



APPLIED PSYCHOLOGY | RESEARCH ARTICLE

Preventing halo bias in grading the work of university students

John M. Malouff, Sarah J. Stein, Lodewicka N. Bothma, Kimberley Coulter, Ashley J. Emmerton

Cogent Psychology (2014), 1: 988937



Received: 01 September 2014
Accepted: 01 November 2014
Published: 03 December 2014

*Corresponding author: John M. Malouff,
Department of Psychology, University
of New England, Armidale, NSW 2351,
Australia
E-mail: jmalouff@une.edu.au

Reviewing editor:
Peter Walla, Webster Vienna Private
University, Austria and University of
Newcastle, Australia

Additional information is available at
the end of the article

APPLIED PSYCHOLOGY | RESEARCH ARTICLE

Preventing halo bias in grading the work of university students

John M. Malouff^{1*}, Sarah J. Stein², Lodewicka N. Bothma³, Kimberley Coulter⁴, Ashley J. Emmerton¹

Abstract: Experts have advocated anonymous marking as a means of minimizing bias in subjective student assessment. In the present study, 159 faculty members or teaching assistants across disciplines were randomly assigned (1) to grade a poor oral presentation of a university student, (2) to grade a good oral presentation of the same student, or (3) not to grade any oral presentation of the student. All graders then assessed the same written work by the student. A linear-contrasts analysis showed that, as hypothesized, the graders assigned significantly higher scores to written work following the better oral presentation than following the poor oral presentation, with intermediate scores for the written work of the student whose oral presentation was not seen by the graders. The results provide evidence of a halo effect in that prior experience with a student biased the grading of written work completed by the student. The findings suggest that keeping students anonymous, as in the condition with no knowledge of the student's performance in the oral presentation, helps prevent bias in grading.

Subjects: Behavioral Sciences; Education; Social Sciences

Keywords: assessment; bias; grading; halo; marking

1. Introduction

Equitable assessment is important to both educators and students (Brennan, 2008; Houston & Bettencourt, 1999). Biases in grading can reduce equity, unfairly helping some students and unfairly harming others. One type of bias, sometimes conscious and sometimes not, involves the halo effect (see Cooper, 1981), in which prior knowledge of a person creates a positive or negative view of the

ABOUT THE AUTHORS

John M. Malouff earned a PhD in clinical psychology from Arizona State University. He is an associate professor of psychology at the University of New England, Australia, who studies teaching in higher education.

Sarah J. Stein earned a PhD at the Queensland University of Technology. She is a senior lecturer in the Higher Education Development Centre at Otago University in Dunedin, New Zealand, who studies teaching in higher education.

Kimberley Coulter is a second-year student in the master of clinical psychology program at Curtin University. She usually studies romantic relationships.

Ashely J. Emmerton is completing a graduate diploma in teaching at the University of New England, Australia. She usually studies teaching methods.

PUBLIC INTEREST STATEMENT

Experts have advocated anonymous marking as a means of minimizing bias in subjective student assessment. In the present study, 159 faculty members or teaching assistants were randomly assigned (1) to grade a poor oral presentation of a university student, (2) to grade a good oral presentation of the same student, or (3) not to grade any oral presentation of the student. All graders then assessed the same written work by the student. Statistical analysis showed that the graders assigned significantly higher scores to written work following the better oral presentation than following the poor oral presentation, with intermediate scores for the written work of the student whose oral presentation was not seen by the graders. The results provide evidence of a halo bias effect in grading and suggest that keeping students anonymous helps prevent bias in grading.



person. For instance, a student who has appeared intelligent, pleasant, and helpful in class sessions may create a positive halo. A student who has submitted poor work for a prior assessment task may create a negative halo. Bias occurs when these halos influence the grading of student work.

To reduce bias in grading, student groups and some academics advocate that instructors keep students anonymous during subjective grading (e.g. National Union of Students, 2008; Warren Piper, Nulty, & O'Grady, 1996). Although a few universities require instructors to keep students anonymous when possible during grading of some types of work (see e.g. Brennan, 2008; La Trobe University, 2014), keeping students anonymous during subjective grading is not common.

