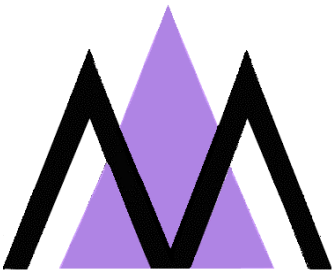

Academy Of Management Journal

January 2, 2007

Mr. David Hekman
Email: hekman@u.washington.edu

Dear Mr. Hekman:

The manuscript that you submitted to the *Academy of Management Journal* entitled, "**An Examination of Race and Sex-Based Biases in Professional Employee Performance Evaluations**" (*AMJ*-06-0696), has now been reviewed by three experts. The reviewers and I think your manuscript deals with an interesting and important topic.

However, the reviewers and I also see some major shortcomings that cast doubt on your paper's appropriateness for publication in *AMJ*. In general, the reviewers have concerns with your study's theoretical contribution, the appropriateness of its methods, and the overall contribution to management theory, research, and practice.

The degree of agreement among the reviewers is moderately high. In the end, two reviewers recommend rejection, while the remaining reviewer expresses a bit more optimism that a major revision opportunity can address his/her concerns. Based on my own reading of your manuscript, I am afraid that I am in agreement with the more pessimistic reviewers. Therefore, I regret to inform you that we shall not be publishing your manuscript, nor shall we be asking you to submit a revision to the *Journal* for further consideration.

I realize that my decision will be greatly disappointing to you. Obviously, you have put a tremendous amount of effort into conducting and writing up this study. However, your reviewers have offered a variety of excellent comments that I hope will be highly useful to you in your future efforts.

Because a primary aim of the *Journal* is to provide developmental reviews regardless of the decision outcome, I have sifted through the reviewers' comments and tried to highlight some of the more salient issues. I also read your manuscript and provided my own comments independently from reviewer reactions. I offer these comments with the objective of helping you

to bring this manuscript to publication in another journal. I hope you find these comments helpful in your further work with the paper.

Major Concerns

1. ***Introduction and study framing***. I think you could do a bit more to bring readers up to speed on the potential contribution of your study. For example, in any quality introduction, you should: (a) specify the domain of interest; (b) indicate to what particular aspect of the domain you intend to contribute; and (c) spell out why your study will add value to the existing literature (primarily by pointing out a few shortcomings or omissions in the literature to date).

   I think you have done a good job with (a) and (b) above, but not (c). Readers will want answers to such questions as what theory or theories are supported, altered, or refuted by your findings and how should readers think differently about this area of research after reading your paper? While I do think your research question appears interesting, I am unsure as to the potential added value *theoretical* contribution.

   I am also a little confused as to the research you cite in the introduction that there is little support for the effects of biases on performance evaluation. For example, on pp. 3-4, you state, "…researchers have found it difficult to reach definitive conclusions about how demographic characteristics might influence the performance evaluation process…" Your statement made me go back to one of my own articles (Kirkman, Tesluk, & Rosen, 2004) that deals with rater-ratee performance evaluation effects at the team level of analysis. In that article (on p. 340), we state:

   "…dyadic rater-ratee similarity was associated with higher performance ratings in a meta-analysis (Kraiger & Ford, 1985). Similar effects, albeit smaller and more inconsistent, were found across various performance dimensions in another meta-analysis (Pulakos, Oppler, White, & Borman, 1989). Minority managers, reporting to a white supervisor, reported poorer fit with their work group than non-minority managers (Kirchmeyer, 1995). Ethnic heterogeneity was also marginally negatively associated with supervisors' liking of subordinates (Tsui & O'Reilly, 1989). Other findings support the idea that African-Americans are rated lower than Caucasians by supervisors (Greenhaus, Parasumraman, & Wormley, 1990; Lefkowitz, 1994; Sackett, DuBois, & Noe, 1991)."

   Based on the research cited above, I think the conclusions you draw in your introduction that "…researchers have found it difficult to reach definitive conclusions about how demographic characteristics might influence the performance evaluation process…" may be a bit strong.

   See:

   Kirkman, B.L., Tesluk, P.E., & Rosen, B. 2004. The impact of demographic heterogeneity and team leader-team member demographic fit on team empowerment and effectiveness. *Group & Organization Management*, 29: 334-368.

2. ***Potential value added theoretical contribution***. The reviewers and I are concerned about the potential your study has to make a ***substantive*** theoretical contribution to the existing literature. Indeed, as indicated in *AMJ's Information for Contributors, AMJ* has a long-standing requirement that all publications must make a substantial contribution to management *theory*. I'm afraid that the reviewers and I think your paper currently falls short on this criterion.

