Gender Differences in the Accuracy of Self-Evaluations on Chemistry, English, and Art Questions.

A study assessed gender differences in confidence for three academic areas of varying gender-type. The study used questions about chemistry (masculine), English (feminine), and art (neutral). Participants were general psychology students (n=361: 229 females, 132 males) who received course credit for their participation. Findings revealed that females underestimated their performance more than did males on chemistry and English questions. (Contains 49 references, 3 notes, and 1 table.)
Gender Differences in the Accuracy of Self-Evaluations on Chemistry, English, and Art Questions.

Sylvia Beyer
Monica Riesselmann
Tiffany Warren
Gender Differences in the Accuracy of Self-evaluations on Chemistry, English, and Art Questions

Sylvia Beyer, Monica Rieselmann, & Tiffany Warren
University of Wisconsin-Parkside

Abstract
Gender differences in confidence in three academic areas of varying gender-type were assessed. We used questions about Chemistry (masculine), English (feminine), and Art (neutral). Females underestimated their performance more than did males on Chemistry and English questions.

Self-evaluation of performance is a fundamental task of self-regulation. However, people use some latitude when evaluating their abilities, making self-evaluations of performance at least somewhat inaccurate for many people (for a review see Beyer, 1998).

Recently there has been an attempt to discover the moderating and mediating variables which affect the accuracy of self-evaluations. Beyer’s (1990, 1998, 1999a, 2002; Beyer & Bowden, 1997) work has focused on participants’ gender, the gender-type of the tasks on which they are working, and expectancies.

She compared participants’ post-task self-evaluations of performance to their actual performance and found that females underestimated their performance on tasks participants considered “masculine” (tests of knowledge of sports figures, politics, or mathematics). Males either overestimated or accurately evaluated their performance. No gender differences in the accuracy of self-evaluations were found for tests of knowledge of movie and TV stars or English (considered feminine gender-typed by participants) or neutral tests of common knowledge, character detection, practical questions, anagrams, or history and geography.

This gender difference in self-evaluations was mediated, in part, by self-consistency: Pre-task expectancies of performance affected post-task self-evaluations of performance. Two individuals who performed equally well but had dissimilar expectancies evaluated their performance after task completion very differently: The individual with the higher pre-task expectancies was more likely to overestimate his/her performance after task completion than the individual with low initial expectancies. Thus, the gender difference in the accuracy of self-evaluations on masculine tasks can be attributed, in part, to their low expectancies which depress their post-task self-evaluations (Beyer, 1990, 1998, 2002; Beyer & Bowden, 1997).

The Goals of this Research
The present research examines gender differences in the accuracy of self-evaluations for three subject areas of varying gender-type: Chemistry (masculine), English (feminine), and Art (neutral). It was hypothesized, in accordance with Beyer’s (1990, 1998, 2002; Beyer & Bowden, 1997) previous research, that women would evaluate their performance less accurately than men in Chemistry, but not English and Art.

Rather than presenting test questions on paper, as was done in the past, we used a modified methodology for this research. Test questions were presented auditorily via a computer. The advantage of this presentation mode is that participants’ reaction times to questions can be recorded.

A study of gender differences in performance on the game show Jeopardy found that women were less likely than men to be able to select question types because of their slower response times (Brownlow, Whitener, & Rupert, 1998). Although slower reaction times may reflect a male advantage in either knowledge or psychomotor speed, an alternative interpretation is that women’s lower confidence in their ability to answer questions correctly manifested itself in slower reaction times. We were interested in whether participants’ reaction times to test questions would reveal gender differences in self-perceptions of ability for different subject areas.

In addition, this research adds to our understanding of the self-evaluation process by investigating the effect of dysphoria. So far, the relation between gender and dysphoria has remained largely unexplored. The present research aims to fill this void in the self-perception literature.

Depressive Realism
According to work on depressive realism, mildly depressed individuals are more accurate self-perceivers than their overly optimistic nondysphoric counterparts.

