Routledge
Taylor & Francis Group

Social Epistemology

A Journal of Knowledge, Culture and Policy

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/tsep20

The Influence of Disciplinary Origins on Peer Review Normativities in a New Discipline

Kacey Beddoes, Yu Xia & Stephanie Cutler

To cite this article: Kacey Beddoes, Yu Xia & Stephanie Cutler (2022): The Influence of Disciplinary Origins on Peer Review Normativities in a New Discipline, Social Epistemology, DOI: 10.1080/02691728.2022.2111669

To link to this article: https://doi.org/10.1080/02691728.2022.2111669

SOCIAL EPISTEMOLOGY
https://doi.org/10.1080/02691728.2022.2111669

Routledge
Taylor & Francis Group

ARTICLE

The Influence of Disciplinary Origins on Peer Review Normativities in a New Discipline

Kacey Beddoes[a], Yu Xia[b] and Stephanie Cutler[c]

[a]College of Engineering Dean's Office, San Jose State University, San Jose, CA, USA; [b]Instructional, Design, and Technology program, Teachers College, Emporia State University, Emporia, KS, USA; [c]Leonhard Center for the Enhancement of Engineering Education, Penn State, University Park, PA, USA

ABSTRACT

STS scholarship has produced important insights about relationships between the roles of peer review and the social construction of knowledge. Yet, barriers related to access have been a continual challenge for such work. This article overcomes some past access challenges and explores peer review normativities operating in the new discipline of Engineering Education. In doing so, it contributes new insights about disciplinary development, interdisciplinarity, and peer review as a site of knowledge construction. In particular, it draws attention to an aspect of peer review not previously discussed – how peer review normativities are shaped by disciplinary origins. A content analysis of peer review documentation revealed that a hyperfocus on methods, which can be traced back to disciplinary origins, continues to be a guiding normativity. However, interviews with editors revealed that they do not acknowledge that normativity. Implications of those findings and their misalignment are discussed, as are contrasts with the history of other disciplines.

KEYWORDS

Peer review; discourse; engineering education; methods; interdisciplinary

Peer review is part of the social construction of knowledge and the functioning of scientific communities, and, as such, has been a subject of interest among STS scholars. What scholarship gets funded and published in scientific outlets is the result of social processes and deliberations that reflect how communities think and what they value (Hirschauer 2010; Huutoniemi 2012; Jasanoff 1985; Langfeldt 2001; Mallard, Lamont, and Guetzkow 2009; Roy 1985). As such, those processes and deliberations and the resulting published artifacts merit examination as sites of knowledge construction. Peer review has been examined in the contexts of academic publishing (Beddoes 2014b; Fitzpatrick 2010; Lipworth and Kerridge 2011; Myers 1990; Origgi 2010; Paltridge 2014; Siler and Strang 2017; Tobin and Roth 2002), as well as funding grant proposals (Abrams 1991; Lamont 2009; Mallard, Lamont, and Guetzkow 2009; Porter and Rossini 1985; Travis and Collins 1991), among other sites.

That prior research has produced important insights about power dynamics, interdisciplinarity, the conservative tendencies of peer review, and conceptualizations of fairness, among other topics, across a wide range of disciplines. However, it has also been limited by several notable empirical research gaps and challenges related to access (Hirschauer 2010; Siler and Strang 2017). First, it has been limited in terms of access to the decision letters that authors receive and correspondence with editors that resolve contradictions and communicate directions for revision (Hirschauer 2010, 73; Simon, Bakanic, and McPhail 1986). Second, it has been limited in terms of access to manuscripts that

CONTACT Kacey Beddoes kacey@sociologyofengineering.org College of Engineering Dean's Office, San Jose State University, San Jose, CA, USA

are rejected rather than published (Siler and Strang 2017). Third, it has been limited in terms of the number of communities that have been open to being studied because usually those involved in peer review decisions do not want to be investigated (Chubin and Hackett 1990). For these and other reasons the "operative nucleus of peer review processes has largely remained a 'black box' to analytical empirical research" (Hirschauer 2010, 73–74).

The fact that peer review is largely black-boxed has implications for individuals who want to participate in a field, as well as for understandings of the social construction of knowledge. For individuals who want to participate in a field, be they less experienced scholars or scholars from one field trying to participate in a different field, the inability to see the micropolitics and normativities of peer review constrains their ability to publish in that field. In the case of the latter (scholars from one field trying to participate in a different field), interdisciplinarity is therefore circumscribed. For STS scholars who seek to understand the social construction of knowledge, the black-boxing of peer review limits understandings of disciplines' epistemologies, boundary work, innovation or maintenance work, and changes thereto over time. For both groups (those wanting to participate in a field and STS scholars) the black-boxing of peer review limits their ability to identify, critique, and/or challenge the biases, power dynamics, and problematic normativities manifested in peer review processes.

This article makes novel contributions to the peer review research landscape in several ways: first by addressing the gaps in empirical research identified in prior scholarship, second by exploring a new discipline, and third by highlighting a topic that has not been the subject of previous enquiries, namely the influence of disciplinary origins on peer review. In contrast to science education research, which has shown much interest in interrogating its own peer review practices, (Eisenhart 2002; Roth 2002a, 2002b, 2002c, 2004; Scantlebury 2002; Tobin 2002), engineering education research has not received the same level of critical attention, and this article therefore contributes some of the first insights about peer review in that discipline.

Our project ultimately aimed to identify and better understand the topical, methodological, theoretical, and epistemological normativities (i.e. beliefs, ideologies, and practices) of engineering education research (EER) and how they are enacted through peer review processes.[1] This mixed-methods article presents findings from peer review documents and interviews with editors. The documents included reviews, decision letters, and correspondence with editors for both published and rejected articles, thus overcoming common limitations encountered during peer review research.

The remainder of the article proceeds as follows—first, salient background information about the construction and emergence of the discipline of Engineering Education is summarized. Specifically, the role that a discourse of methodological rigor played in its construction is explained. Next, the two-phase mixed methods study design is described. Following that, the peer review document content analysis findings are presented, revealing that a preoccupation with methods was the dominant theme to emerge from the documentation. The next section then describes interviews conducted with editors and presents findings on misalignment between the documentation and the editor interviews. The discussion and conclusion sections elaborate on implications of that misalignment, as well as implications of the preoccupation with methods, and differences from other disciplines.

