social sciences

MDPI

Article

Testing Bernard's Thesis of His and Hers Marriage Using Marital Satisfaction Research Published Between 1929 and 1950

Walter Richard Schumm *, Stephan R. Bollman and Anthony P. Jurich

Department of Applied Human Sciences, Kansas State University, Manhattan, KS 66506-1403, USA; bollman@ksu.edu (S.R.B.)
* Correspondence: schumm@ksu.edu

Abstract

Results from 17 studies published between 1929 and 1950 with 36 effect sizes for marital adjustment, happiness, or satisfaction were subjected to a meta-analysis to test Jessie Bernard's thesis about "his" and "hers" marriage as it may have applied to early family research in the USA. While the meta-analysis did not support the hypotheses that wives would report lower levels of or greater variation in marital adjustment, it was found that wives accounted for a much larger percentage of low satisfaction scores at the lowest levels of marital adjustment or when there were large differences in marital adjustment, using paired data between them and their husbands. Thus, this research, despite being derived from much older data, provides partial support for Bernard's thesis, largely in agreement with more recent research and meta-analyses concerning the same hypotheses. However, even though some recent meta-analyses have found stronger support for all three of Bernard's hypotheses, most meta-analyses on this topic did not control for marital social desirability, nor did they consider research published between 1951 and 1969, leaving room for future research to further verify or reject Bernard's thesis.

Keywords: Bernard; marital adjustment; marital satisfaction; gender; "his and her" marriage; meta-analysis

1. Background

The National Council on Family Relation's decade in review for the 1970s (Spanier and Lewis 1980, p. 98) cited Jessie Bernard's statement that "There is by now a very considerate body of well-authenticated research to show that there really are two marriages in every union and that they do not always coincide" (Bernard 1972, p. 4; also see Bernard 1982, p. 5) as an emerging topic of interest during that decade, although Bernard had noted that decades earlier, in 1897, Durkheim had argued that marriage was "a less auspicious relationship for women than for men" (Bernard 1983, p. 209). Rhyne (1981) cited research in the 1970s in support of Bernard's (1972) hypotheses, concluding that "One of the few consistent findings is that men tend to be more satisfied with their marriages than women" (p. 941). Fowers (1991) aptly stated that "Bernard's (1972) provocative thesis that marriage is better for men than it is for women has generated a good deal of study and discussion" that "ought to be reflected in higher marital satisfaction for husbands than wives" (Fowers 1991, p. 209). Bernard (1998) continued to insist that "there really are two marriages in every marital union, and that they do not always coincide." (p. 450); furthermore, "his is better than hers" (p. 456). As recently as 2002, Bernard's thesis was repeated from her 1972/1982 book, to the effect that "The psychological costs of marriage, in brief, seem

Check for updates

Academic Editor: Steven Sheldon

Received: 10 March 2026
Revised: 22 June 2026
Accepted: 25 June 2026
Published: 6 July 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

to be considerably greater for wives than for husbands and the benefits considerably fewer" (p. 210, an exchange theory perspective), that their mental health condition could be summed up as "dismal" (p. 210), that they adapt more to their husband's expectations than vice versa (p. 212), that they have to settle for a "fairly low emotional diet" (p. 213), and that housework is "menial", "dead-end", and "low-status" (p. 214). Later, she says that "the housewife syndrome might well be viewed as Public Health Problem Number One" (p. 215). Even wives' self-reported happiness was an effect of societal imposition of an inferior status they were obliged to accept as normal (Bernard 2002).

2. Literature Review

While a number of tests of Bernard's thesis were performed between 1970 and 2013, the results were mixed (e.g., Amato et al. 2007; Brkljacic et al. 2019; Broman 2005; Chang et al. 1994; Corra et al. 2009; Fowers 1991; Gager and Sanchez 2003; Green et al. 1998; Hyun and Shin 2010; Kamp Dush et al. 2008; Kurdek 1992, 2005; Lowe and Smith 1987; Schumm et al. 1985, 1997, 1998b, 1998a; Schumm and Silliman 1996; Shek 1995; Stevenson and Wolfers 2009; Taniguchi and Kaufman 2013).

Meta-analyses. Then, in 2014, Jackson et al. reported results of a meta-analysis "to empirically test the widely held assumption that women experience lower marital satisfaction than men" (p. 105). Their study examined results from 226 samples (173 reports), including 173 samples using data from dyads and about 25 from clinical reports. They found that their results varied based on the type of sample: for nondyadic data, wives reported lower marital happiness (effect size: Hedges' g = 0.10, $p < 0.001$); for dyadic data, wives were happier (g = −0.01, n.s.); wives from clinical samples were substantially less happy (g = 0.23, $p < 0.001$). The date range of their included studies was from 1970 to 2009, nearly four decades. They did not consider earlier studies on the assumption that the nature of marriage had been quite different prior to 1970, something even Bernard (2002, p. 217) suggested, focusing on Terman's research from previous decades. They concluded that "at least from the perspective of marital satisfaction, our results fail to support Bernard's (1972) widely held dictum that "his" marriage is better than "hers" marriage and that wives consistently report lower marital satisfaction than husbands" (p. 118).

Later, Buhler et al. (2021) did not find a statistically significant gender difference in relationship satisfaction over a lifetime in their meta-analysis based on 95 studies published between 2003 and 2019. Yet, in partial contrast, Whisman and Balzert (2024) conducted a meta-analysis of 161 studies (between 2004 and 2023) concerning marital satisfaction as a function of gender, finding that wives tended to be less satisfied, with an effect size (Cohen's d) of 0.10 ($p < 0.001$), and reported greater variation (exponentiated effect size of 1.19, $p < 0.001$) in marital satisfaction compared to husbands. More recently, Whisman et al. (2025) used a multinational sample of 33 countries and the Kansas Marital Satisfaction Scale (Schumm et al. 1983; Schumm and Bollman 2026) to test Bernard's hypotheses via a meta-analysis. Whisman et al. (2025) found that women reported lower marital satisfaction than men (d = −0.17, $p < 0.001$), and that by Levene's test for homogeneity of variance, women reported higher levels of variance in marital satisfaction ($p < 0.001$); they also found a female-to-male variance ratio of 1.25. At the same time, men were more likely to score the maximum marital satisfaction score, while women were more likely to score in the lower 1%, 5%, or 10% of marital satisfaction scores. None of the four meta-analyses measured or controlled for marital social desirability (Broderick 1971, 1974; Edmonds 1967; Schumm 2015, p. 40); some did not use dyadic data from both spouses.

Bernard's Explanation. Bernard argued that the husband's marriage was better than his wife's (1982, p. 14). Even when marriage seemed beneficial to both men and women, women reported having more marital problems, greater frustrations, more negative feel-

ings (p. 26), were more likely to consider divorce or separation, more regret about their marriage, were more likely to initiate divorce proceedings, reported less consensus with their husbands on important family issues (p. 27), had more mental health problems, were more likely to consider suicide, and to incur greater psychological costs (p. 28). Bernard pinned many of these discrepancies on the inferior status of being a housewife, a role based on unfulfilling manual labor (p. 43), a condition making wives "sick" (p. 48). Wives were expected to make more of the adjustments in marriage (p. 40), to submit to domination under the husband's greater authority (p. 10), and to accept the shock of, among other changes, of shifting from being catered-to to caterer-to (p. 38). Although our focus will be on marital relationships here, Bernard's theory can be applied to romantic relationships in general, as some suggest that men gain the rewards of romantic relationships while women bear the costs or burdens (e.g., Curley et al. 2026).

Some factors may influence marriage for both men and women positively [e.g., how love is desired and expressed (Chapman [1992] 2015; Ince and Isik 2022; Swihart 1978), effective communication and conflict resolution, dedication commitment (Li and Fung 2011), joyful sexual interaction; emotional, social, sexual, intellectual, and recreational intimacy (Harper et al. 2000; Voss et al. 2026); mutual respect, meaning in life (Pence and Sauerheber 2026), differentiation of self (Ince and Isik 2022); positive memories, attachment styles, self-compassion, mutual forgiveness, non-toxic religiosity (Majzoobi and Forstmeier 2022)] or negatively (constraint commitment (Li and Fung 2011), toxic criticism, disrespect, inadequate financial support, value disagreements, destructive levels of stress or daily hassles (Harper et al. 2000), and neuroticism or psychopathology (Chis 2022).

How spousal gender may interact with such factors is of substantial interest but is beyond the scope of this report. Rather, our goal is to investigate the extent to which gender differences, as discussed by Bernard, in marital satisfaction, happiness, or adjustment were observed between 1925 and 1950 and how those findings might compare, in general, to much later but similar research, despite the social upheavals that occurred be tween those early and much later decades in the United States. In other words, we sought to find if there *were or were not* gender differences in the early 20th century rather than to identify the specific *causes* of any such possible gender differences. We think it should be of interest to family scholars whether Bernard's theory, which she fully developed in the 1970s was applicable in the 1930s and 1940s when she was a young sociologist herself (e.g., Bernard 1933, 1934) studying and researching marital success in that much earlier era.

3. Research Problem, Questions, and Hypotheses

Problem. Bernard's ideas about gender and marriage have been confirmed, at least in part, in some recent research; however, the question of whether her ideas were also found in earlier research (1929–1950) remains. A great deal of societal change in the United States occurred between 1950 and 1972 and more recently. In particular, conditions associated with marriage for women did change substantially between 1929 and later decades. Bernard (1976, p. 211) reported how the proportion of married women in the labor force increase from 4.8% in 1890 to 16.4% (1940), 22.5% (1950), 30.5 (1960), 40.5 (1970), to 43.0% in 1974; for wives with only school age children, the percentage in 1974 reached 51.2%, citing the change as a major change in societal roles (p. 208). By 1978, 55.4% of married women were in the labor force (Bernard 1981a, p. 54). Most married women were primarily stay-at-home housewives between 1929 and 1950 while among later cohorts, most married women were employed outside their home. In addition to changes in employment, there were changes in divorce rates, nonmarital childbearing, and cohabitation before marriage (Amato et al. 2003; Zeiss et al. 1980). The relative importance of marriage for men and women may have changed (Kaufman and Goldscheider 2007). Rice (1983), Mace (1972), and Mace and Mace

(1974) argued that marriages had become "more intrinsic than in the past, with greater emphasis on love, open communication, and companionship" (Copeland et al. 1984, p. 107), a shift away from an institutional form of marriage (Hicks and Platt 1970) It would be remarkable if Bernard's ideas were valid, not only in recent decades but also prior to 1951. If so, then we would have nearly 100 years of research supporting her ideas across multiple cohorts and dramatic social upheavals, an impressive result, suggestive of a fundamental aspect of heterosexual marriages in the United States and perhaps other Western nations. If not, then conditions may have evolved in favor of her ideas, even if they were not valid for her own era.

Thus, our problem is whether Bernard's thesis about gender effects and marital outcomes in terms of marital satisfaction, adjustment, or happiness would be validated in research published between 1925 and 1950 (even though the earliest data we found were from 1929). To deal with this problem, our objective was to obtain as much data as possible from that time frame and to evaluate the effect sizes. As far as we know, no one has yet tested Bernard's thesis with such data in that time frame. Perhaps, despite different cultural/historical conditions, Bernard's thesis would be the same prior to 1951; or perhaps, the societal/historical changes that occurred after 1950 might have led to a greater or lesser support for her thesis.

Research Questions. Research concerning Bernard's thesis would involve three general questions, involving gender associations with (1) marital satisfaction, adjustment, or happiness, (2) variation in SAH, and (3) whether wives predominate at lower levels of marital SAH. If Bernard's thesis were correct, across both early and later decades, not only might (1) marital satisfaction levels be lower for wives, but it would be plausible that (2) the variance in marital satisfaction might be higher for wives due to very low satisfaction scores for some wives compared to husbands. It could also be that (3) for independent samples of husbands and wives, the percentage of wives with very low marital satisfaction scores might be higher than for husbands; and (4) in paired samples of husbands and wives (married couples), when spouses had disparate scores, it might be more common for wives to be the spouse with the lower marital satisfaction score.

Hypotheses. These concerns led to the following four null, two-tailed hypotheses:

Hypothesis 1. *When marital satisfaction/adjustment/happiness (SAH) scores of husbands and wives are compared, there will be no significant differences and effect sizes will be less than 0.15.*

Hypothesis 2. *When variances (standard deviations) of marital SAH scores are compared, variances for wives will not differ from those for husbands.*

Hypothesis 3. *When extremely low marital SAH scores are compared, wives and husbands (not within couple) will be equally likely to report such low scores.*

Hypothesis 4. *When within-couple SAH scores of husbands and wives are very different, wives and husband, will be equally likely to report such low scores.*

Scope of hypothesis testing. While it would be of great interest, as suggested by reviewers of this report, to compare factors, including gender, that predicted marital satisfaction across the two eras of time, and to assess how different social factors might explain marital satisfaction differently for men and women, the earlier studies did not consistently use the same variables as predictors and often did not use much more than demographic variables as predictors (e.g., age, religion, employment, number of children, rural or urban location, etc.). While a variety of theories (social exchange, feminist, symbolic interaction, developmental, family systems, intersectional, etc., as well as numerous mid-range theories)

(Adamsons et al. 2022; White et al. 2019) and numerous variables (oxytocin levels, sexual issues, adverse childhood events, provider roles (Bernard 1981b), societal discrimination, personality traits, communication patterns, forgiveness, health, emotional intelligence, commitment levels, levels of investment in a relationship, interactional patterns, alternatives, fairness of division of labor in the home, etc.) have been used to explain how marriage may differ between men and women, those issues, other than gender, are beyond the scope of this report. Here the focus will be on the extent to which, if any, Bernard's ideas about gender and marriage were confirmed in research between 1929 and 1950 in terms of gender differences in marital satisfaction, adjustment, or happiness, given the substantial societal differences that occurred between 1920 and 1950 and 1972–2026 in the United States.

Requirements for hypothesis testing. In order to test Bernard's thesis, at a minimum, a study must report the means and standard deviations for the measure(s) of marital satisfaction used as well as the numbers of husbands and wives; this allows for calculation of effect sizes, which are more useful for testing Bernard's thesis than significance levels alone. However, there has been much debate about how large an effect size should be for it to be deemed a minimum effect of interest (MEI) or significant effect size of interest (SESOI)[1]. Our review of the literature finds recommendations for an SESOI ranging from 0.10 to 0.21, with more leaning toward 0.15 or greater. Therefore, we will consider effects as small as 0.15 (husbands' marital satisfaction levels higher than their wives') to be supportive of Bernard's thesis. In other words, the smallest effect size of interest (SESOI) for our study will be 0.15, which is a lower SESOI than used in other research (Schumm 2026).

All four hypotheses regarding Bernard's thesis will be examined by the relative number of studies/effect sizes that favor Bernard's thesis, by paired-sample *t*-tests, and by meta-analyses. For Hypotheses 3 and 4, based on previous research, we would expect the relative number of very unhappy wives compared to very unhappy husbands to be about 2:1, with significantly greater odds of wives being in very unhappy levels of SAH. If such data points are not reported by a study, then such a study cannot be used to test Bernard's thesis, and it may be that not all studies would have reported the results needed to test our hypotheses.

4. Methods

Samples/Studies. The authors began collecting research reports that permitted testing of the four hypotheses as early as 1985, though the work intensified around 1995. This part of the effort relied upon manual searching, at a university library, of family-related journals (American Sociological Review, American Journal of Sociology, Marriage and Family Living) and books (Burgess and Cottrell 1939; Hamilton 1929; Terman 1938), including one dissertation (Williams 1938) as well as one article (Ort 1950) found by the senior author while researching the topic of role conflicts (Bollman et al. 1988, 1997). This search was performed before PRISMA (Page et al. 2021) was a formal method for finding topical articles.