1 Poster presented at the annual meeting of the American Psychological Society in New Orleans, LA, June 2002.
2 To whom correspondence should be addressed at Department of Psychology, University of Wisconsin-Parkside, Kenosha, WI 53141. e-mail: beyer@uwp.edu
However, research on depressive realism has ignored gender differences in the accuracy of self-evaluations. An innovation of the present research is that it explores the interrelation between gender and dysphoria in terms of the accuracy of self-evaluations.

Dysphorics are more accurate than nondysphorics in their self-evaluations of social competence (Lewinsohn, Mischel, Chaplin, & Barton, 1980; McNamara & Hackett, 1986), recall of their toddlers’ negative behaviors (Lovejoy, 1991), predictions about the future (Alloy & Ahrens, 1987; Crocker, Alloy, & Kayne, 1988), assessments of the degree of control over external stimuli (Abramson & Alloy, 1981; Alloy & Abramson, 1979, 1982; Alloy, Abramson, & Kossman, 1985; Alloy, Abramson, & Viscusi, 1981; Dobson & Franche, 1989; Glass, McKnight, & Valdimarsdottir, 1993; Martin, Abramson, & Alloy, 1984), and they are more sensitive to changes in reward contingencies (Rosenfarb, Burker, Morris, & Cush, 1993).

However, research on depressive realism has been criticized for failing to use an objective standard for accuracy (Ackermann & DeRubeis, 1991). Furthermore, the depressive realism hypothesis has been challenged by a substantial number of studies which found no evidence for its existence (Benassi & Mahler, 1985; Buchwald, 1977; Campbell & Fehr, 1990; DeMonbreun & Craighead, 1977; Dobson & Shaw, 1981; Gotlib, 1981, 1983; Vazquez, 1987; Wener & Rehm, 1975). For example, dysphorics’ thoughts about themselves are quite negative and their predictions about the future more pessimistic than those of nondysphorics (e.g., Pietromonaco & Markus, 1985). But this could be a function of dysphorics’ more negative past experiences rather than representing self-derogation or inaccurate self-perceptions (Koenig, Ragain, & Harrow, 1995; Shrauger, Mariano, & Walter, 1998). Thus, the question is: Are dysphorics’ negative self-perceptions accurate? Or are they, like females’ self-evaluations on masculine tasks, underestimations of their abilities, popularity, etc.?

This research fills a gap insofar as it links the separate literatures on gender and self-evaluations on the one hand and dysphoria and self-perceptions on the other hand. Such research may provide both theoretical insights and be of practical value for intervention programs. The fact that the literature on depressive realism has paid so little attention to gender differences in the accuracy of self-perceptions is surprising given that women are twice as likely to suffer from depression as men (Nolen-Hoeksema & Girgus, 1994). Therefore, according to the depressive realism hypothesis, women should be more accurate self-evaluators than men. However, this prediction has rarely been explicitly tested.

Hypotheses

To summarize, this study addresses the following hypotheses. (1) Gender differences in the accuracy of self-evaluations of performance will be found for a “masculine” subject area (Chemistry) but not for feminine or neutral subjects (English and Art). (2) Dysphorics will not be more accurate self-evaluators than nondysphorics. (3) Gender differences in reaction times to different subject areas will be found even when knowledge of the subject area is controlled. This would reveal gender differences in self-perceptions of ability in different subject areas.

Method

Participants

Participants were 361 General Psychology students (229 females and 132 males). Participants received course credit for their participation.

Materials

Participants were presented with 90 True/False (T/F) questions from three subject areas. Thirty questions each pertained to Chemistry (masculine domain), English grammar (feminine domain), and Art (neutral domain). The three subjects were interspersed across the 90 questions. A pretest of a greater pool of questions was used to select the final 30 questions for each subject. This yielded three tests of roughly equal difficulty.

Procedure

Participants filled out the Beck Depression Inventory (BDI; Beck, Ward, Mendelson, Mock, & Erbaugh, 1961) and were presented with a general description of the task including information on the number of T/F questions, and the amount of time available. Participants were informed that their reaction times would be recorded. Participants were run through two sample questions for each subject to familiarize them with the procedure.

