CAN MARS AND VENUS SIGNIFICANTLY AFFECT RATING SCALES IN MARKET RESEARCH?

Schnurr, Norman V.
Robert Morris University

Klentzin, Jacqueline C.
Robert Morris University

ABSTRACT
Are men from Mars and women from Venus (Gray, 1992) when it comes to marketing evaluations scales? The thought came as one of the authors was compiling a study that compared ratings by two student populations from different cultures. Could the differences in rating questions be due largely to the way these cultures perceive what a 1 to 10 scaling means? If so, what does this mean for other marketing segments like gender? The authors (a male and a female) wanted to explore if gender had any bias for scaling questions. After conducting a review of the literature, they examined several existing research studies to see if ratings are consistently higher for one gender or the other.

INTRODUCTION
When asked by a doctor what pain level are you experiencing on a scale of 1 to 10 where 1 is very little and 10 is extremely painful – how do you decide along the scale? How do judges for gymnast and diving events determine their scores? While those making these kinds of evaluations are often provided with direction and/or training, the final rating can, nonetheless, be quite subjective.

So too with customer evaluations in marketing research where most questionnaires often have scaling questions and respondents are providing with little or no explanation regarding the differences along the rating sequence. This begs the following questions:

- Do demographic differences impact how participants discern the differences along the scale?
- Do different sub-groups or sub-cultures set-up scaling differently?
- Does one group tend to give higher ratings or lower ratings more consistently than the others?

An impromptu verbal survey of marketing colleagues revealed great interest but very different experiences, especially related to the demographic variable of gender.
In the segmentation of customers, companies do a great deal of targeting by gender and any input regarding gender bias in rating would prove to be very rewarding to these organizations. Example - If males or females automatically rate products significantly higher than the other, the results of scaled may be misinterpreted by the researchers, which could result in an ineffective advertising, etc. The truth may be that these seemingly higher ratings and lower ratings are not significantly different and, as a result, companies would benefit more from a different type of advertising campaign, marketing strategy, etc.

Therefore, the purpose of this study is to determine any consistent pattern among consumer ratings by gender. It is not intended to provide any psychological, cognitive, or cultural understanding regarding gender differences in evaluations, if any such difference is determined. In order to address the issue, the authors first conducted a review of the existing literature, which will be presented first. Based upon the information uncovered, existing research articles were examined to test the hypothesis based on the Schmidt, et. al. study (2012) that women consistently tend to rate products and services more positively than men. These research articles presented are referred to in this document as “cases.” These cases were then compared and contrasted in order to uncover any gender-based evaluative patterns.

DEMographics AND MARKet SEGmentation

The use of demographics in social science research is important in understanding how different segments of the population behave in certain circumstances. When applied to consumer behavior research in particular, this nuanced understanding of demographic-based behavior is especially helpful during product and service evaluations, which assist businesses in improving and better positioning products and services for consumer consumption:

Demographic segmentation consists of dividing the market into groups on the basis of demographic variables such as age, sex, family size, family life cycle, income, occupation, education, religion, race, and nationality. Demographic variables are the most popular bases for distinguishing customer groups. One reason is that customer wants, preferences, and usage rates are often highly associated with demographic variables. Another is that demographic variables are easier to measure than most other types of variables. (Kotler, 1984, p. 255)

In order to be considered an effective unit of segmentation, a demographic must conform to four conditions: measurability, accessibility, substantiality, and actionability:

- Measurability, the degree to which the size and purchasing power of the segments can be measured.
- Accessibility, the degree to which the segments can be effectively reached and served.
- Substantiality, the degree to which the segments are large and/or profitable enough.
- Actionability, the degree to which effective programs can be formulated for attracting and serving the segments. (Kotler, 1984, pp. 264-265)
Interestingly, some demographics appear to be more popular with researchers than others. For instance, the influence of national origin on product/service evaluation has been a robust topic of study and evidence of its impact has been widely reported in the literature (Crotts & Erdman, 2000; Hsieh & Tsai, 2009; Seock & Lin, 2011; Tsang & Ap, 2007; Winsted, 1997). Research focusing on gender is less abundant (Laroche, 2003; Schmidt et al, 2012) despite that “among a relatively small number of variables, gender is one that meets all four requirements for effective segmentation” and that “one of the most important avenues of gender research would be to assess possible differences in how men and women evaluate goods and services” (Laroche, 2003, pp. 246-247).