Disciplinary Origins

A discourse of methodological rigor played an instrumental role in constructing the discipline of Engineering Education (Beddoes 2014a). While institutional structures, such as university departments, Ph.D. programs, conferences, and national centers, marked the creation of Engineering Education as a new discipline, it was a discourse of methodological rigor that did the boundary work necessary to distinguish engineering education research from what came before it. Beddoes (2014a) describes how 'beginning around 2003, *rigor* increasingly became a boundary word meant

to demarcate and legitimize the new field of EER' (299).[2] The discipline-building boundary work being done by the discourse of rigor was evident in quotations such as:

> The field of engineering education is in the process of reinventing itself and the January 2005 special issue of the *Journal of Engineering Education* was a milestone event in this transition . . . The four most recent guest editorials have documented this reinvention and have suggested shifts that are needed to establish engineering education as a serious and rigorous research-based discipline . . . . (Streveler and Smith 2006, 103)

> Calls for increased rigor can be understood in terms of the on-going development of engineering education as a new discipline . . . It might be said that engineering education now has the infrastructure but not the research consensus to be called a distinct discipline. In this case, calls for rigor would be an appropriate next step to developing the field of engineering education. (Borrego 2007, 5–6)

The discourse of rigor served not only to demarcate the new discipline, but also to give it legitimacy among engineers, who comprised nearly everyone in the discipline (Borrego and Bernhard 2011), and who were trained to 'equate rigor with quantitative results and traditional notions of objectivity' (Beddoes 2014a, 300). In this way, EER followed in the footsteps of psychology and epidemiology, which also vied to establish their legitimacy by emulating the methodologies of the 'hard sciences' as a means to legitimacy (Amsterdamska 2005; Sherif 1987). Despite the fact that in some other fields traditional notions of objectivity and belief in the greater rigor of quantitative research have long been critiqued as inaccurate, normative, gendered, and raced (Harding 1987, 2004, 2006; Harding and Norberg 2005; Longino 1990; Sprague 2005), such beliefs have persisted in engineering (Riley 2008; Slaton 2010) and impacted the development of EER. Later, the discourse of rigor would fade somewhat, as some of its most prominent mobilizers abandoned the term, and a discourse of methodological *diversity* and *quality* became more prominent. While the particular contours of the focus shifted, and favored terms changed, what remained was a disciplinary zeitgeist preoccupied with methods.

Today, Engineering Education as a discipline is flourishing, by some metrics, as new centers, departments, Ph.D. programs, conferences, and journals continue to be created. Currently, there are over ten journals dedicated specifically to the subject of engineering education. They vary in terms of topical and geographical foci, audience, requirements, and reputation. One commonality across the journals, however, is that the interdisciplinary nature of EER's participants and contents, as well as its relative newness, mean that there is no consensus regarding what qualifies as EER and how the discipline should develop. This is as true today as it was when first noted by Borrego and Bernhard in 2011. The journal that is the subject of this analysis is generally recognized as quite difficult to publish in (high rejection rate), and is more on the researcher end of the *researcher-to-practitioner* spectrum. It is very well-established and contains a majority of publications from U.S.-based authors.

Methods

This study utilized a two-phase, mixed-methods design. In order to better understand the discipline's current normativities, we first conducted a content analysis of peer review documentation. The documentation came from scholars who first participated in an interview about their peer review experiences with a leading engineering education journal (Beddoes, Croninger, and Cutler 2020; Cutler, Beddoes, and Croninger 2019a, 2019b; Cutler, Xia, and Beddoes 2021, 2022). The documentation included the actual reviews participants received as well as correspondence from editors.[3] As noted, these methods enabled us to overcome several common limitations of past peer review analyses, namely lack of access to data about decision-making (Hirschauer 2010) and lack of access to information about papers that have been rejected (Siler and Strang 2017).

We received ten sets of reviews for nine different articles that were rejected by the journal.[4] Four of these were quantitative studies, two were qualitative studies, two were mixed methods, and one was a systematic review article. We received twelve sets of reviews for six articles that were eventually published by the journal. Three of these were quantitative studies, two were qualitative studies, and one was a systematic review. The specific methods utilized were similar across the rejected and

accepted groups, and while some methods were more common than others in EER at that point in time (and/or less complex), both groups of papers contained more and less common and more and less complex methods. It is also worth noting that EER is essentially based exclusively on empirical research, such that articles that are not based on some kind of empirical data are not part of the discipline. The topics of the articles in the dataset were for the most part common topics regularly seen in EER, including but not limited to: teamwork, diversity, self-efficacy, motivation, assessment, competencies, and pedagogy. All articles were submitted to the journal between 2013 and 2018.

We utilized content analysis methods (Krippendorff 2019) to analyze the dataset of review documentation. Using NVivo coding software, all reviewer and editor comments were analyzed comment by comment using open coding (Strauss and Corbin 1998). From the reviews provided to us, we counted and coded 422 distinct units of comments. Multiple codes could emerge in one unit depending on the content, which is why the code counts in Tables 1–3 are higher than 422. The emergent coding underwent an iterative process among the research team, which led to the final codes of: 1) relevance and significance of the topic, 2) audience, 3) data collection or analysis, 4) transparency, 5) theoretical framework, 6) research context, 7) validity and reliability, 8) research design, 9) overall alignment, and 10) journal fit. Theoretical framework units were then further subdivided into units inextricably linked to methods concerns, and units not linked to methods. Examples of theoretical framework units that were not linked to methods were those that said the authors misunderstood, misapplied, or did not engage deeply enough with the theory, that choice of theory needed to be better justified, or that a different theory entirely should have been chosen. Lastly, codes were grouped into three larger thematic categories. These thematic categories were *Topic*, *Methods*, and *Other*.

The second data source informing this analysis was hour-long interviews with six of the journal's editors.[5] All editors who had been with the journal for longer than one year were sent email invitations describing the study and inviting them to participate in an interview. Six agreed to participate. The editor interviews were conducted in Spring 2019 after interviews with the participants who supplied the peer review documentation analyzed above. The interviews were conducted virtually, audio recorded and subsequently transcribed. Among other topics, the interviews asked the editors if they agreed with certain perceptions of authors who had participated in phase one of the study.[6] For the purposes of this article, the one perception of interest was: 'Some authors perceived that the peer review process is primarily focused on methods over other aspects of the manuscript'. Simple counting of those who agreed and those who disagreed, as well as an analysis of their explanation, allowed us to compare the documentation evidence with editors' perceptions to determine the extent of (mis)alignment.

Content Analysis Findings

Overview of Themes

Reviewer comments fell into three main themes, summarized in Table 1. Each theme was divided into sub-themes to more specifically characterize the nature of the comments. As seen in Table 1, 400 comments were related to methods, 128 comments related to the topic of the article, and 93 comments were categorized as other, which included theoretical framework, alignment, and fit with the journal. While code counting is not necessarily the ultimate indicator of importance or weight given to comments, the overwhelming prevalence of methods comments is striking. This documentation evidence therefore supports what has previously been called a *hyperfocus* on methods in EER (Beddoes forthcoming, 2011; Beddoes and Schimpf 2015).[7]

While the hyperfocus on methods spanned all types of studies, studies that utilized approaches with very specific and prescribed sets of methods rules, namely phenomenology and systematic reviews, faced particular challenges when it was perceived that they did not follow those rules. Phenomenological studies faced challenges about their research design, data collection procedure,

Table 1. Overview of themes.