PRISMA methods are more of a challenge for older materials, at least at the Kansas State University library that was our primary source for references. The Social Science Citation Index there dates back only to 1956. Sociological Abstracts dates to only 1963. Social Science Abstracts dates only to 1980. PUBMED dates only to 1950. Proquest Research Library dates only to 1971. The Psychology and Behavioral Sciences Collection dates to 1930 but when we searched our three terms and husbands and wives, no peer-reviewed references were found between 1929 and 1950. Most of what we had found was in sociological sources, hence we did not try PSYINFO even though it included our range of dates. Google Scholar, however, apparently dates further back than 1929, so it still seems to have been a good choice for searching our terms and included both psychological and sociological research.

As other research priorities emerged (e.g., military families, research methods, LGBTQ+ research, as well as the mobilization of the senior author for a military deployment, our project was delayed until recently. Further articles were found in Blumel's (1992) chapter (Terman and Buttenwieser 1935; Adams 1946; Reed 1947); notably Blumel only found 27 family research articles from 1930 to 1939 and 16 from 1940 to 1949, as well as four from 1950. In contrast, Blumel found 59 research articles from 1950 to 1959, 83 from 1960 to 1969; 168 from 1970 to 1979 and 384 from 1980 to 1989. A search for additional studies was performed in 2025 by searching Google Scholar, as noted above, for research published between 1920 and 1950 using the terms marital satisfaction, marital happiness, and marital adjustment and "husbands and wives", reviewing the first 150 results for each of the three outcome measures (SAH); it did not make sense to cross check our Google Scholar findings with databases that did not cover the period 1925 to 1950.

While most of the relevant articles found through Google Scholar were not new to us, new articles and books found from searching Google Scholar included Locke and Klausner (1948) and Terman and Oden (1947). If Blumel's totals were correct, then suitable articles composed over 35% (17/48) of those available. After eliminating duplicates and articles whose titles and abstracts did not appear to report gender differences in SAH (e.g., Terman and Johnson 1939), we arrived at a total of 73 potentially relevant articles, of which 49 did not measure SAH at all and 7 that did not measure gender, report standard deviations, or had serious methodological problems (e.g., measuring SAH with only 2 or 3 levels or reporting SAH data for only one gender), leaving us with 17 relevant articles suitable for comparing SAH scores by gender. Of the 56 articles we downloaded and printed that were not suitable for our study, 5 were from the 1920s, 26 from the 1930s, 21 from the 1940s, and 4 from 1950, of which 41 had not been cited by Blumel (1992). From this perspective, 23.3% (17/73) of the articles we downloaded and printed were useful for our study. Data were obtained from the 17 useful sources for 5253 husbands and 5179 wives, not counting cases from subsamples of primary studies.

Again, to test Bernard's thesis, for both husbands and wives, average marital adjustment/happiness scores were needed, as well as sample sizes and standard deviations; studies that did not include this information had to be excluded. Numerous studies were not selected because they either reported data for only one spouse or they did not report standard deviations for their SAH variables (or it was not possible to estimate standard deviations). Table 1 includes the basic characteristics of the 17 studies found through our literature search that met our criteria and permitted testing of the hypotheses. A few studies included the original data as well as means and standard deviations for SAH variables. When original data were available, data sets were recreated and analyzed. On the other hand, some studies would have allowed for paired-sample *t*-tests, but original data were not reported. One study that was not included was by Burgess and Cottrell (1936) because virtually the same marital adjustment scores and couples were used in their 1939 book; notably, the 1936 report featured one additional couple not used in their 1939 report, possibly because the omitted couple featured a very happy wife and a very unhappy husband, an outlier (Burgess and Cottrell 1936, p. 741). Another study (Kelly 1941) was not included because only the range of spousal scores was reported (i.e., no means or standard deviations), though the low end of the range was lower for wives (13) than for husbands (18), even though for both spouses the highest score was 62, a trend that would support Bernard's thesis with respect to Hypotheses 3 and 4.

Table 1. Characteristics of the 17 studies and their samples and measures.

Authors (Study Number)	Year	Sample Size	Sample	Measures	Response Rate	Remarks
Hamilton (1)	1929	200 currently or formerly married persons: 100 women, 100 men	The study began in 1924 and lasted four years. Most of the participants were from New York and under the age of 40. The author interviewed the participants in his private consulting room, asking 334 questions for men and 357 to 372 for women, depending on their pregnancy history. Each session was limited to two hours, but in total times ranged from just over two hours to over thirty hours, with an approximate average of eight hours. Of the initial 203 interviews, 3 were dropped due to various problems. A total of 78 of the men and 46 of the women had graduated from college.	Satisfaction scores from 0 to 14 based on 13 individual questions selected from 13 tables of multiple questions; one item from Table 12 counted for 2 points (p. 59)	Unknown	Of the 100 couples, 15 women were divorced/separated compared to 8 men.
Hamilton (2)		55 intact, matched heterosexual couples	Couples were instructed to not talk about the questions until both had completed their interviews.	Same as above.	Unknown	
Bernard (3)	1933	115 H, 137 W from 146 marriages	The participants were from urban areas of St. Louis, Seattle, and Los Angeles. Over 70% of the sample had at least some college education. The group had higher income but fewer children than the average person from St. Louis. The sample included 31–35 clinical cases. Bernard's husband secured almost half of the cases for this study. Eleven graduate students from Washington University also helped locate respondents. Removing the 31 clinical couples left a sample of 221 spouses, including 101 husbands and 120 wives. However, Bernard did not report scores for the sample of 221.	The marital success scale runs from 0 to 100. The method involves three sheets with 100 traits each. The first sheet is checked as to one's spouse's traits. On the second sheet, the same traits are checked if desirable, while on the third sheet, undesirable traits are checked.		The formula for the scale is $[(a + d) - (b + c)/2] \times 100$, where a = desirable traits assigned to spouse, b = desirable traits not assigned to spouse, c = unfavorable traits assigned to spouse, d = unfavorable traits not assigned to spouse (Bernard 1933).
Terman & Buttenwieser (4)	1935	345 couples	Participants had to be not over 60 years of age and American-born Whites. Many came from an area between San Jose and Redwood City, California, including Palo Alto, Menlo, Mountain View, and Napa.	Questions were provided by Dr. Burgess of the University of Chicago, used items 3–10, scores from 0 to 62.	71.7%	
Terman and Buttonwieser (5)		343 couples	Same sample as above except at least one husband and one wife did not answer the single item used to measure marital happiness	Ten items on marital happiness, with responses from very unhappy to very happy.		
Kirkpatrick (6)	1937	58 couples (×2)	100 students from the University of Minnesota mailed 200 surveys to couples they thought were well-adjusted or not (200 couples, 400 individuals). At least one spouse returned part of their survey for a total of 284 at least partial returns. Only 210 returns included surveys completed by both husband and wife. Only 120 matched surveys (mailed by the same student) were returned.	The Family Interests Scale (FIS) used 60 items that were checked as enjoyed by oneself vs. with one's spouse. The percentage checked with one's spouse vs. oneself was the score, with a range from 0 to 100.	52.5% (210/400)	See Frumkin (1953) for more details on the FIS.
Kirkpatrick (7)		70 H, 82 W (×2)	These were the individuals who returned surveys whether their spouses did or not.		71% (284/400)	

Table 1. *Cont.*

Authors (Study Number)	Year	Sample Size	Sample	Measures	Response Rate	Remarks
Kirkpatrick (8)		30 H, 30 W (×2)	These are the cases where the student's mailed surveys with at least partial data were returned by all four potential respondents contacted by the same student.		30% (120/400)	
Williams (9)	1938	200 couples	Surveys were distributed through college alumni. The couples had been married between one and up to just under ten years. Only data from couples living in rural areas with towns of less than 2500 population were used. 192/200 couples were from rural areas of the state of New York. Husbands' ages ranged from 21 to 62, wives' from 19 to 38 years. 14 husbands were older than 38 years. Median ages were 31 and 28 for husbands and wives, respectively. A total of 11 couples were separated (6) or divorced (5), while 15 other couples had considered separation or divorce. 85 of the husbands were farmers. 65 couples had no children, 67 had one child, the rest had 2–4 children. 90% of the participants were Protestant. Few of the wives were employed outside the home.	A modification of the Burgess and Cottrell adjustment scale was used; from its 27 items, items 9, 12, 13, and 14 were used, though item 9 had 11 items, with 5 more added in this study. Possible scores were from 0 to 190.		
Terman (10)	1938	792 couples	Respondents from middle and upper-middle classes in urban and suburban California, mainly from within 50 miles of San Francisco, although many were from the Los Angeles area. More than a third of respondents were college graduates. Spouses were prevented from colluding on their answers. Data were collected between December 1934 and May 1935. Average age of husbands was 38.84 (SD = 9.00), for wives 35.76 (SD = 8.70). The average number of years married was 11.40 (SD-7.42).	Marital happiness scale with nine items weighted such that final score could be from 0 to 87. One of the items was a single-item rating of happiness, with seven responses.		
Terman (11)		792 couples	Same as above.	Happiness item with 7 responses, from extremely unhappy to extraordinarily happy.		
Terman (12)		902 H, 644 W	Respondents were from a variety of professional and social backgrounds. Less than 3 percent of those invited to participate did not do so. This approach seemed to yield respondents who were less happily married than in the group of 792 couples. Very little demographic data reported for this group.	Same as above.		

Table 1. *Cont.*

Authors (Study Number)	Year	Sample Size	Sample	Measures	Response Rate	Remarks
Burgess and Cottrell (13)	1939	251 couples	Data were collected mainly between 1931 and 1933 in Illinois as students and others handed out surveys to nearly 7000 couples. About 1300 responded; only 550 were both residents of Illinois and had been married between one and six years; however, 24 responses were not usable, reducing the sample to 526. The median age at marriage was 27.2 years for husbands, 23.1 for wives. The average duration of marriage was 3 years, 1 month. The majority of the respondents had some college education (58.9% husbands, 54.9% wives), were white, Protestant (54.7% husbands, 61.2% wives), and were from Chicago (69.2%) or its suburbs (12.9%). 55.7% had no children.	One-item scale on marital happiness with responses of very happy, happy, average, unhappy, and very unhappy, with scores from 1 to 5.	18.6% (individuals) or 3.6% (couples)	Their 1936 report featured 252 couples; this report omitted one couple in which the husband rated the marriage as very unhappy, the wife as very happy.
Burgess and Cottrell (14)		66 couples	No group descriptions were provided regarding demographics.	Twenty-six items, two rated at 15 points, fourteen at 10 points, one at 7 points, two at 5 points, and seven at 1 point, for a total score from 0 to 194.		
Burgess and Cottrell (15)		70 couples	No group descriptions were provided regarding demographics.	This was a premarital prediction scale, with scores from 0 to 790, mean of 516.0, and SD of 98.8. A total of 41 items were weighted from 10 to 40 points.		None of the participants reported scores of 780 or higher.
Burgess and Wallin (16)	1944	505 couples	505 of 1000 couples completed engagement adjustment scales and then were re-surveyed three years after marriage. Of the other 495, 123 broke their engagements, 26 separated or divorced, 8 lost a member to death, 40 couples filled out the marital adjustment scale after the data were analyzed, and 292 had not responded. The initial surveys were from convenience samples in the Chicago area. Over two-thirds were college-educated, about half were Protestants; all were White. Most were in their twenties and thirties.	The Burgess–Cottrell marital adjustment scale was used. Scores ranged from 90 to 189. Table 3 on p. 330 lists responses from both men and women (not paired but from the same 505 couples).		
Adams/ Terman (17)	1946	100 couples	100 married couples who had been among nearly 4000 who had participated in pre-marriage surveys at Penn State University, 1939 to 1945, offered by Dr. Adams.	Terman's modification of the Burgess and Cottrell Index of Marital Adjustment.	2.5%	All 100 couples responded to all three of the instruments used.

Table 1. *Cont.*

Authors (Study Number)	Year	Sample Size	Sample	Measures	Response Rate	Remarks
Adams/ Hamilton (18)	1946	100 couples	The average age at marriage was 24 for husbands, 22 for wives. The average age of husbands was 26.35 years, of wives 24.13 years, according to the husbands.	Hamilton's 13 questions on marital adjustment.	2.5%	Only three husbands had contemplated separation; in contrast twelve wives had contemplated separation, of whom six had contemplated divorce.
Adams/ Burgess and Cottrell (19)	1946	100 couples	The wives reported their average age to be 24.30, their husbands was 26.18. Most were college graduates. The average duration of marriage was 2.36 years.	Burgess and Cottrell Index of Marital Adjustment.	2.5%	
Locke (20)	1947	200 couples	A random sample of married persons recommended "happily married" couples for the survey, leading to 204 such couples, of whom four were dropped for having reported very low adjustment scores. The average duration of marriage was 16 years; the median level of education was 9.5 years. The data were collected in 1940 from Monroe County, Indiana. 15 spouses had been previously divorced.	Burgess and Contrell marital adjustment scale, with maximum score of 194.		How the random sample was obtained was not explained.
Reed (21)	1947	860 couples	Data were collected from couples in Indianapolis, Indiana between April 1941 and January 1942. The sample was restricted to Whites, Protestants, married between 1927 and 1929, with the wife under 30 and husband under 40 at time of marriage, neither previously married, eight or more years of residence in a city of 25,000 or more since marriage, and both at least elementary school graduates.	One item, "Everything considered, how happy has your marriage been?" Responses were extremely unhappy, decidedly less happy than average, somewhat less happy than average, about average, somewhat more happy than average, decidedly more happy than average, and extremely happy.		While the sample was inflated to 1444, we used N = 860 for our analyses.
Terman and Oden (22)	1947	556 couples	The study began in 1921–1922 with a nonrandom sample of pre-high school (N = 1070) and high school children (N = 458) with IQ scores of 135 (top 1%) or higher from California. In the autumn of 1939, a follow-up study was begun with research performed in 1940. The couples in that study were married between age 18 and 36 (men) and 16 and 34 (women). The average age at marriage was 25.2 for men and 23.4 for women (SD = 3.1, both). The average age was just under 30 years. Husbands were an average of 4.0 years older than their wives (SD = 4.0). A majority of the respondents were college graduates. Of the 556 couples, 310 were gifted husband couples, 243 were gifted wife couples, and in 3 couples both were gifted.	52 questions were used to form a marital adjustment scale with scores from 0 to 100.	697 spouses, 391 husbands and 306 wives, completed the full marital adjustment test, which would be a response rate of 43.6%, although discounting those who died between 1922 and 1940 might raise that rate.	

Table 1. *Cont.*

Authors (Study Number)	Year	Sample Size	Sample	Measures	Response Rate	Remarks
Terman and Oden (23)		567 couples	The sample here included 317 gifted husbands and 250 gifted wives and their spouses. There were eleven couples who did not fill out the entire marital adjustment scale but did fill out the single item for marital happiness.	A single item was used for the rated happiness of the marriage, with seven responses, from extremely unhappy to extraordinarily happy.	Of the 697 couples who responded, 567 couples answered the single item, giving a response rate of 81.3%.	
Terman and Oden (24)		317 couples	317 gifted husbands and their wives.	Marital adjustment scale.		
Terman and Oden (25)		250 couples	250 gifted wives and their husbands.	Marital adjustment scale.		
Locke and Klausner (26)	1948	31 H, 33 W	64 never-divorced spouses were surveyed in the Los Angeles area by students in university sociology classes. About 60% were Protestants, the median for years of education was about 14. Of the 350 surveys distributed, 127 were returned (36.3%) but 16 were not used for a variety of reasons (missing data, non-white race, not married, or did not answer with respect to current marriage).	The Burgess–Cottrell marital adjustment scale, with scores from 4 to 194. Standard deviations were estimated from critical ratios and standard errors of the mean difference.		As low as 18.3%. Of those eligible, possibly 19.2%.
Locke and Mack-eprang (27)	1949	44 couples	Subset from 404 happily married persons from survey in Monroe County, Indiana, begun in 1938; all wives employed outside the home.	Burgess–Cottrell marital adjustment scale, maximum score of 194 points.		
Locke and Mack-eprang (28)		110 couples	Same as above except wives were not working outside the home.	Same as above.		
Locke and Mack-eprang (29)		41 couples	Convenience sample of couples with wives employed outside the home from Los Angeles, CA. Groups of 41 and 51 roughly matched on wife's education; both husbands and wives matched on age, education, length of time married, and husband's income. No significant differences found between husbands and wives.	Burgess–Cottrell marital adjustment scale, minus items 3, 7, 8, 19, and 21–27; maximum score of 167. Same as above, with addition of seven questions from previous research by Terman and Locke; maximum score of 235.		
Locke and Mack-eprang (30)		41 couples	Same as above.			
Locke and Mack-eprang (31)		51 couples	Same as above except wives were not employed outside the home.			
Locke and Mack-eprang (32)		51 couples	Same as above.			