Various studies have produced evidence relating to halo-based bias in student evaluation. See Malouff, Emmerton, and Schutte (2013) for a review. This meta-analysis focuses on experimental studies.

Experimental studies have shown bias against (1) female students (Spear, 1984; Wen, 1979), (2) male students (Martin, 1972; Roen, 1992), (3) black students (Piche', Michellin, Rubin, & Sullivan, 1977), (4) white students (Fajardo, 1985; Wen, 1979), (5) less attractive students (Landy & Sigall, 1974), (6) less liked students (Cardy & Dobbins, 1986), (7) students with less appealing surnames (Erwin & Calev, 1984; Harari & McDavid, 1973), (8) students with more common names when evaluating creative work of the person (Lebuda & Karwowski, 2013), and (9) students who are not described by another person as gifted when others are described as gifted (Babad, 1980; Babad, Mann, & Mar-Hayim, 1975).

However, not every experimental study looking for bias finds it. For instance, Batten, Batey, Shafe, Gubby, and Birch (2013) found no significant evidence of grading bias resulting from written reports given to academics on the reputation of university students. Fogel and Nelson (1983) found no significant evidence of marking bias against primary students identified as having learning disabilities. It is possible that at least some of the graders in these studies saw through the experimental design, identified the research hypothesis, and intentionally assigned higher grades than they might have otherwise to students who had some negative characteristic associated with them. However, there is no empirical support for this possible explanation. Many other explanations might be as likely.

One experimental study (Malouff et al., 2013) examined bias in psychology academics and their teaching assistants and found that they assigned higher grades for *written* work to a university student if she appeared more competent and well-groomed while giving an *oral* presentation on another topic. That was the first experiment-based finding of bias in university graders. However, the results left unclear whether the halo-based bias would generalize to graders in other disciplines and whether keeping the student anonymous would prevent bias from influencing the grading.

The purpose of the present study was to fill these gaps. The hypothesis was that, across university disciplines, good and poor performances in a prior assessment would lead to biased high and low scores in marking of unrelated written work, with graders who had no prior knowledge about the student assigning intermediate scores. The bias could be based on expectations from appearance, manner, or prior performance.

2. Method

2.1. Overview

We randomly assigned graders to one of three conditions: (1) grade the oral video presentation of a student who gives a poor presentation and is not well-groomed; (2) grade no oral presentation of the student; and (3) grade the oral video presentation of a student giving a good oral presentation and looking well-groomed. The student was the same person in both oral presentation videos. Then, all graders assessed unrelated written work of the same student. The written work was the same for all graders. The aim of the method was to determine whether scores on the written work would be

biased by performance on the oral presentation, such that the student who gave the better oral presentation received higher scores on the written work than the student who gave the poorer oral presentation, with scores intermediate for the student whose oral presentation was not seen by the graders. We did not tell the graders the aim of the study or that different graders might grade different oral presentations or no oral presentation.

2.2. Participants

We recruited participants by sending emails, some individual and some en masse, to academics and graduate students to invite them to participate in a study of grading, with a reward of \$15 for completing the study. Almost all the persons invited were affiliated with one of three universities in Australia or one in New Zealand. The invitation stated that participants previously must have graded at least 20 written works of university students. We set that minimum standard to help ensure that participants were actual graders. The 159 participants included 92 women and 67 men, with a total mean age of 39.40, $SD = 12.13$. Among the graders were 95 academics, 48 higher degree research (graduate) students, 5 fellows, and 10 individuals who described themselves as “other.” The participants stated the discipline in which they did most of their grading, with the top discipline, psychology, including 26 individuals, followed by education with 14, and chemistry/biochemistry with 10. Other participants mentioned a total of 60 other disciplines, ranging from accounting to zoology, usually with one or two representatives. The graders had previously graded a mean of 152 oral presentations, with a range from 0 to 3,000, and a mean of 1,328 written assignments, with a range from 20 to 10,000.

