

# Male and Female Pronoun Use in U.S. Books Reflects Women's Status, 1900–2008

Jean M. Twenge · W. Keith Campbell · Brittany Gentile

© Springer Science+Business Media, LLC 2012

**Abstract** The status of women in the United States varied considerably during the 20th century, with increases 1900–1945, decreases 1946–1967, and considerable increases after 1968. We examined whether changes in written language, especially the ratio of male to female pronouns, reflected these trends in status in the full text of nearly 1.2 million U.S. books 1900–2008 from the Google Books database. Male pronouns included he, him, his, himself and female pronouns included she, her, hers, and herself. Between 1900 and 1945, 3.5 male pronouns appeared for every female pronoun, increasing to 4.5 male pronouns during the postwar era of the 1950s and early 1960s. After 1968, the ratio dropped precipitously, reaching 2 male pronouns per female pronoun by the 2000s. From 1968 to 2008, the use of male pronouns decreased as female pronouns increased. The gender pronoun ratio was significantly correlated with indicators of U.S. women's status such as educational attainment, labor force participation, and age at first marriage as well as women's assertiveness, a personality trait linked to status. Books used relatively more female pronouns when women's status was high and fewer when it was low. The results suggest that cultural products such as books mirror U.S. women's status and changing trends in gender equality over the generations.

**Keywords** Language use · Status of women · Feminism · Cultural products · Change over time · Generations

---

J. M. Twenge (✉)  
Department of Psychology, San Diego State University,  
5500 Campanile Drive,  
San Diego, CA 92182-4611, USA  
e-mail: jtwenge@mail.sdsu.edu

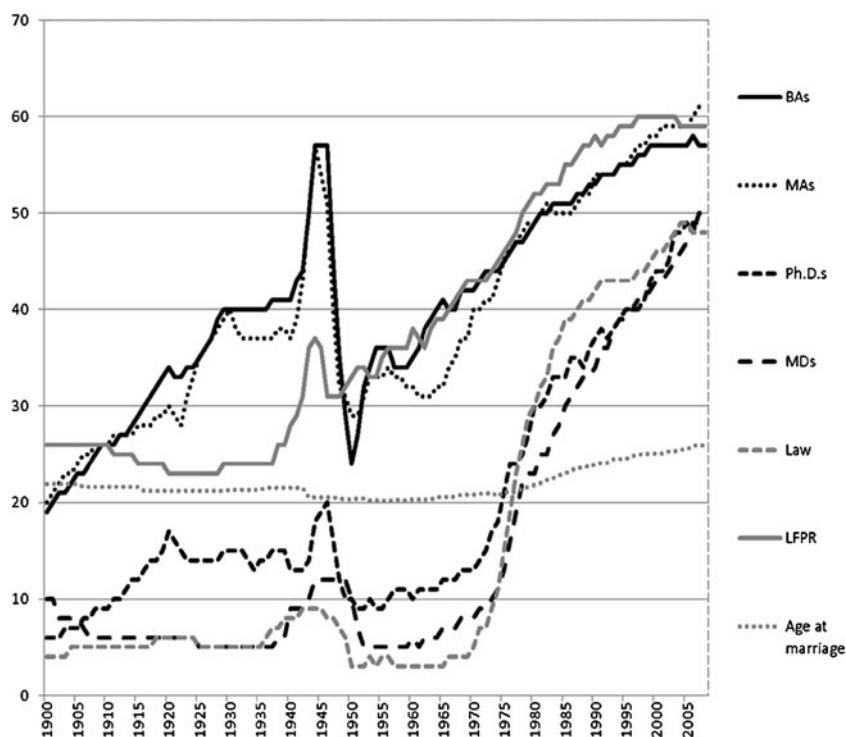
W. K. Campbell · B. Gentile  
University of Georgia,  
Athens, GA, USA