Table 1. *Cont.*

Authors (Study Number)	Year	Sample Size	Sample	Measures	Response Rate	Remarks
Landis et al. (33)	1950	212 couples	Two graduate students handed out surveys to 228 couples living in student housing at Michigan State College. Seven refused cooperation and nine sets of returns were not usable, leading to a final sample of 212 couples. Couples were selected on the basis of living in student housing and having had a first child within the past 2.5 years. Only 10.5% of the husbands and 6.6% of the wives were married at age 26 or older. Most of the couples (76.6%) had been married between 1.5 and 3.5 years, with only 2.9% married for more than 6.5 years.	A single-item scale was used to rate marital happiness, with responses from very unhappy, unhappy, average, happy, and very happy. Only 12.8% of the wives and 15.2% of the husbands had scores of average or below.		The spouses may have colluded on their answers since the surveys were answered in their apartments.
Ort (34)	1950	50 couples	Private interviews conducted with 50 male students, 33 undergraduates and 17 graduate, along with their wives. Men were 21 to 38 years old, wives 20 to 39. Median ages were 25.5 for men, 24 for women. While 12 wives had only a high school education, the rest had at least some university training.	Each spouse ranked the marriages of ten friends from best (10) to worst (1) and then picked the one closest to their own (scores of 1 to 10).		Students most likely from Wabash College in Indiana.
Terman (35)	1950	591 couples	In 1940, respondents were asked 15 questions about their marriage. The 643 couples were part of Terman's study on gifted persons. Eight years later, the same couples were contacted for information on their marital status.	A marital happiness scale of 15 items weighted to yield scores for each spouse from 0 to 100. Nine items, including eight items from the Burgess–Cottrell marital adjustment scale, had been used earlier by Terman (1938) for 792 other couples, while six items were new.		
Terman (36)	1950	52 couples	Same study as above. Fifty-two couples had broken up, with all but two or three of the couples having divorced.	Same as above.		

Note: The numbers in parentheses after the authors' names represent the number for the 36 effect sizes that could be calculated from the raw data.

Measurement. Other measures were developed to permit more detailed assessment of potential methodological moderator variables, which were evaluated through meta-regression (SPSS 31.0), predicting effect sizes from each potential moderator variable that might influence our meta-analyses. Measures of study quality included whether or not standard deviations had to be estimated (yes = 0, no = 1), if the sample size was greater than 100 or not (no = 0, yes = 1), if sufficient data were present to assess a ratio of unhappiness of wives compared to husbands (no = 0, yes = 1); if the study had data from one site only (0), more than one site within a state (1), or multiple sites across different states (2); was a response rate discussed (no = 0, yes = 1), were more than two demographic characteristics of the sample mentioned (no = 0, yes = 1), or were multiple items used to measure marital adjustment (no = 0, yes = 1). Cronbach's alpha was used to assess the internal consistency reliability of these seven items but was only 0.281. Therefore, the weakest items were deleted until adequate reliability was achieved (0.75), deleting items based on response

rate, demographics, and multiple items. However, including the demographics item only reduced the Cronbach's alpha to 0.65; nevertheless, we used the sum of the four items to form a measure of study quality (QUALITY4), with a Cronbach's alpha of 0.75.

Aside from study quality, we also measured other aspects of each study as potential moderator variables to allow for comparison of effect sizes across different subsamples. We checked whether the study stated that divorced spouses were included (yes/no) in the sample, since that could bias results. We assessed whether the study reported multiple samples within one journal article, since such samples often overlapped in terms of having many of the same couples in each sample. We assessed whether the sample(s) consisted entirely of couples married to each other or whether husbands and wives were included who were not married to each other. We also coded effect sizes from the results as positive (supported Bernard's thesis) or negative (did not support Bernard's thesis). Our data set also included the authors of each study and the date of publication. Unfortunately, the year(s) of data collection were often not mentioned in the articles; likewise, many demographics such as average ages at marriage, ages of spouses, educational attainment of spouses, number of children in the family, religious affiliations, and race/ethnicity were often not reported. Table 1 includes descriptions of each study; Table 2 includes the basic data from each study used in our meta-analyses, to permit replication or re-analysis by other scholars if desired.

Table 2. Tests of Bernard's thesis for early studies (1929–1950) for differences in marital satisfaction/adjustment/happiness measures.

Author	Date (Case Number)	Sample N	Mean (SD) Husbands	Mean (SD) Wives	*t*-Test Results +	Effect (Size) g/SE	Lower Satisfaction Ratio for Wives vs. Husbands
Hamilton (1929)	p. 79 (1)	100 husbands 100 wives	6.58 (3.70)	5.92 (3.82)	$t(198) = 1.24$ $p = 0.216$	0.175 0.1417	16:7 (scores of ≤1); data not provided
	p. 79 (2)	(15 wives/ 8 husbands divorced or separated)	6.91	6.36	$t(108) = 0.96$	0.179	
		55 heterosexual couples	[3.00]	[3.00]	$p = 0.339$, paired-sample data not provided	0.1911	
Bernard (1933, 1934)	p. 197 (3)	115 husbands 137 wives	79.70 (17.0)	77.90 (17.00)	$t(250) = 0.403$ $p = 0.403$	0.106 0.1266	6:4 (scores ≤ 38.50)
Terman and Buttenwieser (1935)	p. 162 (4)	345 husbands 345 wives	46.20 (12.46)	46.44 (12.71)	$t(688) = -0.248$ p = n.s.	−0.019 0.0761	5:2 (scores of ≤5)
	p. 167 (5)	343 husbands 343 wives	4.17 (0.89)	4.21 (0.88)	$t(684) = -0.655$ p = n.s.	−0.045 0.0764	11:9 (scores of ≤2)
Kirkpatrick (1937)	pp. 136–37 (6)	58 husbands 58 wives	82.2 [12.75]	80.8 [9.95]	$t(114) = 0.66$ p = n.s.	0.122 0.1859	Data not provided
	pp. 136–37 (7)	70 husbands 82 wives	82.1 [12.75]	79.4 [9.95]	$t(150) = 1.47$ $p = 0.145$	0.237 0.1633	Data not provided
	pp. 136–37 (8)	30 husbands 30 wives	76.3 [4.3]	79.0 [9.95]	$t(58) = 1.13$ $p = 0.265$	−0.348 0.2603	Data not provided
Williams (1938)	p. 48 (9)	200 heterosexual couples	151.25 (24.91)	153.30 (23.38)	$t(199) = 1.59$ $p = 0.114$ r = 0.715 ***	−0.085 0.1000	8:4 cases (≥40 point difference)
Terman (1938)	p. 63 (10)	792 husbands 792 wives	68.40 (17.35)	69.25 (18.75)	$t(1582) = -0.936$ $p = 0.349$	−0.047 0.0503	13:4 (scores < 10) 6:1 (extremely unhappy)
	p. 76 (11)	792 husbands 792 wives	5.71 (1.20)	5.81 (1.25)	$t(1582) = -1.62$ $p = 0.106$	−0.082 0.0503	
Terman (1938)	p. 78 (12)	902 husbands 644 wives	5.23 (1.62)	5.11 (1.82)	$t(1544) = 1.42$ $p = 0.154$	0.070 0.0516	46:36 (scored lowest level of satisfaction, 1)

Table 2. *Cont.*

Author	Date (Case Number)	Sample N	Mean (SD) Husbands	Mean (SD) Wives	*t*-Test Results +	Effect (Size) g/SE	Lower Satisfaction Ratio for Wives vs. Husbands
Burgess and Cottrell (1939)	p. 39 (13)	251 couples	4.20 (1.03)	4.25 (1.03)	$t(250) = -1.21$ $p = 0.226$ r = 0.815 ***	−0.048 0.0893	6:3 (if difference is 2 points, wife less happy)
	p. 70 (14)	66 couples	144.55 (38.01)	142.73 (39.68)	$t(65) = 0.78$ $p = 0.439$ r = 0.882 ***	0.047 0.1741	9:4 (if difference is 30 or more points, wife less happy)
	p. 285 (15)	70 couples	538.86 (99.32)	537.71 (100.70)	$t(69) = 0.199$ $p = 0.843$ r = 0.884 ***	0.011 0.1690	5:4 (difference of 80 points or more, wife less happy)
Burgess and Wallin (1944)	p. 330 (16)	505 heterosexual couples	160.45 (20.04)	163.14 (18.76)	$t(1008) = -2.21$ $p = 0.028$	−0.138 0.0630	4:4 (scores < 100)
Adams (1946)	p. 187 (17)	100 husbands	T: 76.85 (9.81)	75.95 (13.20)	$t(198) = 0.547$ $p = 0.585$	0.077 0.1415	Data not provided
	p. 187 (18)	100 wives	H: 10.55 (2.70)	10.22 (3.00)	$t(198) = 0.818$ $p = 0.415$	0.115 0.1415	Data not provided
	p. 187 (19)	100 husbands; 100 wives	BC: 169.60 (19.00)	163.75 (21.75)	$t(198) = 2.03$ $p = 0.044$	0.285 0.1422	Data not provided
Locke (1947)	p. 188 (20)	200 couples	167.3 (12.7, est.)	165.6 (12.7, est.)	$t(398) = 1.34$ $p = 0.182$	0.134 0.1001	4.5%/2.5% of men/women had lowest scores
Reed (1947)	p. 387 (21)	860 couples	5.34 (1.36)	5.39 (1.37)	$t(1718) = -0.76$ $p = 0.448$	−0.037 0.0482	38:42 (scores ≤ 3)
Terman and Oden (1947)	p. 241 (22)	556 husbands; 556 wives	63.15 (19.76)	66.34 (18.90)	$t(1110) = -2.75$ $p = 0.006$	−0.165 0.0601	2:3 (scores ≤ 10)
	p. 244 (23)	567 husbands; 567 wives	5.93 (1.11)	6.10 (1.10)	$t(1132) = -2.69$ $p = 0.007$	−0.154 0.0595	12:7 (scores ≤ 2)
	p. 245 (24)	317 couples with gifted husbands	59.50 (20.36)	66.66 (19.30)	$t(632) = -4.54$ $p < 0.0001$	−0.361 0.0801	Data not provided
	p. 245 (25)	250 couples with gifted wives	66.90 (18.80)	64.86 (18.71)	$t(498) = 1.22$ $p = 0.225$	0.109 0.0895	Data not provided
Locke and Klausner (1948)	p. 100 (26)	31 husbands 33 wives	159 (25.4)	151 (25.4)	$t(62) = 1.26$ $p = 0.213$	0.311 0.2517	Data not provided
Locke and Mackeprang (1949)	p. 536 (27)	44 couples, wives employed	168.2 (28.0)	167.8 (34.6)	$t(86) = 0.06$ $p =$ n.s.	0.013 0.2132	Data not provided
	pp. 537–38 (28)	110 couples, wives not employed	168.7 (28.0)	164.10 (34.6)	$t(218) = 1.08$ $p = 0.280$	0.146 0.1350	Data not provided
	pp. 537–38 (29)	41 couples, wives employed	137.8 (27.05)	139.5 (17.31)	$t(80) = -0.34$ $p =$ n.s.	−0.074 0.2209	Data not provided
	pp. 537–38 (30)	41 couples, wives not employed	191.4 (40.53)	193.6 (34.55)	$t(80) = -0.26$ $p =$ n.s.	−0.058 0.2209	Data not provided
	pp. 537–38 (31)	51 couples, wives not employed	136.4 (27.05)	142.8 (17.31)	$t(100) = -1.42$ $p = 0.158$	−0.280 0.1990	Data not provided
	pp. 537–38 (32)	51 couples, wives not employed	189.7 (40.53)	196.5 (34.55)	$t(100) = -0.91$ $p = 0.364$	−0.179 0.1984	Data not provided
Landis et al. (1950)	p. 767 (33)	212 couples	4.19 (1.03)	4.34 (1.00)	$t(422) = -1.48$ $p = 0.139$	−0 0.148 0.0973	27:32 (scores ≤ 3)
Ort (1950)	p. 695 (34)	50 couples	8.46 (1.43)	8.90 (1.47)	$t(98) = -1.51$ $p = 0.133$	−0.301 0.2012	5:3 (scores ≤ 6)
Terman and Wallin (1949) Terman (1950)	p. 52 (35) p. 503 p. 52 (36)	591 couples	63.81 (16.2)	66.68 (16.2)	$t(1180) = -3.05$ $p = 0.0024$	−0.177 0.0583	Data not provided
		52 couples	49.8 (22.4)	48.7 (21.7)	$t(102) = 0.25$ $p =$ n.s.	0.050 0.1961	Data not provided

+ Inclusion of a correlation coefficient (r) indicates use of a paired-sample *t*-test. Nonparametric test results included only if different from *t*-test results. Two-tailed tests of statistical significance are used. Positive *t*-tests and Hedges' g indicate higher marital satisfaction for husbands; negative for wives. Hedges' g's are reported in Figure 1 for the studies. When $p > 0.50$, p is reported as n.s. (not significant). Brackets [ ] around standard deviations indicate that they were estimated, usually from the standardized difference in means or from *t*-test results not involving husband/wife comparisons. Adams (1946) used three different measures of marital adjustment: Terman's (T), Hamilton's (H), and Burgess and Cottrell's (BC). The numbers in parentheses within the citation column refer to the sequence number of the tested effect sizes, as was used in Table 1. *** $p < 0.001$.