   For example, Reviewer #1 (Point #1) states, "Your two hypotheses are interesting, but they are derived directly from status characteristics theory. There is no extension of the theory, or any challenge to its basic tenets. I don't think anyone who is familiar with the SCT literature would disagree with your hypotheses. On the contrary, we've seen a lot of evidence to support these hypotheses in the past. This doesn't mean that your work is not interesting—it just means that it may not be well suited for *AMJ*, a journal whose mission is to publish important new theoretical and empirical insights" and (Point #2), "In revising this paper, I would encourage you to highlight what you think the unique contribution of your research is. If it's the link between objective and subjective performance measures, you'll have to do a better job of explaining why that is important. If it's examining the main predictions of SCT in a new context, you'll have to do more to explain why it's important to study this particular context. Finally, if it's your surprising finding – that certain types of customer-centered behaviors were *negatively* related to customer evaluations of performance for women and non-whites – then you'll need to do more to account for this."

   Given the research I pointed to in Major Point #1 above, I too am unsure as to exactly what you are attempting to add to the existing literature. Previous research shows that raters have biases that play into their ratings of others. What are you adding beyond this broad and consistent stream of research?

3. ***Ruling out alternative explanations***. Another major concern among the reviewers is the lack of attention to ruling out alternative explanations for the gap between objective and subjective performance.

   For example, Reviewer #1 (Point #3) states, "One of the concerns I had with your analysis was the selection of your control variables. In SCT studies that examine the impact of *diffuse* status characteristics, such as race or sex, researchers must be careful to control for *specific* status characteristics (e.g., ksa indicators, prior performance) that may be loosely correlated with the diffuse status characteristics that are of interest. I felt there were several missing control variables in your analysis that might offer a different explanation for why race or sex differences produced a bigger gap between subjective and objective performance. First, the quality of the physician's education and training would likely be a critical indicator of actual performance. I doubt that women or minorities have worse educational experience than do white males, but they may receive less qualified training because they are discriminated against in the matching process. It would be important to include these data in order to rule this out. Second, I'd like to know more about how Prohealth coded the minority variable.

Specifically, I'm curious to know how many of the minority physicians were foreign born and had noticeable accents. Accents tend to worsen others' impressions even though they may not be related to actual performance outcomes. Adding other demographic variables such as these could really help strengthen your case."

Similarly, Reviewer #2 (Point #1) states, "The biggest problem I have with the study is its basic assumption that the differences (based on physicians' gender and race) in the relationship between the objective criteria of the physician performance and the customer subjective rating of professional performance quality, is due solely to perceptual bias against women or non-white physicians. There may be other potential explanations that were not addressed in the study. For example, as Table 1 indicates, women were significantly younger, less tenured, and were by far more likely to work part time than full time (a correlation of .63). It may be that these factors led patients to perceive women, as a group, as being less trained and experienced, and therefore as performing less well relative to men. This, in turn, may have influenced the patients' perception about their own personal physician, if this physician was a woman. In other words, one can speculate that patients' bias towards women physicians is based on the above objective job related differences between men and women that give support to the idea that men physicians were better than women, rather than on inherent biases against women in general. Concerning race, the number of non-white physicians is very small, but the correlation matrix still shows a tendency for white physicians to have more tenure (experience) with Prohealth than non-white physicians, which again may have contributed to the perceptual differences based on race. This does not negate the possibility that inherent gender and race related biases may have played a role in how patients rate their physicians. However, the data the study is based on cannot provide an answer to the degree in which people's perceptions of their physicians' performances were based on inherent gender and racial biases versus other job related factors. Using direct measures on the respondents' attitudes and perceptions towards women and minorities would have helped to sort out this issue. Unfortunately, it does not seem that such measures were part of the study."

Finally, Reviewer #3 (Point #1) states, "One of the main theoretical frameworks underlying the hypotheses is one of stereotyping. Although years of research have demonstrated the existence of stereotypes, it would be useful to know to what extent the customers/patients in the current sample hold these stereotypes. Otherwise, the mechanism that links sex/race to performance outcomes is assumed to exist but not shown to exist. This link is needed to rule out alternative explanations for the findings and to provide stronger support for your theoretical rationale."

Clearly, everyone believes that while your findings are interesting, much more needs to be done to convince readers that you have adequately ruled out the many other plausible predictors of the gap between objective and subjective performance.

