

Laughter and the Chair: Social Pressures Influencing Scoring During Grant Peer Review Meetings

Elizabeth L. Pier¹, Joshua Raclaw², Molly Carnes, MD, MS^{3,4,5}, Cecilia E. Ford^{6,7}, and Anna Kaatz⁸

¹Education Analytics, Madison, WI, USA; ²Department of English, West Chester University, West Chester, PA, USA; ³Department of Medicine, University of Wisconsin-Madison, Madison, WI, USA; ⁴Department of Psychiatry, University of Wisconsin-Madison, Madison, WI, USA; ⁵Department of Industrial & Systems Engineering, University of Wisconsin-Madison, Madison, WI, USA; ⁶Department of English, University of Wisconsin-Madison, Madison, WI, USA; ⁷Department of Sociology, University of Wisconsin-Madison, Madison, WI, USA; ⁸Center for Women's Health Research, University of Wisconsin-Madison, Madison, WI, USA.

J Gen Intern Med 34(4):513-4
DOI: 10.1007/s11606-018-4751-9
© Society of General Internal Medicine 2018

INTRODUCTION

During NIH peer review meetings (“study sections”), scientists discuss and assign “priority scores” to grant applications that largely determine funding outcomes. Although the final priority score is an average of each panelist’s score, their individual score is anchored to the scores declared publicly by those scientists (usually three) assigned to review and report on the grant application in detail. We have identified “score calibration talk” (SCT), a discourse practice where a study section member discusses and interprets the scoring rather than the content of a grant application. We found two forms: self-initiated SCT, when a panelist provides commentary about their own scoring (e.g., “So I gave it a four, which was probably generous”); and other-initiated SCT, when a panelist challenges the scoring of an assigned reviewer (e.g., “Yeah, that was generous.”). Only other-initiated SCT correlated with changes from the initial to the final score among the assigned reviewers of NIH R01 applications.¹ To gain insights into which interactional patterns accompanying SCT influence score change, we examined in detail the five cases of SCT followed by immediate declaration of a score change.

METHODS

Meticulously adhering to NIH practices, we constructed and video-recorded four study sections where groups of 8–12 oncology researchers evaluated the same 25 R01 grant applications previously submitted to NIH. Pier et al. provides a full description of our methodology.¹

Participant-reviewers evaluated R01 applications previously reviewed by study sections within NIH’s National Cancer Institute. Applications were donated by Principal Investigators identified using NIH’s public access database, then de-

identified and re-identified by the research team. All applications had been funded either on the first submission or after revision between October 1, 2012 and September 31, 2015.

We transcribed all discourse with attention to verbal and non-verbal actions, timing of pauses, marking of overlaps, and details of sound production, including laughter. We identified the speaker who initiated SCT, the assigned reviewer who was the target, and whether or not the target announced a score change. We coded and analyzed data as a case series,² treating score change as the “condition” under study and calculating the odds ratio—given “exposure” to identified patterns of discourse associated with SCT.²

RESULTS

We identified 15 cases of other-directed SCT, five featuring immediate, public score change (Table 1). Each announced score change was preceded either by shared laughter among study section members ($N = 4$) or SCT by the chair of the study section ($N = 1$). There were no cases of immediate score change when a panelist initiated SCT but no shared laughter ensued (7/15). No instances of score change in conjunction with SCT occurred without either laughter or the chair’s initiation of SCT.

Odds ratios indicated that score change was significantly more likely when laughter was present during SCT than when it was absent (OR = 16; 95% CI: 1.09, 234.25, $p = .04$). Score change was also significantly more likely when the chair initiated SCT than when another panelist initiated SCT (OR = 36; 95% CI: 1.77, 731.56, $p = .02$). It appears that both the presence of shared laughter and the role of the chair during SCT are independently important to reviewers’ score change, because the odds ratio between laughter and chair initiation was not significant (OR = 3.5; 95% CI: 0.37, 32.97, $p = .27$).

Table 1 Episodes of Score Calibration Talk (SCT) ($N = 15$): with ($N = 5$) and Without ($N = 10$) Declaration of Score Change by Reviewer

| SCT initiated by chair with shared laughter | | SCT initiated by chair without shared laughter | | SCT initiated by panelist with shared laughter | | SCT initiated by panelist without shared laughter | |
|---|-----------------|--|-----------------|--|-----------------|---|-----------------|
| Score change | No score change | Score change | No score change | Score change | No score change | Score change | No score change |
| 3 | 0 | 1 | 1 | 1 | 2 | 0 | 7 |

DISCUSSION

This was a small sample, but a unique window into the importance of social interactions in determining the scoring outcomes of research grant applications in NIH study sections. Prior research finds laughter within multi-party institutional settings serves a variety of interpersonal functions, including changing topics,³ reflecting status differences within a meeting⁴ or managing delicate interactions.⁵ Interpersonal pressure from the chair's SCT or SCT-invoked group laughter could introduce bias in an R01's final priority score by altering the range of scoring for all members of the study section.⁶ Our findings should alert grant review participants, particularly chairs, to SCT and its potential consequences. Chairs should be cautious when initiating SCT, checking to be certain their influence does not precipitate score change. While laughter can have the pro-social function of supporting solidarity, laughter should be refrained from during the serious work of SCT, where reasoned scoring is the explicit goal.

Acknowledgements: During the time this research was conducted, Elizabeth Pier was a postdoctoral fellow and Anna Kaatz was an Associate Scientist at the Center for Women's Health Research at the University of Wisconsin-Madison.

Corresponding Author: Molly Carnes, MD, MS; Department of Industrial & Systems Engineering University of Wisconsin-Madison, Madison, WI, USA (e-mail: mlcarnes@wisc.edu).

Funding The research reported in this paper was supported by the National Institute of General Medical Sciences of the National Institutes of Health under Award Numbers R01GM111002 and R35GM122557.

Compliance with Ethical Standards:

Conflict of Interest: The authors declare that they do not have a conflict of interest.

Publisher's Note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

REFERENCES

1. Pier EL, Raclaw J, Kaatz A, Brauer M, Carnes M, Nathan MJ, et al. 'Your comments are meaner than your score': score calibration talk influences intra- and inter-panel variability during scientific grant peer review. *Res Eval.* 2017;26(1):1–14.
2. Kooistra B, Dijkman B, Einhorn TA, Bhandari M. How to design a good case series. *J Bone Joint Surg Am.* 2009;91 Suppl 3:21–6.
3. Holt E. The last laugh: shared laughter and topic termination. *J Pragmat.* 2010;42 (6):1513–25.
4. Glenn P. Interviewees volunteered laughter in employment interviews: a case of "nervous" laughter? In: Glenn P HE, ed. *Studies of Laughter in Interaction.* London: Bloomsbury Academic; 2013:255–75.
5. Raclaw J, Ford CE. Laughter and the management of divergent positions in peer review interactions. *J Pragmat.* 2017;113:1–15.
6. Pier EL, Brauer M, Filut A, Kaatz A, Raclaw J, Nathan MJ, et al. Low agreement among reviewers evaluating the same NIH grant applications. *Proc Natl Acad Sci U S A.* 2018;115(12):2952–7.