## Introduction

Women's status and roles have changed considerably over the last century in the United States. In U.S. society, status largely comes from education and work (e.g., Eagly 1987). Thus status can be measured quantitatively by indicators such as educational attainment (the percentage of higher education degrees granted to women, including BAs, MAs, Ph.Ds, MDs, and law degrees), women's labor force participation rate (LFPR), and median age at first marriage, which is younger during eras when women focus less on education and work (Stewart and Healy 1989; Twenge 2001; all of the cited studies are on U.S. samples unless otherwise noted). By these measures, U.S. women's status has followed a cubic pattern, with gains before and during World War II, decreases postwar, and increases after the late 1960s (see Fig. 1). Women's assertiveness, a personality trait linked to status (e.g., Diekmann and Eagly 2000), also displays a cubic pattern during the 20th century (Twenge 2001). Qualitative reviews of U.S. women's history also point to a cubic trend in women's status (Chafe 1972; Coontz 2000; Evans 1989; Friedan 1963; Honey 1984). In addition, several studies have documented the rise in attitudes more favorable to gender equality after the late 1960s (e.g., Koenig et al. 2011; Thornton and Young-DeMarco 2001; Twenge 1997; for a review, see Twenge 2006, Chapter 7).

However, it is relatively unknown if these trends in women's status are reflected in the products of U.S. culture – for example, in song lyrics, TV shows, movies, and books. Studying cultural products is one of the best ways to quantify culture at the group level, capturing the general cultural viewpoint (Lamoreaux and Morling 2012; Morling and Lamoreaux 2008). Cultural products reflect the individualism and collectivism of regional cultures (Morling and Lamoreaux 2008). For example, American

**Fig. 1** Indicators of U.S. women's status, 1900–2008. NOTES: 1. BAs, MAs, Ph.D.s, MDs, and Law indicate the percentage of these degrees granted to women. 2. LFPR = Women's labor force participation rate. 3. Age at marriage = Median age at first marriage for women



advertisements focus more on standing out, whereas Korean advertisements emphasize fitting in (Kim and Markus 1999). Cultural products can also be used to examine cultural change within a nation. A recent study found that U.S. popular song lyrics became more self-centered and antisocial between 1980 and 2007 (DeWall et al. 2011). In this study, we examine whether the use of gendered pronouns such as “he” and “she” in the cultural product of U.S. books mirrors women’s status during the 20th and early 21st centuries.

The Mutual Constitution Model (MCM) in cultural psychology (Markus and Kitayama 2010) posits that cultures shape individuals and individuals shape culture, with changes in cultural products often stronger than generational shifts among individuals. Lamoreaux and Morling (2012) contend that cultural products are important for at least three reasons. First, culture includes the context as well as the person, and cultural products capture culture “outside the head.” Second, cultural products are not subject to the biases that plague self-report measures such as reference group and social desirability effects. Third, and perhaps most important, cultural products shape individuals’ ideas of cultural norms and “common sense,” a central source of information about gender roles (Lamoreaux and Morling 2012). People’s behavior is often influenced by their beliefs about what others in their culture believe and do, even if these assumptions are erroneous (e.g., Zou et al. 2009, in a study using participants from the U.S., Poland, and China). Thus cultural products may be one of the main sources from which individuals learn about gender inequality: If men are mentioned much more than women, this suggests that women

are lower in status. The MCM predicts that culture and individuals influence each other in a dynamic system (Markus and Kitayama 2010), so cultural products should reflect women’s status measured at the individual level (such as through educational attainment, labor force participation, and median age at first marriage).

A small number of empirical studies have explored changes in the portrayal of women in U.S. cultural products such as TV commercials (Bretl and Cantor 1988) and magazine articles (Zube 1972). However, practical considerations have limited these studies to a few decades of data on a very small number of cultural products. Fortunately, the recent advent of the Google Books ngram viewer has made it possible to analyze language use over time in the full text of a corpus of 5 million books, 4 % of the books ever published (Michel et al. 2010). The corpus is so large that it would take 80 years for someone to read all of the books for the year 2000 alone (Michel et al. 2010). In the present study, we used the Google Books database to examine changes in women’s status in the U.S. between 1900 and 2008 ( $N = 1.2$  million books) by analyzing trends in the use of gendered pronouns (e.g., he and she) in U.S. books.

