1988; Schneider et al., 1997).

In addition to the various versions of the main SEQ (an 18–24 item inventory to be filled out by women), there are other versions (see Appendix A). At the very least, there is now a short version (SEQ-s) containing eight items (Cortina et al., 1998), which is “. . . an abbreviated version of the original Sexual Experiences Questionnaire” (p. 424). There were no citations for information on the shortening of the scale and its reliability analyses in that article. This shorter version was said to measure the three broad categories as listed above. There is also a revised version designed for use by women and men in the Armed Forces (SEQ-DoD) (Donovan & Drasgow, 1999; Hay & Elig, 1999), which consists of 26 items (of which 23 are used in one set of analyses [Fitzgerald et al., 1999] and 26 used in another set of analyses [Donovan & Drasgow, 1999]). Stark, Chernyshenko, Lancaster, Drasgow, and Fitzgerald (2002) proposed a 16-item version called the SEQ-DoD-s. Although this version is shorter than its predecessor, the authors offer evidence that the reduction in items results in minimal loss of potential information. At least three other versions, only slightly different from each other (Waldo, Berdahl, & Fitzgerald,

1998), have also been developed to be filled out by men. More recently, a situation-specific version of the SEQ (Mazzeo, Bergman, Buchanan, Drasgow, & Fitzgerald, 2001) and a 16-item Latina version have also been developed (Cortina, 2001). For references to the different versions used in various articles, see Appendices A, B, and C.

Complicating the problem of identifying all of the various versions of the SEQ is the fact that the same samples are used in multiple publications. For example, results for a female version of the SEQ from a public utility company sample are reported in several articles (see Appendix A). SEQ results from an agribusiness sample are reported in Glomb et al. (1997), Waldo et al. (1998; male version), and Magley et al. (1999). SEQ data for study of faculty and staff at a Midwestern university collected in 1994 was reported in Waldo et al. (male version), Glomb et al. (1999; female version), Schneider et al. (1997; female version), and Magley et al. (female version) (again, see Appendix A). In each of these publications, there was no mention of the fact that SEQ results have already been reported in previous publications. An exception to this rule is a set of papers all published in the same issue of *Military Psychology* in 1999. In that case, it is clear that the papers all come from the same data set, 22,372 female and 5,924 male military personnel who completed a survey in 1995.

VARIATIONS OF THE SEQ

When psychologists speak of a particular measure of a construct, they are likely to mean a stable and fixed number of items designed according to recognized standards for writing survey items that together assess a well-defined concept and that meet acceptable standards of reliability and validity. In that regard, it is instructive to examine important properties of the SEQ: the extent to which it is a standardized measure, the wording of the questions and response options, and how reliable and valid it is.

Standardization

Ghiselli, Campbell, and Zedeck (1981, p. 428) define a test as “. . . a standardized measure of a sample of behavior.” A standardized measure is one with a fixed set of questions asked the same way each time, with the same response options. Standardization is what allows one to compare results of one study using that measure with the results of other studies using the same measure. Without standardization, generalizability of findings becomes strained if not impossible. The SEQ fails this principle of a psychological measurement; there is really no SEQ per se, but rather a family of related questionnaires, confusingly all labeled under the rubric of the SEQ. This is not obvious because published works involving the SEQ rarely delineate the details of the measures actually used. In fact, no version of the SEQ was published until recently (for published versions, see Cortina, 2001; Waldo et al., 1998). Appendices A, B, and C show some of the ways the SEQ has differed across publications.

Number and Wording of Questions and Response Options

The number of questions that comprise the SEQ varies (see Appendix A). Adding to the confusion about the number of questions contained in the SEQ is the fact that the global measure (“Have you ever been sexually harassed?”) is sometimes considered a part of the SEQ and sometimes not. The SEQ-DoD includes in the 26-item measure a question assessing whether any of the other items assesses sexual harassment (“Do you consider ANY of the behaviors (in No. 1–25) to have been sexual harassment?” [Donovan & Drasgow, 1999, p. 273]).

In addition, the number of items reported in the SEQ is not always the same as the number used in the analyses (e.g., Fitzgerald et al., 1999; for others, see Appendix A). Sometimes items are dropped because of lack of response variance (e.g., Fitzgerald et al. 1999; Gelfand et al., 1999). In their comparison of male and female versions of the SEQ-DoD, Donovan and Drasgow (1999) recommended dropping four items from their analyses of the 26-item SEQ-DoD that function differently across genders.

The various versions of the SEQ use lead-ins that differ in time frame and in source of SEQ experience (see Appendix B). Different time frames are reflected in the following lead-ins: Respondents were asked to “report only situations they had experienced in the target organization in the past 2 years” (Fitzgerald, Drasgow, et al., 1997) or “Have you ever been in a situation where . . .” (Barak, Pitterman, & Yitzhaki, 1995; Fitzgerald et al., 1988). In other cases, the SEQ assessed behaviors that occurred during the past year only (Fitzgerald et al., 1999; Stark et al., 2002). When others have used the various versions of the SEQ, they have not used a consistent time frame (e.g., Barak et al. [1995] used a 1-year time frame; O’Connell & Korabik [2000] who worked with a “slightly amended” SEQ, used a 2-year time frame because, they reported, that is the time frame over which the SEQ had been most frequently used).

While many of the female versions focus on behavior initiated by men only: (e.g., “During the past 2 years while you have been employed, has any of your male supervisors or coworkers where you worked . . .” [Glomb et al., 1997]), some male versions assess behavior initiated by either sex: “During the past 2 years at work, have any of your supervisors or co-workers . . .” (Waldo et al., 1998).

We know that there may be theoretical reasons for using different time frames and different initiators of SEQ behavior. In addition, in field research, participating organizations may dictate some of the conditions of the research. But deviations from standard procedures should be noted in the paper so that readers will know that results from one study are not directly comparable to another.

Respondents filling out the SEQ first read the lead-in statement at the top of the page and then read each of the items, each of which refers to a type of behavior, such as “Made you uncomfortable by staring at you (for example, at your breasts),” “Kept on asking you out even after you have said ‘no,’” or “Often told dirty stories or jokes.” It is worth noting that this last item is sometimes worded “often [or repeatedly] told” dirty stories or jokes (e.g., Fitzgerald et al., 1999; Waldo et al., 1998) and sometimes worded “. . . told” dirty stories or jokes (Gelfand et al., 1995). Obviously, these are different questions.

Determining exactly how much the wording of the questions changes from one administration of the SEQ to another is difficult because not even examples of questions from the SEQ are included in several articles and the article does not cite a publication where they are available (e.g., Fitzgerald, Drasgow, et al., 1997; Glomb et al., 1997; Magley et al., 1999; Wasti, Bergman, Glomb, & Drasgow, 2000).

It is clear that some of the items do not describe sexual behavior. A male version includes, "Given a job to a less qualified woman rather than you;" a female version includes, "Said things to put women down, for example, that women don't make good supervisors"; and the DoD version includes, "Treated you 'differently' because of your sex." (Donovan & Drasgow, 1999, p. 272). Barak et al. (1995), gave the following examples from the version of the SEQ they used: "Have you ever been in a situation where . . . a man made crude and offensive sexual remarks toward you?" and "Have you ever been in a situation . . . where a man treated you 'differently' because you're a woman?"

In early versions of the SEQ, respondents had a choice of three options: *never*, *once*, and *more than once* (see Appendix B). Fitzgerald et al. (1995, p. 427) noted that the latter two categories (*once*, *more than once*) ". . . were frequently collapsed for purposes of scoring." The response options listed in Fitzgerald et al. (1997, p. 580) were described as follows: "Participants respond on a 5-point scale, ranging from 1 (*never*) to 5 (*most of the time*)."⁴ The male versions of the SEQ use the following categories: *never*, *once or twice*, *sometimes*, *often*, and *most of the time* (Waldo et al., 1998), as does the Latina version (Cortina, 2001). In at least one survey, the fifth category was *many times* (Magley et al., 1999) and on another, the fifth category was *very often* (Mazzeo et al., 2001) instead of *most of the time*. Thus, in some versions, the response options are a mixture of both specific and general quantities (e.g., "*once or twice*" versus "*most of the time*").