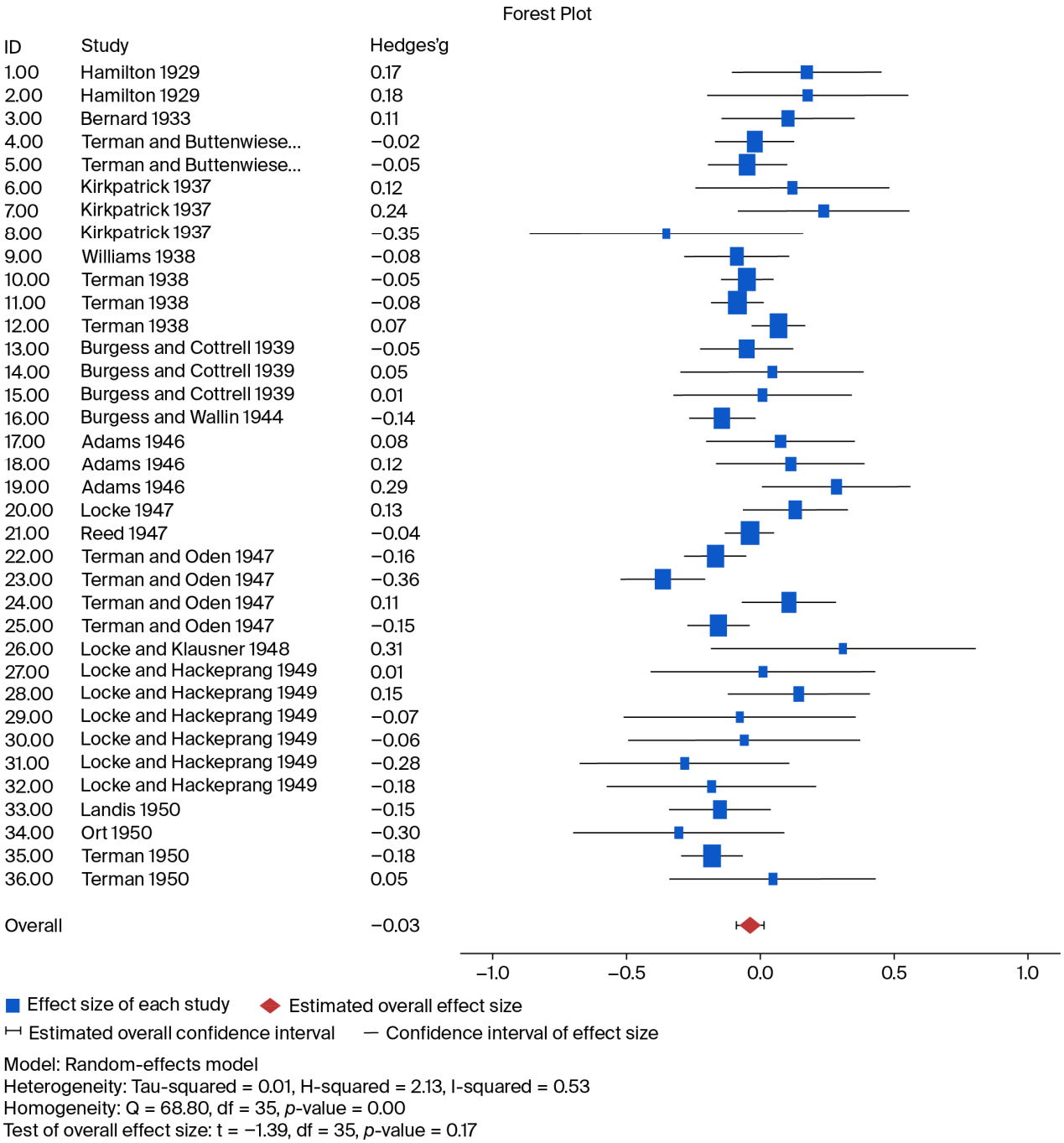


Figure 1. Forest plot for meta-analysis of 17 studies (36 effect sizes) published from 1929 to 1950 on marital adjustment/happiness/satisfaction as related to gender of spouse.

Analyses. For Hypothesis 1, meta-analysis was used to assess differences in mean SAH scores for husbands and wives. The data needed for this type of meta-analysis included the mean scores, standard deviations, and sample sizes for both husbands and wives. SPSS was used to calculate results based on raw data, continuous outcomes, Hedges' g to measure effect size, random- and fixed-effects models, inverse-variance weights, restricted maximum likelihood (REML) estimation methods, and standard errors adjusted by the truncated Knapp–Hartung method. Borenstein et al. (2021, p. 25) accept the inclusion of both independent-sample-based studies and paired-sample-based studies (effect sizes), as we used here, in meta-analyses.

For Hypothesis 2, meta-analysis was used to compare the variation in SAH scores for husbands and wives. The data needed for this type of meta-analysis included computations of the variance of husbands and wives and standard errors for each of the 36 cases. For Hypothesis 2, we used SPSS selections: continuous outcomes with pre-calculated effect sizes, as well as an REML estimation method, inverse-variance weights, and the truncated Knapp–Hartung standard error adjustment, to determine the average effect size in terms of the natural log of the ratio of wives' to husbands' variances (see Whisman and Balzert 2024; Barker et al. 2021, for more details regarding this methodology, although Whisman et al. 2025, used merely the ratio of variances).

For Hypothesis 3, meta-analysis was used to compare the odds for husbands and wives when (a) spouses reported SAH scores below selected low cut-off points (Table 1).

For Hypothesis 4, meta-analysis was also used when spouses had very different SAH scores. The data needed for these types of meta-analysis (for Hypotheses 3 and 4) included the number of "successes" (how many spouses had very low SAH scores compared to their "failures" (how many in each sample did not have such low scores, where the total of successes and failures = sample size) or the number of failures and successes based on how many wives vs. husbands had lower scores than their spouse when SAH differences were large. For the 3rd and 4th hypotheses, SPSS was used to calculate results based on raw data, binary outcomes, log odds ratios, with random- or fixed-effects methods, with inverse-variance weights with additional methodological details.[2]

Supplemental analyses. For each selected study, independent-sample or paired-sample *t*-tests were used to compare husband and wife SAH scores, depending on whether paired data were available. If data were paired but not sufficient to permit paired-sample *t*-tests, independent-sample *t*-tests were used. Effect sizes (g) and their standard errors were calculated for each study (Table 2). SPSS version 31.0 was used for statistical computations. These analyses are included as separate endnotes[3] for further tests of each of our four hypotheses; they differ from meta-analytic approaches in that the cases are weighted equally rather than with weights adjusted by sample size.

5. Results

Hypothesis 1. *Differences in mean scores for marital adjustment?*

Table 3 presents the overall meta-analysis results obtained using a variety of methods. Regardless of the method used, the overall effect sizes were negative (wives reporting higher marital satisfaction than husbands, contrary to Bernard's thesis, except where she took that situation as one of several signs of poor mental health on the part of wives, Bernard (1975), p. 601; Bernard (1972), pp. 54–57) and not statistically significant except for the fixed-effects analysis. The prediction intervals were wider than the confidence intervals as expected. There seemed to be some publication bias in all the models and methods, with the trim and fill procedure (Borenstein et al. 2021, p. 322) leading to the addition of six cases, resulting in larger effect sizes that were statistically different from zero, in a more negative direction; in other words, if there was publication bias, it obscured or overlooked studies that featured an apparently greater marital adjustment of wives compared to husbands. As another check on our methods, we either used only the 17 studies that featured one effect size or used the effect size most favoring Bernard's thesis when there were multiple effect sizes reported for a study. For this second meta-analysis, we obtained g = 0.027 (SE = 0.0330), with $t(16) = 0.817$ ($p = 0.426$). However, four additional studies were imputed, which changed the estimated effect size to −0.004 ($p = 0.906$, two-tailed).

Table 3. Meta-analytic results using a variety of methods for Hypothesis 1 on gender differences in marital satisfaction/adjustment/happiness scores.

Model	Effects Model Type	Estimation Method	Effect Size Type	SE Adj?	Knapp–Hartung SE Adj? [a]	Effect Size	SE Effect Size	Z/t	*p*	CI [b]	Imputed Trim/Fill Effect Size	SE	Z/t	*p*	CI [c]
1	Fixed	N/A	Hedges' g	No	No	−0.054	0.0150	−0.362	<0.001	−0.084 −0.025 N/A	−0.069 n = 42	0.0146	−4.70	<0.001	−0.098 −0.040
2	Random	REML	Hedges' g	Yes	Yes	−0.035	0.0250	−1.39	0.173	−0.086 0.013 −0.239 0.170	−0.063 n = 42	0.0265	−2.39	0.022	−0.117 −0.010
3	Random	DS-L	Hedges' g	Yes	Yes	−0.037	0.0245	−1.50	0.144	−0.086 0.013 −0.226 0.153	−0.064 n = 42	0.0263	−2.43	0.020	−0.117 −0.011
4	Random	DS-L	Hedges' g	No	No	−0.036	0.0240	−1.51	0.130	−0.083 0.011 −0.226 0.154	−0.064 n = 42	0.0246	−2.60	0.009	−0.112 −0.016
5	Random	REML	Hedges' g	No	No	−0.035	0.0250	−1.38	0.166	−0.084 0.014 −0.240 0.171	−0.063 n = 42	0.0256	−2.47	0.013	−0.113 −0.013

Negative effect sizes indicate that wives reported greater marital satisfaction than did husbands; positive effect sizes, the opposite (which would support Bernard's thesis). [a] Truncated Knapp–Hartung standard error adjustment. [b] First two numbers represent the 95% confidence interval for the effect size; next two numbers represent the prediction interval calculated for random estimation models (Borenstein et al. 2021, pp. 123–24). [c] This is the 95% confidence interval for the combined 36 observed studies and the 6 imputed studies. For all models, df for Q = 35. For all versions of Model 1, results were approximately the same, with Q = 69.18, $p < 0.001$; $I^2 = 49.4\%$; $p = 0.027$ for Egger's regression-based test. For Model 2, Q = 68.80, $p < 0.001$; $I^2 = 53.0\%$; $p = 0.093$ for Egger's regression-based test. For Model 3, Q = 68.80, $p < 0.001$; $I^2 = 49.1\%$; $p = 0.091$ for Egger's regression-based test. For Model 4, Q = 68.80, $p < 0.001$; $I^2 = 49.1\%$; $p = 0.087$ for Egger's regression-based test. For Model 5, Q = 68.80, $p < 0.001$; $I^2 = 53.2\%$, $p = 0.092$ for Egger's regression-based test. REML = restricted maximum likelihood method; DS-L = DerSimonian and Laird method.

Table 4 presents the results for the moderation effects obtained by splitting the sample by several variables. Bernard's thesis (more positive effect sizes) was supported more often in non-dyadic samples (possibly some selection bias towards less happy wives?), samples that included divorced spouses (usually more divorced wives than divorced husbands), smaller samples, samples restricted to only one city, and samples from lower-quality studies. However, none of the pro-Bernard splits yielded significant results compared to all but one of the other splits ($p < 0.10$, two-tailed, even for that one split). In every moderation test reported in Table 4, the methodological change led to a change from a positive effect size to a negative effect size, flipping the observed results in favor of a result contrary to Bernard's thesis. Using study quality as a continuous variable in meta-regression, $t(34) = -1.97$ ($p < 0.06$), with a regression coefficient of −0.036. None of the results were significant for splits based on actual or estimated standard deviations; on the earlier half of the sample set vs. the more recent half; on whether a response rate was or was not reported; nor on whether a single item or multiple items were used to measure marital adjustment (though for the latter split, the effect size was −0.094 ($p = 0.095$) for the studies using single-item measures vs. −0.009 for the studies using multiple-item measures). Differences between the compared subgroups were not significant only for divorce and demographic comparisons (Table 4). The year of publication was significant when used as an interval variable but not when coded as a binary variable (early/later). When we performed a meta-analysis on only the studies whose results favored Bernard's thesis, the overall effect size was 0.11 ($p = 0.002$), which suggested that under the most favorable of circumstances, Bernard's thesis might be weakly supported but with what most scholars would consider to be a "small" effect size, though below our SESOI of 0.15. On the other hand, the effect size for the other studies was −0.117 ($p < 0.001$), also below our SESOI, such that the magnitudes of the effect sizes were essentially equal, though the direction of effects was opposite (also see note 1). Since most of our results were either not statistically significant or were below our SESOI, calculating fail-safe N's (Orwin 1983) did not seem appropriate.

Table 4. Tests of moderation effects for selected variables for Hypothesis 1.

How Samples Were Split (n)	Effect Size	SE (Effect Size)	*t*	*p*	95% CI	Prediction Interval
Not dyadic (5)	0.103	0.0430	2.39	0.075	−0.017 to 0.222	−0.034 to 0.239
Dyadic only (31)	−0.061	0.0249	−2.46	0.020	−0.112 to −0.010	−0.239 to 0.120
Difference	−0.192	0.0689	−2.78	0.009	−0.332 to −0.051	Q(34) = 53.70, p = 0.017, I^2 = 41.8%
Divorced included (6)	0.038	0.0368	1.05	0.343	−0.056 to 0.133	−0.064 to 0.141
No divorced included (30)	−0.049	0.0284	−1.72	0.097	−0.107 to 0.009	−0.266 to 0.168
Difference	0.085	0.0634	1.35	0.186	−0.043 to 0.214	Q(34) = 61.21, p = 0.003, I^2 = 49.3%
Sample size equal or less than 100 (18)	0.056	0.0424	1.33	0.203	−0.033 to 0.146	−0.034 to 0.146
Sample size greater than 100 (18)	−0.065	0.0289	−2.25	0.038	−0.126 to −0.004	−0.267 to 0.137
Difference	−0.115	0.0544	−2.12	0.042	−0.226 to −0.005	Q(34) = 61.08, p = 0.003, I^2 = 46.6%
Two or fewer demographics reported	0.010	0.0337	0.28	0.780	−0.061 to 0.080	−0.188 to 0.207
Three or more demographics reported	−0.082	0.0367	−2.22	0.042	−0.160 to −0.003	−0.302 to 0.139
Difference	−0.093	0.0483	−1.92	0.063	−0.191 to 0.005	Q(34) = 62.99, p = 0.002, I^2 = 49.0%
Sample from one city (25)	0.006	0.0293	0.22	0.831	−0.054 to 0.067	−0.150 to 0.162
Sample from more than one site (11)	−0.091	0.0388	−2.34	0.041	−0.178 to −0.004	−0.326 to 0.144
Difference	−0.101	0.0463	−2.18	0.036	−0.195 to −0.007	Q(34) = 57.96, p = 0.006, I^2 = 44.4%
Lower-QUALITY4 studies (21)	0.033	0.0426	0.76	0.454	−0.056 to 0.121	−0.207 to 0.272
Higher-QUALITY4 studies (15)	−0.073	0.0291	−2.52	0.024	−0.136 to −0.011	−0.262 to 0.116
Difference	−0.104	0.0506	−2.06	0.047	−0.207 to−0.001	Q (34) = 64.93, p = 0.001, I^2 = 49.8%
Only positive effect size studies (17)	0.110	0.0296	3.73	0.002	0.048 to 0.173	0.047 to 0.173
Only negative effect size studies (19)	−0.117	0.0221	−5.31	<0.001	−0.164 to −0.071	−0.229 to −0.005
Difference	−0.228	0.0372	−6.14	<0.001	−0.304 to −0.153	Q(34) = 27.07, p = 0.795, I^2 = 10.1%

Differences between each of the pairs of groups are assessed by meta-regression, predicting effect sizes from the binary variables used to create the pairs of groups. Year of publication was assessed as a continuous covariate, with two digits: g = −0.008, SE = 0.0040, t(34) = −2.04, p = 0.049, Q(34) = 58.91, p = 0.005, I^2 = 46.4%. However, when assessing year of publication as a binary (early/late) variable, the effect size was not significant: −0.074 (SE = 0.0487), t(34) = −1.52, p = 0.137, Q(34) = 62.55, p = 0.002, I^2 = 49.7%. All tests were two-tailed.

Our results for Hypothesis 1 support Ferguson and Heene's (2012) claim that meta-analyses can yield statistically significant results for effect sizes below the SESOI. Most of our results for Hypothesis 1 were below an SESOI of 0.15. Only selectively using studies that favored Bernard's thesis did we find effect sizes >0.10, but usually below 0.15. The higher the methodological quality of the study, the less likely it was that results trended in the direction of Bernard's thesis. Looking at the raw data as a function of low vs. high

QUALITY4, for the former, the range of effect sizes was from −0.348 (with 7 negative effect sizes) to 0.311 (with 14 positive effect sizes). For the latter, the range was from −0.361 (12 negative effect sizes) to 0.109 (3 positive effect sizes). Using Levene's test for equality of variances, comparing standard deviations of 0.191 and 0.119 for lower- and higher-quality studies, respectively, without weighting, p was equal to 0.053. Smaller-sample studies, studies that did not use paired spouses, and studies including divorced persons (more often wives than husbands) were more likely to find effect sizes favoring Bernard's thesis. In general, null Hypothesis 1 cannot be rejected; Bernard's thesis is not supported, particularly for higher-quality studies.