Category	Code & Count	Description
Topic	Relevance and significance of the topic (101)	How the topic of the article was relevant to engineering education or not, or the importance or significance of the topic. Or, the reviewer asked for the implications of the study or a more in-depth interpretation to be presented in the conclusions.
	Audience (27)	The audience of the journal or article, usually how the article did not fit with or benefit the journal audience.
Methods	Data collection or analysis (136)	Specific issues/problems about sample size, comparability, incorrect or inappropriate analytic procedures, lower than expected effect size, specific problems with data analysis procedure, how a specific method process might have limitations that were not addressed by the authors.
	Transparency (125)	Missing information about methods or certain aspects of methods. That is, instead of critiquing certain parts, reviewers ask for more information.
	Theoretical framework (31)	The theoretical framework in terms of how the understanding of theoretical framework or construct often negatively affected the data analysis or research design. That is, the comments critique the theoretical underpinning of the adopted methods.
	Research context (42)	Issues about research contexts. For example, requesting more information on research contexts and/or how the contexts should affect how authors interpreted the findings and implications, etc. Asking for more contextual information (i.e. curriculum, research contexts, research settings, participants in the settings, learning contexts, experimental procedure).
	Validity and reliability (43)	Raised questions concerning the reliability or validity of the measurement instruments or specific analytic methods/procedure.
	Research design (23)	Raised questions about the overall study design or research design.
Other	Theoretical framework (49)	The theoretical framework was not strong enough or there were questions about how the authors approached key theoretical or conceptual constructs.
	Overall alignment (30)	Sections of the article were not well aligned, e.g. connections between sections, the overall argument of the paper, or consistency between sections.
	Fit (14)	The article did not fit the journal's scope for a variety of reasons related to methods, writing, and other aspects of the article.

and coding process. Comments indicated that authors' use of phenomenology did not usually align with what editors and reviewers understood about phenomenology. There were repeated concerns about 'how your enactment of phenomenology aligns with the major traditions of the methodology'. The comments on phenomenological studies, along with some other qualitative approaches, such as autoethnography, surfaced on-going boundary work between qualitative and quantitative normativities, which is further discussed elsewhere (Beddoes, Croninger, and Cutler 2020).

Rejected Articles

Table 2 shows the frequency of code counts for the group of rejected papers. Overall, 267 comments were about methods, 78 were about the topic, and 75 were about other issues. All the rejected papers received comments on their potential contributions to EER (SIG), and some reviewers questioned the value to the audience or the fit with the journal. However, even when a topic was thought to be useful to EER, the rejected papers encountered severe methodological challenges concerning their data collection, data analysis, the use of theoretical framework in data analysis, research contexts, research design, and/or the validity and reliability of instruments used or analysis conducted. While it was not uncommon to receive decisions of major revision or reject based on issues of topical importance, rejection was more commonly linked to perceived quality of methods and concerns related to unmet methods expectations for publishing in the journal.

For the rejected articles, there were 87 comments that revealed perceived problems with data collection or analysis procedures. Examples of such comments included:

> A paired-sample *t*-test should not be used because the sample groups include different students. This test is only used to compare groups of the same students. The authors needed to use an independent sample *t*-test instead.

Table 2. Frequency of codes for rejected articles.

	Topic		Methods						Other		
Review	SIG*	AUD	DCA	TRAN	TF	CON	VR	DES	TF	ALI	FIT
R1	7	0	3	2	2	0	2	0	0	2	0
R2	5	3	3	7	0	1	0	0	1	0	2
R3	5	1	8	17	0	3	8	2	1	0	3
R4.1	3	0	3	7	2	1	1	4	6	1	0
R4.2	3	0	18	8	7	2	4	4	16	5	0
R5	6	2	4	7	0	3	2	3	0	2	0
R6	6	0	20	12	10	5	2	4	7	10	0
R7	11	2	20	16	7	1	2	5	5	3	1
R8	11	5	3	4	0	1	1	0	1	1	1
R9	7	1	5	6	1	1	2	1	3	3	1
Total	64	14	87	86	29	18	24	23	40	27	8
		78						267			75

*SIG = significance; AUD = audience; DCA = data collection or analysis; TRAN = transparency; TF = theoretical framework; CON = research context; VR = validity and reliability; DES = research design; ALI = alignment.

> Your approach for comparing the distributions is incorrect. You should look into a Chi-Square test.

> As in any quasi-experimental study, when you cannot randomly assign students to different conditions, you should check that the two populations, face-to-face and synchronous chat, were similar in all relevant attributes: demographics, SAT/ACT scores, high school GPA, etc.

Data collection and analysis (DCA) comments also included perceived problems with participant selection (e.g. 'The selection of the interview participants does not match this empirical focus . . . '), and specific decisions made during data analysis (e.g. 'a sizeable 35.5% of the codes were on competence, a percentage that should not be ignored'.). Besides the comments on specific analyses, there were less frequent yet critical problems that were thought to lessen the potential contribution or threaten the overall rigor and quality of the study. For example, a misalignment between the theoretical underpinning, analysis, and interpretation of findings led to serious questions about the overall credibility.

The other reoccurring issue under the DCA code was concern about the sample size. These concerns surfaced for both quantitative and qualitative articles. Sample size was often linked to the generalizability of the study, as shown in the following two comments: 'Number of respondents is very limited (N = 56). In my opinion, results of this study cannot be generalized to all Engineering Mechanics Classrooms' and 'Given the size of your sample, I am unsure how generalizable results are for undergraduate engineering'. As another example, one qualitative study elicited the following comment:

> . . . I consider the sample itself to be a fatal flaw for the current research. The authors were successful in recruiting only 17 of 216 students in the population, which is a response rate of less than 8%. Moreover, the sample is likely not to be representative of the population . . .

One comment about a quantitative article did open up the space for justification and arguments given the fact of low sample size, even though the article was rejected, stating that:

> The low N is also not adequately addressed. I would expect some justification regarding the statistical strength of the study and how that low N impacts the interpretation of the findings beyond simply stating there is a low N in the limitations section.

That is, a larger sample size might not be the only way to contribute knowledge, especially generalizable knowledge, to the discipline. Beyond generalizability, small sample size also led reviewers to question the claims made in the article. For instance: 'One of the bigger challenges I have with this article is the small sample size from the survey used to support the results, conclusions, and implications of the article'.