4. ***Methods***. The reviewers and I also have concerns about some of the methodological aspects of your study including sample and measures. For example, there are some concerns that the lack of diversity in your sample on race may lead to inaccurate conclusions. Reviewer #2

(Point #3) states, "The sample of non-whites is too small, and thus the results may be open to chance. Therefore you may consider running a 'jackknifing' analysis, in which the regression analysis is repeated a number of times, each time with the exclusion of one individual from the group of non-whites." Reviewer #3 (Point #5) adds, "Although race is a very important demographic characteristic to consider, it is unclear why it is being considered in your study with such a small sample size. Although 12.1% is representative of the national average, is a sample of 13 physicians large enough to draw meaningful and generalizable conclusions about race? This point dilutes the contribution of the study and its external validity."

The reviewers are also concerned about some of the measures in your study. For example, Reviewer #2 (Point #7) states, "Concerning Physician productivity, the authors indicated that intensity of each visit 'was measured by the Relative Value Units (RVUs), which are coded by physicians at the end of each visit according to national coding guidelines' (p. 14). It seems to me that this self-report instrument provides the physician the latitude to inject his/her subjective evaluation on the treatment. Can you discuss this issue?"

Reviewer #3 (Point #4) states, "…your measure of customer perception of quality consists of attention provider paid, thoroughness and competence, and opportunity to ask questions. It could very well be the case that a physician does pay attention to the patient, is thorough and competent, and provides the opportunity to ask questions during the actual visit while at the same time prescribing fewer tests and prescriptions and sending fewer emails. It is unclear that the objective and subjective forms of evaluations as operationalized in your study should be strongly related. They may be getting at different facets of physician performance and not the same facet of performance."

5. ***Discussion***. I suggest a re-organization of the Discussion section. First, I would create an introductory paragraph that briefly summarizes your key findings. More importantly, in that same paragraph, please indicate (in two or three sentences) how these findings *extend* previous research in this area.

   Second, create a sub-heading labeled, "Theoretical Implications." In this section, elaborate on ways in which your study adds value to the existing literature and extends theory.

   Third, create another sub-heading labeled, "Managerial Implications." In this section, explain how your findings influence managerial behavior in specific ways. One of the aims of the *Journal* is to provide managers with concrete actions that they can take in order to have a demonstrable impact on the organizations they lead. What is missing here is *how* managers should change what they do, based on your findings. These practical implications are not apparent to me.

   Finally, include a section on limitations and future research.

   By breaking up your Discussion into this generally accepted (and expected) format, you allow readers to more quickly digest the key implications of your study and improve the

clarity and readability of the Discussion. Readers will spend less time trying to locate your implications, and more time digesting them.

Minor Concerns

1. On p. 3, you state, "As Rotundo and Sackett (1999) write, 'There is no definitive way of determining whether the criterion used in a validity study is biased. Thus, there is no current method of establishing whether there is bias in performance ratings (Rotundo & Sackett, 1999: 816).'"

   This can be simplified to: "As Rotundo and Sackett (1999: 816) write, 'There is no definitive way of determining whether the criterion used in a validity study is biased. Thus, there is no current method of establishing whether there is bias in performance ratings.'"

2. I suggest you remove as many direct quote as possible in your paper, particularly the introduction. You can paraphrase these with similar effect.

The reviewers also provide constructive advice regarding several alternative avenues you could consider in developing this work. Some of this advice is explicit and much more is implicit in their various comments. As you go forward with further attempts to get this work published, I sincerely hope that you will find the reviews to be helpful. This may be a fine contribution waiting to be made. Therefore, I hope you find the proper outlet for your revision of this work.

Again, I appreciate the disappointing nature of the reviews and this letter. However, it is our intention that the review process be constructive and developmental for all authors. I sincerely hope you accept our feedback in this spirit. Thank you for the opportunity to review your work, and I hope you will continue to consider the *Journal* as a major outlet for your best theory-driven research.

Sincerely,

Brad

Bradley L. Kirkman, PhD
Associate Editor
Associate Professor and Mays Research Fellow
Texas A&M University

Feedback for the Author(s)

Reviewer # 1 ***Manuscript #AMJ-06-0696***

I really enjoyed reading your paper. It is an easy and enjoyable read with several strong features. I was really impressed by the field data you collected and your clear and careful analysis. That said, I think there are some weaknesses that might negate this work's potential impact. In particular, I'm concerned that the conceptual contribution is not sufficient to warrant publication. I'll elaborate on this concern and other more minor concerns below.