Language use in books could reflect cultural change in several ways. First, language use shows the viewpoints of book authors, capturing changes in the values and attitudes of an influential portion of the population. Books may also reflect a market-driven assessment of what consumers want to read. Last, the language in books may reflect the larger body of written and spoken language at a particular time, as authors are likely to use language currently in vogue.

Pronouns are especially useful for examining cultural changes in language (Brown and Gilman 1960; Pennebaker 2011) as they are a well-defined group of words in English that has not changed substantially for at least a century (e.g., Walker 2007). In particular, the use of gendered pronouns may reflect the status of men and women in society. Eras with higher women's status should see an increasing use of female pronouns, and a decreasing use of male pronouns. However, these trends could also be influenced by general trends in the use of pronouns or third person pronouns.

The best indicator of women's status through language might be the ratio of male to female pronouns, demonstrating the relative use of pronouns referring to males and females. In cultures and at times when women are lower in status and fade into the background, the male to female ratio of gendered pronouns should be high. At times when women's status is higher and women are more visible members of society, the ratio should be lower. This may occur for several reasons. First, books may incorporate more female topics or female characters. Second, times with lower status for women may use the universal "he" for all people, while times with more gender equality may alternate the use of "he" and "she" or use the constructions "he/she," "s/he," or "he or she." Attitudes toward gender-inclusive language are correlated with more progressive attitudes toward women's roles (Parks and Robertson 2004), and attitudes toward women became more progressive during the late 20th century (Thornton and Young-DeMarco 2001; Twenge 1997). Thus, cultural change in women's status may be reflected in the use of gendered pronouns in books.

## Hypotheses

Previous research and theory suggests that women's status followed a cubic pattern during the 20th century in the U.S., rising before WWII, declining afterward, and then increasing substantially after about 1968 (Stewart and Healy 1989; Twenge 2001; see Fig. 1).

Our goal in the present research is to test this pattern of change in gender roles in cultural products, specifically the use of gendered pronouns in U.S. books. We thus make three basic predictions:

1. We hypothesize a cubic pattern for the ratio of male to female pronouns in U.S. books with year that follows the cubic pattern of women's status shown in Fig. 1, using the year turning points in women's status identified in Twenge (2001) based on the indicators of women's status: increases 1900 to 1945, decreases 1946 to 1967, and increases 1968 to 2008 (*Hypothesis 1*). This will be tested in two ways. First, we will perform a regression equation that predicts the ratio of gendered pronouns (male pronouns divided by female pronouns)

from year and includes cubic and quadratic terms, to test the overall pattern between 1900 and 2008. We predict that the cubic term will be significant. The second test will examine the linear correlation between year and pronoun ratio during the specific time periods (1900–1945; 1946–1967; 1968–2008). We predict the gender pronoun ratio will decrease (with fewer male pronouns used relative to female pronouns) as women's status increased between 1900 and 1945; to increase with the decrease in women's status from 1946 to 1967; and to decrease with the increase in women's status from 1968 to 2008.

2. We hypothesize that the male to female pronoun ratio will be negatively correlated with indicators of women's status, specifically: educational attainment (the percentage of BAs, MAs, Ph.Ds, MDs, and law degrees granted to women), women's labor force participation rate (LFPR), and median age at first marriage (*Hypothesis 2*). This will be tested by examining the linear relationship between the ratio of gendered pronouns and the status indicators.
3. We hypothesize that the male to female pronoun ratio will be negatively correlated with women's assertiveness over time (*Hypothesis 3*). Assertiveness is a personality trait closely linked to status (e.g., Diekmann and Eagly 2000); thus, these scores provide a view of status as expressed by individual women in their responses to a personality measure (Twenge 2001). This correlation is likely to be smaller, however, as it measures individual personality rather than population-level indicators (Markus and Kitayama 2010).