A similar number of comments (86) on transparency issues reflected a concern that the reviewer could not properly evaluate the study due to a lack of information. Typical problems included the following: little information was provided about groups of participants, especially when the differences in participant groups should play a critical part in interpreting the results; a specific yet critical analysis step was unclear (e.g. ' . . . unclear as to how the three data sets are merged . . . '); concepts used in measurement were not defined; a lack of information on the measurement as a whole (e.g. 'The design of the quizzes and rubrics for projects was not well articulated in the paper'.); and/or a lack of justification for the appropriateness of the methods. However, though the number of transparency problems was high, comments in this category were not the most critical factor in making the final rejection decision because reviewers and editors usually expected the author(s) would be able to provide such information after revising the article.

The code of 'theoretical framework' under the methodological theme was applied to cases when a comment about theory was specifically linked to data analysis, distinguishing it from when theory was mentioned in ways not specifically linked to methods (in the Other category). In this case, a reviewer or editor questioned the theoretical soundness of data analysis due to the incorrect operationalization of constructs or concepts from the theory, immature understanding of the theory, or the alignment between constructs and data analysis. Example comments in this category included: 'Also be careful to follow the philosophical assumptions of phenomenology - phenomenology is different from constructivism. And methodologically saturation is not a concept typically used in phenomenology', 'There is a lack of theoretical clarity in the data analysis', and ' . . . it is theoretically unsound to do the same for introjected regulation . . . '. Another example that critiqued the alignment between theoretical framework and interpretations of findings stated: 'I still do not think you present a strong argument for mapping findings to Boix Mansilla, and I think it could actually be detrimental to your findings'.

Other comments on methods were made about the research context, validity and reliability of measurements and analyses, and the overall research design. Comments about the research contexts asked for clarification on the specific roles of researchers or instructors, on the learning settings or education programs pertinent to the study. For example, one comment requested more information about the setup of the study:

> Add a little detail about the number of different instructors of the course and different time slots and how these were used to distribute the treatment, e.g. if instructors have multiple sections, was that conflated with the treatment?

In another comment, similar contextual information was requested about the program reported in that study: 'I would also want to know more about the life of the program – how many years it has been in existence and how many faculty have participated over time'.

Validity and reliability comments primarily reflected concern with specific instruments used in the measurement of constructs or knowledge assessment. For example, one reviewer had ' . . . some concerns about some of the validity measures that were used such as interrater reliability and bracketing'. Validity and reliability comments on the overall research design included that a reviewer ' . . . questioned the research design, particularly the rationale for the single-interview approach and the potential problems resulting from self-selection of students into the study'. And a concern that ' . . . the data was collected for a different purpose and thus doesn't align well with the research design of this study'. Other examples included:

> You set out to use an exploratory design, yet seem to be using a deductive dominant approach in this work, which in many ways is set out by having a theoretical framework and a priori codes. There is tension between these two ideas that needs to be resolved in how you discuss your exploring of the data and presenting of themes.

> Finally, the discussion section raises the point that the students in the control and treatment group were in fact different in terms of their academic backgrounds. Since this is the case, it raises many questions about whether

> observed differences can actually be attributed to the treatment, as this sample difference raises several threats to internal validity.
>
> If you really used mixed methods, then for the qualitative phase, you should explain how you coded the data and ideally, how you promoted trustworthiness (credibility, dependability, transferability).
>
> Researchers have shown that expectancy and value components are separate (see the bullets below for some empirical evidence), which leads me to question the validity of combining these constructs. What does this combined expectancy/value construct measure? What is this construct called? Have other researchers combined the constructs in this manner? This is very confusing.

Taken together, the overall pattern of the types of comments coded indicate that methods issues were perceived to greatly affect the potential contribution that the study could make and were the most likely to negatively shape the perceived quality of the study, leading to rejection. Besides major concerns with methodological rigor, other issues concerning theoretical framework, writing, and argumentation also contributed to rejection decisions, as did being perceived as too practical or not sufficiently research-oriented. In sum, the coding of comments for the group of rejected articles revealed a preoccupation with methods over other aspects of the work.

Published Articles

By comparison, as shown in Table 3, articles that were eventually published received 133 comments related to methods, 50 related to topic, and 18 related to other issues. In all categories, there were fewer comments than the respective numbers of comments for the rejected articles. This comparison gives a broad idea that published articles generally received fewer comments, especially for the codes within the Methods category.

Not only was the overall number of methods comments much smaller (133 comments compared to 267 in reviews for rejected articles), but the reviews also started to include positive comments, and the majority went into nuances rather than noting the overall methods were inappropriate. For instance, 'The methods are very well thought out', and 'the information about effect sizes could be removed as long as it's stated what ES is used. However, meta-analyses generally reported Hedge's g rather than Cohen's d'.

Notably, compared to the rejected group, comments for the published group were more often posed as questions rather than statements about unfixable problems that warranted rejection.

For example, one *transparency* comment asked:

Table 3. Frequency of codes for eventually published articles.

	Topic		Methods						Other		
Review	SIG*	AUD	DCA	TRAN	TF	CON	VR	DES	TF	ALI	FIT
P1.1	3	1	7	4	0	0	2	0	0	0	2
P1.2	0	0	1	1	0	0	0	0	0	0	0
P1.3	2	0	4	7	0	6	0	0	0	0	0
P1.4	5	1	0	0	0	0	0	0	0	0	0
P2	3	0	4	12	0	5	2	0	1	0	0
P3	7	4	2	2	0	0	1	0	1	0	1
P4.1	3	0	3	1	0	5	3	0	3	1	1
P4.2	3	3	3	2	2	2	2	0	4	1	2
P5.1	5	0	5	5	0	5	4	0	0	0	0
P5.2	0	1	6	1	0	0	1	0	0	0	0
P6.1	3	3	13	3	0	1	3	0	0	1	0
P6.2	3	0	1	1	0	0	1	0	0	0	0
Total	37	13	49	39	2	24	19	0	9	3	6
		50						133			18

*SIG = significance; AUD = audience; DCA = data collection or analysis; TRAN = transparency; TF = theoretical framework; CON = research context; VR = validity and reliability; DES = research design; ALI = alignment.

> Where is the analysis section? It is particularly important to explain Mahalanobis distance (MD) method which is not a commonly used method. What was included in the model? Why was this method used? Where the three theories included in one model? Any concerns about multicollinearity? The results are unclear without this information.

Similarly, a *validity and reliability* comment asked:

> The authors discuss the reliability of the rubric, but there is no discussion of the validity of the rubric. How do we know that the rubric is an accurate measure of the task at hand? What process was used to develop the rubric and to ensure its validity across the several dimensions (content, face, etc.).