If the time period is short, then "*once or twice*" could be the same as "*sometimes*." It would be good survey research practice to construct response options that make sense given the length of time to be considered in answering each question. If all respondents are considering the same length of time (say 24 months), then "0, 1–2 times, 3–5 times, etc." might be more appropriate. If respondents are considering different lengths of time, i.e., "during the time you worked at . . .," it would be more appropriate to use response options to reflect frequency per given unit of time, e.g., more than once a week, about once a week, about once a month, about once every 6 months, once a year, no more than once every 5 years, never. Another possibility is to use all general responses (i.e., *never*, *occasionally*, *frequently*, *very frequently*). The items could also be reworded to accommodate a true Likert-type scale (i.e., *strongly disagree* to *strongly agree*).

As the SEQ is currently written, the response options do not always make sense given the item stem. Most confusing are the items that contain "often" in the item stem and also as one of the response options, for example, "often told dirty stories or jokes," "*never*," "*once or twice*," "*sometimes*," "*often*," "*most of the time*." What

⁴Although the response options were described as ". . . a 5-point Likert scale" (Fitzgerald, Gelfand, et al., 1995), Likert scales involve the expression of agreement or disagreement, i.e., "strongly agree" to "strongly disagree" (see, for example, Kidder & Judd, 1986, pp. 210–214).

if a male supervisor or coworker occasionally told a dirty story or joke; would the respondent answer “never” to the question since the behavior did not occur “often”?

Reliability and Validity of the SEQ

Appendix C provides specific information about the reliability and validity of the SEQ as reported in publications by its developers. In their claims that the SEQ is “psychometrically sound,” many researchers (see Cortina et al., 1998, p. 424; Fitzgerald et al., 1999, p. 245; Fitzgerald, Gelfand, et al., 1995, p. 428, Magley et al., 1999, p. 394; Schneider et al., 1997, p. 404) consistently cite an article (Arvey & Cavanaugh, 1995) that bases its information solely on the Fitzgerald et al. (1988) article. In citing this article, which does not contain the SEQ, Arvey and Cavanaugh (p. 49) noted that Fitzgerald et al. “. . . evaluated reliability (e.g., internal consistency, test-retest stability, and split-half reliability coefficients for the five scales) and validity (content and criterion) for two separate samples.” In other words, Arvey and Cavanaugh noted that Fitzgerald and her colleagues assessed the reliability and validity of early versions of the SEQ, but Arvey and Cavanaugh did not evaluate whether or not the SEQ was “psychometrically sound.” They may not have even seen a copy of the SEQ (if their only exposure to it was the single reference they cited in their article, i.e., Fitzgerald et al., 1988).

Contrary to some claims, the reliability and validity of the SEQ are not especially strong. We first review claims about the SEQ’s reliability and follow with claims about its validity. Of course, a measure can be reliable (get the same results each time) but not valid (actually measure what it is supposed to measure). Psychometric soundness is the result of a measure having both of these qualities.

Reliability

Several types of reliability can be assessed: internal consistency, test-retest, and parallel forms. The developers of the SEQ claim that they have demonstrated its internal consistency reliability (items all contribute to the measure of the construct) as well as its test-retest reliability (i.e., one would observe the same responses if the scale were administered again under the same circumstances) (Fitzgerald et al., 1988). They have not made any claims of reliability based on parallel forms (two more or less equivalent forms of a measure should be highly correlated).

Various versions of the SEQ have been analyzed for their internal consistency reliability (see Appendix C). The complete versions (e.g., 18, 26 items) generally showed acceptable reliability, varying from the high .70s to the low .90s. The reliabilities reported for the subscales vary but are generally lower. Fitzgerald, Gelfand, et al. (1995, p. 427) wrote that for the original version of the SEQ described in the 1988 article “Corrected split-half reliability coefficients for the five subscales ranged from .62 to .86 and averaged .75. . . .” For the revised version of the SEQ that was based on their tripartite model and was described in that 1995 article, coefficient alpha measures of internal consistency were .82 for gender harassment, .85 for unwanted sexual attention, but only .42 for the 5-item subscale of sexual coercion. The low reliability for the sexual coercion subscale is partly due to the fact that very few

people experience sexual coercion so there is little variance in the responses, thus little covariance between items.⁵

Fitzgerald et al. (1988) also claimed that the version of the SEQ reported in that article had acceptable test-retest reliability. In their 1995 discussion of the SEQ they cited the earlier report (1988) regarding the test-retest reliability of the earlier version of the SEQ. A careful reading of the articles published subsequent to the 1988 article (e.g., Fitzgerald et al., 1999; Fitzgerald, Drasgow, et al., 1997; Fitzgerald, Gelfand, et al., 1995; Gelfand et al., 1995) reveals that virtually all of the evidence for the SEQ's test-retest reliability comes from the sample of 46 graduate students from the first study (see Appendix C). We know of no other short-term test-retest reliability assessments of any version of the SEQ.

Validity

A psychological measure is valid if it measures what it purports to measure. In the case of the SEQ, Fitzgerald, Swan, et al. (1997) describe it as a measure of sexual harassment although they claim it measures *psychological* sexual harassment but not *illegal* sexual harassment (a distinction that is discussed later).

Nunnally and Bernstein (1994) delineate three central kinds of validity: content validity, construct validity, and predictive validity. Fitzgerald and colleagues (Fitzgerald et al., 1988; Fitzgerald, Gelfand, et al., 1995) claim that the SEQ meets standards for three kinds of validity: content validity, construct validity, and criterion validity (synonymous with predictive validity).

Content Validity

High content validity is obtained if the items used to measure a construct are in total representative of that construct (Nunnally & Bernstein, 1994). For example, a measure of sexual harassment should tap all the different facets of sexual harassment and not tap any additional constructs. Fitzgerald and colleagues (Fitzgerald et al., 1988; Fitzgerald, Swan, et al., 1997) claim that the SEQ has content validity because they wrote questions to assess five kinds of behavior identified by Till (1980). This is a weak justification for content validity.

The Till report is a product of the National Advisory Council of Women's Educational Programs, established by Congress in the Women's Educational Equity Act of 1974. The Council wrote a letter, called a "Call for Information" and mailed over 8,000 copies of the "Call" in the fall of 1979, "... of which 6,000 were addressed to administrators and student governments on each of the nation's campuses. This direct contact effort was supplemented by mailings to campus women's centers, state and national student organizations, advocate groups, and professional education associations" (Till, 1980, part II, p. 2). The "Call" itself is reproduced in the report; it is a one page letter which includes the following: "As part of the study, the Council is requesting information from former and present victims about their experiences, and

⁵Little variance in sexual coercion items suggests these items are not particularly suited to statistics that depend on responses that are normally distributed.

from any others who may have knowledge of such harassment” (Till, Appendix A, p. 31). The 8,000 plus mailings yielded only 259 responses: 116 from victims, 43 from secondary sources, and 7 from researchers. The other respondents are not identified.

Based on their reading of the 259 comments, the Advisory Council developed five categories of sexual harassment. They are: generalized sexist remarks or behavior; inappropriate and offensive but essentially sanction free sexual advances; solicitation of sexual activity or other sex-linked behavior by promise of rewards; coercion of sexual activity by threats of punishment; and sexual assaults (Till, 1980, pp. 7–8). In a footnote, Till (1980, p. 7) noted that some people who read a pre-publication report thought that the categories were too broad and did not cover sexual harassment, per se, but included sex bias and/or stereotyping. Furthermore, when Fitzgerald and Hesson-McInnis (1989) examined whether students (24 graduate and 4 undergraduate students) perceived each of the five categories identified in the Till (1980) report as constituting sexual harassment, gender harassment items were generally not viewed as sexual harassment. Thus, the Till categories are suspect as an indicator of types of sexual harassment. Surprisingly, Till’s report (Till, 1980) is described by Fitzgerald et al. (1988, p. 157) as follows: “The first step in instrument development was the generation of an initial item pool, based on the five levels of sexual harassment identified by Till (1980) through content analysis of his national survey of college women.” This is a distortion of the Till report since he did not conduct a national survey of college women. A further distortion of the Till report is in Arvey and Cavanaugh’s critique of measures of sexual harassment (Arvey & Cavanaugh, 1995, p. 49): “. . . Till’s five categories (Till, 1980) of behavior . . . were based upon a content analysis of responses from a national sample of college women who were asked an open-ended question regarding experiences of sexual harassment.”