Participants were run in groups of two on Macintosh SE computers. The T/F questions were auditorily presented via computer to avoid individual differences in reading speed. Prior to each question the prompt “READY?” appeared. Participants had to press a key to activate the presentation of the next question. This ensured that participants were ready to listen to the next question. Half of the participants listened to a female, the other half to a male voice. Participants pressed one of two keys to indicate whether they believed the answer
to the question was True or False (for half of the participants the T/F keys were reversed). Approximately half of the participants were informed prior to each question which of the three subjects would be presented next, the other half of the participants were unaware of the subject area of the next question. Response times were recorded by the computer.

The questions were broken into three blocks. Prior to each block of 30 T/F questions, participants were asked how many questions from each domain they expected to answer correctly (expectancies). After each block, participants were asked to estimate how many questions from each domain they had answered correctly (self-evaluations of performance). After answering all 90 questions, participants estimated how many questions of each type they answered correctly.

Results

All significance tests are two-tailed.

Gender Differences in Expectancies, Performance, and Self-evaluations

Males answered significantly more Chemistry questions correctly, $F(1, 358) = 5.14, p < .03$, expected to answer more questions, $F(1, 318) = 7.68, p < .006$, and thought they had answered more questions than did females, $F(1, 353) = 18.05, p < .0001$. For English there was neither a gender difference in performance, $F(1, 352) = 2.24, p < .14$, nor a gender difference in expectancies, $F(1, 311) < 1$, but males thought they had answered more questions than did females to a marginal degree, $F(1, 339) = 3.46, p < .07$. For Art there was neither a gender difference in performance, $F(1, 355) = 2.61, p < .11$, nor a gender difference in expectancies, $F(1, 313) < 1$, but males thought they had answered more questions than did females, $F(1, 349) = 4.16, p < .05$.

Dysphoria Status Differences in Accuracy of Self-evaluations

Hypothesis 1 predicts that gender differences in the accuracy of self-evaluations will be significant only for Chemistry. Hypothesis 2 predicts that dysphorics are not more accurate self-evaluators than nondysphorics.

Accuracy of self-evaluations was calculated by subtracting performance from post-test self-evaluation scores. Positive accuracy scores indicate overestimations of performance, negative scores underestimations, while scores approaching 0 indicate accurate evaluations.

For each of the three subject areas, the accuracy of self-evaluations was regressed on participant gender, score on the BDI (as a continuous variable), and their interaction. Table 1 lists the estimated means.

<table>
<thead>
<tr>
<th></th>
<th>Females</th>
<th>Males</th>
</tr>
</thead>
<tbody>
<tr>
<td>English Questions Nondysphorics</td>
<td>-6.2</td>
<td>-4.9</td>
</tr>
<tr>
<td></td>
<td>Dysphorics</td>
<td>-7.3</td>
</tr>
<tr>
<td>Art Questions   Nondysphorics</td>
<td>-7.4</td>
<td>-6.8</td>
</tr>
<tr>
<td></td>
<td>Dysphorics</td>
<td>-8.1</td>
</tr>
<tr>
<td>Chemistry Questions Nondysphorics</td>
<td>-6.7</td>
<td>-4.9</td>
</tr>
<tr>
<td></td>
<td>Dysphorics</td>
<td>-7.4</td>
</tr>
</tbody>
</table>

Gender was a significant predictor of the accuracy of self-evaluations for Chemistry, $F(1, 349) = 9.33, p < .002$, and English, $F(1, 330) = 6.67, p < .01$, where males underestimated their performance less than did females. There was no gender difference in the accuracy of Art self-evaluations, $F(1, 342) = 1.25, p < .27$.

For Art and Chemistry, BDI score did not predict the accuracy of self-evaluations, $F(1, 342) < 1; F(1, 349) = 1.35, p < .25$. Only for English did BDI score predict the accuracy of self-evaluations, $F(1, 330) = 5.67, p < .02$. Dysphorics underestimated their English performance more than nondysphorics did. Thus, contrary to the depressive realism hypothesis, but as predicted, dysphorics were not more accurate self-evaluators than nondysphorics.

Gender Differences in Reaction Times

For each of the three subject areas, reaction times were regressed on participant gender, performance score, score on the BDI (as a continuous variable), and the interaction terms. Performance score was included to control for differences in knowledge.