2.3. Materials

2.3.1. Bias manipulation videos

We used for the study two three-minute videos, originally created for the study by Malouff et al. (2013), one to create a more positive image of the student’s ability, effort, and attractiveness, and the other to create a less positive image of these factors. We chose these differences to maximize bias, on the basis of earlier findings that prior experience with a student and the physical attractiveness of a student can lead to assessment bias in non-professionals (Babad et al., 1975; Landy & Sigall, 1974). The two videos each showed the same student giving a three-minute presentation in answer to this question: “Suppose that you attended a counseling agency meeting and you were invited to give suggestions about what to do to help a developmentally delayed man who intentionally swallows sharp objects such as tacks and thorns. You have 3 to 5 min to speak.” In the poorer-performance video, the student had her hair in a pony tail and wore casual clothes and no makeup. The content of her answer was unsophisticated, and her presentation did not flow well and tended to be repetitive. In the better-performance video, the student had her hair down and wore dressier clothes and makeup. She gave a presentation that provided detailed information while making good eye contact and using more varied speech tones and fewer filler sounds such as “um.”

2.3.2. Written material for grading as a dependent variable

We used the same written material as used in the study of Malouff et al. (2013). Graders saw the following question and answer described as being provided by the same person as in the video:

Question: Suppose that you work for a counseling agency and you receive the following email. Please respond within 5 minutes, using good writing style.

Hi Cathy. I have a new client coming to see me in a few minutes. This is a young woman who is terrified of contracting meningococcal disease. Every time she sees something new on her skin, such as a bruise, she fears that she has the disease and rushes home 60 kms away to her mother (a nurse) or her father to be reassured that it is nothing. What do you suggest that I do in the first session with her?

Cathy’s written response, an email exactly as she wrote it: In the first session it is most important to build rapport so that the client knows she can trust you. Firstly I’d start with training for maladaptive cognitions. Ask her what the value is in her worrying about contracting the disease. She will likely say that she wishes to avoid the disease in which case you could point out that this will only be effective if the symptoms she identifies are actually

those of meingococcal. You could provide her with specific information on the signs and symptoms of the disease so that she will know which marks on her skin and so on, to just ignore and which may indicate a real problem. Also help her to develop more effective ways to deal with fear of the disease such as self talk. For instance if she sees a mark on her skin and becomes fearfull she could say to herself, “It’s just a mark. This is not a symptom of the disease.” The most important thing would be to ensure she does not let the fear affect her life.

2.4. Procedure

Using online research materials, we randomly assigned graders to view one of two video oral presentations of “Cathy” or to a condition in which they were informed that Cathy had submitted a video oral presentation that had been graded by someone else. Participants in the two video conditions assigned a grade for the oral presentation. The grading instructions stated that the student gave the oral presentation as part of a third-year undergraduate psychology course and had three minutes to prepare before beginning. After completing assessment of the oral presentation or learning that someone else had graded the presentation, graders assigned a score for the written answer, following these instructions: “Assume for the sake of marking that the content is fine. Assign a mark on the basis of writing quality, including clarity, sentence structure, paragraph structure, punctuation, and spelling, in the context of having 5 min to answer by email in class.” All participants graded the same writing. We wanted graders to focus on writing quality rather than correctness of the response to avoid variance relating to differing levels of grader knowledge relating to the question. Written instructions for each assessment task indicated that grades could range from 0 to 100, with letter grades equal to below 50 = fail, 50 to under 65 = pass, 65 to under 75 = credit, 75 to under 85 = distinction, and 85 and over = high distinction. This is a common grading scale in Australian universities.

3. Results

3.1. Grader assignment to condition

Random assignment led to 57 graders viewing the student give the poor oral presentation, 54 not viewing any oral presentation, and 67 graders viewing the student give the good presentation. The numbers were not equal because some individuals who accessed the study web site and who were thus assigned to condition did not complete the study.

3.2. Comparison of graders in the three conditions

The graders assigned to the three conditions had no significant differences, as shown by χ^2 or ANOVA, with regard to age, sex, number of oral presentations previously graded, number of written assignments previously graded, or whether they were in psychology or not. Table 1 shows the characteristics of the three groups and the results of statistical comparisons of the groups.