In one sense, what the developers of the SEQ did is similar to what other researchers interested in sexual harassment have done. They tried to find out what kinds of experiences people had that might be considered sexual harassment and then wrote questions to try to capture those experiences, but no one else has claimed to have established content validity on the basis of their procedures.⁶

⁶The questions used by the U.S. Merit Systems Protection Board to measure sexually harassing behaviors were generated in much the same way, but without relying on someone else’s categorization—and without claiming to have established “content validity.” Three experienced and published psychologists (Sandra Tangri, Martha Burt, and Leonor Johnson) conducted the research. In addition, prior to the development of the first USMSPB survey, the research team put together an advisory panel of scholars including many who had conducted research on sexual harassment, including the first author. That group had substantial input to the development of the questions. Justification for asking particular questions on social-sexual behavior studied by Gutek is included in *Sex and the Workplace* (Gutek, 1985, pp. 41–42). “An operating assumption at the beginning was that the experiences of sexuality at work, including sexual harassment, are subjective. It would be difficult, if not impossible, to make a comprehensive list of events that are necessarily experienced as sexual harassment. Furthermore, at the time of the survey, final legal guidelines about harassment had not been established. Thus, there were no legal or psychological limits to the definitions of sexual harassment. Furthermore, I was interested not in sexual harassment alone but in the range of social-sexual interactions between the sexes at work. The particular categories of behavior studied were developed from two sources. One was our interest in sampling a spectrum of social-sexual behaviors that might be considered sexual harassment. (See Gutek, Nakamura, Gahart, Handschumacher, and Russell, 1980, for details on the rationale for selecting categories.) The second source was the responses obtained in two pilot studies, which led to the addition of two more categories.” Other measures have also been developed to study sexual harassment. For example, Lott, Reilly, and Howard (1982; Reilly, Lott, & Gallogly, 1986) developed a 10-item measure, the Tolerance for Sexual

Construct Validity

According to Nunnally and Bernstein (1994) constructs are only approximately measured by the aggregation of observations, as measures of a particular construct correlate with the construct but do not wholly define it. Construct validity cannot be established by any single analysis. Nunnally and Bernstein (p. 86) contend that there “are three major aspects of construct validation: (1) specifying the domain of observables related to the construct, (2) determining the extent to which these observables measure the same thing or different things, and (3) performing subsequent studies to determine whether the observables behave as they are expected.” Research by the developers of the SEQ provides some evidence that the measure has construct validity. The developers of the SEQ have provided evidence that higher scores on the SEQ are positively associated with outcomes expected if one had experienced social-sexual behaviors at work or had been sexually harassed (e.g., Fitzgerald et al., 1999; Fitzgerald, Drasgow, et al., 1997). Many of these studies consist of cross-sectional data (as indeed is true of almost all of the published research on sexual harassment), and such data make inferences about the direction of causality problematic. So, for example, it is plausible that respondents who have SEQ experiences perceive the work environment as tolerant of sexual harassment. (Otherwise, they would not have had that experience.) But, these findings have most often been interpreted as support for the relationship being in the opposite direction, namely that tolerance for sexual harassment predicts SEQ score (Fitzgerald, Drasgow, et al., 1997).

Perhaps the strongest evidence for the SEQ’s construct validity is shown in Glomb et al. (1997). In this study, each respondent was assigned a score for “ambient sexual harassment,” constructed from the mean score of the others in her work group. The researchers found that indirect exposure within the context of the work group (as indicated by their ambient sexual harassment score) was negatively associated with job satisfaction and psychological well-being.

Here we consider three additional kinds of evidence in evaluating the construct validity of the SEQ: (1) we compare the developers’ definition of sexual harassment with their measurement of sexual harassment; (2) we examine factor analysis results; and (3) we address the argument that between-version item variance in the SEQ is acceptable and does not undermine its value as a psychological measure.

Definitions of Sexual Harassment. In order to specify a domain of observables, one needs a definition of the construct. Fitzgerald, Swan, et al. (1997, p. 15) define psychological sexual harassment as “. . . unwanted sex-related behavior at work that is appraised by the recipient as offensive, exceeding her resources, or threatening her well-being.” In that article, they describe sexual harassment as a form of psychological stress. However, SEQ items are not consistent with this definition of psychological sexual harassment. Wording of the items does not allow one to conclude that the respondent who answers other than “never” to any of the SEQ items is overwhelmed in an attempt to cope with the behavior or that her well-being is threatened.

The description of the SEQ is also inconsistent with the definition of psychological sexual harassment. The SEQ is described as measuring “. . . offensive, sex-related

Harassment Inventory (TSHI), yielding three factors, labeled “flirtations are natural,” “provocative behavior,” and “feminist beliefs.”

behavior that is unwanted, unwelcome, and unreciprocated” (Fitzgerald, Swan, et al., 1995, 1997). There is, however, no evidence from the material provided about the SEQ that all of the items are worded to assess only behavior that is unwanted, unwelcome, or unreciprocated. None of the items assess reciprocation and only some include words that explicitly assess or imply unwelcomeness (e.g., made offensive remarks, unwanted sexual attention). Furthermore, while an examination of the short phrases provided for the items in Fitzgerald, Gelfand, et al. (1995, Appendix B) suggests that many are probably unwanted, some such as the item abbreviated as “. . . told suggestive stories” does not necessarily imply behavior that is unwanted, unwelcome, or unreciprocated, although it certainly could be. (Reciprocation does not necessarily imply that the behavior was welcome. Reports of women who respond to obscene language and sexual jokes by reciprocating say they sometimes do so in an attempt to be accepted, as Fitzgerald among others have discussed elsewhere [see Fitzgerald, Swan, & Fischer, 1995].)

Factor Analysis Results. Internal consistency across items is a necessary but not sufficient condition to establish construct validity. Internal consistency across items may be examined, in part, through factor analysis, a technique that shows how many different types or components there are to the construct being measured. If repeated factor analyses of the same measure taken from different samples yield the same factor structure, with items consistently loading on the same factors, such results would contribute to the internal consistency aspect of construct validity.

Based on factor analytic results, the construct validity of the SEQ is questionable. Fitzgerald et al. (1988) wrote questions to reflect the five categories of sexual harassment identified by Till in his report. If there are five kinds of sexual harassment and questions were written to measure each of these kinds, then a factor analysis should uncover five factors corresponding to the five different types of harassment. Further, each item should load on only one of the factors—the one for which it was intended. This is a psychometric trait known as simple structure (Thurstone, 1947). The SEQ has not fulfilled this ideal.

The number of factors that the SEQ is measuring varies across studies (e.g., five in Fitzgerald et al., 1988; three in Fitzgerald, Gelfand, et al., 1995; four in Fitzgerald et al., 1999). Fitzgerald et al. (1988) asserted that the SEQ consisted of five components described by Till (1980). As discussed above, the fact that the SEQ was written to reflect the five categories of sexual harassment identified by Till was considered support for the content validity of the SEQ. Yet, both the initial article and the other important article describing the development of the SEQ (Fitzgerald et al., 1988 p. 169; Fitzgerald, Gelfand, et al., p. 429) noted that a previously conducted study (Fitzgerald & Shullman, 1985) “. . . cast doubt on the five-dimensional structure, suggesting that only three dimensions are necessary to explain variation in SEQ data.”