There was no significant gender difference in reaction times for English, $F(1, 340) < 1$. There were borderline significant gender differences in reaction times for Art, $F(1, 337) = 3.68, p < .06$, and Chemistry, $F(1, 343) = 3.74, p < .06$, with females having faster reaction times than males.

Dysphorics had slower reaction times than nondysphorics for English, $F(1, 340) = 7.60, p < .006$, and Art, $F(1, 337) = 3.21, p < .08$. For Chemistry the triple interaction between gender, BDI score and performance was significant, $F(1, 343) = 7.09, p < .
.008, with low-performing nondysphoric females having the fastest reaction times and high-performing dysphoric males the slowest reaction times.

Discussion

Gender Differences in the Accuracy of Self-evaluations

There was a significant gender difference in the accuracy of self-evaluations for Chemistry and English with males underestimating their performance significantly less than females did. The results for Chemistry are in line with findings that girls, but not boys, underestimate their grades and ability in mathematics (Bornholt, 2000; Frome & Eccles, 1998; Hannover, 1991; Tiedemann, 2000). Furthermore, the gender stereotype regarding males’ superiority in chemistry is still pervasive (Beyer, 1999b). This suggests that chemistry is still an arena fraught with self-doubt for females. This replicates the gender difference in the accuracy of self-evaluations for mathematics (Beyer, 2002) and sports (Beyer, 1990, 1998, 1999a; Beyer & Bowden, 1997). The fact that women underestimated their performance for English more than men did was not anticipated.

The consequences of inaccurately negative self-perceptions have been discussed in detail by Beyer (1998). Briefly, positive self-perceptions are related to self-esteem and psychological health (e.g., Kurman & Sriram, 1997; Taylor & Armor, 1997), whereas negative self-evaluations are related to depression (e.g., Glass et al., 1993). Positive self-perceptions are also intimately tied to aspirations, preference for challenging tasks, curiosity, intrinsic motivation, and persistence and thus have desirable effects on performance (see Beyer, 1995 for an overview; Greene, DeBacker, Ravindran, & Krows, 1999). This suggests that the gender difference in the accuracy of self-evaluations for Chemistry may have deleterious affective and behavioral consequences.

Depressive Realism

The depressive realism hypothesis received no support from the present research. A significant effect of dysphoria was found only for English, where dysphorics were less accurate than nondysphorics. A similar result was obtained by Beyer (2002). These results are in line with research which found that dysphorics show biases involving slowed processing of positive information and faster processing of negative information, whereas nondysphorics possess the reverse orientation to the self (Moretti, Segal, McCann, Shaw, Miller, & Vella, 1996). In fact, dysphorics tend to activate negative self-reference information relatively automatically and unintentionally (Bargh & Tota, 1988).

Nondysphorics have been called “cognitive misers” (Fiske & Taylor, 1991) who “process information in trivial situations casually, with self-serving optimistic biases such as the illusion of control” (Pacini, Muir, & Epstein, 1998, p. 1057). Maybe so, but dysphorics are cognitive misers as well, processing information with a negative bias.

Thus, the present study adds to a growing body of research suggesting problems with the depressive realism hypothesis. The question of whether dysphorics or nondysphorics are more accurate self-evaluators is too simplistic. The reason for the now-you-see-it-now-you-don’t pattern of results for depressive realism may be that too little attention has been paid to discovering moderating variables which affect the accuracy of self-perceptions of dysphorics and nondysphorics. For example, the depressive realism effect is most likely obtained in artificial laboratory situations, whereas nondysphorics may be more accurate self-evaluators in nontrivial, realistic situations (Dobson & Franche, 1989; Dunning & Story, 1991; Pacini et al., 1998). A similar challenge to the generality of the depressive realism effect was obtained by Shrauger et al. (1998). Dysphorics were more accurate than nondysphorics when making predictions about undesirable events in their lives, whereas the reverse was true for desirable events. In all likelihood specifics about the task at hand, such as task difficulty, or the relative ease with which one can judge one’s performance, are of importance.