**DIFFERENCES IN PRODUCT AND SERVICE EVALUATIVE RATINGS**

The limited numbers of studies that have attempted to address this issue are problematic in nature. At first glance, there appears to be no discernible pattern in how gender demographics directly impact product/service evaluation.

Few studies in the literature have hypothesized about and then tested evaluation differences in the ratings of tangible/intangible products between males and females and those that do exist show that females tend to rate products more favorably than their male counterparts (Williams, 2002; Schmidt et al, 2012). Explanations as to why this is the case vary among researchers and include, but are not limited to, the complexity of women’s information processing when compared to men (Schmidt et al, 2012), women attaching more importance to evaluative criteria than men (Williams, 2002), and men and women using/communicating via measurement scales differently – “for example, a response of a “6” on an agree-disagree scale does not mean the same thing to men and women” (Schmidt et al, 2012, p. 95).

This phenomenon does not hold true when evaluating services where female customers maintain higher expectations of service quality (Yelkur & Chakrabarty, 2006), tend to rate service quality lower when compared to their male study counterparts (Dimitriades, 2007; Lee et al., 2011), and use different evaluative criteria (Iacobucci & Ostrom, 1993; Lee, et al, 2011). Possible reasons for this phenomenon include, but are not limited to, the different genders focusing on different aspects of the service encounter (Iacobucci & Ostrom, 1993; Lee, et al, 2011; Lin, et al, 2001) and the gender of the person providing the service (Poria, 2008).

**DEMOGRAPHIC WRINKLES**

Taking a wider view, some researchers have questioned the usefulness of the gender demographic for market segmentation in product/service evaluations because it is inherently problematic. These complications include the concept that gender is much more complex than the binary sex designation of male/female (Fischer & Arnold, 1994) and that the influence of the gender demographic is secondary because it is often tempered by other demographics like national origin (Mattila, 2000).

**HYPOTHESIS**

One gender consistently evaluates and rates products and services more positively than the other gender.
METHODS
To determine if gender caused significant differences in response patterns for the rating of products or services, the researchers first retrieved evaluative studies that included results differentiated by gender to act as cases. These cases were discovered in the following databases: ProQuest Central, Robert Morris University Library’s Discovery services which retrieves journal articles from a variety of EBSCOhost databases as well many smaller publish-based products, and Taylor & Francis Online. Key search phrases used included, “Gender differences in rating”, “male product rating”, “gender bias rating”, “leniency bias”, “product evaluation and gender”, “product evaluation and sex”, “impact of gender differences on the evaluation of products”, “product evaluation and gender,” and “consumer product evaluations.” The articles were also read in order to find research studies that measured significant differences (often t-tests) between male and female responses, specifically rating scale questions, which were often Likert scales. Only studies met these criteria were selected for inclusion. Additionally, e-mails were also sent to some colleagues by the authors to see if any of their studies would meet the criteria to be included. This outreach proved to be unsuccessful.

Once a collection of cases was developed, the authors analyzed the male and female evaluative findings, which are presented below, and examined the findings for any discernible patterns.

CASE DESCRIPTIONS
The following is a list of 13 case descriptions, which includes pertinent information used for the analysis. Complete citations for the cases can be found in the References section at the end of this document.

CASE 1
Title: Destination USA by Croatian and Serbian University Students (2012)
First Author: Stanko Racic
Sample: University students in Croatia (2008) (n= 74 males, 287 females) and (2010) (n= 46 males, 75 females) and Serbia (2010) (n= 89 males, 185 females)
Scale: 1 low/negative to 10 high/positive, 18 items rated
Example of question: USA is presently a great place to live
Results: In Croatia ’08 study, males gave ratings that were significantly higher on 5 of the 18 questions asked while females gave significantly higher ratings on 4 of the 18 areas. Croatia ’10 study in 2 of 18 items males responded significantly higher while females were higher in 9 of 18 questions. In the Serbia ’10 study, females rated items significantly higher in 4 of 18 and there was no significant difference in the other 14 items.
Net Results: Croatia ’08 – mixed result, Croatia ’10 – females generally rated higher (9 females higher; 2 males higher), Serbia’10 – females rated higher

CASE 2
Title: Hispanic Consumers’ Shopping Orientation and Apparel Retail Store Evaluation Criteria (2008)
First Author: Yoo-Kyoung Seock
Sample: Convenience sample of Hispanics in Georgia (n= 138 males, 168 females)
Scale: 1 not important at all – 4 very important, 3 items rated
Example of question: Rate customer service
Results: In all 3 store evaluation questions, females rated it significantly higher in importance.
Note: This study was not rating characteristics but giving the importance