Gelfand et al. (1995) reported on the factor structure of the SEQ using several of the samples also used in Fitzgerald et al. (1988) as well as a sample from Brazil. On the same samples ($N = 1,746$ and 307) plus the Brazilian sample of university students ($N = 389$), three factors instead of five were found for the combined sample as well as for the three individual samples. The military version of the SEQ (SEQ-DoD) is reported to contain four factors: sexist hostility, sexual hostility, unwanted sexual attention, and sexual coercion (Fitzgerald et al., 1999; see Appendix C). The

SEQ-DoD-s (Stark et al., 2002) uses this four-factor structure as a framework. A Latina version of the SEQ is reported to reflect only three factors: sexist hostility, sexual hostility, and unwanted sexual attention (Cortina, 2001).

Factor analyses of the SEQ reveal yet another problem. The correlations among the factors are quite high. Using a three-factor model, correlations among the factors averaged .75 according to Fitzgerald, Gelfand, et al. (1995) and .70 according to Magley et al. (1999). Such high correlations call into question the assertion that the SEQ is measuring three separate constructs. Perhaps it is because of high inter-correlations among the SEQ factors that many studies use an aggregated SEQ score instead of component scores (e.g., Schneider et al., 1997).

The SEQ as a Sampling of Items From a Domain. Keep in mind there is no SEQ per se, but rather a family of different scales comprised of overlapping items and related response options. In this instance the SEQ's between-version item variations (different versions being composed of different items with different response options, etc.) have been defended by drawing a parallel to other metrics, most notably the Scholastic Aptitude Test (SAT). The SAT measures the latent construct of scholastic aptitude by drawing items from a well-defined and empirically validated pool of items. It is possible that two versions of the SAT will be comprised of completely different items yet still have psychometric validity as well as intercomparability of aggregate results. Thus, so the argument goes, the variation between the SEQ versions is not problematic because these different versions are all developed to measure the same hypothetical construct.

This argument and analogy are invalid for the following reasons. First, the SAT (and other high-stakes tests like it) has been extensively empirically validated. Each of the potential items that may be included in a SAT has been thoroughly pretested and rigorously analyzed, each with a well-defined item response curve. The SEQ has not been exposed to the same rigor. It was only recently (Stark et al., 2002) that item response theory was brought to bear on the SEQ, and then with the intention of shortening the questionnaire, not the validation of equivalence between different previously used versions. To our knowledge, there has not been a published effort to empirically validate the contention that different versions of the SEQ are equivalent. Secondly, the purveyors of the SAT have a clear motivation to develop multiple different versions of the SAT in order to maintain test security. Conversely, the authors and users of the SEQ do not have a clear reason or justification to repeatedly modify the questionnaire. As Stark et al. rightly point out, the lack of a "standard method" of measurement has undermined sexual harassment research to date.

Criterion Validity

Nunnally and Bernstein (1994) consider a measure to have predictive validity if the measure is functionally related to a particular criterion that is external to the measure. A measure with predictive validity can be related to a criterion that already happened or existed in the past (postdiction), a criterion that currently exists (concurrent), or a criterion that will exist in the future (prediction).

A major problem in the establishment of predictive validity is selecting an appropriate criterion. The measure of the criterion may itself not be without bias or

error. Typically, the criterion is measured independently of the predictive measures. For example, job satisfaction assessed by a questionnaire may be used to predict absenteeism, measured by days absent from work. If a common method of measurement is used to assess both the construct of interest and the criterion, a method bias could be introduced that inflates the appearance of predictive validity.

Fitzgerald, Gelfand, et al. (1995) reported that they found evidence of the SEQ's criterion validity. Some of the versions of the SEQ include the question: "Have you ever been sexually harassed?" (see Fitzgerald, Gelfand, et al., 1995) or "I have been sexually harassed." (Fitzgerald et al., 1988). That question, a self-report item like the items in the SEQ, was used as the criterion (e.g., Cortina et al., 1998; Fitzgerald, Drasgow, et al., 1997; Fitzgerald, Gelfand, et al., 1995; Glomb et al., 1997). In other words, it has been argued that the SEQ shows criterion validity if the SEQ score is related to a single-item global assessment of sexual harassment obtained by self-report from the same participants who completed the SEQ.

In general, very weak relationships have been found between components of the SEQ and the question about whether or not the respondent had ever been sexually harassed. Correlations varied from .15 for gender harassment to .37 for sexual threats (Fitzgerald, Gelfand, et al., 1995, p. 428). Thus, only 2% of the variance in the "criterion" is predicted by the SEQ's component called gender harassment, which is also the most commonly occurring of the three components of the SEQ. If this criterion measure was a good criterion (i.e., an accurate measure of sexual harassment) one might simply ask that single question and forget about the 18, 23, or 33 items of the SEQ. But it is not a good criterion for several reasons. The criterion question uses the wording "have you ever" but the SEQ lead-in usually specifies a time period. Thus, some respondents report that they have "ever" been sexually harassed but do not report any SEQ items where the SEQ experiences are limited to, say, the past 24 months (see, for example, Magley et al., 1999, p. 396, footnote 5). The differences in time periods that respondents are asked to consider may partially explain the low correlations between the SEQ or its components and the global "criterion" measure of sexual harassment. However, it does not explain why the criterion is differentially related to the various components (i.e., $r = .37$ [for the criterion and sexual threats] vs. $r = .15$ [for the criterion and gender harassment]). If the single global item were to be considered an adequate criterion, then the subscale called gender harassment (or sexual hostility and sexist hostility in the SEQ-DoD survey [Fitzgerald et al., 1999]) would not be considered to measure sexual harassment because of the low correlation between that subscale and the criterion.

As Nunnally and Bernstein (1994, p. 99) note, "to use predictive validity . . . is to assume that the criterion is appropriate." Paradoxically, even the SEQ's developers argued that the sexual harassment criterion question itself is not a good measure of sexual harassment. In research, it is typically found that fewer women answer in the affirmative when asked if they have been sexually harassed than are considered sexually harassed based on their responses to the SEQ. Some researchers consider this to indicate that sexual harassment is frequently not acknowledged (Fitzgerald, Swan, et al., 1995, 1997; Magley et al., 1999). Recently, they have used the single global item they call the "criterion" as an indicator that the victim has labeled her experience sexual harassment (Magley et al., 1999).

In conclusion, the available evidence suggests that the single global assessment is not an appropriate criterion because it shares common method variance with the SEQ, the correlations between the components of the SEQ and the criterion are low, and the percentage of women who are considered sexually harassed by the single-item criterion question is always much lower than the percent of women who are considered sexually harassed based on SEQ scores.⁷

CONSEQUENCES OF USING THE SEQ

The various features of the SEQ described above have implications for researchers interested in the topic of sexual harassment and for forensic experts interested in documenting the prevalence of sexual harassment in an organization charged with sexual harassment. First, the unusual features of the SEQ described above severely limit its utility. Second, using the SEQ as a measure of sexual harassment has the effect of distorting important substantive findings about sexual harassment.

A test is “. . . a standardized measure of a sample of behavior” (Ghiselli et al., 1981, p. 428). The SEQ is not standardized. It is continuously evolving. As a result, the SEQ has few, if any, of the advantages of a standardized measure. First, there is no base rate score against which changes can be measured. Based on the set of studies using the SEQ, we cannot draw any conclusions about whether the amount of social-sexual behavior is increasing or decreasing (although, of course, individual researchers who use the same version of the SEQ could establish their own base rate and examine changes over time). This problem has also been noted by Stark et al. (2002). They note a “major shortcoming” of sexual harassment research is “the absence of a standard method of assessing sexual harassment” (p. 49). They address this shortcoming by using item response theory to pare down the SEQ-DoD into the SEQ-DoD-s, a 16-item, 4-factor measure. Whether this revision of the SEQ scale will become the gold standard for future military and civilian research remains to be seen.