## Method

We examined the American English corpus from the Google Books database, which includes books published in the United States between 1800 and 2008. The corpus is more reliable after 1900; in addition, previous empirical research on changes in women's roles focused on the 20th century and later, and most indicators of women's status such as higher education degrees are only available since the 20th century. Thus we examined the 1,182,400 books in the database published between 1900 and 2008. Results after 2000 should be interpreted with caution as Google Books was instituted in that year, introducing small changes to the selection of books (Michel et al. 2010).

The corpus contains 4 % of books published since the 1800s. These books were likely not truly randomly selected (Michel et al. 2010); however, we assume these books were not selected in a way dependent on pronoun use frequency that also varied systematically with year. In addition, the Google Books database (found at <http://books.google.com/>

grams) is by far the largest database available of digitized books. Google used 100 sources such as university libraries and publishers to generate a comprehensive catalog of books. The books were digitally scanned and the corpus was winnowed of serial publications, multiple editions, and books with poor print quality, unknown publication dates, or miscoded language (e.g., a book listed in the library catalog as written in English that was not actually in English). Country of publication (in this case, the United States) was determined by 100 bibliographic sources (Michel et al. 2010). If the books are representative of all titles published in the U.S. in 2002 (the most recent statistics available), 87 % are nonfiction and 13 % are fiction (U.S. Bureau of the Census 1925–2011).

The database reports usage frequency by dividing the number of instances of the word in a given year by the total number of words in the corpus in that year, thus correcting for changes in the number of published works and their length. Our unit of analysis was the frequency of the use of a pronoun in a specific year; we then added these frequencies together within pronoun categories (e.g., third person singular female consisted of she, her, hers, and herself). We then tested for changes in those frequencies over time by examining the correlation between year and frequency. Our results thus refer to the annual change in the frequency of the use of pronouns. The gender pronoun ratio was calculated by dividing male pronouns by female pronouns (so that a higher ratio indicates more male pronouns).

We also matched the pronoun data with indicators of women's status for each year. These included educational attainment (the percentage of BAs, MAs, Ph.D.s, MDs, and law degrees granted to women); women's labor force participation rate; and the median age at first marriage for women, obtained from the Statistical Abstract of the United States (U.S. Bureau of the Census 1925–2011). We also included the standardized scores of 25,783 college women on measures of assertiveness between 1931 and 1993, matched by year, from the Twenge (2001) meta-analysis of change over time in assertiveness, a personality trait linked with status (e.g., Diekmann and Eagly 2000; Gilroy et al. 1981). The measures of assertiveness, all commonly used, valid, and reliable, were the dominance scale of the Bernreuter Personality Inventory, the dominance scale of the California Personality Inventory, the dominance scale of the Edwards Personal Preference Schedule, the College Self-Expression Scale, and the Rathus Assertiveness Schedule.

## Results

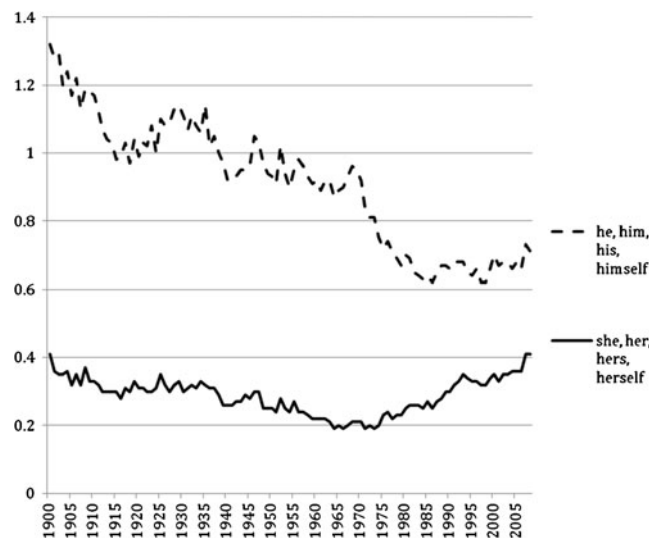
The means by year for male and female pronouns and their ratio are displayed in Figs. 2 and 3.