Table 1. Characteristics of graders in each of three conditions

Characteristic	Condition			χ^2/IF	p
	Poor oral presentation	No oral presentation	Good oral presentation		
Women/men	32/25	33/21	27/21	$\chi^2 = 0.35$	0.93
Psychologist/not	9/48	10/44	10/38	$\chi^2 = 0.45$	0.80
Age mean (SD)	39.54 (11.32)	40.72 (11.42)	37.75 (13.78)	$F = 0.77$	0.47
Mean number of oral presentations graded before	124.02 (218.21)	190.26 (467.72)	142.63 (326.16)	$F = 0.52$	0.60
Mean number of written assignments graded before	1352.09 (2258.88)	1568.58 (2071.78)	1033.19 (1242.35)	$F = 0.95$	0.39

Notes: Not all participants had data on all variables. No group differences were statistically significant, as evaluated by χ^2 for nominal data and ANOVA for interval-level data.

Table 2. Means and standard deviations of conditions on grades assigned for oral presentation (manipulation-check variable) and written task (dependent variable)

Variable	Condition		
	Poor oral presentation	No oral presentation	Good oral presentation
Oral presentation grade	60.51 (8.47)*	None	71.50 (10.38)*
Written task grade			
Full sample (N = 159)			
Winsorized	62.19 (9.56)*	64.53 (9.59)	65.60 (9.18)*
Unwinsorized	62.28 (9.80)	64.53 (9.59)	65.35 (10.24)
Non-psychologists only (N = 130)			
Winsorized	61.83 (10.16)*	64.97 (10.01)	65.47 (9.66)*
Unwinsorized	61.94 (10.42)	64.97 (10.01)	65.16 (10.92)

*Group difference statistically significant at $p < 0.05$ as determined by ANOVA.

3.3. Manipulation check

The distribution of assigned oral presentation scores was within the limits of normality, with no outliers. Table 2 shows the mean scores assigned to the student in the two conditions in which participants graded an oral presentation. The good student oral presentation received a significantly higher mean oral presentation score than that of the other student, $t(96) = 5.76, p < 0.001$. The mean difference was greater than one standard deviation, indicating that the graders found the “good” presentation much better than the other one.

3.4. Halo bias results

To test for the existence of a halo-based grading bias, we used a one-tailed linear-contrasts analysis. That test examined whether there was a linear progression in mean writing-task scores from the condition with the poor oral presentation to no viewed oral presentation to the good oral presentation. We used a linear-contrasts analysis because our hypothesis involved linear effects from a negative bias situation to no bias to positive bias. We used a one-tailed test because our hypothesis met the standards of Kimmel (1957) for use of a one-tailed test in that (1) the hypothesis was a unidirectional application of a theory, that of halo effects, with no theory suggesting findings in the opposite direction, (2) differences in the opposite direction would be meaningless, and (3) under no circumstances would findings in the opposite direction be used as a basis for a course of action.

The assigned writing-task scores created a roughly normal distribution by visual inspection of the graphed data, but we found three outliers in the scores. We identified outliers by using a box plot of scores showing the 25th and 75th percentiles. The difference between these two percentiles is called the interquartile range (IQR). Following standard criteria (Dawson, 2011), we considered scores farther than 1.5 IQR from the nearest edge of the IQR to be outliers. We completed the main analysis first with the assigned scores winsorized (changed to the nearest non-outlier score), and then with the scores unwinsorized. Following the suggestion of Field (2013), we winsorized the scores because they seemed to be real scores more extreme than one would expect in a normal distribution. With both winsorized and unwinsorized data-sets, the linear-contrasts analysis was statistically significant.

With the winsorized data, the linear-contrasts analysis showed a significant effect, $t(156) = 1.90, p = 0.03$, indicating that mean scores for the written work rose linearly from the condition of poor prior experience with the student to no prior experience to positive prior experience. Table 2 shows the means for each group with the winsorized and unwinsorized data. Partial η^2 was 0.02.

The linear contrasts with the unwinsorized data also showed a significant effect, $t(156) = 1.67, p = 0.049$. Partial η^2 was 0.02.