Considering current SEQ findings, there are no norms that might inform researchers or others (managers, attorneys, judges) whether any particular firm, work group, occupational setting, etc. is above or below average on social-sexual behavior. SEQ results from one study cannot be compared with results from any other study unless one is sure that the same version of the SEQ was used in both. It does not allow one to conclude that a firm being sued has more or less sexual harassment (or social-sexual behavior) than average, or that an organization has little or no illegal or psychological sexual harassment.⁷ In short, it does not allow one to draw conclusions

⁷A measure of sexual harassment based on the EEOC’s definition reveals only a small gap between the percentage of people who are labeled harassed via that instrument and who answer the single-item global question in the affirmative (Guttek et al., 2002). Analyses show that there is very little “underreporting” of sexual harassment based on this new measure (Experiences of Sexual Harassment), suggesting that perhaps the SEQ overreports sexual harassment rather than that the global question substantially underreports it.

about whether the amount of sexual harassment in a particular setting is high or low. Providing SEQ scores is not meaningful in the absence of established base rates or norms or scoring procedures. Indeed the authors frequently do not report mean scores, or if they do, they are not discussed.⁸

Because of these deficiencies, the SEQ is not an appropriate tool for providing information in legal proceedings (see also *EEOC v. Dial Corporation*, 2002).⁹

Erroneous Substantive Conclusions

Because of its limitations, widespread use of the myriad versions of the SEQ have led to incorrect substantive conclusions about important aspects of sexual harassment. Here we discuss four of those areas. We contend that the SEQ overreports the prevalence of sexual harassment, it overestimates the gap between the experiences of harassment and labeling the behavior harassment, it underestimates the rate of reporting of sexual harassment, and it distorts the type and distribution of targets' responses to harassment.

Overreporting the Prevalence of Sexual Harassment

Using the SEQ measure results in overreporting of occurrences of sexual harassment if sexual harassment is considered an illegal behavior, if the developers' own definition of harassment is used, or if the public's perceptions are considered.

An examination of what information there is in the various publications using a version of the SEQ confirms the contention that the SEQ assesses behavior that

⁸In principle, one might conclude no illegal or psychological sexual harassment occurred if everyone filling out the SEQ instrument scored "never" on every item, but we know of no case where that has happened. Depending on the time frame, means on the sexual coercion scale have been very low suggesting that little or no illegal or psychological sexual coercion occurred (see Barak et al., 1995; O'Connell & Korabik, 2000). The determination of whether or not someone has been psychologically sexually harassed depends, however, on reporting any of the SEQ experiences at least once.

⁹Glomb et al. (1999) report an SEQ score of 20.69, $SD = 3.79$ for T1 (sum of 18 items; sample is women at a large Midwestern university in 1994), and an SEQ score of 19.99 for T2 in 1996 (sum of 18 items; same sample). Magley et al. (1999) report an SEQ score of 22.93 (sum of 19 items) for the regulated utility sample and 23.80 (sum of 22 items) for the agribusiness survey. In contrast to the SEQ developers and their colleagues and students who tend to report sums across the varying number of SEQ items, making it exceptionally difficult to compare results from one study to another, other researchers frequently provide more easily interpretable mean scores. For example, O'Connell and Korabik (2000) used the SEQ to inquire about sexual harassment from workers higher in the hierarchy, at equal levels, and at lower levels (in the previous 24 months). For gender harassment, on a scale of 1–5, they reported means of 1.52, 1.47, and 1.28 and for unwanted sexual attention, they reported means of 1.12 (from employees at a higher level), 1.11 (from employees at an equal level), and 1.04 (from employees at a lower organizational level). Barak et al. (1995) reported SEQ scores of 1.31 for a sample of urban women in Israel and 1.31 for women living in Kibbutzim. (In this study, women were asked about experiences in the last year and the SEQ was scored on a 3-point scale, not a 5-point scale.) The Barak et al. study provides another example of why the scale scores on the SEQ are so difficult to interpret; early versions of the SEQ used 3-point scales instead of 5-point scales. (However, even the 5-point scales are confusing because sometimes those scores [as well as the 3-point scales] are collapsed to reflect the presence or absence of the SEQ experience [e.g., Magley et al., 1999]). A final reason for confusion is in the way the response options are coded. Most of the articles describe it as a 5-point scale ranging from 1 to 5 (e.g., Fitzgerald, Drasgow, et al., 1997). However, more recently, the scale has been described as ranging from 0 to 4 (Magley et al., 1999).

would not meet the legal definition of sexual harassment and the developers of the SEQ have written repeatedly that the SEQ does not measure illegal sexual harassment (for example, Fitzgerald, Drasgow, et al., 1997, p. 580). (However, in some publications, only the legal definition of harassment is given, not a psychological definition [e.g., Schneider et al., 1997]).

In addition, the wording of many of the items used in at least some versions of the SEQ do not allow one to conclude that the SEQ behavior reported is unwanted, unreciprocated, or unwelcome. Nor are the SEQ items worded in such a way to suggest that the respondent who answers other than “never” to any of the SEQ items is overwhelmed in an attempt to cope with the behavior or that her well-being is threatened. Finally, the SEQ clearly does not measure respondents’ definition either, because more people report SEQ experiences than respond affirmatively to the global measure of “Have you ever been sexually harassed?”

Overestimating the Gap Between Reported Harassment and its Identification

The gap between the percent experiencing at least one SEQ behavior at least once or twice and the percent who report they have been sexually harassed is very large. For example, Barak et al. (1995) found according to SEQ results that 38.5% of a sample of urban women and kibbutzim women in Israel experienced verbal sexual harassment within the last year, but only 8.4% of the urban women and 5% of the kibbutzim women reported that they had ever been sexually harassed in their workplace. There is no gap, of course, if the global item rather than the SEQ is considered the “real” measure of sexual harassment. Instead, the gap is typically interpreted as indicating that many women who are sexually harassed do not label the behavior harassment. Thus, the SEQ score is interpreted as measuring sexual harassment and the global item is considered a measure of labeling. More likely the SEQ overreports the amount of sexual harassment at least to some extent. A new measure focusing on the EEOC’s legal definition of sexual harassment (EEOC, 1980, 1997) suggests that the gap between experiencing sexual harassment and labeling it as such still exists, but is quite small (Guttek, Done, Swindler, & Stockdale, 2002).¹⁰

Underestimating the Percent Reporting Sexual Harassment

Because anyone who has experienced one or more of the SEQ behaviors during the time frame assessed is considered to have been sexually harassed, underreporting is assessed against a very generous empirical definition of sexual harassment. If the score on the SEQ is used as the base upon which underreporting is calculated, then it is the case that most victims of sexual harassment do not report it. This effect may be exacerbated by the fact that, as noted above, items exhibiting little variance are sometimes dropped from the analysis of the SEQ. These dropped items are likely to be the most severe behaviors because they occur relatively less frequently. Although most people who experience severe harassing behaviors also report experiencing

¹⁰In this same study, 94% of women and 99% of men were considered to have been sexually harassed based on their answers to the SEQ, i.e., 99% of men and 94% of women ever experienced at least one SEQ behavior at least “once or twice” at work.

less severe behaviors, the practice of dropping items may eliminate the behaviors about which people are most likely to complain (see Gruber & Bjorn, 1982). If a more stringent indicator of sexual harassment is used and all items are included in the analyses, then the incidence of failure to report harassment will be much smaller. This is suggested by the low correlation between the gender harassment and the single-item global item used to assess sexual harassment (Fitzgerald, Gelfand, et al., 1995).

Other studies of social-sexual behavior show that a relatively high proportion of people who experience at least one such behavior did not think it serious enough to report. For example Gutek (1985, p. 72) noted that among women who reported experiencing sexual touching or were expected to date or engage in sexual behavior as part of the job, a significant majority (82%) checked *no real need to report it*. Sixty-six percent did not want to hurt the initiator, 60% thought nothing would be done, 60% thought they might be blamed, 32% said reporting took too much time and effort, and 31% said they were too embarrassed to report it. Ten years later, in a 1995 U.S. Merit Systems Protection Board publication on sexually harassing behaviors in the Federal workforce, of those who did not report their harassing behavior, 50% checked that it *was not serious enough to report*. These findings together suggest that the use of the SEQ to measure sexual harassment underestimates the percent of harassed women who report the harassment. Note that this criticism of the SEQ does not imply that many harassed women do not report it, only that treating the SEQ as a measure of sexual harassment overstates the amount of underreporting. As evidenced by concerns about being blamed for complaining, it is highly likely that many women who are harassed do not complain—and for reasonable reasons.

