

Gender Equity in HDR Confirmation of Candidature Milestones:

Report on findings for the Vice-Chancellor's Gender Equity Fund

April 2026

Associate Professor Connar McShane and Dr. Sophie Walker

Contents

Contents	1
Summary	2
Introduction	3
Method	4
Results.....	7
Discussion	17
Conclusions and recommendations	19
References.....	21

Summary

This document reports on findings based on a research project conducted as part of the Vice-Chancellor’s Gender Equity Project fund. This fund aimed to support a small number of targeted projects that can inform work on JCU’s institutional goals. This project examines gender equity in higher degree research (HDR) experiences of Confirmation of Candidature (CoC) milestone. Specifically, this research examines effect of the candidate gender and college level differences in Questions and Answer (Q&A) portions of CoC presentations as well as quantitative ratings of candidates on CoC forms.

The analysis revealed consistent, though not statistically significant, gender differences in Q&A dynamics during Confirmation of Candidature presentations across colleges. Female students generally experienced shorter Q&A sessions, fewer interactions, and less speaking time compared to male students, with notable variation between colleges. CASE demonstrated the most balanced and robust discussions, while CSE had shorter Q&A but afforded students more speaking time, albeit with less polite framing of questions. In contrast, CBLG showed a significant imbalance, with question-askers dominating discussion time, particularly for female students, raising concerns about a potentially “chilly climate” that could undermine equity and empowerment. Additionally, male question-askers were more likely to pose challenging questions to both male and female students and less likely to frame questions politely toward female students. These patterns suggest risks of gendered norms influencing academic experiences and highlight the need for further research and targeted interventions to ensure equitable and supportive environments.

Recommendations based on these findings were divided into short-term and long-term actions.

Short-term Recommendations:

1. Improve recording quality for CoCs to ensure reliable records are kept.
2. Chair Responsibilities to include setting expectations and actively moderating question-askers if perceived to be unduly adversarial or dominating speaking-time.
3. Develop a post-CoC pulse survey for students to track inequities in experiences going forwards.

Long-term Recommendations:

1. Expansion of research program
 - a. Conduct an inductive thematic analysis of the Q&A transcripts to better inform quality of experience by gender.
 - b. Conduct an in-depth qualitative study using a phenomenological framework to investigate the experiences of students and panel members.
 - c. Extend the investigation beyond gender and investigate intersectionality of experiences by minority groups including international students and where English is not the first language.
 - d. Increase the sample size of the current study through extracting more recordings from the GRS. Bring on an AI researcher to develop a framework for analysing the recordings and therefore provide the basis for an auditing tool.

Introduction

This project aims to assess the extent of the ‘chilly climate’ that might surround candidature milestones. A chilly climate is a commonly used term in gender equity in academia research and was coined to encompass the overt or subtle ways men and women might be treated differently within institutions that contributes to a general unwelcoming atmosphere (Hall & Sandler 1982). Candidature milestones are reflections of wider institutional norms and cultures and experiences of sexism at these sorts of events are predictors of whether a woman might want to leave academia (Biggs et al., 2018). Differences in level of professional support, mentoring and resource provision between men and women have been seen to create a general unfavorable departmental climate (Greene et al., 2010). Academic staff often have different expectations of women and men and respond differently to them in academic settings (Janz & Pyke, 2000). This might take the form of assumptions of incompetence, making sexist remarks or requiring a female student outperform a male student in order to be taken seriously (Palmgren et al., 2012). Remarks or interactions that are gendered might manifest itself in addressing women in a less polite way when asking questions (Jarvis et al 2022). The way in which language is used can either be empowering or enforce gendered stereotypes (Hassan and Zahid 2024). It might be direct use of language or interactional patterns which contribute to creating a ‘chilly climate’ for women (Lee and McCabe 2020). The persistence of a ‘chilly climate’ despite increasing parity in undergraduate and post graduate completions is one of the reasons given for the stall in progress towards gender equity in higher education institutions (Lee and McCabe 2020). It is therefore important to identify whether an unfavorable or unwelcome atmosphere exists in candidature milestones so that recommendations can be put in place. These may take the form of resilience training, social cohesion initiatives, education/training about the forms this differential treatment can take or developing supports for female students (Sandler et al 1996, Walton et al 2015).

A desktop review of research could not identify specific research published on HDR candidature gendered experiences in milestone assessments. There is also a lack of reference to gender equity in milestone and progression policies in Australia (including JCU). Victoria University was the only education provider that specifically states in their milestone guidelines that consideration should be given to diversity in composition of the panel. This guideline is a common recommendation and action in increasing gender equity in academic promotions processes and recruitment panels (for example Universities Australia Gender Equity Toolkit- Hamilton et al 2022). Despite this lack of research, there has been some research that has explored the gendered experiences of presenters during question time at academic conference proceedings. For example, audio transcriptions of questions during medical conferences have been used to demonstrate that gender is a significant factor in experiences of women at conferences (Salem et al 2021). Other analysis of live Q&A recordings found that men spoke more during question time and women are much more likely to fear backlash asking and answering questions (Jarvis et al 2022).