Hypothesis 1 predicted a cubic function for the use of gendered pronouns across time, with the gender pronoun



**Fig. 2** Changes in the ratio of male to female pronouns in U.S. books, 1900–2008

ratio increasing 1900–1945, decreasing 1946–1967, and increasing 1968–2008. Consistent with this hypothesis, in a regression equation predicting the pronoun ratio from centered year, year squared, and year cubed, the year cubed was significant (see Table 1). However, the quadratic term was also significant. The ratio of male to female pronouns followed the hypothesized pattern in two out of three eras (see Table 2). The ratio increased, with more male pronouns relative to female, in the post WWII era (1946 to 1967) and then decreased markedly after 1968, with female pronouns increasing relative to male pronouns. Thus female pronouns



**Fig. 3** Changes in male and female pronouns in U.S. books, 1900–2008. NOTE: The y-axis represents the percentage of words in books for that year (e.g., .4 = .4 % of words)

**Table 1** Regression equations predicting the male to female pronoun ratio from year, year squared (quadratic), and year cubed (cubic)

	Beta	F	R <sup>2</sup>
Cubic model		95.44***	.72
Year	-.05		
Year squared	-.61***		
Year cubed	-.56***		
Quadratic model		113.47***	.68
Year	-.56***		
Year squared	-.61***		
Bivariate with year	-.56***	48.54***	.31

df=108, the number of years

were used progressively less often in the postwar era (1946 to 1967) when women's status declined or stagnated, and more often after 1968 when women's status rose considerably (see Table 2 and Figs. 1, 2, 3). The gendered pronoun ratio did not change 1900–1945, however (we had predicted that it would decrease in response to women's increased status during that era).

Hypothesis 2 predicted that more female pronouns (relative to male) would be used in eras when women's status was high. Consistent with this hypothesis, the gender pronoun ratio (high numbers = more male pronouns relative to female) was negatively correlated with indicators of women's status including educational attainment, labor force participation, and age at first marriage (see Table 3). Thus, U.S. books used relatively more female pronouns when women earned a higher percentage of higher education degrees, participated in the labor force, and married later.

Hypothesis 3 predicted that more female pronouns (relative to male) would be used when women scored higher in assertiveness, a personality trait linked to status. This hypothesis was confirmed, as U.S. college women scored higher on measures of assertiveness at times when relatively more female pronouns appeared in books (see Table 3).

**Table 2** Correlations between year and gendered third person pronouns, 1900–2008

	1900–1945	1946–1967	1968–2008
Male to female ratio	.00	.90***	-.91***
Third person singular male (he, him, his, himself)	-.73***	-.68***	-.62***
Third person singular female (she, her, hers, herself)	-.70***	-.88***	.97***

1. The male to female ratio is male pronouns divided by female pronouns. Higher numbers indicate more male pronouns

2. \*\*\*= $p < .001$

3. Degrees of freedom are the number of years (1900–1945=46; 1946–1967=22; 1968–2008=41)

**Table 3** Correlations between indicators of women's status and the ratio of male to female pronouns in American books, 1900–2008

	r
% BAs awarded to women	-.64***
% MAs awarded to women	-.75***
% Ph.D.s awarded to women	-.89***
% MDs awarded to women	-.89***
% Law degrees awarded to women	-.91***
Women's labor force participation	-.67***
Median age at first marriage for women	-.91***
Women's assertiveness (1931–1993)	-.30*

1. The male to female ratio is male pronouns divided by female pronouns. Higher numbers indicate more male pronouns

2. \*= $p < .05$  \*\*\*= $p < .001$

3. BAs = bachelor's degrees (college degrees). MAs = master's degrees. Ph.D. = Doctor of Philosophy degrees. MDs = Medical doctor degrees