Distorting the Range and Distribution of Coping Responses

Although reporting sexual harassment to someone in authority is one possible response to sexual harassment, there are others such as doing nothing or ignoring the behavior or the perpetrator. Several attempts to represent classes of responses to social-sexual behavior have been proposed (e.g., Gruber, 1989; Gutek & Koss, 1993; Knapp, Faley, Ekeberg, & DuBois, 1997). One of these classifications is based on a survey using one of the early versions of the SEQ. Fitzgerald (1990, discussed in Fitzgerald, Swan, et al., 1995) proposed a number of internally focused responses (i.e., denial, endurance, detachment, illusory control, reattribution) and externally focused responses (i.e., avoidance or assertive responses), based on coding responses of people who apparently experienced one or more SEQ behaviors.

When responses to the SEQ are used to indicate sexual harassment, it is likely that the proportion of respondents who engage in certain responses (like avoidance or detachment) is likely to be overestimated whereas the proportion of those who file formal complaints or respond assertively to the harassment is likely to be underestimated. For example, the percent who “do nothing” in response to a single sexual joke is likely to be much higher than the percent who “do nothing” in response to sexual touching or groping. Use of the SEQ to measure sexual harassment may have

the effect of underestimating just how assertive victims of sexual harassment might be.¹¹

CONCLUSIONS

By all appearances, the SEQ is still a work in progress and there is no single measure that can be called “the SEQ”. Instead many versions exist, each of which represents an ongoing modification of some unpublished “original” version.

Although we are critical of the SEQ, we do not mean to imply that it is easy to measure sexual harassment, that the research on sexual harassment is generally weak or poor research, that the SEQ is useless, or that there are other superior measures that researchers should use. We conclude, however, that the SEQ is a flawed instrument and that its positive features have been greatly exaggerated. It does not seem to measure anyone’s definition of sexual harassment, including that of its own developers. Even if the SEQ measures only psychological harassment that raises another important question: what is psychological sexual harassment? As reviewed above, there is one definition of sexual harassment, another definition of what the SEQ measures, and then there is the SEQ itself which is not consistent with either the definition of sexual harassment or the explanation of what the SEQ measures. The broader term, social-sexual behavior (see, for example, Gutek, 1985; Wiener & Hurt, 2000), does not fit either because some SEQ items have no sexual content. Perhaps a better description of the SEQ is that it assesses nonwork related behavior that would probably be considered inappropriate if it is unwelcome.

We believe it is a mistake to consider the SEQ (in its various incarnations) a measure of sexual harassment. If one claims to be measuring sexual harassment, it needs to be operationalized in a manner that is consistent with someone’s definition of sexual harassment, whether that is a legal or a lay definition of sexual harassment, or even the researcher’s own definition. We agree with Frazier, Cochran, and Olson (1995, p. 33) who note, “Sexual harassment needs to be defined more clearly in future research.” If the general public (and the legal system) consider sexual harassment to be illegal behavior, saying one is measuring “psychological sexual harassment” is unlikely to clarify the issue for them.

Finally, there is the issue of the use of the SEQ in forensic work. In research, the SEQ in its various versions has been used primarily to understand the relationships among concepts, for example, whether particular kinds of work environments result in more or fewer SEQ type behaviors. In forensic work, where such information is useful to assist the finder of fact, citing research using the SEQ is appropriate. However, administering the SEQ to employees in an organization being investigated for sexual harassment in order to determine the level of harassment in that organization is problematic. The SEQ is particularly ill-suited when the plaintiff and defendant want to be able to compare the amount of sexual harassment in one organization with some norm or standard, with the amount of sexual harassment in another organization, or with an “average” amount of harassment.

¹¹It is important to note that the fact that the SEQ may have the effect of underestimating assertive responses to harassment does not imply that the majority or even a sizable minority of women respond assertively.

APPENDIX A

Characteristics in Development of the SEQ Measure

| Article | SEQ Version | SEQ included | Number of items | | Sample size | |
|-----------------------------------|-----------------------------|-------------------|-----------------|-------------|---|--|
| | | | In scale | In analysis | Female | Male |
| Fitzgerald, et al., 1988 | S1 ^a ; SEQ | Abbr ^b | 28 | 24 | S1: 658g ^c , 1088ug ^c | 853ug ^c |
| Fitzgerald, Gelfand, et al., 1995 | S2 ^a ; SEQ2 | | 33 | 28 and 19 | S2: 307u ^d | |
| | SEQ-W | Abbr ^b | 20 | 17 | 448p ^e | |
| Gelfand et al., 1995 | SEQ-E or SEQ-W ^f | Abbr ^b | 25 in common | 18 | s1 ^g : 658g ^c , 1089ug ^c s2 ^g : 389s ^h s3 ^g : 307u ^d | |
| Fitzgerald, Drasgow, et al., 1997 | SEQ-R | No | 18 | 18 | 357p ^e | |
| Glomb et al., 1997 | SEQ | No | 18 | 18 | 455p ^e , 295f ⁱ | |
| Schneider et al., 1997 | SEQ | No | 18 | 18 | s1 ^g : 447p ^e s2 ^g : 300u ^d | |
| Cortina et al., 1998 | SEQ-s | No | 8 | 8 | 651ug ^j , 386g ^j | s1 ^g : 378p ^e s2 ^g : 209u ^d |
| Waldo et al., 1998 | SHOM | Yes | Varies | Varies | | s3 ^g : 420f ⁱ 5,855 ^k |
| Donovan and Drasgow, 1999 | SEQ-DoD | Abbr ^b | 26 | 26 | 22,399 ^k | |
| Fitzgerald et al., 1999 | SEQ-DoD | Abbr ^b | 26 | 23 | 298u ^d | Refer to Hay and Elig (1999) |
| Glomb et al., 1999 | "SEQ" | No | 18 | unclear | | |
| Magley et al., 1999 | "SEQ" | No | Varies | Varies | s1 ^g : 459p ^e s2 ^g : 300u ^d | |
| Wasti et al., 2000 | "SEQ" | No | 19 | 19 | s3 ^g : 419f ⁱ | |
| Cortina, 2001 | SEQ-L | Yes | 20 | 20 | s1 ^g : 355f ⁱ s2 ^g : 455p ^e | |
| Stark et al., 2002 | SEQ-DoD-s | Yes | 23, 16 | 23, 16 | 476 ^m 22,035 ^k | 5,904 ^k |

^aS1 = Study 1. S2 = Study 2.
^bAbbr = Abbreviated form (usually without lead-in) in tables only.
^cug = undergraduate students, g = graduate students from two medium-sized universities.
^du = employees from a university (faculty, administrators, and staff).
^ep = employees from west coast public utility company.
^fNote: Gelfand et al. identify the two forms they use as SEQ-E and SEQ-W. However, they cite Fitzgerald et al., 1988 as the developers of those forms. They were called SEQ and SEQ2 in Fitzgerald et al., 1988. They apparently became SEQ-E and SEQ-W after a revision that took place after Gelfand et al. was written (Fitzgerald et al., 1995).
^gs1 = Sample 1. s2 = Sample 2. s3 = Sample 3.
^hs = university students from Brazil.
ⁱf = employees from a food processing company; also known as the agribusiness sample.
^jug = undergraduate students, g = graduate students from a large Midwestern university.
^k1995 Department of Defense gender issues survey.
^lEmployed urban women from Turkey.
^mLatinas in adult schools or training centers.