Gender Differences in Reaction Times

For Art and Chemistry women had faster reaction times than men. Thus, women’s low perception of performance in Chemistry obtained by using self-evaluations of performance was not mirrored by more conservative reaction times. This suggests that reaction times are not a good measure of confidence. Interestingly, dysphoria clearly manifests itself in slowed reaction times, even when equating for performance.

Conclusions

The present study has replicated the findings of Beyer (1990, 1998, 1999a, 2002; Beyer & Bowden, 1997) in the area of chemistry. The results indicate that females harbor self-doubts regarding their chemistry ability.

No evidence for depressive realism was obtained. Dysphorics were not more accurate self-evaluators than nondysphorics. Because females are twice as likely to suffer from depression than males (Nolen-Hoeksema &
Girgus, 1994), more research on the interrelations among gender, depression, and the accuracy of self-perceptions is warranted.

References


Title: Gender differences in the accuracy of self-evaluations on Chemistry, English, and Art questions

Author(s): Sylvia Beyer, Monica Riesselmann, Tiffany Warren

Corporate Source: University of Wisconsin-Parkside

Publication Date: June 2002

I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce and disseminate this document as indicated above. Reproduction from the ERIC microfiche or electronic media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries.

Sylvia Beyer
Assoc. Prof.
Univ. of Wisc. Parkside, Kenosha, WI
53148

Date: 6-16-02

I. DOCUMENT IDENTIFICATION:

Title: Gender differences in the accuracy of self-evaluations on Chemistry, English, and Art questions

Author(s): Sylvia Beyer, Monica Riesselmann, Tiffany Warren

Corporate Source: University of Wisconsin-Parkside

Publication Date: June 2002

II. REPRODUCTION RELEASE:

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, Resources in Education (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic media, and sold through the ERIC Document Reproduction Service (EDRS). Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to the document.

If permission is granted to reproduce and disseminate the identified document, please CHECK ONE of the following three options and sign at the bottom of the page.

The sample sticker shown below will be affixed to all Level 1 documents

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

Level 1

X

Check here for Level 1 release, permitting reproduction and dissemination in microfiche or other ERIC archival media (e.g., electronic) and paper copy.

The sample sticker shown below will be affixed to all Level 2A documents

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE, AND IN ELECTRONIC MEDIA FOR ERIC COLLECTION SUBSCRIBERS ONLY, HAS BEEN GRANTED BY

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

Level 2A


Check here for Level 2A release, permitting reproduction and dissemination in microfiche and in electronic media for ERIC archival collection subscribers only.

The sample sticker shown below will be affixed to all Level 2B documents

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

Level 2B


Check here for Level 2B release, permitting reproduction and dissemination in microfiche only.

Documents will be processed as indicated provided reproduction quality permits. If permission to reproduce is granted, but no box is checked, documents will be processed at Level 1.
### III. DOCUMENT AVAILABILITY INFORMATION (FROM NON-ERIC SOURCE):

If permission to reproduce is not granted to ERIC, or, if you wish ERIC to cite the availability of the document from another source, please provide the following information regarding the availability of the document. (ERIC will not announce a document unless it is publicly available, and a dependable source can be specified. Contributors should also be aware that ERIC selection criteria are significantly more stringent for documents that cannot be made available through EDRS.)

<table>
<thead>
<tr>
<th>Publisher/Distributor:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Address:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Price:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
</tbody>
</table>

### IV. REFERRAL OF ERIC TO COPYRIGHT/REPRODUCTION RIGHTS HOLDER:

If the right to grant this reproduction release is held by someone other than the addressee, please provide the appropriate name and address:

<table>
<thead>
<tr>
<th>Name:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Address:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
</tbody>
</table>

### V. WHERE TO SEND THIS FORM:

Send this form to the following ERIC Clearinghouse:

**ERIC/CHESS**
2805 E. Tenth Street, #120
Bloomington, IN 47408

However, if solicited by the ERIC Facility, or if making an unsolicited contribution to ERIC, return this form (and the document being contributed) to:

**ERIC Processing and Reference Facility**
4483-A Forbes Boulevard
Lanham, Maryland 20706

Telephone: 301-552-4200
Toll Free: 800-799-3742
FAX: 301-552-4700
e-mail: ericfac@inet.ed.gov
WWW: http://ericfacility.org