We examined differences between pairs of the three individual conditions by using one-tailed t tests. With the winsorized data, only the difference between the condition with the poor oral presentation and the condition with the good oral presentation was significant, $t(103) = 1.85$, $p = 0.04$, $d = 0.36$. With the unwinsorized data, this same comparison was not significant, $t(103) = 1.57$, $p = 0.06$, $d = 0.31$, and the other comparisons of conditions with the unwinsorized data were nonsignificant.

The grading content pertained to psychology, and some of the graders were psychologists. In order to evaluate whether the above results applied to graders who were not psychologists, we repeated the above analyses with a reduced sample of 130 non-psychologists. The pattern of results was essentially the same as with the full sample. Table 2 shows the means for each non-psychologist group with the winsorized and unwinsorized data. With the winsorized data, the linear-contrasts analysis showed a significant bias effect, $t(127) = 1.68$, $p = 0.048$. Partial η^2 was 0.02. The linear contrasts with the unwinsorized data did not show a significant effect, $t(127) = 1.42$, $p = 0.08$. Partial η^2 was 0.02. Using just the non-psychologist graders, we examined differences between pairs of the three individual conditions by using one-tailed t tests. With the winsorized data, only the difference between the condition with the poor oral presentation and the condition with the good oral presentation was significant, $t(84) = 1.69$, $p = 0.048$, $d = 0.36$. With the unwinsorized data, this same comparison was not significant, $t(84) = 1.39$, $p = 0.048$, $d = 0.30$, and the other comparisons of conditions with the unwinsorized data were nonsignificant.

4. Discussion

In support of the research hypothesis, the results provide experiment-based evidence of bias across disciplines in academics and their grading assistants assessing university-student work. The finding of a significant linear-contrasts analysis indicates that bias occurred linearly from prior negative experience with the student to no experience to positive prior experience. The bias generated by the *good* and *poor* oral presentations resulted in a 3.4-point mean difference in scores on the written work, with the condition of no prior experience with the student showing an intermediate mean score for the written work. The differences between no prior experience and each prior-experience condition were not significant, but they were in the expected direction, as indicated by the significant linear-contrasts result.

The partial η^2 of 0.02 for the linear-contrasts analysis indicates that the linear-contrasts effect size was small. The lack of a statistically significant difference between no prior experience with a student and either good or poor prior experience also suggests that the bias effect was small. The results do not clearly show either positive or negative bias occurring from prior knowledge *compared* to no prior knowledge. Rather, the results show the potential for positive prior-experience bias compared to negative prior-experience bias.

The pattern of results was essentially the same with the full sample and with only the non-psychologist graders. That is important because the grading content pertained to psychology, and a prior similar study (Malouff et al., 2013) used only psychologists for grading. The present results suggest that bias occurred with non-psychologists, as well as with a mixed sample of graders from over 60 disciplines, including psychology.

The gap (3.4 points) in the winsorized data between mean writing-task scores for the two conditions in which graders first evaluated an oral presentation was somewhat smaller than that in the study of Malouff et al. (2013) that used only these two conditions and used only graders in the discipline of psychology (4.2 points). This difference might reflect a difference in grading standards in the two study samples or something else about the multidisciplinary grader sample of the present study. Nevertheless, the pattern of differences was the same in both studies.

The results of the present study add to prior findings showing a halo effect influencing judgments of various sorts (Cooper, 1981). The results suggest that keeping students anonymous, as in the condition of no prior grading of work by the student, can help prevent bias in grading. Hence, the

present findings provide additional empirical support for recommendations of university-student groups and education experts that instructors keep students anonymous during grading (e.g. National Union of Students, 2008; Warren Piper et al., 1996).

The results are consistent with findings of prior studies with high school and primary students using experimental designs that showed bias relating to sex (Martin, 1972; Roen, 1992; Spear, 1984), race (Fajardo, 1985; Piche' et al., 1977), liking of a student (Cardy & Dobbins, 1986), liking of a student's surname (Erwin & Calev, 1984; Harari & McDavid, 1973), physical attractiveness (Landy & Sigall, 1974), and thinking that a student is gifted (Babad, 1980; Babad et al., 1975). The present results extend these prior findings of bias to university grading across disciplines by individuals with, on average, extensive grading experience. On average, the graders in the study had assessed 1,328 written assignments of university students prior to the study.