APPENDIX B

Development and Changes to the SEQ

| Article | SEQ Version (items) | Response options | Lead-in | Time frame |
|-----------------------------------|---|------------------|---|--|
| Fitzgerald et al., 1988 | S1 ^a : SEQ (28) S2 ^a : SEQ2 (33) | 3 3 | S1: Unknown S2: "Have you ever been in a situation where a male coworker . . . ?" | Ever Ever |
| Fitzgerald, Gelfand, et al., 1995 | SEQ-W (20) | 5 | "Have you ever been in a situation where . . . ?" | Ever |
| Gelfand et al., 1995 | SEQ-E or SEQ-W (25+) | 3 | "Have you ever been in a situation where . . . ?" | Ever |
| Fitzgerald, Drasgow, et al., 1997 | SEQ-R (18) | 5 | Not specific; experiences within organization | Past 2 years |
| Glomb et al., 1997 | SEQ (18) | 5 | Not specific; experiences from male coworkers or supervisors | Past 2 years |
| Schneider et al., 1997 | SEQ (18) | 5 | Not specific; experiences within organizations | Past 24 months |
| Cortina et al., 1998 | SEQ-s (8) | 5 | Not specific; experiences from a male professor or instructor during time at university | Unspecified |
| Waldo et al., 1998 | SHOM (varied) | 5 | "During the past 2 years at work, have any of your supervisors or coworkers": | Past 2 years |
| Donovan and Drasgow, 1999 | SEQ-DoD (26) | 5 | Not specific; experiences in the military during the previous 12 months | Previous 12 months |
| Fitzgerald et al., 1999 | SEQ-DoD (26) | See above | See above | See above |
| Glomb et al., 1999 | SEQ (18) | 5 | t1 ^b : Not specific t2 ^b : Not specific | s1 ^b : previous 2 years t2 ^b : since last interviewed |
| Magley et al., 1999 | SEQ (varied) | 5 | Not specific; experiences at their organization | Past 24 months |
| Wasti et al., 2000 | SEQ (19) | 5 | Not specific; experiences from male coworkers or supervisors | Previous 2 years |
| Cortina, 2001 | SEQ-L (20) | 5 | Experiences from any man in their current workplace | Previous 2 years |
| Stark et al., 2002 | SEQ-DoD-s (16) | 5 | Not specific; experiences in the military during the previous 12 months | Unspecified, assumed to be previous 12 months |

^aS1 = Study 1. S2 = Study 2.^bt1 = Time 1. t2 = Time 2.

APPENDIX C

Psychometric Properties of the SEQ

| Article | SEQ Version | Test-retest | Reliability | | Factors |
|-----------------------------------|---|--|---|-------------|---|
| | | | Overall internal consistency | | |
| | | | Number | Alpha range | |
| Fitzgerald et al., 1988 | S1 ^d : SEQ (28) S2 ^a : SEQ2 (33) | .86 (2-week period; N = 46 female g ^b) | S1: $\alpha = .92$ S2: $\alpha = .86$ | 5 5 | .62-.86 nr ^c |
| Fitzgerald, Gelfand, et al., 1995 | SEQ-W (20) | Refer to Fitzgerald, et al., 1988 | nr ^c | 3 | .42-.85 |
| Gelfand et al., 1995 | SEQ-E or SEQ-W (25+) | Refer to Fitzgerald, et al., 1988 | s2 ^d : $\alpha = .75$ | 3 | nr ^c |
| Fitzgerald, Drasgow, et al., 1997 | SEQ-R (18) | Refer to Fitzgerald, et al., 1988 | $\alpha = .86$ | 3 | .41-.81 |
| Glomb et al., 1997 | SEQ (18) | nr ^c | $\alpha = .87$ | 3 | nr ^c |
| Schneider et al., 1997 | SEQ (18) | nr ^c | s1 ^d : $\alpha = .86$ s2 ^d : $\alpha = .79$ | 3 | nr ^c |
| Cortina et al., 1998 | SEQ-s (8) | nr ^c | $\alpha = .85$ | 3 | nr ^c |
| Waldo et al., 1998 | SHOM (varied) | nr ^c | nr ^c | 5 | s1 ^d : .60-.95 s2 ^d : -. -.92 s3 ^d : .62-.85 |
| Donovan and Drasgow, 1999 | SEQ-DoD (26) | nr ^c | nr ^c | 4 | nr ^c |
| Fitzgerald et al., 1999 | SEQ-DoD (26) | nr ^c | nr ^c | 4 | M ^e : -.97 W ^e : .83-.95 |
| Glomb et al., 1999 | SEQ (18) | nr ^c | t1 ^f : $\alpha = .77$ t2 ^f : $\alpha = .81$ | 3 | nr ^c |
| Magley et al., 1999 | SEQ (varied) | Refer to Fitzgerald, et al., 1988 | s1 ^d : $\alpha = .86$ s2 ^d , s3 ^d : "virtually identical" | 3 | nr ^c |
| Wasti et al., 2000 | SEQ (19) | nr ^c | s1 ^d : $\alpha = .82$ s2 ^d : $\alpha = .86$ | 3 | nr ^c |
| Cortina, 2001 | SEQ-L (20) | nr ^c | $\alpha = .96$ | 3 | .88-.95 |
| Stark et al., 2002 | SEQ-DoD-s (16) | nr ^c | W ^e : $\alpha = .92$ M ^e : $\alpha = .91$ | 4 | W ^e : .83-.92 M ^e : .78-.94 |

^aS1 = Study 1. S2 = Study 2.

^bg = graduate students from a Midwestern university.

^cnr = not reported.

^ds1 = Sample 1. s2 = Sample 2. s3 = Sample 3.

^eM = male subjects. W = female subjects.

^ft1 = Time 1. t2 = Time 2.

ACKNOWLEDGMENTS

The authors wish to thank Peggy Stockdale, Ken Koput, and Karen Korabik for their comments on an earlier version of this manuscript.

REFERENCES

- Arvey, R. D., & Cavanaugh, M. A. (1995). Using surveys to assess the prevalence of sexual harassment: Some methodological problems. *Journal of Social Issues, 51*(1), 39–52.
- Barak, A., Fisher, W. A., & Houston, S. (1992). Individual difference correlates of the experience of sexual harassment among female university students. *Journal of Applied Social Psychology, 22*, 17–37.
- Barak, A., Pitterman, Y., & Yitzhaki, R. (1995). An empirical test of the role of power differential in originating sexual harassment. *Basic and Applied Social Psychology, 17*, 497–517.
- Brooks, L., & Perot, A. R. (1991). Reporting sexual harassment: Exploring a predictive model. *Psychology of Women Quarterly, 15*, 31–47.
- Cortina, L. M. (2001). Assessing sexual harassment among Latinas: Development of an instrument. *Cultural Diversity and Ethnic Minority Psychology, 7*(2), 164–181.
- Cortina, L. M., Swan, S., Fitzgerald, L. F., & Waldo, C. (1998). Sexual harassment and assault: Chilling the climate for women in academia. *Psychology of Women Quarterly, 22*, 419–441.
- Donovan, M. A., & Drasgow, F. (1999). Do men's and women's experiences of sexual harassment differ? An examination of the differential test functioning of the Sexual Experiences Questionnaire. *Military Psychology, 11*(3), 265–282.
- Equal Employment Opportunity Commission. (1980). Guidelines on discrimination because of sex. (Sect 1604.11). *Federal Register, 45*, 74676–74677.
- Equal Employment Opportunity Commission. (1993). Policy guidelines. 29 C.F.R. 1609.1, Washington DC: US Government Printing Office.
- Equal Employment Opportunity Commission. (1997). Guidelines on discrimination because of sex. 29, (Sect 1604.11) (b). *Federal Register, 62*, 12033–12051.
- Equal Employment Opportunity Commission v. Dial Corporation. (2002). Northern District, Ill., No. 99 C 3356.
- Fitzgerald, L. F. (1990). *Assessing strategies for coping with harassment: A theoretical/empirical approach*. Paper presented at the Midwinter Conference of the Association for Women in Psychology, Tempe, Arizona (cited in Fitzgerald et al., 1995).
- Fitzgerald, L. F., Drasgow, F., Hulin, C. L., Gelfand, M. J., & Magley, V. J. (1997). Antecedents and consequences of sexual harassment in organizations: A test of an integrated model. *Journal of Applied Psychology, 82*, 578–589.
- Fitzgerald, L. F., Gelfand, M. J., & Drasgow, F. (1995). Measuring sexual harassment: Theoretical and psychometric advances. *Basic and Applied Social Psychology, 17*(4), 425–445.
- Fitzgerald, L. F., & Hesson-McInnis, M. (1989). The dimensions of sexual harassment: A structural analysis. *Journal of Vocational Behavior, 35*, 309–326.
- Fitzgerald, L. F., Magley, V. J., Drasgow, F., & Waldo, C. R. (1999). Measuring sexual harassment in the military: The Sexual Experiences Questionnaire (SEQ-DoD). *Military Psychology, 11*, 243–263.
- Fitzgerald, L. F., & Shullman, S. (1985). *The development and validation of an objectively scored measure of sexual harassment*. Paper presented at the annual meeting of the American Psychological Association, Los Angeles. (cited in Fitzgerald, et al., 1988 and Fitzgerald, Gelfand, et al., 1995).
- Fitzgerald, L. F., Shullman, S. L., Bailey, N., Richards, M., Swecker, J., Gold, Y., et al. (1988). The incidence and dimensions of sexual harassment in academia and the workplace. *Journal of Vocational Behavior, 32*, 152–175.
- Fitzgerald, L. F., Swan, S., & Fischer, C. (1995). Why didn't she just report him? The psychological and legal implications of women's responses to sexual harassment. *Journal of Social Issues, 51*, 117–138.
- Fitzgerald, L. F., Swan, S., & Magley, V. J. (1997). But was it really sexual harassment? Legal, behavioral, and psychological definitions of the workplace victimization of women. In W. O'Donohue (Ed.), *Sexual harassment: Theory, research, treatment* (pp. 5–28). Boston: Allyn & Bacon.
- Frazier, P., Cochran, C. C., & Olson, A. M. (1995). Social science research on lay definitions of sexual harassment. *Journal of Social Issues, 51*, 21–38.
- Gelfand, M. J., Fitzgerald, L. F., & Drasgow, F. (1995). The structure of sexual harassment: A confirmatory analysis across cultures and settings. *Journal of Vocational Behavior, 47*, 164–177.