1. Your two hypotheses are interesting, but they are derived directly from status characteristics theory. There is no extension of the theory, or any challenge to its basic tenets. I don't think anyone who is familiar with the SCT literature would disagree with your hypotheses. On the contrary, we've seen a lot of evidence to support these hypotheses in the past. This doesn't mean that your work is not interesting—it just means that it may not be well suited for *AMJ*, a journal whose mission is to publish important new theoretical and empirical insights.

2. In revising this paper, I would encourage you to highlight what you think the unique contribution of your research is. If it's the link between objective and subjective performance measures, you'll have to do a better job of explaining why that is important. If it's examining the main predictions of SCT in a new context, you'll have to do more to explain why it's important to study this particular context. Finally, if it's your surprising finding – that certain types of customer-centered behaviors were *negatively* related to customer evaluations of performance for women and non-whites – then you'll need to do more to account for this.

3. One of the concerns I had with your analysis was the selection of your control variables. In SCT studies that examine the impact of *diffuse* status characteristics, such as race or sex, researchers must be careful to control for *specific* status characteristics (e.g., ksa indicators, prior performance) that may be loosely correlated with the diffuse status characteristics that are of interest. I felt there were several missing control variables in your analysis that might offer a different explanation for why race or sex differences produced a bigger gap between subjective and objective performance. First, the quality of the physician's education and training would likely be a critical indicator of actual performance. I doubt that women or minorities have worse educational experience than do white males, but they may receive less qualified training because they are discriminated against in the matching process. It would be important to include these data in order to rule this out. Second, I'd like to know more about how Prohealth coded the minority variable. Specifically, I'm curious to know how many of the minority physicians were foreign born and had noticeable accents. Accents tend to worsen others' impressions even though they may not be related to actual performance outcomes. Adding other demographic variables such as these could really help strengthen your case.

4. I tried to think of some novel hypotheses you can test with these data. One recommendation I would make is to look at zip code, assuming that these doctor-patient interactions are taking place in many different geographic areas. Some of these areas may be more racially diverse

than others or have higher percentages of Democrats, etc. Perhaps the gap is bigger in populations that have more or fewer minorities, for example. A tough test, but it might be interesting to try.

5. It also might be worth considering assertiveness and backlash effects in doctor-patient interactions. I'm sure you've seen the work showing that doctor assertiveness (or a lack of agreeableness) leads to a greater likelihood that the patient will sue for malpractice holding constant the quality of care. Perhaps race and sex is a moderator – people don't mind such assertiveness from white male doctors, but they hate it from their female and minority doctors. If there is any way you can get a measure of doctor agreeableness, that would be nice.

Good luck!

Feedback for the Author(s)

Reviewer # 2 ***Manuscript #AMJ-06-0696***

The paper is dealing with an important and timely issue of bias in performance appraisal at work (this time by customers). The authors should be also commended for their field study in which both objective and perceptual measures of performance were used to assess the hypotheses on perceptual biases based on race and sex. However, the study has a number of limitations that weaken its contribution to the literature.

1. The biggest problem I have with the study is its basic assumption that the differences (based on physicians' gender and race) in the relationship between the objective criteria of the physician performance and the customer subjective rating of professional performance quality, is due solely to perceptual bias against women or non-white physicians. There may be other potential explanations that were not addressed in the study. For example, as Table 1 indicates, women were significantly younger, less tenured, and were by far more likely to work part time than full time (a correlation of .63). It may be that these factors led patients to perceive women, as a group, as being less trained and experienced, and therefore as performing less well relative to men. This, in turn, may have influenced the patients' perception about their own personal physician, if this physician was a woman. In other words, one can speculate that patients' bias towards women physicians is based on the above objective job related differences between men and women that give support to the idea that men physicians were better than women, rather than on inherent biases against women in general. Concerning race, the number of non-white physicians is very small, but the correlation matrix still shows a tendency for white physicians to have more tenure (experience) with Prohealth than non-white physicians, which again may have contributed to the perceptual differences based on race. This does not negate the possibility that inherent gender and race related biases may have played a role in how patients rate their physicians. However, the data the study is based on cannot provide an answer to the degree in which people's perceptions of their physicians' performances were based on inherent gender and racial biases versus other job related factors. Using direct measures on the respondents' attitudes and perceptions towards women and minorities would have helped to sort out this issue. Unfortunately, it does not seem that such measures were part of the study.

Other issues:

2. Were there any significant differences (e.g., demographic and occupational differences) between those who participated in the study and those who decided not to participate?