The aim of this project was to assess if the manner of questioning and candidate contribution expectations are different dependent on candidate gender. Differences in the quality of questioning might take the form of style of questioning (such as challenges to the candidate's skill set), interruptions, interactions or time allowed for answers. Outcomes of this project will contribute to JCU's understanding of potential drivers of attrition of women across the academic career pipeline. Specifically, this exploratory project will provide insight into potential gender biases in the evaluation of early career researchers and highlight how gendered performance expectations may shape the broader academic culture. Given the high attrition rate of candidates

and the difficulty in retaining, recruiting and promoting women into senior academic positions, this project will assist in identifying potential gender bias at an early and important career milestone stage.

Method

Confirmation of Candidature (CoC) is the first milestone during a student's candidature, generally taking place in the first 6-12 months. CoC's generally last for 1 hour, with 30 minutes of presentation by the student followed by questions and discussion of student performance amongst the committee. The committee is made up of the candidate's advisory panel, a chair (responsible for managing the session) and an independent academic (responsible for assessing student capacity and providing constructive criticism on the project). Some level of questioning by audience members during a CoC is expected as it forms part of the criteria on which the panel evaluates the readiness of the candidate to proceed in their degree. This project evaluated potential gender differences in experiences of students by assessing the recording of the question-and-answer (Q&A) time at the end of the CoC presentation as well as the quantitative assessment of student performance reported within the CoC assessment form.

Sample

Thirty-four Video CoC recordings and forms were provided by the Graduate Research School (GRS), with a target of 3 per gender (male/female) by college¹. This could be met in all cases except for the CMD where there were insufficient numbers of recorded male CoC's. In this college 2 videos/forms were analysed. Videos and forms were analysed by a single coder (SW) with several checked for indicator agreement by CM, except for the coders' college affiliation whereby SW coded CHS and CM coded CSE.

Each student and question-asker were given a reference code for anonymity. These codes were not unique to individual question askers. For example, if the same question asker appeared in multiple CoC sessions they were given a unique identifier each time. The JCU Chair of the Human Research Ethics Committee and JCU Privacy Officer were both consulted at the start of this project to ensure the safety and confidentiality of student data.

Materials

The analysis framework and data source can be found in Table 1. The analysis framework mirrors similar research conducted by Jarvis et al. (2022) on academic conference session Q&A.

Gender of Question-Asker

Presenting-gender make-up of the panel was inferred from staff profiles on the JCU Research Portal. Advisor academic level was extracted from the form where provided, where it was not included on the form this was also inferred from staff profiles on the JCU Research Portal (or general web searches for external advisors). In some cases, the academic level or gender was inferred from the CoC presentation video. Best efforts were made by the researcher to not assume gender, and where there was a lack of clarity then gender was not coded. A ratio score (*Gender of Question-Asker Ratio*) was calculated whereby male question-askers were coded as

¹ College of Arts, Society, & Education (CASE), College of Business, Law, & Governance (CBLG), College of Healthcare Sciences (CHS), College of Medicine & Dentistry (CMD), College of Public Health, Molecular and Veterinary Sciences (CPHMVS), College of Science and Engineering (CSE)

-1 and female question-askers were coded as +1. Scores were then computed as a summed variable whereby the closer to positive one, the greater number of female questioners, the closer to negative one the greater number of male questioners, and a zero-score indicating equal numbers of male versus female question askers.

Table 1: Summary of data points and sources used in this report

Data point	Source
College	CoC form
Gender make-up of panel	CoC form, CoC video recording, JCU Research Profiles Online
Academic stage of panel members (optional dependent on quality of data – i.e. may be high number of missing data points)	CoC form, CoC video recording, JCU Research Profiles Online
Advisor level (e.g. primary or secondary)	CoC form, CoC video recording
Quantitative rating on CoC form (adequate/inadequate)	CoC form
Qualitative assessment on CoC form (flagged – coding to be determined)	CoC form
Gender presentation of Student	CoC form, CoC video recording
Gender presentation of Question-asker	CoC video recording
Is Question-asker an Advisor	CoC video recording
Construct of Question Positive (complimenting) or Negative (confronting)	CoC video recording
Total Number of Interactions for QandA	CoC video recording
Time for each interaction	CoC video recording
Total time and average time of interaction	CoC video recording
Per interaction – time of questioner speaker versus time of student speaking (can calculate difference)	CoC video recording
Is it a question or a comment	CoC video recording
Challenging (Challenges did not include suggestions for improving research. Rather, this category focused on whether the question asker indicated or implied that the speaker was wrong about something): (a) questioning the validity of a claim made by the speaker, (b) questioning the speaker’s knowledge or expertise, (c) questioning the integrity or reliability of the presented research.	CoC video recording
Politeness (a) thanking the speaker for the talk, (b) beginning their remarks with a positive comment about the talk (e.g., “That was an interesting talk”), (c) thanking the speaker for their response.	CoC video recording