Moreover, the tone sounded more like negotiations or suggestions rather than negative critique. For instance, one *research context* comment said:

> Cost Benefit Analysis: they might want to mention that while the video never changes, the yearly interactions will change and should be more responsive to the knowledge of the individual students. The authors might suggest a time scale for remaking the video, causing the COB estimate to no longer be linear.

Generally, the Methods comments for the eventually published articles recognized the quality of the work before going into nuanced considerations. Additionally, codes such as *theoretical framework* and *research design* reflected overall alignment and quality. There were no comments around *research design* for the published articles and only two comments that mentioned theoretical framework. One of those comments could have been addressed by simply changing the title of the section to something other than 'Theoretical Framework':

> . . . what you've currently termed the Theoretical Framework does not, at least in the manuscript, function as a theoretical framework (and, in fact, it is not a single framework but rather a synthesis of the literature to identify the characteristics of ill-structured problems).

The other theoretical framework comment in Methods asked for a justification for a specific approach chosen under the larger theoretical framework, rather than pointing out anything was wrong or unclear.

Interview Findings – Misalignment with Content Analysis

When asked if they agreed with the author perception that the journal's peer review process is primarily focused on methods over other aspects of the manuscript all six editors said they disagreed with that perception. They said this is a misperception on the part of authors. Given the findings above, those statements are not aligned with what the peer review content analysis revealed.

At the same time, most editors explicitly pointed out that methods are critical to the quality of research and should be credible, valid, clearly explained and well-justified. The editors believed that methods are the core to a study and have a large influence on how arguments are shaped and supported. Some editors also recognized that the journal has expected ways of talking about validity, reliability, trustworthiness, and meaningfulness in a methods section and expects 'empirically, methodologically robust research'. Papers that do not meet the standards should seek other publication venues. Editors' perception was that there is no rigid rule for which methods will be accepted, but the method should meet the journal's standards of credibility, validity, trustworthiness, etc. As discussed in the next section, however, there are other ways in which 'quality' can conceptualized, and methods do not have to be the 'core' of a study.

One editor offered an explanation for why authors might have the perception that there is a hyper-focus on methods. The possible explanation goes that the methods section tends to receive more concrete and actionable comments from reviewers and editors, which can create a long list around methods, which other parts of the paper, such as literature review or discussion, do not receive. In other words, this editor believed that reviewers and editors more commonly point out specific information about what is missing in the methods or the appropriateness of data analysis, but do not commonly articulate how authors should frame the overall argument or reorganize their

literature review or discussion with as much detail. Rather than countering our argument about the hyperfocus on methods, however, this could be seen as supporting it.

It is possible that editors were not considering everything we coded as 'Methods' in their conceptualization of that term when answering this question. Their conceptualization of 'Methods' was not discussed during the interviews. Nonetheless, the overwhelming number of Methods code counts seen in Tables 2 and 3 indicate that even if that were the case, there would still be misalignment with the content analysis evidence to some extent.

Discussion of Implications

While a preoccupation with methods as the dominant metric of quality appears natural, inevitable and desirable to many within Engineering Education (so much so that it is not even 'seen' or acknowledged), comparisons to other disciplines reveal that it is far from universal or inevitable. It appears to be an artifact of Engineering Education's particular disciplinary origins rather than a universal feature of all disciplines. For instance, bell hooks has identified – and problematized – 'theoretical elitism' as the dominant metric of quality within gender studies (hooks 1989). Hooks' point is that a preoccupation with convoluted and inaccessible theory dominates the field, such that it inhibits the growth and impact of the field. Shapin (2005) has levied similar critiques against dominant metrics of quality in STS scholarship, problematizing its inaccessibility, jargon, and valuation of critique alone. The field of organization studies stands as another contrasting example. Siler and Strang (2017) have quantitatively documented how the peer review process in organization studies focuses much more heavily on theory than on methods. In their study of the journal *Administrative Science Quarterly*, the vast majority of reviewer comments were about theory, and changes related theory dominated the changes made during the revision process. Indeed, moving toward theory development and abstraction has long been a common route disciplines take to establish their status and prestige (Abbott 1981, 2001; Jesiek, Borrego, and Beddoes 2010). Thus, the preoccupation with methods in EER could be otherwise and can be seen as a reflection of the particular origins from which the discipline arose (i.e. that a discourse of 'rigorous' methods was central to its origins).

It is important to note that the peer review documentation analyzed in this article represents a relatively small proportion of all submissions made to the journal between 2013 and 2018, and both the content analysis and the interviews represent only one journal in Engineering Education. Nevertheless, these findings reveal new insights that contribute to understandings of peer review, interdisciplinarity, and disciplinary development. First, this analysis reveals a continued hyperfocus on methods that was first documented a decade ago (Beddoes 2011) – indicating that that preoccupation, likely rooted in concerns over legitimacy, has not changed, even as the discipline has evolved in other ways. What we see manifesting in peer reviews today reflects the origins of Engineering Education as a discipline needing legitimacy among engineers who were skeptical about education research. As noted, the discipline was constructed through a discourse of methodological rigor. Our findings indicate that even as overt references to 'rigor' have faded, the discourse continues to hold sway among reviewers and editors and is enacted through the peer review process. In this way, the findings provide evidence of links between disciplinary origins and current peer review norms.

Aside from contributing new insights to the history of disciplinary development by showing how discursive origins and epistemological traditions influence peer review normativities, our findings also have implications for EER and other social science fields with similar tendencies. When methods are *the* metric of quality scholarship, it narrows the field of vision to the exclusion of other issues and values. For instance, it can mean that larger questions about research directions, problematic assumptions therein, and theory engagement and development do not receive the same attention as methods (Beddoes forthcoming; Beddoes and Borrego 2011). In other words, what is focused on

are micro-issues rather than what could be considered more macro-issues. That could be a problem for a new discipline whose impact and future are far from certain.

For example, it has been pointed out that EER diversity, equity and inclusion research has a tendency to *study down*, rather than *study up*, meaning that it predominantly problematizes students, particularly students in non-dominant groups, rather than problematizing institutions, structures, policies, and people in positions of power (Beddoes 2017; Beddoes and Schimpf 2021). Rather than challenging this tendency, however, the peer review process perpetuates it. This was clearly evident in our content analysis, for instance, when an article about students' self-efficacy was commended by reviewers for addressing an important topic (i.e. a topic that studies down), but the paper was rejected because of methods concerns with analytic process, small sample size, validity of instruments used, and validity of qualitative analysis, along with other issues. What we see in this case then is a preoccupation with methods and an absence of any concern about the problematic assumptions guiding the larger research agenda of self-efficacy. Theory development and engagement is another example of a macro-issue that has been neglected in favor of methods micro-issues, and many critiques have been made of the EER's lack of engagement with theorizing (Beddoes forthcoming; Beddoes and Borrego 2011; Beddoes, Schimpf, and Pawley 2014; Beddoes and Schimpf 2015).