3. The sample of non-whites is too small, and thus the results may be open to chance. Therefore you may consider running a "jackknifing" analysis, in which the regression analysis is repeated a number of times, each time with the exclusion of one individual from the group of non-whites.

4. In addition, you may consider to also analyze separately the Asian or pacific Islanders who are 10 of the 13 non-whites. Conceptually, you may want to address the question if people

are biased against all non-whites equally. One may challenge this idea, because of the success and reputation Asians have gained in this country in education and science.

5. It will be interesting to explore the effect of race and gender composition of the patients' panel on satisfaction level from the physician.

6. You indicated in the Method section that you used patient age and chronic sickness ascovariates because they may increase or decrease patients' satisfaction with the physician. Do you have any theoretical direction for the potential effect of these variables on the dependent variable? Such theoretical rationale is needed to justify the use of covariates.

7. Concerning Physician productivity, the authors indicated that intensity of each visit "was measured by the Relative Value Units (RVUs), which are coded by physicians at the end of each visit according to national coding guidelines" (p. 14). It seems to me that this self-report instrument provides the physician the latitude to inject his/her subjective evaluation on the treatment. Can you discuss this issue?

8. Concerning Physician quality, I am curious as to why the average prescription rate of Stains and ACE inhibitors for cardiovascular disease patients at Prohealth is 50%, despite the fact that Prohealth administrators promote the prescription of this medicine.

9. You indicated that that perceptual bias on one's performance is more likely to occur when the objective performance criteria is ambiguous. However, in the present case of physicians, the objective criteria seem largely clear and measurable, but still you argue for gender and race based evaluation biases. You should clarify this point.

Feedback for the Author(s)

Reviewer # 3 **Manuscript #AMJ-06-0696**

The author(s) seek to study the relationship between objective and subjective measures of performance by gender and racial groups. They find that objective and subjective measures of performance demonstrate stronger relationships for white males than for non-white or female employees. My comments follow.

1. One of the main theoretical frameworks underlying the hypotheses is one of stereotyping. Although years of research have demonstrated the existence of stereotypes, it would be useful to know to what extent the customers/patients in the current sample hold these stereotypes. Otherwise, the mechanism that links sex/race to performance outcomes is assumed to exist but not shown to exist. This link is needed to rule out alternative explanations for the findings and to provide stronger support for your theoretical rationale.

2. Page 5 (2nd full paragraph). The purpose of this paragraph needs to be carefully considered and reviewed. The statements made in this paragraph contradict each other and subsequent statements made later in the paper. For example, the 2nd sentence (By shifting the focus…) states that "shifting from subjective performance evaluations to customer rather than supervisory ratings", needs to be reworded. Customer and supervisory ratings are both subjective. Furthermore, customer satisfaction and customers' perceptions of the quality of care are two very different constructs. Yet these two labels are used interchangeably throughout the paper.

3. In the same paragraph as noted above, the authors then state that customers do not have prior knowledge about the types of physician behaviors that the organizations value. A stronger justification is needed for this statement for two reasons. First, presumably HMOs acquire reputations for being more/less patient focused. In fact, there are numerous advertisements by HMOs for this very purpose. Hence, the statement that customer-raters have NO knowledge is an awfully strong and potentially inaccurate statement. Second, one of the main assumptions that a reader of this paper needs to be convinced of is that the HMO's measure of physician performance (email, health procedures performed, prescription rates), which you assert is the objective measure, is in fact a valid measure of physician performance.

4. Furthermore, your measure of customer perception of quality consists of attention provider paid, thoroughness and competence, and opportunity to ask questions. It could very well be the case that a physician does pay attention to the patient, is thorough and competent, and provides the opportunity to ask questions during the actual visit while at the same time prescribing fewer tests and prescriptions and sending fewer emails. It is unclear that the objective and subjective forms of evaluations as operationalized in your study should be strongly related. They may be getting at different facets of physician performance and not the same facet of performance.

5. Although race is a very important demographic characteristic to consider, it is unclear why it is being considered in your study with such a small sample size. Although 12.1% is

representative of the national average, is a sample of 13 physicians large enough to draw meaningful and generalizable conclusions about race? This point dilutes the contribution of the study and its external validity.

Minor suggestions:

6. Page 3 (1st full paragraph, line 3). Should the sentence "employees' rated performance scores can be viewed…" read "cannot" be viewed?

7. Relabel figures 1-4 with specific patient-centered behavior in title (e.g., quality, accessibility).