Hypothesis 2. *Differences in standard deviations for marital adjustment?*

Our meta-analysis comparing variances in marital adjustment across gender across the 36 effect sizes resulted in an overall effect size of 0.014 (SE = 0.0346) (z = 0.407, p = 0.684, prediction interval was from 0.769 to 1.338), trivial by virtually any definition, although finding wives to have slightly higher variation in marital adjustment. Using only the 17 studies with one effect size each, our meta-analysis yielded g = 0.025 (SE = 0.0434), $t(16)$ = 0.584 (p = 0.567).

One reviewer cautioned that our estimated standard deviations, being equal for husbands and wives, might bias our results in favor of the null hypothesis. For the 32 cases of unequal standard deviations, our meta-analysis yielded an effect size of 0.014 (SE = 0.0454) and a non-significant t-test (p = 0.758) (also see note 3). Therefore, since our results were both non-significant and below our SESOI, we did not calculate a fail-safe N (Orwin 1983) and null Hypothesis 2 cannot be rejected; Bernard's thesis is not supported.

Hypothesis 3. *Differences in marital adjustment when husbands and wives score in the lowest categories of marital adjustment, using independent data.*

Only 18 of our 36 effect sizes (Table 2) involved studies that reported information on these two issues related to Hypotheses 3 and 4. The cases in which husbands or wives scored below cut-off points could be treated as a binary "success" and the remaining cases as a binary "fail", with the successes and fails adding up to the sample size for that study. We performed a random-effect (with truncated Knapp–Hartung adjustment for SE) meta-analysis of binary data for the 18 cases to test Hypothesis 3. The resulting effect size was 0.396 (SE = 0.1344), with $t(17)$ = 2.95, p = 0.009, with an exponentiated effect size of 1.49 and a 95% prediction interval from 0.799 to 2.763. Q(17) = 17.8 (p = 0.40), with I^2 = 23.6%. A trim-and-fill analysis added four imputed cases, reducing the effect size to 0.276 (SE = 0.1262), $t(17)$ = 2.19 ($p < 0.05$), with an exponentiated effect size of 1.32, with a 95% confidence interval of 1.014 to 1.71. The Egger's regression-based test SE was not significant (p = 0.114, two-tailed). For the individual studies, the effect sizes ranged from −0.407 to 1.79, while the exponentiated effect sizes ranged from 0.111 to 1.182. Tests for moderation did not find significant results as a function of divorce, QUALITY4, or joint pairs. For contrast, we also conducted a fixed-effects meta-analysis, which yielded an effect size of 0.338 (SE = 0.1060), z = 3.19 (p = 0.001, two-tailed), with an exponentiated effect size of 1.40 (95% CI, 1.14 to 1.73), with Q(17) = 17.80, p = 0.402 and I^2 = 4.5%. Trim-and-fill calculations added four imputed studies, which changed the effect size to 0.251 (SE = 0.1006), z = 2.50 (p = 0.013) with an exponentiated effect size of 1.29 (95% CI, 1.06 to 1.57).[4]

Hypothesis 4. *Major differences in husband/wife scores, when using paired data?*

There were 4 studies of the 18 (Table 2) that reported husband and wife scores jointly, so that within-couple differences in marital adjustment could be determined. For these four studies, when husbands and wives differed by larger numbers in their marital adjustment scores, of the 43 cases reported, 28 (65.1%, close to a 2:1 ratio) involved the wife reporting lower marital adjustment than her husband. For these four cases the percentages of such cases of the total of all wives were 6.79% (SD = 4.97%, median = 5.57%) for wives and 1.49% (SD = 2.50, median = 3.86%) for husbands, with $t(3) = 2.01$, $p = 0.138$, two-tailed, 0.069, one-tailed, with Cohen's d = 1.01, Hedges' g = 0.73. A meta-analysis of the data for these four cases yielded an effect size of 0.654 (SE = 0.3301) with $t = 1.98$ ($p = 0.142$), with an exponentiated effect size of 1.92 and an exponentiated prediction interval of 0.465 to 7.96, with no imputed studies. The means and medians might lead us to overlook that half of the four wives had percentages of 7.14 and 13.64 compared to only one husband with a percentage above six percent (6.06%). The results for the other 14 cases, where the issue was relative percentages of husbands and wives reporting very low scores on marital adjustment, were 5.08% (median = 4.08%) for wives and 3.59% (median = 2.56%) for husbands, with $t(13) = 2.10$, $p = 0.056$, two-tailed, 0.028, one-tailed, with Cohen's d = 0.56 and Hedges' g = 0.53. $Q(13) = 0.521$ ($p = 0.914$), while $I^2 = 0\%$. The mean and median can be limited because they do not reveal that 3 of those 14 studies featured some cases of percentages between 10.0 and 16.0 for unhappy wives compared to only one such level (15.09) for husbands. A meta-analysis of these 14 cases yielded an effect size of 0.360 (SE = 0.1498), with $t(13) = 2.40$ ($p = 0.032$), with an exponentiated effect size of 1.43 (95% prediction interval of 0.713 to 2.88). $Q(13) = 16.26$ ($p = 0.236$), while $I^2 = 29.4\%$; trim and fill imputed two more studies and reduced the effect size to 0.283 (SE = 0.1462), $t(15) = 1.94$ ($p = 0.072$) with an exponentiated effect size of 1.33 (95% confidence interval of 0.972 to 1.81). Most of our tests for the third hypothesis clearly supported Bernard's thesis, even though some of the results, limited by lower sample sizes, were of borderline significance even when the effect sizes were of a medium-to-large size, well above an SESOI of 0.15.

6. Discussion

A narrative review of the literature assessed here would have found divided and contradictory evidence, with 17 effect sizes in favor of Bernard's thesis and 19 against (regardless of statistical significance), an almost even split, a nearly ideal situation for meta-analysis, which Bartos, Maier, Shanks, Stanley, Sladekova, and Wagenmakers assert is "widely regarded as the best way to combine and summarize seemingly conflicting evidence across a set of primary studies" (Bartos et al. 2023, p. 2).

Hypothesis 1. Even though some meta-analyses may overestimate effect sizes (Bartos et al. 2023, p. 2; Ferguson et al. 2025), in our case, even if our analyses overestimated a negative effect size, the real effect size being closer to zero, the results of our analyses for hypothesis one would nevertheless *not* support Bernard's thesis, which would have anticipated a positive (>0), and presumably statistically significant, effect size. Meta-analyses that found otherwise (Whisman et al. 2025, $g = -0.17$, $p < 0.001$) benefitted from a large number of cases (greater statistical power) along with the use of the exact same instrument to measure marital satisfaction. With all cases responding to the same instrument with the same response categories (while the most common response categories for the Kansas Marital Satisfaction Scale range from 1 to 7, as used in Whisman et al. 2025, others have used 1–4, 1–5, or 1–9), it is much easier to compare percentages of husbands and wives responding to the highest and lowest levels of satisfaction.

Hypothesis 2. Here, the null hypothesis could not be rejected, as variances in SAH did not differ significantly by gender. It is possible that prior to 1951, social norms were so supportive of a traditional, instrumentally oriented marriage that both husbands and wives

were accepting of less intimate marital conditions. The 1929–1950 cohort of marriages included those of whom many had survived the Great Depression and/or World War 2; merely surviving those disasters may have been interpreted as family success by both husbands and wives. Because fewer wives were working outside the home for pay, many marriages were likely more traditional in orientation and stricter gender roles between husbands and wives were probably more normative, more socially acceptable.

Hypotheses 3 and 4. However, our results for Hypotheses 3 and 4 tended to support Bernard's thesis as well as the results of Whisman and Balzert (2024) and Whisman et al. (2025), in that (a) at very low levels of marital satisfaction, women tended to outnumber men in terms of percentages, or (b) when there were major differences in marital adjustment between spouses, the wife was more likely to be the spouse reporting a lower marital adjustment score. The meta-analytic results featured a larger effect size for the within-couple data than for the between-couple data, although only in four cases were the results not significant. What is causing these differing results? Many studies have found that when a woman is not happy with her relationship, she is more likely to initiate leaving or divorce, whether married (Kalmijn and Poortman 2006; Parker et al. 2022; Zeiss et al. 1980) or not (Hill et al. 1981, 1976; Rubin et al. 1981), which seems especially true if she can support herself financially (Sayer et al. 2011).

Women, whether married or not, are probably more sensitized to relationship deterioration and less willing to accept the painful risks of continuing in such situations, leading them to acknowledge the dissatisfactions of their relationships earlier and thus report some of the lowest levels of marital satisfaction or adjustment compared to their male partners/husbands or compared to other happier couples. For example, married women are more likely to initiate divorce or separation than married men (Amato and Previti 2003; Joyner et al. 2017, p. 2370). Within some couples, some husbands seem relatively clueless about the relative differences in marital adjustment. Gender differences with respect to lower levels of marital adjustment may well reflect higher levels of violence against women, unfair distribution of domestic chores, and many other forms of discrimination and mistreatment experienced by women (Whisman and Balzert 2024, p. 3). It may also be that the marriage gradient—women of higher value tend to remain single while men of lower value are more prone to marry—is partly responsible for these results (i.e., lower-value men frustrate their higher-value wives, leading to greater unhappiness for the latter).

At the same time, the consistency of results for Hypotheses 3 and 4 over nearly one hundred years of marriage in the USA is remarkable in our opinion, making those early (1929–1950) results almost prophetic. After all, those results have remained despite large changes in wives' outside employment, changing divorce rates, changing ages at first marriage, more frequent remarriages, dramatic increases in the use of technology within households (television, cell phones, computers, etc.), cycles of boom and bust, etc.

Consistent with that idea, for the six studies that included divorced spouses in their samples, the results, though not strong, did lean in the expected direction of Bernard's thesis (g = 0.038) compared to the studies without divorced spouses (g = −0.049). For the five effect sizes (including Hamilton 1929) that included spouses who were not necessarily partnered with each other, the results also leaned in Bernard's direction (g = 0.103, $p = 0.075$) compared to the other 31 effect sizes that involved couples married to each other (g = −0.06, $p = 0.02$). On the other hand, even when only the 17 studies with results favoring Bernard's thesis were analyzed, the effect size (g = 0.11, $p = 0.002$) remained lower than our SESOI (g = 0.15). Thus, even under the most ideal circumstances, support for Bernard's thesis for hypothesis 1 was small, even though Funder and Ozer (2019) might argue that even that small effect (>0.10) could have important implications for society at large. However, the results of 15 effect sizes from our higher-quality studies favored a rejection of Bernard's

thesis (g = −0.07, p = 0.024) compared to 21 effect sizes from our lower-quality studies (g = 0.03, p = 0.454). Perhaps the best way to summarize our conflicting results would be that, on average, husbands and wives are equally happy, but women more quickly (than their male partners) recognize when their relationship has been deteriorating and thus report lower satisfaction scores than their partners at the lowest levels of marital adjustment. Thus, the toxic aspects of a "his" and "hers" marriage would seem to occur mainly for the most unhappy of wives and may reflect that, on average, a woman recognizes her relationship as problematic some time before her male partner.

Limitations. Our results should be considered in view of several limitations of our research. Some might question our cut-off at the year 1950. One reason for this is our attempt to avoid biasing our results against Bernard's thesis. Karlsson's (1951, pp. 100–1) study of marital adjustment among 90 Swedish husbands (134.80, SD = 11.40) and 91 Swedish wives (153.60, SD = 12.73) yielded an effect size of −1.56 (t_{179} = 10.46; $p < 0.0001$); the difference in Karlsson's standard deviations was not significant (p = 0.15, one-tailed), with F(90, 89) = 1.25. The ratio of scores below 110 was 1:7, with husbands being more likely to have such low scores. Inclusion of such data in our meta-analysis might have shifted our results even further against Bernard's thesis for all four of our hypotheses. Given our standpoint of married men vs. women (married or not), including such a strong effect-sized study against Bernard's thesis (for all four of our hypotheses) might have seemed to confirm any concerns of others regarding our biases, including confirmation bias (Schumm 2021). Thus, restricting our studies to 1950 or earlier avoided the bias that might have been associated with the results of Karlsson's (1951) research in Sweden (also, see Locke and Karlsson 1952), which it could be argued would also include bias from having been conducted outside the USA. The fact that our 36 studies were almost evenly divided between pro-Bernard and anti-Bernard results would also seem to suggest less bias on our part. Our marital studies included a great variety of measures of adjustment, happiness, and satisfaction, rather than relying upon only one measure (Whisman et al. 2025).

Because many of the studies in our analysis did not report much demographic information, moderation effects of specific demographic variables could not be determined, unlike meta-analyses of more recent studies that often report a fuller set of demographics. Many of the studies here did not report the periods over which their data were collected, unlike many more recent studies. A further limitation of our study is that almost all of the studies we used were published either during the Great Depression, World War 2, or in the post-war period, challenging times for U.S. society in general. Divorce rates were lower during the Great Depression but spiked in 1946 right after World War 2. Furthermore, a lack of social psychological predictor variables denied us a chance to investigate how such factors might be differentially associated with gender or gender's relationship with SAH outcomes, as requested by multiple reviewers. Most of the studies in our meta-analyses involved samples of younger, more educated, and higher socioeconomic status than typical for that time period (see Table 1), although some studies did not report much in the way of demographics. One might expect higher levels of marital adjustment among young, more educated, and higher SES couples (Anderson et al. 1983; Veroff and Feld 1970). Certainly our results are limited by a relative absence of people of color as well as those who were elderly, less educated, and of lower socioeconomic status, which limits comparisons to meta-analyses of data from more recent decades.

Other limitations include the fact that none of the 36 studies measured or controlled for marital social desirability (Bergen and Labonte 2020; Blake et al. 2006; Bobbio and Manganelli 2011; Broderick 1971, 1974; Edmonds 1967; Kim et al. 2026; King 2022; King and Bruner 2000; Lanz et al. 2022; Moorman and Podsakoff 1992; Miller 2021; Olson 1969; Paulhus 1991, 2002; Perinelli and Gremigni 2016; Reynolds 1982; Spanier and Cole 1976;

Steenkamp et al. 2010; Strahan and Gerbasi 1972; Vesely and Klockner 2020; Schumm 2015), or dealt with marital satisfaction or adjustment in collectivist cultures (Kazim and Rafique 2021) or other cultures outside of the USA (Dobrowolska et al. 2020). Comparing women and men who both reported low levels of marital social desirability would probably yield results more favorable to Bernard's thesis because of what we found for Hypotheses 3 and 4 (the very dissatisfied spouses would represent a higher percentage of a sample if some spouses reporting high levels of marital happiness but also reporting high levels of marital social desirability were to be excluded from analysis). This important limitation in terms of marital social desirability also applies to other similar and recent meta-analyses (Buhler et al. 2021; Jackson et al. 2014; Whisman and Balzert 2024; Whisman et al. 2025). However, there are several ways to try to control for marital social desirability, and these might lead to different conclusions with regard to Bernard's hypotheses (Kim et al. 2026). In general, higher-quality research tended to reject Bernard's thesis in terms of Hypotheses 1 and 2, an indication that the quality of research should be considered in future tests of Bernard's thesis, though better measures of such quality should be more easily constructed with more recent research. Another limitation is that this research was begun before meta-analysis was widely known and ours has its limitations when compared to ideal approaches (Ioannidis 2016).

A final limitation is that, to the best of our knowledge, no one has performed a meta-analysis of marital adjustment scores for studies published between 1951 and 1969, a period when Bernard's publications and popular books might have sensitized women to their relationship predicaments to a greater extent, and therefore meta-analysis results during those decades might prove more favorable for gender differences along the lines of our four hypotheses. Blumel's (1992) chapter would be a good starting point for such research.