- Ghiselli, E. E., Campbell, J. P., & Zedeck, S. (1981). *Measurement theory for the behavioral sciences*. New York: W. H. Freeman.
- Glomb, T. M., Munson, L. J., Hulin, C. L., Bergman, M. E., & Drasgow, F. (1999). Structural equation models of sexual harassment: Longitudinal explorations and cross-sectional generalizations. *Journal of Applied Psychology, 84*, 14–28.
- Glomb, T. M., Richman, W. L., Hulin, C. L., Drasgow, F., Schneider, K. T., & Fitzgerald, L. F. (1997). Ambient sexual harassment: An integrated model of antecedents and consequences. *Organizational Behavior and Human Decision Processes, 71*, 309–328.
- Gruber, J. E. (1989). How women handle sexual harassment: A literature review. *Sociology and Social Research, 74*, 3–9.
- Gruber, J. E., & Bjorn, L. (1982). Blue-collar blues. *Work and Occupations, 9*, 271–298.
- Gutek, B. A. (1985). *Sex and the workplace*. San Francisco: Jossey-Bass.
- Gutek, B. A., & Done, R. (2001). Sexual harassment. In R. K. Unger (Ed.), *Handbook of the psychology of women and gender* (pp. 367–387). New York: Wiley.
- Gutek, B. A., Done, R., Swindler, S., & Stockdale, M. S. (2002). *A short measure of sexual harassment*. Manuscript in preparation, University of Arizona.
- Gutek, B. A., & Koss, M. P. (1993). Changed women and changed organizations: Consequences of and coping with sexual harassment. *Journal of Vocational Behavior, 42*, 28–48.
- Gutek, B. A., Nakamura, C. Y., Gahart, M., Handschumacher, I., & Russell, D. (1980). Sexuality in the workplace. *Basic and Applied Social Psychology, 1*, 255–265.
- Hay, M. S., & Elig, T. W. (1999). The 1995 Department of Defense Sexual Harassment Survey: Overview and methodology. *Military Psychology, 11*(3), 233–242.
- Houston, S., & Hwang, N. (1996). Correlates of the objective and subjective experiences of sexual harassment in high school. *Sex Roles, 34*, 189–204.
- Kidder, L., & Judd, C. M. (1986). *Research methods in social relations* (5th ed.). New York: Holt Rinehart & Winston.
- Knapp, D. E., Faley, R. H., Ekeberg, S. E., & DuBois, C. L. Z. (1997). Determinants of target responses to sexual harassment: A conceptual framework. *Academy of Management Review, 22*, 687–729.
- Lengnick-Hall, M. L. (1995). Sexual harassment research: A methodological critique. *Personnel Psychology, 48*, 841–864.
- Lott, B., Reilly, M. E., & Howard, D. (1982). Sexual assault and harassment: A campus community case study. *Signs, 8*, 296–318.
- Magley, V. J., Hulin, C. L., Fitzgerald, L. F., & DeNardo, M. (1999). Outcomes of self-labeling sexual harassment. *Journal of Applied Psychology, 84*, 390–402.
- Mazzeo, S. E., Bergman, M. E., Buchanan, N. T., Drasgow, F., & Fitzgerald, L. (2001). Situation-specific assessment of sexual harassment. *Journal of Vocational Behavior, 59*, 120–131.
- Murray, B. (1998). Psychology's voice in sexual harassment law. *American Psychological Association Monitor, 29*(8). (We use quotations from forensic psychologist Laura Brown published in this article).
- Nunnally, J., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York: Wiley.
- O'Connell, C. E., & Korabik, K. (2000). Sexual harassment: The relationship of personal vulnerability, work context, perpetrator status, and type of harassment to outcomes. *Journal of Vocational Behavior, 56*, 299–329.
- O'Donohue, W. (Ed.). (1997). *Sexual harassment: Theory, research, treatment* (pp. 5–28). Boston: Allyn & Bacon.
- O'Hare, E. A., & O'Donohue, W. (1998). Sexual harassment: Identifying risk factors. *Archives of Sexual Behavior, 27*, 561–580.
- Reilly, M. E., Lott, B., & Gallogly, S. M. (1986). Sexual harassment of college students. *Sex Roles, 15*, 333–358.
- Schneider, K. T., Swan, S., & Fitzgerald, L. F. (1997). Job-related and psychological effects of sexual harassment in the workplace: Empirical evidence from two organizations. *Journal of Applied Psychology, 82*(3), 401–415.
- Stark, S., Chernyshenko, O. S., Lancaster, A. R., Drasgow, F., & Fitzgerald, L. F. (2002). Toward standardized measurement of sexual harassment: Shortening the SEQ-DoD using item response theory. *Military Psychology, 14*(1), 49–72.
- Stockdale, M. S. (Ed.). (1996). *Sexual harassment in the workplace*. Newbury Park, CA: Sage.
- Thurstone, L. L. (1947). *Multiple factor analysis*. Chicago: University Press.
- Till, F. J. (1980). *Sexual harassment: A report on the sexual harassment of students*. Washington, DC: National Advisory Council of Women's Educational Programs, 35 pages.
- United States Merit Systems Protection Board. (1995). *Sexual harassment in the federal workplace: Trends, progress, continuing challenges*. Washington, DC: U.S. Government Printing Office.

- Waldo, C. R., Berdahl, J. L., & Fitzgerald, L. F. (1998). Are men sexually harassed? If so, by whom? *Law and Human Behavior, 22*, 59–79.
- Wasti, S. A., Bergman, M. E., Glomb, T. M., & Drasgow, F. (2000). Test of the cross-cultural generalizability of a model of sexual harassment. *Journal of Applied Psychology, 85*(5), 766–778.
- Welsh, S. (1999). Gender and sexual harassment. *Annual Review of Sociology, 24*, 169–190.
- Wiener, R. L., & Hurt, L. (2000). How do people evaluate social sexual conduct at work? A psycholegal model. *Journal of Applied Psychology, 85*(1), 75–85.