**Net Results:** Females rated higher

**CASE 3**  
**Title:** Gender Differences in Patients’ Perceptions of Inpatient Care (2012)  
**First Author:** Marc N. Elliott  
**Sample:** Patients from 3,830 acute care hospitals (n= 823,714 males 1,147,918 females)  
**Scale:** 0 worst possible – 10 best possible, 10 items rated  
**Example of question:** What number (0-10) would you use to rate your hospital stay?  
**Results:** In 9 out of 10 areas rated, males were significantly more positive in their ratings. Females gave significantly higher ratings only for “Doctor Communication”. Authors also noted that women studied were generally older on average and suggested that may have caused rating differences.  
**Net Results:** Males rated higher

**CASE 4**  
**Title:** Evaluating Gender Bias in Ratings of University Instructors’ Teaching Effectiveness (2009)  
**First Author:** Suzanne Young  
**Sample:** Undergraduate and graduate students enrolled in a medium sized university in the western U.S. (n= 246 males, 519 females)  
**Scale:** 1 not at all descriptive – 9 very descriptive, 25 items rated – put into 3 factors  
**Example of question:** Adapt to student needs  
**Results:** The data was factored loaded into 3 areas: interpersonal, pedagogical, course content characteristics. No significant difference in the interpersonal characteristics but in the other two males rated their male professors significantly higher and females rated their female professors higher.  
**Net Results:** Mixed result

**CASE 5**  
**Title:** Gender Differences in Rating the Teaching of Economics (1997)  
**First Author:** Kathryn H. Anderson  
**Sample:** 2,185 introductory macroeconomics students in 80 different classes and 2,408 introductory microeconomics students in 87 classes at 53 different colleges (n=did not break number of male vs. females)  
**Scale:** 1-low to 5 – high, 9 areas rated  
**Example of questions:** Preparation for class  
**Results:** The authors found that women tend to give instructors higher scores on preparation, grading standards, and English skills. There were no gender differences when rating enthusiasm of the instructor.  
**Net Results:** Females rated higher

**CASE 6**  
**Title:** Gender Bias in the Evaluation of New Age Music (2003)  
**First Author:** Ann Colley  
**Sample:** sixty four undergraduate students (n=32 males 32 females)  
**Scale:** 0 not at all – 5 neutral – 10 very/very much, 10 items rated
Example of questions: Artistic merit  
Results: Females gave significantly higher ratings on 9 out of 10 items rated.  
Net Results: Females rated higher

CASE 7  
Title: Gender Bias in Customer Evaluations of Service Quality: an Empirical Investigation (2006)  
First Author: Robin L. Snipes  
Sample: Students at 6 teaching-oriented colleges (n = 3,640 males, 5,027 females)  
Scale: 1 much worse than I expected liked -7 much better than I expected, 27 items rated  
Example of question: The organization of this class  
Results: Females rated many items lower especially the fairness of the course. The authors had suggested that the results that 2/3 of men are considered thinking types and 2/3 of women are considered feeling types were the possible cause.  
Net results: Mixed result

CASE 8  
Title: University Students’ Perceptions of Two Countries: Turkey and U.S.A. (2009)  
First Author: Norman V. Schnurr  
Sample: 3rd and 4th university students at a U.S. university and a Turkish university (n=76 Turkey, USA 471 - females/males not available)  
Scales: I low/negative - 10 high/positive, Number of items: 20 variables  
Example of question: I am proud to be an American/Turkish  
Results: Both male and female Turkish students rated all but 2 of the 18 variables with no difference. Females rated Turkey’s present leadership significantly lower than Turkish males while males rated “Turkish people are hardworking” slightly lower than females. In the U.S. students’ survey females rated 3 of the 20 variables higher than U.S. males.  
Net results: Turkish students – mixed results; USA students – females generally rated higher

CASE 9  
Title: Best and Worst Professors: Gender Patterns in Students’ Choices (2000)  
First Author: Susan A. Basow  
Sample: Students at a small private liberal arts college in the northeast U.S. (n= 47 males, 61 females)  
Scales: 1 never or almost never true – 7 always or almost always true, 30 items rated  
Example of question: Knowledgeable  
Results: Female students were disproportionately more likely to rate female faculty member “best” while male students were disproportionally less likely.  
Net results: Rating of female faculty – females rated higher; Rating of male faculty – mixed results

CASE 10  
Title: Gender Differences in Self-Ratings of Abilities and Skills (1990)  
First Author: Jane L. Swanson  
Sample: 112 introductory psychology students (n= 53 men, 59 women)  
Scales: 1 definitely could learn - 5 definitely could not learn, 14 items  
Example of question: Manual dexterity
**Results:** Men rated themselves higher on 3 (numerical, manual dexterity, mechanical) of 14 general abilities. When rating their own skills, women rated themselves higher on social skills, men rated themselves higher on realistic but neither was significantly different than their counterpart’s ratings.