7. Implications

In terms of future research, studies should obtain data from both spouses and measure and control for marital social desirability. With data from both spouses, it can be determined if wives and husbands as pairs are reporting large discrepancies in marital adjustment, happiness, or satisfaction. Controlling for marital social desirability may permit a more valid test of Bernard's thesis by reducing the effects of marital social desirability response bias on statistical outcomes. There may be larger data sets available that have both within-couple data and measures of marital social desirability that would meet these concerns, even though they probably date to later than 1929 to 1950. There are other implications for researchers: everyone may partake in confirmation bias (Schumm 2021) and see the world as we would prefer (Marks 2012). Future research should include a wider variety of studies, not only from the USA, but from a wide variety of other nations and cultures, a suggestion taken by Whisman et al. (2025), which found strong support for all three of Bernard's hypotheses. Effect sizes should be assessed in addition to levels of statistical significance, given that future research may be more likely to include much larger samples, for which trivial effect sizes, even in meta-analyses, might attain statistical significance. Issues of common methods variance need further investigation (Podsakoff et al. 2024).

Most of the earliest studies on marital adjustment featured results that favored Bernard's thesis, including Bernard's own 1933 research. Ioannidis (2005) has argued that most of this early research is incorrect for a variety of reasons, including less available research literature, weaker theory, poor measurement, and less complex analysis. Here, too, Ioannidis appears to have been correct, for the most part. In addition, Bernard's own life involved a marital breakup coming out of a dual relationship (student/professor), a type of relationship that often involves higher risks due to the power differentials. Regardless, her stance that every marriage is susceptible to "his" and "hers" different perspectives is

no doubt true, even if those differences do not always lead to lower marital satisfaction or poorer adjustment for the wife. The point is that we should be careful to not "buy" into early research in its totality, nor should we be eager to dismiss it in its entirety either. Even if 80% may be unfounded, the other 20% might have some validity, no matter how much we might like to dismiss it all, or vice versa. Humility is increasingly being recognized as an important aspect of good basic (Bland 2025; Hannah 2025; Jensen 2025; Krumrei-Mancuso et al. 2025; O'Connor et al. 2025; Schumm 2021; Schumm and Crawford 2023; Van Tongeren et al. 2019; Van Tongeren 2024; Wolgemuth et al. 2023) and applied science (McMillan 2023; Michalec et al. 2024; Naumova 2023).

Depending on the validity of Bernard's thesis, there might be a variety of implications for women and/or husbands. At the surface level, if marriage really is bad for women, perhaps women ought to cohabit contingently rather than marry permanently, or at least delay marriage and childbearing. If parenting stress is too much for women or men, perhaps fewer children should be conceived or born. Perhaps women should not marry men unsuitable from a feminist perspective, i.e., men who believe in more traditional sex roles or those who believe that a husband's role is to maintain authority over his wife. Even if marital satisfaction scores do not differ much as a function of gender, perhaps women who find themselves in a marriage in which their happiness is much lower than their husband's should consider either change, therapy, or leaving that situation. If large numbers of women were to take Bernard's thesis as correct, marriage rates and birth rates might fall while divorce rates might increase over time, with substantial implications for society. However, here our results for hypotheses one and two indicate that average levels of marital satisfaction and even variation in marital satisfaction are similar for husbands and wives, even though Whisman et al.'s (2025) results differ. Our results were more supportive of hypothesis 3 and 4, as were Whisman et al.'s, but mainly applied to very unhappy wives (<6%), a small percentage within our samples.

8. Conclusions

Although this analysis was based on gender and marital adjustment research published between 1925 and 1950, it yielded similar results to those of more recent meta-analyses in terms of gender differences in mean scores between husbands and wives (Jackson et al. 2014; Buhler et al. 2021; but not similar to the results found by Whisman and Balzert (2024) and Whisman et al. (2025)) and in terms of differences in variances (Whisman and Balzert 2024; Whisman et al. 2025). However, as noted by Whisman et al. (2025), "a combination of higher central tendency in one group [husbands] and greater variance in the other group [wives] indicates that the second group [wives] will be overrepresented in the lower tail of the distribution [of marital satisfaction scores]" (p. 4), which may remain correct even when the scores of the two groups are approximately equal. In general, Bernard's thesis that women are less satisfied with marriage than men was not supported in our study, with mixed results in these other meta-analyses, with effect sizes from near zero to as much as −0.17, which is remarkable given the societal changes in the past one hundred years. These conclusions are limited, though, by an absence of controls for marital social desirability. However, since women generally recognize difficulties in their relationships before their male partners and are often the first to decide to leave these relationships, we did find that at the lowest satisfaction levels of marital adjustment, or when there are major differences in marital adjustment within couples, women were represented twice as often as men, a result that supports Bernard's thesis and agrees with the findings of Whisman and Balzert (2024) and Whisman et al. (2025). This situation could easily lead to greater variance in wives' marital satisfaction scores than husbands', especially with larger samples than ours. Thus, there is mixed evidence for Bernard's first two hypotheses

and consistent evidence for her third and fourth hypotheses. Our research confirms that even the earliest available data on marital adjustment in the United States support at least part of Bernard's hypotheses about gender differences in marriage, a remarkably consistent finding for nearly 100 years of research on heterosexual marriage.

Author Contributions: Conceptualization, W.R.S., S.R.B. and A.P.J.; methodology, W.R.S., S.R.B. and A.P.J.; formal analysis, W.R.S.; investigation, W.R.S., S.R.B. and A.P.J.; resources, S.R.B.; data curation, W.R.S.; writing—original draft preparation, W.R.S.; writing—review and editing, W.R.S. and S.R.B.; project administration, W.R.S. and S.R.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Because no human subjects were solicited in this report, being based on secondary data, human subject approval was not needed.

Informed Consent Statement: Not applicable.

Data Availability Statement: The raw data used in this report are contained in the tables. Upon request, an SPSS file will be supplied that contains the data file used here.

Conflicts of Interest: The authors declare no conflicts of interest.

Notes

[1] Funder and Ozer (2019) suggested Cohen's d = 0.10. Anvari and Lakens (2021) found that effect sizes had to be at least 0.11 to be detectable subjectively. Ferguson and Heene (2021), Ferguson (2013), Ferguson et al. (2022), and Ferguson and Smith (2024) have argued that d = 0.20 (r = 0.10) "should be regarded as basement cut-off under which an effect should not be interpreted as hypothesis supportive" (p. 624) while "exceeding r = 0.10 or even 0.20 is not a guarantee that an effect is "real"" (p. 624) rather than mere statistical noise. More specifically, Ferguson and Smith (2024) accepted d = 0.20, or perhaps 0.21 (Ferguson 2025), as an "evidentiary threadhold" (p. 201). Brydges (2019) argues for a Hedges' g of 0.17 or higher, below which effects in psychosocial research should not be interpreted as meaningful (which would require a sample size of 538 to have statistical power of 0.809). Other scholars prefer a much stricter standard of d or g > 0.20 (Larzelere et al. 2024), while Seok and Kim (2024) prefer 0.15. Acceptance of very small effect sizes may lead to unjustified belief in "our beloved theories" (Heene and Ferguson 2017, p. 35). One problem is that if one were to accept an effect size below 0.10 as meaningful, it might make the falsification of a hypothesis impossible, which may turn science in pseudoscience (Heene and Ferguson 2017, p. 35). Unfortunately, there are many examples in meta-analysis in which very small effect sizes have led to (unjustified) claims of needed changes in public policy and ad hominem attacks on those who have argued against such use of results featuring very small effect sizes (Ferguson et al. 2025; Ferguson 2015). In particular with respect to meta-analyses, Ferguson (2013) argues that "Publication bias, the tendency to publish statistically significant findings at the expense of null studies, can result in spuriously high effect sizes in meta-analysis" (p. 23), an effect that may be stronger if the literature features extensive p-hacking (Friese and Frankenbach 2020). Because of our human tendencies to adopt theories that reflect how we wish the world were rather than how it actually is (Ferguson and Heene 2012, p. 559), it seems that meta-analytic results seldom change minds (p. 558), resulting in what Ferguson and Heene call "undead theory" (p. 559), a theory that survives in spite of extensive contrary evidence. With respect to meta-analysis, part of the problem is that "most meta-analyses are statistically significant no matter how trivial the average effect size" (p. 558), even though even trivial effects may be magnified by an absence of published null findings in the literature.

[2] Estimating missing standard deviations: Several of the studies did not report standard deviations but reported critical ratios. Since CR = difference in means divided by the standard error of the mean differences, it was possible to determine the standard error of the mean differences (Schmid 1946). Using a *t*-test calculator and assuming that the standard deviations were equal, it was possible to identify a standard deviation that was close to the average of the standard deviations of the husbands and wives, which allowed us to calculate an approximate effect size. In some situations, CRs were reported for husbands alone and for wives alone, allowing us to estimate standard deviations for husband and wife scores separately. Random-effects modeling was primarily used in our meta-analyses since fixed-effects modeling would assume that gender differences would be identical throughout many different groups, which we estimated was unlikely. Borenstein et al. (2021, p. 77) recommended using random-effects modeling for most research applications when studies are conducted by independent investigators and at least ten studies are available for analysis; likewise, Mikolajewicz and Komarova (2019, p. 9) also recommended random-effects modeling. At least three of four recent meta-analyses (Buhler et al. 2021; Jackson et al. 2014; Whisman and Balzert 2024; Whisman et al. 2025) that we are using for comparison purposes used random-effects modeling, so it mades sense to retain the same analytic

approach. Even so, we used some fixed-effects modeling for comparison purposes. As noted by Borenstein et al. (2021, p. 79), "Study weights are more balanced under the random-effects model than under the fixed-effects model. Large studies are assigned less relative weight and small studies are assigned more relative weight as compared with the fixed-effect model". Confidence intervals will be wider with random-effects models than with fixed-effects models (i.e., dispersion intervals will be wider than confidence intervals). Since we expected standard deviations to be different as part of Bernard's hypothesis (i.e., if we used the wife's SD, then—on average if Bernard's thesis was correct and wives had larger SDs—there would be bias against significance for differences in mean scores on marital adjustment; if we used the husband's SD, then we would be omitting an important aspect of Bernard's thesis and probably biasing results in terms of statistical significance, whether her thesis was correct or not). Effect sizes were adjusted for Hedges' g by the truncated Knapp–Hartung method for standard error adjustment; Hedges' g was used instead of Cohen's d to measure effect size because Cohen's d tends to overestimate effect sizes in small samples, and small samples comprised most of our 36 studies. The restricted maximum likelihood estimator (REML) was used. We also varied our analyses by not using standard error adjustment and/or by using the DerSimonian and Laird method for comparison purposes to see if changes in methodology led to any changes in apparent results. Regardless of the use of fixed-effect or random-effects models and differing levels of heterogeneity in the latter, with 25 or more studies included in a meta-analysis, levels of power for detecting an effect size of 0.20 should approach or exceed 90% (Borenstein et al. 2021, pp. 307–9). Mikolajewicz and Komarova (2019, p. 13) reported that N < 10 might be too small and N > 50 might be too large for meta-analyses, so our meta-analysis, with 17 studies with 36 effect sizes, falls within the appropriate range. Other recent social science meta-analyses have assessed different numbers of studies, ranging from as low as 4 to as high as 73, with an average of 33.25 (Bresin et al. 2023; de Lange et al. 2022; Goldbach et al. 2014; Durrbaum and Sattler 2020). As an alternative to meta-analysis, we also computed paired-sample *t*-tests, treating the 36 effect sizes as paired husband and wife (group) scores. The average marital adjustment scores for husbands (93.48, SD = 100.23)) and wives (93.61, SD = 100.05) did not differ significantly (Hedges' g = −0.042, $p = 0.799$, two-tailed). For the 17 individual studies with one effect size each, selecting results that favored Bernard's thesis, our paired-sample *t*-test yielded means of 86.79 (SD = 68.16) for husbands and 85.33 (SD = 67.07) for wives, with $t(16) = 2.22$, $p = 0.042$, Hedges' g = 0.512, a result favoring Bernard's thesis. Had that data been analyzed with an independent-sample *t*-test, Hedges' g = 0.022, a trivial effect. Repeating the same analysis but selecting results that did not favor Bernard's thesis, our paired-sample *t*-test yielded means of 68.78 (SD = 64.23) for husbands and 69.46 (SD = 64.18) for wives, with $t(16) = -0.835$ ($p = 0.416$), Hedges' g = 0.268. With an independent-sample *t*-test, the resulting effect size would have been g = −0.011, again a trivial effect.

[3] Our paired-sample *t*-test for all 36 effect sizes, contrasting standard deviations for husbands and wives, obtained scores of 17.60 (SD = 18.37) for husbands and 17.26 (SD = 18.27) for wives, a non-significant result ($p = 0.563$, two-tailed), with an effect size of −0.097 (d) or −0.095 (g). We also performed the same paired-sample *t*-test for the 32 cases of unequal standard deviations, obtaining scores of 18.77 (SD = 19.12) for husbands and 18.39 (SD = 19.07) for wives, a non-significant difference, with $t(31) = 0.622$ ($p = 0.539$, two-tailed). For our paired-sample *t*-test with those 17 studies, our results were 15.33 (SD = 10.92) for husbands and 15.64 (SD = 11.69) for wives, with $t(16) = -0.628$, $p = 0.539$, although Hedges' g = −0.145.

[4] As another approach, using percentages of spouses below cut-off values, we found, for the 18 effect sizes, the average such percentage was 5.46 (SD = 4.73, median = 4.19) for wives and 3.63 (SD = 3.62, median = 2.56) for husbands, with $t(17) = 2.85$ ($p = 0.011$, two-tailed, Cohen's d = 0.672, Hedges' g = 0.642), a significant difference, favoring Bernard's thesis.

References

Adams, Clifford R. 1946. The prediction of adjustment in marriage. *Educational and Psychological Measurement* 6: 185–93. [CrossRef]

Adamsons, K., A. L. Few-Demo, C. M. Proulx, and K. Roy. 2022. *Sourcebook of Family Theories and Methodologies: A Dynamic Approach*. Cham: Springer Nature.

Amato, Paul R., Alan Booth, David R. Johnson, and Stacy J. Rogers. 2007. *Alone Together: How Marriage in America Is Changing*. Cambridge: Harvard University Press.

Amato, Paul R., and Denise Previti. 2003. People's reasons for divorcing: Gender, social class, the life course, and adjustment. *Journal of Family Issues* 24: 602–26. [CrossRef]

Amato, Paul R., David R. Johnson, Alan Booth, and Stacy J. Rogers. 2003. Continuity and change in marital quality between 1980 and 2000. *Journal of Marriage and Family* 65: 1–22. [CrossRef]

Anderson, Stephen A., Candy S. Russell, and Walter R. Schumm. 1983. Perceived marital quality and family life-cycle categories: A further analysis. *Journal of Marriage and the Family* 45: 127–39. [CrossRef]

Anvari, Farid, and Daniel Lakens. 2021. Using anchor-based methods to determine the smallest effect size of interest. *Journal of Experimental Social Psychology* 96: 104159. [CrossRef]

Barker, Timothy H., Celina B. Migliavaca, Cinara Stein, Veronica Colpani, Maicon Falavigna, Edoardo Aromataris, and Zachary Munn. 2021. Conducting a proportional meta-analysis in different types of systematic reviews: A guide for synthesisers of evidence. *Medical Research Methodology* 21: 189. [CrossRef] [PubMed]

Bartos, Frantisek, Maximilian Maier, David R. Shanks, T. D. Stanley, Martina Sladekova, and Eric Jan Wagenmakers. 2023. Meta-analyses in psychology often overestimate evidence for and size of effects. *Royal Society Open Science* 10: 230224. [CrossRef] [PubMed]

Bergen, Nicole, and Ronald Labonte. 2020. "Everything is perfect and we have no problems": Detecting and limiting social desirability bias in qualitative research. *Qualitative Health Research* 30: 783–92. [PubMed]

Bernard, Jessie. 1933. The distribution of success in marriage. *American Journal of Sociology* 39: 194–203. [CrossRef]

Bernard, Jessie. 1934. Factors in the distribution of success in marriage. *American Journal of Sociology* 40: 49–60. [CrossRef]

Bernard, Jessie. 1972. *The Future of Marriage*. New York: World Publishing Company.