4. Degrees of freedom are the number of years ( $n=108$ )

## Discussion

In the full text of nearly 1.2 million U.S. books, the use of male and female pronouns reflects changes in women's status during the 20th century. U.S. books used a fairly constant ratio of 3.5 male pronouns for every female pronoun between 1900 and 1945; the ratio steadily increased to 4.5 male pronouns per female pronoun by the mid-1960s. Beginning around 1968, the ratio dropped markedly until, by the 21st century, U.S. books used about 2 male pronouns for every female pronoun. This pattern follows the ups and downs of U.S. women's status over time fairly closely, and the ratio correlates in the expected direction with indicators of women's status such as educational attainment and a later age at first marriage. The only discrepancy was a lack of change in the gender pronoun ratio between 1900 and 1945, when most indicators of women's status increased.

These results show a clear link between cultural products (i.e., pronoun use in books), larger cultural and demographic markers of women's status (e.g., percentage of advanced degrees; age of first marriage), and the personalities of individual women (i.e., assertiveness). This is consistent with the Mutual Constitution Model of culture, which posits that the culture affects individuals just as individuals affect culture (Markus and Kitayama 2010). As would be expected from the model, the correlations with population-level status indicators such as education were higher than those for assertiveness, an individual personality trait that is only a proxy for status. Cultural products, demographic markers, and personality appear to operate together in a systemic way, although we want to be clear that the present data cannot demonstrate causal pathways between the variables. It is

possible that greater gender equality in U.S. culture, reflected in the language in books, led to greater status and assertiveness for women. It is also possible that women gained status, leading to more egalitarian language use in books.

### Limitations

Google Books is the most comprehensive database of the full text of books available. However, it does not include every book ever published. We have assumed that the selection of books is not systematically related to both year and gendered pronoun use. We examined gendered pronoun use in U.S. books only, to mirror our focus on women's status in the U.S. When examining change over time in culture, it is important to limit the analysis to one culture to avoid confounding with regional culture. However, this does limit our results on language use and women's status to the U.S. Future research should explore if cultural products in other nations also reflect trends in women's status over time.

### Conclusions

Gendered pronoun use in U.S. books closely follows women's status during the 20th and early 21st centuries. Authors used relatively more male pronouns when women's status was lower during the postwar era (1946–1967), but this ratio was cut in half as women gained status after 1968. These results are a further indication that women's status has changed considerably over the last century.