While the discourse of methodological rigor was undoubtedly successful in establishing Engineering Education as a new discipline, it remains to be seen whether that can sustain a discipline if larger (macro) questions about research assumptions, directions, and questions, as well as theory engagement and development, are not given the same attention. Although the specificities differ, parallels can be drawn between concerns about theory in feminist scholarship being detrimental to the future of that field and concerns about methods in engineering education research. Critiquing the functions that elitist theory had come to serve in feminist scholarship in the 1980s, bell hooks argued that, 'There is a place for theory that uses convoluted language, metalanguage, yet such theory cannot become the groundwork for feminist movement unless it is accessible ... Visionary feminist theory must be articulated in a manner that is accessible if it is to have meaningful impact' (1989, 39). We might likewise ask of disciplines like EER, *Can a continued hyperfocus on methods micro-issues be the groundwork that sustains the growth of a new discipline and facilitates meaningful impact into the future?*

Aside from the implications of documenting a continued hyperfocus on methods, a second major contribution of these findings is that they surface a disconnect between editors' stated perceptions and what was found in the content analysis of reviews. The editors we interviewed either did not see the hyperfocus on methods or did not want to acknowledge it in interviews – we cannot claim to know which. In either case, this disconnect has implications for the community's ability to have a critical conversation about this normativity and the functions it serves. How can a community have a conversation about where to go if it cannot agree on where it is at? It also has implications for authors trying to navigate the peer review process because an unacknowledged normativity could be serving inadvertent gatekeeping functions and causing confusion for authors trying to understand current normativities. For comparative purposes, it would be valuable to have data on the extent to which similar disconnects exist in other disciplines.

Conclusion

By opening the black box of peer review and analyzing 'the context of evaluation – including which standards define and constrain what' a discipline sees as excellent (Lamont 2009, 3), or how it conceptualizes 'quality', peer review analysis allows us to identify similarities and differences across and between disciplines. Those differences can revolve around disciplines' epistemologies, normativities, and boundary work contours. STS scholars are not alone in their interest in such questions.

Engineering education scholars in Australia have shown an interest in studying peer review themselves (Gardner and Willey 2019; Jolly et al. 2011; Willey et al. 2011a, 2011b), and asking questions such as, *Does peer review 'actually work to discriminate against innovation and impede the development of the field'?* (Willey et al. 2011a, 1).

By identifying a dominant normativity shaping EER peer review and its provenance in disciplinary origins, the findings presented in this article reveal differences between EER and other disciplines in terms of how quality is conceptualized, perhaps most notably in terms of how theory fits into those conceptualizations. Further, by revealing a heretofore unknown disconnect between disciplinary gatekeepers and documentary evidence, the findings revealed an important point of reconciliation needed among those in the discipline. Given these findings, it seems that there are indeed peer review normativities impeding the development of EER on some fronts.

Notes

1. This article refers to both *Engineering Education* and *engineering education research*. We intend *Engineering Education* to refer to the discipline comprised of institutional structures (e.g. Ph.D. programs, centers, conferences) and *engineering education research* – abbreviated EER – to refer to the research enterprise that is an integral part of that new discipline.
2. Earlier work referenced in this article more often referred to Engineering Education and EER as a *field* rather than a *discipline*, for various reasons. While not wanting to impose a strict or clear distinction between those two terms, we chose to use *discipline* in this article because Engineering Education does now exhibit (outwardly at least) many of the features that confer disciplinary status.
3. Here, "editors" means the one Editor, as well as other people with "editor" in their title on the masthead, such as Associate Editors.
4. By "sets of reviews" we mean all reviews included along with a decision. One set of reviews typically included comments an Editor, Associate Editor, and an average of three reviewers. For some articles, we had multiple sets of reviews including both the original submission and a revision, which is why the number of sets of reviews is higher than the number of articles. For example, we had 10 sets of reviews for the 9 rejected articles because one of those articles went through two rounds of review before being rejected. The number of sets of reviews for each eventually published article varied.
5. Here the term "editors" refers to people on the masthead below the Editor with "editor" in their title, and we are not more specific than that in order to protect anonymity.
6. These perceptions were identified based on data from interviews with authors during Phase One of the study. Those findings can be found in Cutler, Beddoes, and Croninger (2019b).
7. The term *hyperfocus* in those works was inspired by Shapin (2005) discussion of *hyperprofessionalism* in history of science. As in Shapin's coining, the prefix *hyper* denotes that something not inherently bad has gone too far, such that it has negative implications or outcomes for the discipline.

Acknowledgments

We thank our participants for sharing their time, perspectives, and documentation with us. We also thank Dr. Corey Schimpf for input on this article, as well as Rachel Miriam Vriend Croninger for data collection and Eun Jung Paik for assistance during data analysis. This work is supported by the National Science Foundation through grants EEC #1929728 and EEC #1762436. The contents, opinions, and recommendations expressed are those of the authors and do not represent the views of the National Science Foundation.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work is supported by the National Science Foundation through grants [EEC #1929728 and EEC #1762436].

Notes on contributors

Kacey Beddoes is a Project Director in the College of Engineering Dean's Office at San Jose State University. She holds a Ph.D. in Science and Technology Studies from Virginia Tech. Her current research focuses on gender, interdisciplinarity, and mental wellness in engineering and engineering education. Further information about her work can be found at www.sociologyofengineering.org.

Yu Xia is an Assistant Professor at Instructional, Design, and Technology program in Teachers College at Emporia State University. Her current research interests lie in the fields of learning sciences, computer-supported collaborative learning, technologically enhanced learning, and power and politics in learning.

Stephanie Cutler holds degrees in Mechanical and Industrial Engineering as well as Ph.D. in Engineering Education from Virginia Tech. Dr. Cutler currently works as Director of Assessment and Instructional Support and Associate Research Professor in the Leonhard Center for the Enhancement of Engineering Education at Penn State. Her research interests include the peer review process, the faculty and graduate student experience, as well as educational development, including the adoption of evidence-based instructional practices in the engineering classroom.

ORCID

Yu Xia http://orcid.org/0000-0001-8718-7506

References

Abbott, A. 1981. "Status and Status Strain in the Professions." *The American Journal of Sociology* 86 (4): 819–835. doi:10.1086/227318.

Abbott, Andrew. 2001. *Chaos of Disciplines*. Chicago and London: University of Chicago Press.

Abrams, Peter A. 1991. "The Predictive Ability of Peer Review of Grant Proposals: The Case of Ecology and the US National Science Foundation." *Social Studies of Science* 21 (1): 111–132. doi:10.1177/030631291021001006.