Bernard, Jessie., ed. 1975. *Women, Wives, and Mothers: Values and Options*. Chicago: Aldine Press.

Bernard, Jessie. 1976. Change and stability in sex-role norms and behavior. *Journal of Social Issues* 32: 207–24. [CrossRef]

Bernard, Jessie. 1981a. Facing the future. *Society* 18: 53–59. [CrossRef]

Bernard, Jessie. 1981b. The good-provider role: Its rise and fall. *American Psychologist* 36: 1–12. [CrossRef]

Bernard, Jessie. 1982. The Future of Marriage, *1982 edition*. New Haven: Yale University Press.

Bernard, Jessie. 1983. The family facing the future. In *Family, Self, and Society: Emerging Issues, Alternatives, and Interventions*. Edited by Douglas Gutknecht. New York: University Press of America, pp. 207–25.

Bernard, Jessie. 1998. The two marriages. In *Families in the US: Kinship and Domestic Policies*. Edited by Karen V. Hansen and Anita Ilta Garey. Philadelphia: Temple University Press, pp. 449–57.

Bernard, Jessie. 2002. The husband's marriage and the wife's marriage. In *Gender: A Sociological Reader*. Edited by Stevi Jackson and Sue Scott. London and New York: Routledge, pp. 207–19.

Blake, Brian F., Jillian Valdiseeri, Kimberly A. Neuendorf, and Jacqueline Nemeth. 2006. Validity of the SDS-17 measure of social desirability in the American context. *Personality and Individual Differences* 40: 1625–36. [CrossRef]

Bland, Steven. 2025. Intellectual humility and humbling environments. *Review of Philosophy and Psychology* 16: 653–74.

Blumel, S. R. 1992. Explaining marital success and failure. In *Family Research: A Sixty-Year Review, 1930–1990*. Edited by Stephen J. Bahr. New York: Lexington Books, vol. 2, pp. 1–77.

Bobbio, Andrea, and Anna M. Manganelli. 2011. Measuring social desirability responding: A short version of Paulhus's BIDR 6. *Testing, Psychometrics Methodology in Applied Psychology* 18: 117–35.

Bollman, Stephan R., Walter R. Schumm, Anthony P. Jurich, and Gyung Ja Yoon. 1997. Predicting marital satisfaction from ideal and actual maternal roles. *Psychological Reports* 80: 99–106. [CrossRef]

Bollman, Stephan R., Walter R. Schumm, Margaret A. Bugaighis, and Anthony P. Jurich. 1988. Family members' perceptions of actual and ideal maternal roles as a function of maternal employment. *Perceptual and Motor Skills* 67: 185–86. [CrossRef]

Borenstein, Michael, Larry V. Hedges, Julian P. T. Higgins, and Hannah R. Rothstein. 2021. *Introduction to Meta-Analysis*, 2nd ed. Hoboken: Wiley.

Bresin, Konrad, Julia K. Nicholas, Alexandra L. Cowand, Helena F. Alacha, Alyssa M. Rodriguez, and Dominic J. Parrott. 2023. The effects of sexual and gender minority stress on relationship functioning: A meta-analysis. *Personal Relationships* 30: 1208–31. [CrossRef]

Brkljacic, Tihana, Renata Glavak Tkalic, Lana Lucic, Ines Sucic, and Ljiljana Kaliterna Lipovcan. 2019. A brief scale to measure marital/relationship satisfaction by domains: Metrics, correlates, gender, and marriage/relationship status differences. *Social Research: A Journal of General Social Issues* 28: 647–68.

Broderick, Carlfred B. 1971. Beyond the five conceptual frameworks: A decade of development in family theory. *Journal of Marriage and the Family* 33: 139–59. [CrossRef]

Broderick, Carlfred B. 1974. Editorial. *Journal of Marriage and the Family* 36: 224.

Broman, Clifford L. 2005. Marital quality in Black and White marriages. *Journal of Family Issues* 26: 431–41. [CrossRef]

Brydges, Christopher R. 2019. Effect size guidelines, sample size calculations, and statistical power in gerontology. *Innovation in Aging* 3: igz036. [CrossRef]

Buhler, Janina L., Samantha Krauss, and Ulrich Orth. 2021. Development of relationship satisfaction across the life-span: A systematic review and meta-analysis. *Psychological Bulletin* 147: 1012–53. [PubMed]

Burgess, Ernest W., and Leonard S. Cottrell, Jr. 1936. The prediction of adjustment in marriage. *American Sociological Review* 1: 737–51. [CrossRef]

Burgess, Ernest W., and Leonard S. Cottrell, Jr. 1939. *Predicting Success or Failure in Marriage*. New York: Prentice-Hall.

Burgess, Ernest W., and Paul Wallin. 1944. Predicting adjustment in marriage from adjustment in engagement. *The American Journal of Sociology* 49: 324–30. [CrossRef]

Chang, Li-Wen, Walter R. Schumm, L. Ann Coulson, Stephan R. Bollman, and Anthony P. Jurich. 1994. Dimensionality of brief family interaction and satisfaction scales among couples from eight western and midwestern states. *Psychological Reports* 74: 131–44. [CrossRef]

Chapman, Gary. 2015. *The Five Love Languages: The Secret to Love That Lasts*. Chicago: Northfield Publications. First published 1992.

Chis, Roxana M. 2022. A critical rview of the literature on the relationships between personality variables, parenting and marital satisfaction. *Postmodern Openings* 13: 17–46. [CrossRef]
Copeland, Janette M., Margaret A. Bugaighis, and Walter R. Schumm. 1984. Relationship characteristics of couples married thirty years or more: A four-sample replication. *Lifestyles: A Journal of Changing Patterns* 7: 107–14. [CrossRef]
Corra, Mamadi, Shannon K. Carter, J. Scott Carter, and David Knox. 2009. Trends in marital happiness by gender and race, 1973 to 2006. *Journal of Family Issues* 30: 1379–404. [CrossRef]
Curley, Christine M., Blair T. Johnson, and Emersyn K. Rokos. 2026. Who benefits most from romantic relationships and who bears the burdens? *Behavioral and Brain Sciences* 49: e98. [CrossRef] [PubMed]
de Lange, Jennifer, Laura Baams, van Diana D. Bergen, Henny M. W. Bos, and Roel J. Bosker. 2022. Minority stress and suicidal ideation and suicide attempts among LGBT adolescents and young adults: A meta-analysis. *LGBT Health* 9: 222–37. [CrossRef] [PubMed]
Dobrowolska, Malgorzata, Agata Groyecka-Bernard, Piotr Sorokowski, Ashley K. Randall, Peter Hilpert, Khodabakhsh Ahmadi, Ahmad M. Alghraibeh, Richmond Aryeetey, Anna Bertoni, Karim Bettache, and et al. 2020. Global perspective on marital satisfaction. *Sustainability* 12: 8817. [CrossRef]
Durrbaum, Tina, and Frank A. Sattler. 2020. Minority stress and mental health in lesbian, gay male, and bisexual youths: A meta-analysis. *Journal of LGBT Youth* 17: 298–314.
Edmonds, Vernon H. 1967. Marital conventionalization: Definition and measurement. *Journal of Marriage and the Family* 29: 681–88. [CrossRef]
Ferguson, Christopher J. 2013. In the eye of the beholder: Thin-ideal media affects some, but not most, viewers in a meta-analytic review of body dissatisfaction in women and men. *Psychology of Popular Media Culture* 2: 20–37. [CrossRef]
Ferguson, Christopher J. 2015. Pay no attention to that data behind the curtain: On angry birds, happy children, scholarly squabbles, publication bias, and why betas rule metas. *Perspectives on Psychological Science* 10: 683–91. [CrossRef] [PubMed]
Ferguson, Christopher J. 2025. Do social media experiments prove a link with mental health: A methodological and meta-analytic review. *Psychology of Popular Media* 14: 201–6. [CrossRef]
Ferguson, Christopher J., and Moritz Heene. 2012. A vast graveyard of undead theories: Publication bias and psychological science's aversion to the null. *Perspectives on Psychological Science* 7: 555–61. [CrossRef] [PubMed]
Ferguson, Christopher J., and Moritz Heene. 2021. Providing a lower-bound estimate for psychology's "crud factor": The case of aggression. *Professional Psychology: Research and Practice* 52: 620–26. [CrossRef]
Ferguson, Christopher J., and Sven Smith. 2024. Race, class, and criminal adjudication: Is the US criminal justice system as biased as is often assumed? A meta-analytic review. *Aggression and Violent Behavior* 75: 101905. [CrossRef]
Ferguson, Christopher J., James D. Sauer, Aaron Drummond, Julia Kneer, and Emily Lowe-Calverly. 2022. Does sexualization in video games cause harm in players? A meta-analytic examination. *Computers in Human Behavior* 135: 107341. [CrossRef]
Ferguson, Christopher J., Linda K. Kaye, Dawn Branley-Bell, and Patrick Markey. 2025. There is no evidence that time spent on social media is correlated with adolescent mental health problems: Findings from a meta-analysis. *Professional Psychology: Research and Practice* 56: 73–83. [CrossRef]
Fowers, Blaine J. 1991. His and her marriage: A multivariate study of gender and marital satisfaction. *Sex Roles* 24: 209–21. [CrossRef]
Friese, Malte, and Julius Frankenbach. 2020. p-Hacking and publication bias interact to distort meta-analytic effect size estimates. *Psychological Methods* 25: 456–71. [CrossRef] [PubMed]
Frumkin, Robert M. 1953. The Kirkpatrick Scale of Family Interests as an instrument for the indirect assessment of marital adjustment. *Marriage and Family Living* 15: 35–37. [CrossRef]
Funder, David C., and Daniel J. Ozer. 2019. Evaluating effect size in psychological research: Sense and nonsense. *Advances in Methods and Practices in Psychological Science* 2: 156–68. [CrossRef]
Gager, Constance T., and Laura Sanchez. 2003. Two as one? Couples' perceptions of time spent together, marital quality, and the risk of divorce. *Journal of Family Issues* 24: 21–50. [CrossRef]
Goldbach, Jeremy T., Emily E. Tanner-Smith, Meredith Bagwell, and Shannon Dunlap. 2014. Minority stress and substance use in sexual minority adolescents: A meta-analysis. *Prevention Science* 15: 350–63. [PubMed]
Green, Robert G., Debra Woody, Susan Maxwell, Rachel Mercer, and Susette Williams. 1998. Reliability and validity of the Kansas Marital Satisfaction Scale in a sample of African-American husbands and wives. *Psychological Reports* 82: 255–58. [CrossRef]
Hamilton, Gilbert van Tassel. 1929. *A Research in Marriage*. New York: Albert & Charles Boni.
Hannah, David. 2025. A call for greater humility in management research. *Journal of Management Inquiry* 34: 136–38.
Harper, James M., Bruce G. Schaalje, and Jonathan G. Sandberg. 2000. Daily hassles, intimacy, and marital quality in later life marriages. *The American Journal of Family Therapy* 28: 1–18. [CrossRef]
Heene, Moritz, and Christopher J. Ferguson. 2017. Psychological science's aversion to the null, and why many of the things you think are true, aren't. In *Psychological Science Under Scrutiny: Recent Challenges and Proposed Solutions*. Edited by Scott O. Lilienfeld and Irwin D. Waldman. Hoboken: Wiley, pp. 34–52.

Hicks, Mary W., and Marilyn Platt. 1970. Marital happiness and stability: A review of research in the sixties. *Journal of Marriage and the Family* 31: 553–74. [CrossRef]
Hill, Charles T., Letitia A. Peplau, and Zick Rubin. 1981. Differing perceptions in dating couples: Sex roles vs. alternative explanations. *Psychology of Women Quarterly* 5: 418–34. [CrossRef]
Hill, Charles T., Zick Rubin, and Letitia A. Peplau. 1976. Breakups before marriage: The end of 103 affairs. *Journal of Social Issues* 32: 147–68. [CrossRef]
Hyun, Sangkyu, and Heuikwang Shin. 2010. Korean pastors and their wives' satisfaction and its predicting factors. *Pastoral Psychology* 59: 495–512.
Ince, Sanel, and Erkan Isik. 2022. The mediating role of five love languages between differentiation of self and marital satisfaction. *The American Journal of Family Therapy* 50: 407–23.
Ioannidis, John P. A. 2005. Why most published research findings are false. *PLoS Medicine* 2: e124. [CrossRef] [PubMed]
Ioannidis, John P. A. 2016. The mass production of redundant, misleading, and conflicted systematic reviews and meta-analyses. *The Millbank Quarterly* 94: 485–514. [CrossRef]
Jackson, Jeffrey B., Richard B. Miller, Megan Oka, and Ryan G. Henry. 2014. Gender differences in marital satisfaction. *Journal of Marriage and Family* 76: 105–29. [CrossRef]
Jensen, Todd M. 2025. The power of humility in theory development. *Journal of Family Theory & Review* 17: 189–92.
Joyner, Kara, Wendy Manning, and Ryan Bogle. 2017. Gender and the stability of same-sex and different-sex relationships among young adults. *Demography* 54: 2351–74. [CrossRef] [PubMed]
Kalmijn, Matthijs, and Anne-Rigt Poortman. 2006. His or her divorce? The gendered nature of divorce and its determinants. *European Sociological Review* 22: 201–14. [CrossRef]
Kamp Dush, Claire M., Miles G. Taylor, and Rhiannon A. Kroeger. 2008. Marital happiness and psychological well-being across the life course. *Family Relations* 57: 211–26. [CrossRef]
Karlsson, Georg. 1951. *Adaptability and Communication in Marriage: A Swedish Prediction Study of Marital Satisfaction*. Uppsala: Almqvist and Wiksells.
Kaufman, Gayle, and Frances Goldscheider. 2007. Do men "need" a spouse more than women?: Perceptions of the importance of marriage for men and women. *The Sociological Quarterly* 48: 29–46. [CrossRef]
Kazim, Syed M., and Rafia Rafique. 2021. Predictors of marital satisfaction in individualistic and collectivist cultures: A mini review. *Journal of Research in Psychology* 3: 55–67. [CrossRef]
Kelly, E. Lowell. 1941. Marital compatibility as related to personality traits of husbands and wives as rated by self and spouse. *The Journal of Social Psychology* 13: 193–98. [CrossRef]
Kim, Dongyeon, Hyewon Lee, Dong-Heon Kwak, and Jose Benitez. 2026. Assessing and controlling social desirability bias in self-reported survey research: Common practices, recommendations, and an application in cyberbullying research. *European Journal of Information Systems*. [CrossRef]
King, Bruce M. 2022. The influence of social desirability on sexual behavior surveys: A review. *Archives of Sexual Behavior* 51: 1495–501. [CrossRef] [PubMed]
King, Maryon F., and Gordon C. Bruner. 2000. Social desirability bias: A neglected aspect of validity testing. *Psychology & Marketing* 17: 79–103. [CrossRef]
Kirkpatrick, Clifford. 1937. Community of interest and the measurement of marriage adjustment. *The Family* 18: 133–37. [CrossRef]
Krumrei-Mancuso, Elizabeth J., Philip Parnamets, Steven Bland, Mandi Astola, Aleksandra Cichocka, Jeroen de Ridder, Hugo Mercier, Marco Meyer, Cailin O'Connor, Tenelle Porter, and et al. 2025. Toward an understanding of collective intellectual humility. *Trends in Cognitive Sciences* 29: 15–27. [CrossRef] [PubMed]
Kurdek, Larry A. 1992. Dimensionality of the Dyadic Adjustment Scale: Evidence from heterosexual and homosexual couples. *Journal of Family Psychology* 6: 22–35. [CrossRef]
Kurdek, Larry A. 2005. Gender and marital satisfaction early in marriage: A growth curve approach. *Journal of Marriage and Family* 67: 68–84. [CrossRef]
Landis, Judson T., Thomas Poffenberger, and Shirley Poffenberger. 1950. The effects of first pregnancy upon the sexual adjustment of 212 couples. *American Sociological Review* 15: 766–72. [CrossRef]
Lanz, Lucas, Isabel Thielman, and Fabiola H. Gerpott. 2022. Are social desirability scales desirable? A meta-analytic test of the validity of social desirability scales in the context of prosocial behavior. *Journal of Personality* 90: 203–21. [PubMed]
Larzelere, Robert E., Marjorie L. Gunnoe, Joshua Pritsker, and Cristopher J. Ferguson. 2024. Resolving the contradictory conclusions from three reviews of controlled longitudinal studies of physical punishment: A meta-analysis. *Marriage & Family Review* 60: 395–433. [CrossRef]
Li, Tianyuan, and Helene H. Fung. 2011. The dynamic goal theory of marital satisfaction. *Review of General Psychology* 15: 246–54. [CrossRef]

Locke, Harvey J. 1947. Predicting marital adjustment by comparing a divorced and a happily married group. *American Sociological Review* 12: 187–91. [CrossRef]

Locke, Harvey J., and Georg Karlsson. 1952. Marital adjustment and prediction in Sweden and the U.S. *American Sociological Review* 17: 10–17. [CrossRef]

Locke, Harvey J., and Muriel Mackeprang. 1949. Marital adjustment and the employed wife. *American Journal of Sociology* 54: 536–38. [CrossRef] [PubMed]

Locke, Harvey J., and William Klausner. 1948. Adjustment of the divorced in later marriages. *The Proceedings of the Pacific Sociological Society* 16: 30–32.