The present study explored the effects of a single prior assessment of a student's performance. In actual courses, instructors might have several prior assessments of a student's work, possibly leading to a cumulative bias.

Keeping students anonymous, when feasible, in subjective grading need not create a great deal of extra work for the instructor or reduce any feedback the students receive. Students can simply put their student number in place of their name on their work. Alternatively, graders can cover each student name until grading is complete. Any feedback the grader wants to add specific to that individual student can be added after entering the grade.

In some types of subjective assessment, such as of in-person performances, keeping students anonymous is not feasible. However, with most subjectively scored assessment tasks, such as written answers on exams and most student papers, it is feasible to keep students anonymous during grading.

It remains unclear what specific differences between the two oral presentations produced the biased grading of the writing. The graders may have perceived, consciously or not, differences in appearance, competence, and effort. The purpose of the present study was to determine whether a halo bias would occur rather than to determine the precise factors leading to the bias. Future studies could systematically vary factors in order to identify which ones have a biasing effect.

The participants asked to grade the oral presentation had to enter a grade for the oral presentation before they went on to grade the written work, but that does not mean they actually watched the oral presentation. One of the limitations of the research method used was that it did not assess whether the participants asked to grade the oral presentation actually watched it in its entirety.

The present study focused on the writing quality of work of one female undergraduate student. It is possible that the results would not generalize to other students, other assessment tasks, or other levels of study. Future research could test whether bias affects assessment results in other situations. Further research could also examine whether multiple consistent prior experiences with a student create a larger bias effect than a single experience.

In sum, the results of the present study, in combination with prior findings of bias in grading, support keeping students anonymous during subjective grading when that is feasible.

Funding

The authors received no direct funding for this research.

Author details

John M. Malouff¹
E-mail: jmalouff@une.edu.au
Sarah J. Stein²

E-mail: sarah.stein@otago.ac.nz
Lodewicka N. Bothma³
E-mail: vicole.bothma@research.uwa.edu.au
Kimberley Coulter⁴
E-mail: kimberleyc1@live.com.au
Ashley J. Emmerton¹
E-mail: aemmerto@une.edu.au

¹ Department of Psychology, University of New England, Armidale, NSW 2351, Australia.

² Higher Education Development Centre, University of Otago, Dunedin, New Zealand.

³ Department of Psychology, University of Western Australia, Perth, Australia.

⁴ Clinical Psychology, Curtin University, Perth, Australia.

Citation information

Cite this article as: Preventing halo bias in grading the work of university students, J.M. Malouff, S.J. Stein, L.N. Bothma, K. Coulter & A.J. Emmerton, *Cogent Psychology* (2014), 1: 988937.

Cover image

Source: Christine.