**Net Results:** Mixed results

**CASE 11**
**Title:** Personality, Gender, Age and Logical Overlap in Multi-source Ratings (2002)
**First Author:** Peter Warr
**Sample:** Sales staff in three organizations (n= 131 males, 67 females)
**Scales:** 1 hardly ever – 5 always, 16 attributes
**Example of question:** Relating to customers
**Results:** In the self-rating, there were six attributes rated significantly different. In four, females rated themselves higher, while males had rated themselves higher for two. In rating their supervisors, males rated one aspect - problem solving — higher and females rated on reliability significantly higher.

**Net Results:** Self-rating females generally rated higher; Supervisor rating – mixed results

**CASE 12**
**Title:** The Role of Gender in Teaching Effectiveness Ratings of Faculty  (2006)
**First Author:** Jonathan Kohn
**Sample:** Students at Shippensburg University (n= 458 males, 472 females)
**Scales:** Liked 1 strongly agree – 6 strongly disagree, 6 effectiveness variables
**Example of Question:** The professor was effective teaching this course
**Results:** Female students rated male faculty significantly higher than did male students. Female students also rated faculty effectiveness higher than male

**Net Results:** Females rated faculty higher

**CASE 13**
**Title:** The Influence of Service Quality on Satisfaction and Intention: A Gender Segmentation Strategy (2011)
**First Author:** Jeoung-Hak Lee
**Sample:** Multiple visitors of 5 private golf clubs in S. Korea (n= 375 males, 153 females)
**Scales:** 1 strongly agree – 7 strongly disagree, 21 items put into 5 factors
**Example of question:** Being a customer of XYZ golf course is usually a satisfying experience
**Results:** 5 factors – males rated tangible factors significantly higher, females rated empathy factors significantly higher, the other 3 factors were not significantly different.

**Net Results:** Mixed results

**RESULTS OF THE ANALYSIS**
In the 18 studies (13 cases), females rated items higher than males in six studies and somewhat higher in 3 studies. Males had given higher ratings in only one study and finally, there were mixed results in the remaining eight studies.
<table>
<thead>
<tr>
<th>Gender Based Results</th>
<th>Number of Studies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Females significantly higher</td>
<td>6</td>
</tr>
<tr>
<td>Females somewhat higher</td>
<td>3</td>
</tr>
<tr>
<td>Males significantly higher</td>
<td>1</td>
</tr>
<tr>
<td>Mixed results</td>
<td>8</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>18</strong></td>
</tr>
</tbody>
</table>

CONCLUSIONS

The analysis of the cases did not uncover conclusive evidence that there are gender differences in response to rating scale questions, which bolsters the review of the literature. Although the leaning of the results suggest that women tend to rate things more positively than men (9 studies vs. 1 study), there are enough mixed results (8 studies) to muddy the proverbial waters. Therefore, the authors determined that the hypothesis is not supported. This study, however, calls to attention the difficulty in using gender as a market segment, despite the seemingly easy way it conforms to the Kolter (1984) criteria. Specifically, it appears that other demographic segments like age (Dimitriades, 2007; Elliott, et.al. 2012) impact the evaluation process and must, therefore, be taken into account. Additionally, the product/service under consideration may also influence responses between genders. Specifically, Williams (2002) and Schmidt, et.al (2012) found that women tend to rate products higher than men while others (Yelkur and Chakrabarty, 2006; Iacobucci & Ostrom, 1993; Lee, et.al, 2011; Dimitriades, 2007) found that men tended to rate services higher than women. More in-depth investigations of both product and service evaluations that isolate the gender demographic need to be conducted in order to truly determine any evaluative patterns and possible differences in scaling interpretations.

REFERENCES


Acknowledgement: The authors would like to thank Jessica Lascola, RMU School of Business Graduate Assistant and RMU MBA Student, for her contribution to the success of this study.