Lowe, George D., and Robert R. Smith. 1987. Gender, marital status, and mental well-being: A retest of Bernard's his and her marriage. *Sociological Spectrum* 7: 301–7. [CrossRef]

Mace, David R. 1972. Contemporary issues in marriage. In *Encounter: Love, Marriage, and Family*. Edited by Ruth E. Albrecht and E. Wilbur Bock. Boston: Holbrook, pp. 2–12.

Mace, David R., and Vera Mace. 1974. *We Can Have Better Marriages If We Really Want Them*. Nashville: Abingdon.

Majzoobi, Mohammad R., and Simon Forstmeier. 2022. The relationship between the reminiscence of relationship-defining memories and marital outcomes: A systematic review and meta-analysis. *Journal of Family Theory & Review* 14: 7–27. [CrossRef]

Marks, Loren. 2012. Same-sex parenting and children's outcomes: A closer examination of the American Psychological Association's brief on lesbian and gay parenting. *Social Science Research* 41: 735–51. [CrossRef] [PubMed]

McMillan, John. 2023. Humility. *Journal of Medical Ethics* 49: 227–28. [CrossRef] [PubMed]

Michalec, Barret, Andreas Xyrichis, and Christine Arenson. 2024. "Professional humility": Introducing a new framework to advance interprofessionalism. *Journal of Interprofessional Care* 38: 587–92. [CrossRef] [PubMed]

Mikolajewicz, Nicholas, and Svetlana V. Komarova. 2019. Meta-analytic methodology for basic research: A practical guide. *Frontiers in Physiology* 10: 203. [CrossRef] [PubMed]

Miller, Brian K. 2021. Impact of social desirability and common method variance on two measures of entitlement. *Psychological Reports* 124: 1845–62. [PubMed]

Moorman, Robert H., and Philip M. Podsakoff. 1992. A meta-analytic review and empirical test of the potential confounding effects of social desirability response sets in organizational behaviour research. *Journal of Occupational and Organization Psychology* 65: 131–49. [CrossRef]

Naumova, Elena N. 2023. Intellectual humility in public health training, research, and practice. *Journal of Public Health Policy* 44: 1–5. [CrossRef] [PubMed]

O'Connor, Peter J., Andrew Spark, Yasmin Awan, Jordan Moss, and Jessica Chandra. 2025. Beyond black and white: Reducing dichotomous thinking to foster intellectual humility. *Thinking Skills and Creativity* 57: 101841. [CrossRef]

Olson, David H. 1969. The measurement of family power by self-report and behavioral methods. *Journal of Marriage and the Family* 31: 545–50. [CrossRef]

Ort, Robert S. 1950. A study of role-conflicts as related to happiness in marriage. *Journal of Abnormal and Social Psychology* 45: 691–99. [CrossRef]

Orwin, Robert G. 1983. A fail-safe *N* for effect size in meta-analysis. *Journal of Educational Statistics* 8: 157–59. [CrossRef]

Page, Matthew J., David Moher, Patrick M. Bossuyt, Isabelle Boutron, Tammy C. Hoffman, Cynthia D. Mulrow, Larissa Shamseer, Jennifer M. Tetzlaff, Elie A. Aki, Sue E. Brennan, and et al. 2021. PRISMA 2020 explanation and elaboration: Updated guidance and exemplars for reporting systematic reviews. *BMJ* 372: 1–9.

Parker, Gillian., Kristina M. Durante, Sarah E. Hill, and Martie G. Haselton. 2022. Why women choose divorce: An evolutionary perspective. *Current Opinion in Psychology* 43: 300–6. [CrossRef] [PubMed]

Paulhus, Delroy L. 1991. Measurement and control of response bias. In *Measures of Personality and Social Psychological Attitudes*. Edited by John P. Robinson, Phillip R. Shaver and Lawrence S. Wrightsman. San Diego: Academic Press, pp. 17–59.

Paulhus, Delroy L. 2002. Socially desirable responding: The evolution of a construct. In *The Role of Constructs in Psychological and Educational Measurement*. Edited by Henry I. Braun, Douglas N. Jackson and David E. Wiley. Mahwah: Lawrence Erlbaum, pp. 49–69.

Pence, Cheryl., and Jill D. Sauerheber. 2026. The relationship among and between marital satisfaction, meaning in life, and political orientation. *The Family Journal: Counseling and Therapy for Couples and Families* 34: 112–21.

Perinelli, Enrico., and Paola Gremigni. 2016. Use of social desirability scales in clinical psychology: A systematic review. *Journal of Clinical Psychology* 72: 534–51. [CrossRef] [PubMed]

Podsakoff, Philip M., Nathan P. Podsakoff, Larry J. Williams, Chengquan Huang, and Junhui Yang. 2024. Common method bias: It's bad, it's complex, it's widespread, and it's not easy to fix. *Annual Review of Organizational Psychology and Organizational Behavior* 11: 17–61. [CrossRef]

Reed, Robert B. 1947. Social and psychological factors affecting fertility: VII. The interrelationship of marital adjustment, fertility control, and size of family. *The Milbank Memorial Fund Quarterly* 25: 383–425. [CrossRef]

Reynolds, William M. 1982. Development of reliable and valid short forms of the Marlowe-Crowne Social Desirability Scale. *Journal of Clinical Psychology* 38: 119–25. [CrossRef]
Rhyne, Darla. 1981. Bases of marital satisfaction among men and women. *Journal of Marriage and the Family* 43: 941–55. [CrossRef]
Rice, F. Philip. 1983. *Contemporary Marriage*. Boston: Allyn & Bacon.
Rubin, Zick, Letitia A. Peplau, and Charles T. Hill. 1981. Loving and leaving: Sex differences in romantic attachments. *Sex Roles* 7: 821–35. [CrossRef]
Sayer, Liana C., Paula England, Paul Allison, and Nicole Kangas. 2011. She left, he left: How employment and satisfaction affect men's and women's decisions to leave marriages. *American Journal of Sociology* 116: 1982–2018. [CrossRef]
Schmid, John, Jr. 1946. Calculation of the critical ratio from gross sources. *The Journal of Educational Research* 40: 219–20. [CrossRef]
Schumm, Walter R. 2015. Navigating treacherous waters—One researcher's 40 years of experience with controversial scientific research. *Comprehensive Psychology* 4: 24. [CrossRef]
Schumm, Walter R. 2021. Confirmation bias and methodology in social science: An editorial. *Marriage & Family Review* 57: 285–93. [CrossRef]
Schumm, Walter R. 2026. Sexual and transgender orientations of 62 adolescents adopted by lesbian mothers, gay fathers, and heterosexual parents soon after birth: Associations with parental sexual orientation, adolescent natal gender, parental and child race, and residential locations. *American Journal of Biomedical Science & Research* 30: 1101–10.
Schumm, Walter R., and Benjamin Silliman. 1996. Gender and marital satisfaction: A replication with a sample of spouses from the Christian Church (Disciples of Christ). *Psychological Reports* 79: 496–98. [CrossRef]
Schumm, Walter R., and Duane W. Crawford. 2023. Difficulties with methodology in social science research with controversial issues regarding human sexuality. *The Linacre Quarterly* 90: 194–216. [CrossRef] [PubMed]
Schumm, Walter R., and Stephan R. Bollman. 2026. The history and further development of the Kansas Marital Satisfaction Scale. *American Journal of Biomedical Science & Research* 31: 347–51.
Schumm, Walter R., Anthony P. Jurich, Stephan R. Bollman, and Margaret A. Bugaighis. 1985. His and her marriage revisited. *Journal of Family Issues* 6: 221–27. [CrossRef]
Schumm, Walter R., Connie W. Nichols, Karen L. Schectman, and Cynthia C. Grigsby. 1983. Characteristics of responses to the Kansas Marital Satisfaction Scale by a sample of 84 married mothers. *Psychological Reports* 53: 567–72. [CrossRef]
Schumm, Walter R., Farrell J. Webb, and Stephan R. Bollman. 1998a. Gender and marital satisfaction: Data from the National Survey of Families and Households. *Psychological Reports* 83: 319–27. [CrossRef]
Schumm, Walter R., Gary Resnick, Stephan R. Bollman, and Anthony P. Jurich. 1998b. Gender effects and marital satisfaction: A brief report from a sample of dual military couples from the 1992 Department of Defense Worldwide Survey of Members and Spouses. *Psychological Reports* 82: 161–62. [CrossRef]
Schumm, Walter R., Stephan R. Bollman, and Anthony P. Jurich. 1997. Gender and marital satisfaction: A replication using a seven-point item response version of the Kansas Marital Satisfaction Scale. *Psychological Reports* 81: 1004–6. [CrossRef]
Seok, JjiWoo, and Jung-Dae Kim. 2024. Light therapy for older people with depressive symptoms: Systematic review and meta-analysis. *Journal of Clinical Medicine* 13: 6982. [CrossRef] [PubMed]
Shek, Daniel T. L. 1995. Gender differences in marital quality and well-being in Chinese married adults. *Sex Roles* 32: 699–715. [CrossRef]
Spanier, Graham B., and Charles L. Cole. 1976. Toward clarification and investigation of marital adjustment. *International Journal of Sociology of the Family* 6: 121–46.
Spanier, Graham B., and Robert A. Lewis. 1980. Marital quality: A review of the seventies. *Journal of Marriage and the Family* 42: 96–110. [CrossRef]
Steenkamp, Jan-Benedict E. M, Martijn G. De Jong, and Hans Baumgartner. 2010. Socially desirable response tendencies in survey research. *Journal of Marketing Research* 47: 199–214. [CrossRef]
Stevenson, Betsey, and Justin Wolfers. 2009. The paradox of declining female happiness. *American Economic Journal: Economic Policy* 1: 190–225. [CrossRef]
Strahan, Robert, and Kathleen Carrese Gerbasi. 1972. Short, homogenous versions of the Marlowe-Crowne Social Desirability Scale. *Journal of Clinical Psychology* 28: 191–93. [CrossRef]
Swihart, Judson J. 1978. *How Do You Say 'I Love You'?* Sandycroft: Kingsway.
Taniguchi, Hiromi, and Gayle Kaufman. 2013. Gender role attitudes, troubles talk, and marital satisfaction in Japan. *Journal of Social and Personal Relationships* 31: 975–94. [CrossRef]
Terman, Lewis M. 1938. *Psychological Factors in Marital Happiness*. New York: McGraw-Hill.
Terman, Lewis M. 1950. Prediction data: Predicting marriage failure from test scores. *Marriage and Family Living* 12: 51–54. [CrossRef]
Terman, Lewis M., and Melita H. Oden. 1947. *The Gifted Child Grows Up: Twenty-Five Years' Follow-Up of a Superior Group*. Stanford: Stanford University Press.

Terman, Lewis M., and Paul Buttenwieser. 1935. Personality factors in marital compatibility. *Journal of Social Psychology* 6: 143–71. [CrossRef]

Terman, Lewis M., and Paul Wallin. 1949. The validity of marriage prediction and marital adjustment tests. *American Sociological Review* 14: 497–504. [CrossRef]

Terman, Lewis M., and Winifred B. Johnson. 1939. Methodology and results of recent studies in marital adjustment. *American Sociological Review* 4: 307–24. [CrossRef]

Van Tongeren, Daryl R. 2024. The existential value and challenge of intellectual humility. *Possibility Studies & Society* 2: 214–19. [CrossRef]

Van Tongeren, Daryl R., D0n E. Davis, Joshua N. Hook, and Charlotte.vanOyen Witvliet. 2019. Humility. *Current Directions in Psychological Science* 28: 463–68. [CrossRef]

Veroff, Joseph, and Sheila Feld. 1970. *Marriage and Work in America*. New York: Van Nostrand-Reinhold.

Vesely, Stepan, and Christian A. Klockner. 2020. Social desirability in environmental psychology research: Three meta-analyses. *Frontiers in Psychology* 11: 1396. [CrossRef] [PubMed]

Voss, Raymond P., Jr., Annalise Harrison, Dylan Sczerbowicz, and Michelle Drouin. 2026. Intimacy needs across different relationship types: The role of gender and attachment style. *Sexuality & Culture*.

Whisman, Mark A., and Antonia Balzert. 2024. Gender differences in variability in intimate relationship satisfaction: A secondary analysis and meta-analysis. *Clinical Child and Family Psychology Review* 27: 1167–87. [CrossRef] [PubMed]

Whisman, Mark A., Ashley K. Randall, and Piotr Sorokowski. 2025. Gender differences in variability of intimate relationship satisfaction in a multinational sample. *Journal of Family Psychology* 39: 548–53. [CrossRef] [PubMed]

White, James M., Todd F. Martin, and Kari Adamsons. 2019. *Family Theories: An Introduction*, 5th ed. Thousand Oaks: Sage.

Williams, Edith W. 1938. Factors Associated with Adjustment in Rural Marriages. Doctoral dissertation, Cornell University, Ithaca, NY, USA.

Wolgemuth, J. R., Y. H. Chen, L. Rodriguez-Campos, J. Ferron, E. Kim, R. F. Dedrick, and D. Lamb. 2023. What are we teaching for? Humility and responsibility in social science research. In *Handbook of Teaching and Learning Social Research Methods*. Edited by Melanie Nind. Northhampton: Edward Elgar Publishing, pp. 198–212.

Zeiss, Antonette M., Robert A. Zeiss, and Stephen M. Johnson. 1980. Sex differences in initiation of and adjustment to divorce. *Journal of Divorce* 4: 21–33.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.