Amsterdamska, Olga. 2005. "Demarcating Epidemiology." *Science, Technology, & Human Values* 30 (1): 17–51. doi:10.1177/0162243904270719.

Beddoes, Kacey. 2011. *Practices of Brokering: Between STS and Feminist Engineering Education Research*. Blacksburg, VA: Virginia Tech.

Beddoes, Kacey. 2014a. "Methodology Discourses as Boundary Work in the Construction of Engineering Education." *Social Studies of Science* 44 (2): 293–312. doi:10.1177/0306312713510431.

Beddoes, Kacey. 2014b. "Using Peer Reviews to Examine the Micropolitics and Disciplinary Development of Engineering Education: A Case Study." *Discourse: Studies in the Cultural Politics of Education* 35 (3): 266–277.

Beddoes, Kacey. 2017. "Institutional Influences That Promote Studying Down in Engineering Diversity Research." *Frontiers: A Journal of Women's Studies* 38 (1): 88–99. doi:10.5250/fronjwomestud.38.1.0088.

Beddoes, Kacey. forthcoming. "Reflections on the Use of Theory in Engineering Education Research: Interdisciplinary Challenges and Comparisons." In *Engineering, Social Science, and the Humanities: Has Their Conversation Come of Age?*, edited by Steen Hyldgaard Christensen, Anders Buch, Eddie Conlon, Christelle Didier, Carl Mitcham, and Mike Murphy. Dordrecht, Springer.

Beddoes, Kacey, and Maura Borrego. 2011. "Feminist Theory in Three Engineering Education Journals: 1995-2008." *Journal of Engineering Education* 100 (2): 281–303. doi:10.1002/j.2168-9830.2011.tb00014.x.

Beddoes, Kacey, Rachel M. V. Croninger, and Stephanie Cutler. 2020. "Peer Review in an Emerging Interdisciplinary Field: Identifying Differences in Authors' Experiences and Perspectives." Paper presented at Hawaii International Conference on Education, Honolulu, HI, January 4.

Beddoes, Kacey, and Corey Schimpf. 2015. "Undisciplined Epistemology: Conceptual Heterogeneity in a Field in the Making." Paper presented at American Society for Engineering Education Annual Conference, Seattle, WA, June 14-17.

Beddoes, Kacey, and Corey Schimpf. 2021. "What's Wrong with Grit?- Considerations and Better Alternatives for Engineering Education Research." Paper presented at Research on Engineering Education Symposium/Australasian Association for Engineering Education Annual Conference joint conference, Perth, Australia, December 1.

Beddoes, Kacey, Corey Schimpf, and Alice L. Pawley. 2014. "New Metaphors for New Understandings: Ontological Questions About Developing Grounded Theories in Engineering Education." Paper presented at American Society for Engineering Education Annual Conference, Indianapolis, IN, June 15-18.

Borrego, Maura. 2007. "Development of Engineering Education as a Rigorous Discipline: A Study of the Publication Patterns of Four Coalitions." *Journal of Engineering Education* 96 (1): 5–18. doi:10.1002/j.2168-9830.2007.tb00911.x.

Borrego, Maura, and Jonte Bernhard. 2011. "The Emergence of Engineering Education Research as an Internationally Connected Field of Inquiry." *Journal of Engineering Education* 100 (1): 14–47. doi:10.1002/j.2168-9830.2011.tb00003.x.

Chubin, Daryl E., and Edward J. Hackett. 1990. *Peerless Science*. Albany, SUNY Press.

Cutler, Stephanie, Kacey Beddoes, and Rachel M. V. Croninger. 2019a. "WIP: The Field of Engineering Education Research as Seen Through the Peer Review Process." Paper presented at American Society for Engineering Education Annual Conference, Tampa, FL, June 18.
Cutler, Stephanie, Kacey Beddoes, and Rachel M. V. Croninger. 2019b. "Manuscript Authors' Perspectives on the Peer Review Process of the Journal of Engineering Education." Paper presented at Research in Engineering Education Symposium (REES), Cape Town, South Africa, July 10.
Cutler, Stephanie, Yu Xia, and Kacey Beddoes. 2021. "Peering into the Black Box of Peer Review: From the Perspective of Editors." Paper presented at Australasian Association for Engineering Education and Research on Engineering Education Symposium joint conference, Perth, Australia, December 2.
Cutler, Stephanie, Yu Xia, and Kacey Beddoes. 2022. "A Growth Mindset for Peer Review: Guidelines for Writing Constructive Peer Reviews." Paper presented at Hawaii International Conference on Education, Waikoloa, HI, January 1.
Eisenhart, Margaret. 2002. "The Paradox of Peer Review: Admitting Too Much or Allowing Too Little?" *Research in Science Education* 32 (2): 241–255. doi:10.1023/A:1016082229411.
Fitzpatrick, Kathleen. 2010. "Peer-to-peer Review and the Future of Scholarly Authority." *Social Epistemology* 24 (3): 161–179. doi:10.1080/02691728.2010.498929.
Gardner, Anne, and Keith Willey. 2019. "The Role of Peer Review in Identity Development for Engineering Education Researchers." *European Journal of Engineering Education* 44 (3): 347–359. doi:10.1080/03043797.2018.1500526.
Harding, Sandra. 1987. "Introduction: Is There a Feminist Method?" In *Feminism & Methodology*, edited by Sandra Harding, 1–14. Bloomington: Indiana University Press.
Harding, S. 2004. "Rethinking Standpoint Epistemology: What is 'Strong Objectivity'?" In *Feminist Perspectives on Social Research*, edited by S. N. Hesse-Biber and M. L. Yaiser, 39–64. New York: Oxford University Press.
Harding, S. 2006. *Science and Social Inequality: Feminist and Postcolonial Issues*. Champaign, IL: University of Illinois Press.
Harding, S., and K. Norberg. 2005. "New Feminist Approaches to Social Science Methodologies: An Introduction." *Signs* 30 (4): 2009–2015.
Hirschauer, Stefan. 2010. "Editorial Judgments: A Praxeology of 'Voting' in Peer Review." *Social Studies of Science* 40 (1): 71–103. doi:10.1177/0306312709335405.
hooks, Bell. 1989. *Talking Back: Thinking Feminist, Thinking Black*. Boston: South End Press.
Huutoniemi, Katri. 2012. "Communicating and Compromising on Disciplinary Expertise in the Peer Review of Research Proposals." *Social Studies of Science* 42 (6): 897–921. doi:10.1177/0306312712458478.
Jasanoff, Sheila. 1985. "Peer Review in the Regulatory Process." *Science, Technology & Human Values* 10 (3): 20–32. doi:10.1177/016224398501000303.
Jesiek, Brent K., Maura Borrego, and Kacey Beddoes. 2010. "Advancing Global Capacity for Engineering Education Research (AGCEER): Relating Research to Practice, Policy, and Industry." *Journal of Engineering Education* 99 (2): 107–119. doi:10.1002/j.2168-9830.2010.tb01048.x.
Jolly, Lesley, Keith Willey, Gregory Tibbits, and Anne Gardner. 2011. "Conferences, Reviews and Conversations About Improving Engineering Education." Paper presented at Research in Engineering Education Symposium (REES), Madrid, Spain, January 1.
Krippendorff, Klaus. 2019. *Content Analysis: An Introduction to Its Methodology*. Thousand Oaks, SAGE.
Lamont, Michèle. 2009. *How Professors Think: Inside the Curious World of Academic Judgment*. Cambridge, Harvard University Press.
Langfeldt, Liv. 2001. "The Decision-Making Constraints and Processes of Grant Peer Review, and Their Effects on the Review Outcome." *Social Studies of Science* 31 (6): 820–841. doi:10.1177/030631201031006002.
Lipworth, Wendy, and Ian Kerridge. 2011. "Shifting Power Relations and the Ethics of Journal Peer Review." *Social Epistemology* 25 (1): 97–121. doi:10.1080/02691728.2010.534567.
Longino, H. E. 1990. *Science as Social Knowledge: Values and Objectivity in Scientific Inquiry*. Princeton, Princeton University Press.
Mallard, Grégoire, Michèle Lamont, and Joshua Guetzkow. 2009. "Fairness as Appropriateness: Negotiating Epistemological Differences in Peer Review." *Science, Technology, & Human Values* 34 (5): 573–606. doi:10.1177/0162243908329381.
Myers, Greg. 1990. *Writing Biology: Texts in the Social Construction of Scientific Knowledge*. Madison: University of Wisconsin Press.
Origgi, Gloria. 2010. "Epistemic Vigilance and Epistemic Responsibility in the Liquid World of Scientific Publications." *Social Epistemology* 24 (3): 149–159. doi:10.1080/02691728.2010.499179.
Paltridge, Brian. 2014. "Referees' Comments on Submissions to Peer-Reviewed Journals: When is a Suggestion Not a Suggestion?" *Studies in Higher Education* 40 (1): 106–122. doi:10.1080/03075079.2013.818641.
Porter, Alan L., and Frederick A. Rossini. 1985. "Peer Review of Interdisciplinary Research Proposals." *Science, Technology, & Human Values* 10 (3): 33–38. doi:10.1177/016224398501000304.
Riley, Donna. 2008. *Engineering and Social Justice*. San Rafael, Morgan & Claypool.
Roth, Wolff-Michael. 2002a. "Editorial Power/Authorial Suffering." *Research in Science Education* 32 (2): 215–240. doi:10.1023/A:1016030212572.