**Acknowledgments** We thank Candice Davis, Van Ethan Levy, and Grace Rasmusson for their assistance.

### References

- Bretl, D. J., & Cantor, J. (1988). The portrayal of men and women in U.S. television commercials: A recent content analysis and trends over 15 years. *Sex Roles, 18*, 595–609. doi:10.1007/BF00287963.
- Brown, R., & Gilman, A. (1960). The pronouns of power and solidarity. In T. A. Sebeok (Ed.), *Style in language* (pp. 253–276). Cambridge, MA: MIT Press.
- Chafe, W. H. (1972). *The American woman: Her changing social, economic, and political roles, 1920–1970*. New York: Oxford University Press.
- Coontz, S. (2000). *The way we never were: American families and the nostalgia trap*. New York: Basic Books.
- DeWall, C. N., Pond, R. S., Campbell, W. K., & Twenge, J. M. (2011). Tuning in to psychological change: Linguistic markers of psychological traits and emotions over time in popular U.S. song lyrics. *Psychology of Aesthetics, Creativity, and the Arts, 5*, 200–207. doi:10.1037/a0023195.
- Diekmann, A. B., & Eagly, A. H. (2000). Stereotypes as dynamic constructs: Women and men of the past, present, and future. *Personality and Social Psychology Bulletin, 26*, 1171–1188. doi:10.1177/0146167200262001.
- Eagly, A. H. (1987). *Sex differences in social behavior: A social-role interpretation*. Hillsdale, NJ: Erlbaum.
- Evans, S. M. (1989). *Born for liberty: A history of women in America*. New York: Free Press.
- Friedan, B. (1963). *The feminine mystique*. New York: W. W. Norton.
- Gilroy, F. D., Talierco, T. M., & Steinbacher, R. (1981). Impact of maternal employment on daughters' sex-role orientation and fear of success. *Psychological Reports, 49*, 963–968. doi:10.2466/pr0.1981.49.3.963.
- Honey, M. (1984). *Creating Rosie the Riveter: Class, gender, and propaganda during World War II*. Amherst: University of Massachusetts Press.
- Kim, H., & Markus, H. R. (1999). Deviance or uniqueness, harmony or conformity? A cultural analysis. *Journal of Personality and Social Psychology, 77*, 785–800. doi:10.1037/0022-3514.77.4.785.
- Koenig, A. M., Eagly, A. H., Mitchell, A. A., & Ristikari, T. (2011). Are leader stereotypes masculine? A meta-analysis of three research paradigms. *Psychological Bulletin, 137*, 616–642. doi:10.1037/a0023557.
- Lamoreaux, M., & Morling, B. (2012). Outside the head and outside individualism-collectivism: Further meta-analyses of cultural products. *Journal of Cross-Cultural Psychology, 43*, 299–327. doi:10.1177/00220221110385234.
- Markus, H. R., & Kitayama, S. (2010). Cultures and selves: A cycle of mutual constitution. *Perspectives on Psychological Science, 5*, 420–430. doi:10.1177/1745691610375557.
- Michel, J. B., Kui, S. Y., Presser, A. A., et al. (2010). Quantitative analysis of millions of digitized books. *Science, 331*, 176–182. doi:10.1126/science.1199644.
- Morling, B., & Lamoreaux, M. (2008). Measuring culture outside the head: A meta-analysis of individualism-collectivism in cultural products. *Personality and Social Psychology Review, 12*, 199–221. doi:10.1177/1088868308318260.
- Parks, J. B., & Roberton, M. A. (2004). Attitudes toward women mediate the gender effect on attitudes toward sexist language. *Psychology of Women Quarterly, 28*, 233–239. doi:10.1111/j.1471-6402.2004.00140.x.
- Pennebaker, J. W. (2011). *The secret life of pronouns: What our words say about us*. New York: Bloomsburg Press.
- Stewart, A. J., & Healy, J. M. (1989). Linking individual development and social changes. *American Psychologist, 44*, 30–42. doi:10.1037/0003-066X.44.1.30.
- Thornton, A., & Young-DeMarco, L. (2001). Four decades of trends in attitudes toward family issues in the United States: The 1960s through the 1990s. *Journal of Marriage and the Family, 63*, 1009–1037. doi:10.1111/j.1741-3737.2001.01009.x.
- Twenge, J. M. (1997). Attitudes toward women, 1970–1995: A meta-analysis. *Psychology of Women Quarterly, 21*, 35–51. doi:10.1111/j.1471-6402.1997.tb00099.x.
- Twenge, J. M. (2001). Changes in women's assertiveness in response to status and roles: A cross-temporal meta-analysis, 1931–1993. *Journal of Personality and Social Psychology, 81*, 133–145. doi:10.1037/0022-3514.81.1.133.
- Twenge, J. M. (2006). *Generation me: Why today's young Americans are more confident, assertive, entitled—and more miserable than ever before*. New York: Free Press.
- U. S. Bureau of the Census, *Statistical Abstract of the United States*. (1925–2011). Washington, DC: U.S. Government Printing Office.
- Walker, T. (2007). *Thou and you in early modern English dialogues: Trials, depositions, and drama come*. Philadelphia, PA: John Benjamins Publishing.
- Zou, X., Tam, K., Morris, M., Lee, S., Lau, I., & Chiu, C. (2009). Culture as common sense: Perceived consensus versus personal beliefs and mechanisms of cultural influence. *Journal of Personality and Social Psychology, 97*, 579–597. doi:10.1037/a0016399.
- Zube, M. J. (1972). Changing concepts of morality: 1948–69. *Social Forces, 50*, 385–393. doi:10.2307/2577043.