References

- Babad, E. Y. (1980). Expectancy bias in scoring as a function of ability and ethnic labels. *Psychological Reports*, 4, 625–626. doi:10.2466/pr0.1980.46.2.625
- Babad, E. Y., Mann, M., & Mar-Hayim, M. (1975). Bias in scoring the WISC subtests. *Journal of Consulting and Clinical Psychology*, 42, p. 268. <http://dx.doi.org/10.1037/h0076368>
- Batten, J., Batey, J., Shafe, L., Gubby, L., & Birch, P. (2013). The influence of reputation information on the assessment of undergraduate student work. *Assessment & Evaluation in Higher Education*, 38, 417–435. doi:10.1080/02602938.2011.640928
- Brennan, D. J. (2008). University student anonymity in the summative assessment of written work. *Higher Education Research & Development*, 27, 43–54. doi:10.1080/07294360701658724
- Cardy, R. L., & Dobbins, G. H. (1986). Affect and appraisal accuracy: Liking as an integral dimension in evaluating performance. *Journal of Applied Psychology*, 71, 672–678. doi:10.1037/0021-9010.71.4.672
- Cooper, W. H. (1981). Ubiquitous halo. *Psychological Bulletin*, 90, 218–244. doi:10.1037/0033-2909.90.2.218
- Dawson, R. (2011). How significant is a boxplot outlier? *Journal of Statistics Education*, 19(2), 1–12.
- Erwin, P. G., & Calev, A. (1984). The influence of Christian name stereotypes on the marking of children's essays. *British Journal of Educational Psychology*, 54, 223–227. doi:10.1111/j.2044-8279.1984.tb02583.x
- Fajardo, D. M. (1985). Author race, essay quality and reverse discrimination. *Journal of Applied Social Psychology*, 15, 255–268. doi:10.1111/j.1559-1816.1985.tb00900.x
- Field, A. (2013). *Discovering statistics using IBM SPSS statistics*. Los Angeles, CA: Sage.
- Fogel, L. S., & Nelson, R. O. (1983). The effects of special education labels on teachers' behavioral observations, checklist scores, and grading of academic work. *Journal of School Psychology*, 21, 241–251. doi:10.1016/0022-4405(83)90019-5
- Harari, H., & McDavid, J. (1973). Name stereotypes and teachers' expectations. *Journal of Educational Psychology*, 65, 222–225. doi:10.1037/h0034978
- Houston, H. B., & Bettencourt, L. A. (1999). But that's not fair! An exploratory study of student perceptions of instructor fairness. *Journal of Marketing Education*, 21, 84–96. doi:10.1177/0273475399212002
- Kimmel, H. D. (1957). Three criteria for the use of one-tailed tests. *Psychological Bulletin*, 54, 351–353. doi:10.1037/h0046737
- Landy, D., & Sigall, H. (1974). Beauty is talent: Task evaluation as a function of the performer's physical attractiveness. *Journal of Personality and Social Psychology*, 29, 299–304. doi:10.1037/h0036018
- La Trobe University. (2014). *Assessment procedures*. Retrieved November 23, 2014, from <http://www.latrobe.edu.au/policy/assessment>
- Lebuda, I., & Karwowski, M. (2013). Tell me your name and I'll tell you how creative your work is: Author's name and gender as factors influencing assessment of products' creativity in four different domains. *Creativity Research Journal*, 25, 137–142. doi:10.1080/10400419.2013.752297
- Malouff, J. M., Emmerton, A. J., & Schutte, N. S. (2013). The risk of a halo bias as a reason to keep students anonymous during grading. *Teaching of Psychology*, 40, 233–237. doi:10.1177/0098628313487425
- Martin, W. D. (1972). The sex factor in grading composition. *Research in Teaching of English*, 6, 36–47.
- National Union of Students. (2008). *Mark my words, not my name*. London. Retrieved May 30, 2012, from <http://www.nus.org.uk/en/news/mark-my-words-not-my-name/>
- Piche', G. L., Michlin, M., Rubin, D., & Sullivan, A. (1977). Effects of dialect-ethnicity, social class and quality of written compositions on teachers' subjective evaluations of children. *Communications Monographs*, 44, 60–72. doi:10.1080/03637757709390115
- Roen, D. (1992). Gender and teacher response to student writing. In N. McCracken, & B. Appleby (Eds.), *Gender issues in the teaching of English* (pp. 126–141). Portsmouth, NH: Heinemann.
- Spear, M. G. (1984). The biasing influence of pupil sex in a science marking exercise. *Research in Science & Technological Education*, 2, 55–60. doi:10.1080/0263514840020107
- Warren Piper, D., Nulty, D. D., & O'Grady, G. (1996). *Examination practices and procedures in Australian universities*. Canberra: Australian Government Publishing Service.
- Wen, S. (1979). Racial halo on evaluative rating: General or differential? *Contemporary Educational Psychology*, 4, 15–19. doi:10.1016/0361-476X(79)90022-5



© 2014 The Author(s). This open access article is distributed under a Creative Commons Attribution (CC-BY) 3.0 license.

You are free to:

Share — copy and redistribute the material in any medium or format

Adapt — remix, transform, and build upon the material for any purpose, even commercially.

The licensor cannot revoke these freedoms as long as you follow the license terms.

Under the following terms:

Attribution — You must give appropriate credit, provide a link to the license, and indicate if changes were made.

You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.

No additional restrictions

You may not apply legal terms or technological measures that legally restrict others from doing anything the license permits.