Roth, Wolff-Michael. 2002b. "Peer Review in Science Education: An Introduction." *Research in Science Education* 32 (2): 127–134. doi:10.1023/A:1016082925776.
Roth, Wolff-Michael. 2002c. "Evaluation and Adjudication of Research Proposals: Vagaries and Politics of Funding." *Forum: Qualitative Social Research* 3 (3): 1–29.
Roth, Wolff-Michael. 2004. "Vagaries and Politics of Funding: Beyond 'I Told You So'." *Forum: Qualitative Social Research* 5 (1): 1–15.
Roy, Rustum. 1985. "Funding Science: The Real Defects of Peer Review and an Alternative to It." *Science, Technology & Human Values* 10 (3): 73–81. doi:10.1177/016224398501000309.
Scantlebury, Kathryn. 2002. "A Snake in the Nest or in a Snake's Nest: What Counts as Peer Review for a Female Science Educator in a Chemistry Department?" *Research in Science Education* 32 (2): 157–162. doi:10.1023/A:1016069826685.
Shapin, Steven. 2005. "Hyperprofessionalism and the Crisis of Readership in the History of Science." *Isis* 96 (2): 238–243. doi:10.1086/431535.
Sherif, Carolyn Wood. 1987. "Bias in Psychology." In *Feminism & Methodology*, edited by Sandra Harding, 37–56. Bloomington: Indiana University Press.
Siler, Kyle, and David Strang. 2017. "Peer Review and Scholarly Originality: Let 1,000 Flowers Bloom, but Don't Step on Any." *Science, Technology, & Human Values* 42 (1): 29–61. doi:10.1177/0162243916656919.
Simon, Rita, Von Bakanic, and Clark McPhail. 1986. "Who Complains to Journal Editors and What Happens." *Sociological Inquiry* 56 (2): 259–271. doi:10.1111/j.1475-682X.1986.tb00087.x.
Slaton, Amy. 2010. *Race, Rigor, and Selectivity in U.S. Engineering: History of an Occupational Color-Line*. Cambridge: Harvard University Press.
Sprague, J. 2005. *Feminist Methodologies for Critical Researchers: Bridging Differences*. Walnut Creek, CA: AltaMira Press.
Strauss, Anselm C., and Juliet M. Corbin. 1998. *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory*. 2nd ed. Thousand Oaks: SAGE.
Streveler, R. A., and K. A. Smith. 2006. "Guest Editorial: Conducting Rigorous Research in Engineering Education." *Journal of Engineering Education* 95 (2): 103–105.
Tobin, Kenneth. 2002. "The Multiple Faces of Peer Review in Science Education." *Research in Science Education* 32 (2): 135–156. doi:10.1023/A:1016017809846.
Tobin, Kenneth, and Wolff-Michael Roth. 2002. "The Contradictions in Science Education Peer Review and Possibilities for Change." *Research in Science Education* 32 (2): 269–280. doi:10.1023/A:1016038414389.
Travis, G. D. L., and Harold Maurice Collins. 1991. "New Light on Old Boys: Cognitive and Institutional Particularism in the Peer Review System." *Science, Technology, & Human Values* 16 (3): 322–341. doi:10.1177/016224399101600303.
Willey, Keith, Lesley Jolly, Gregory Tibbits, and Anne Gardner. 2011a. "Investigating Research Quality and the Review Process: Some Suggestions for Improvement." Paper presented at European Society for Engineering Education (SEFI) Annual Conference, Lisbon, Portugal, September 27.
Willey, Keith, Lesley Jolly, Gregory Tibbits, and Anne Gardner. 2011b. "Gatekeeping or Filtering?: Investigating the Connection Between Peer Review and Research Quality." Paper presented at Australasian Association for Engineering Education (AAEE) Annual Conference, Fremantle, Australia, December 5.