Speaking Time

In each Q&A session, the timestamp and duration for each question and student response was recorded. The number of interactions between student and each question-asker was also recorded. This data was used to calculate the total amount of time spoken per question asker. The total amount of time and total number of interactions as well as average time per iteration were also calculated for the CoC Q&A session to provide an overall view per student. Total time speaking for question-askers (*Total Time QA*) and students (*Total Time SR*) were also converted to z-scores for standardised comparison, given variations in question time across colleges. Finally,

a *Time-Talk Ratio* was calculated as proportion of time speaking between question-askers and students (raw total time of question-asker/raw total time of student). The higher the number, the greater proportion of speaking time dominated by the question-asker, with scores closer to one indicating more equitably distributed speaking time. Scores below one indicates a greater proportion of speaking time dominated by the student.

Question Characteristics

Each interaction was recorded as a ‘question’ (required student answer), ‘comment’ (no student response elicited) or ‘question comment’ (portion of interaction taken up with a general comment/asker did not want a response from student).

Deductive Content Analysis using predefined categories were used to analyse the Q&A time in the video recordings. This approach was used to code the content quantitatively and as such provide a quantitative summary of the experiences of students across gender and college. This study focused on several signifiers of politeness versus challenges which were coded based on the content of the question or comment. The definitions and operationalisations of polite or challenging questions/comments were adapted from research by Jarvis et al. (2022). Challenging characteristics were primarily based on instances where the question-asker questioned some aspect of the research or presentation:

- (a) questioning the validity of a claim made by the speaker, e.g. student considered X, why didn't they consider Y?
- (b) questioning the speaker's knowledge or expertise, (how much do you actually know about this theory, context, literature, etc? Justify further rationale points)
- (c) questioning the integrity or reliability of the presented research. (design of research or quality of data)

We would like to stress that “Challenging” does not automatically mean the question or comment was negative. One of the main aims of the CoC is for the student to receive feedback on aspects of their research like study design and to be able to ‘defend’ their choice of theory, for example.

Politeness characteristics were coded when the question-asker added compliments to their question or comment across 3 main categories:

- (a) thanking the speaker for the talk,
- (b) beginning their remarks with a positive comment about the talk (e.g., “That was an interesting talk”),
- (c) thanking the speaker for their response.

Questions or comments could be simultaneously coded as challenging and polite. Frequency of Challenging and Polite questions/comments were summed for total Challenging and Polite scores. Additionally, a *Challenging-Polite Ratio* was calculated by dividing Challenging frequency total by the Polite frequency total. Higher scores indicated more challenging and less polite, with scores closer to one indicating equal proportion of challenging and polite question/comments, and scores less than one indicating more polite than challenging questions/comments.

CoC Assessment Forms

Quantitative ratings on the form were calculated by the number of boxes marked ‘adequate’ or ‘inadequate’ and whether the candidate was passed, passed with support, or failed.

AI methodology

This project also aimed to examine initial feasibility of using Generative AI to audit equity in CoC experiences. CoC forms were used to validate the ability of Gen AI model to accurately extract information from forms.

Student CoC forms were uploaded to Claude (Sonnet 4.5) and Copilot using the following prompt:

“Looking at the seminar evaluation section of the attached forms, summarise in a table the number of boxes marked inadequate versus adequate”

This result was then cross checked with data extracted by SW and a level of agreement between AI analysis and human coder recorded.

Results

Data Analysis

Given the small sample size, the majority of reporting below relies upon descriptive and as such observational reporting and interpretation. For inferential statistics, non-parametric analysis was undertaken. Between group differences for gender of students were analysed using Mann-Whitney U tests. Between group differences for college were analysed using Kruskal-Wallis tests. Spearman’s *rho* correlations were used to identify significant relationships between variables.

Video Recording Confirmation of Candidature QandA

Below is a descriptive summary of the speaking time and proportion of polite and challenging question characteristics for male and female students across the colleges (Table 2). In all, there was more discussion during Q&A sessions for male students than for female students (Total Time), with more interactions and more total time spent per interaction during male student Q&A sessions than for female students. There were large observable differences between colleges in the overall Q&A time, with time ranging from 9.77 total minutes on average for CSE male students to 29.40 total minutes on average for CASE male students. For female students, total average time per student ranged from 6.53 minutes (CSE) to 18.10 minutes (CHS).

Table 2. *Descriptive Summary of Speaking Time and Question Characteristics for Male and Female Students Across Colleges*

		Summed Q&A time overall (min)	Summed Question-asker speaking time (min)	Summed speaking time student (min)	Summed Interactions (Questions/ Comments)	Mean Q&A Time Per Student (mins)	Mean Interactions (Questions/ Comments) per student
Male	CASE	88.20	41.02	39.25	57.00	29.40	19.00
	CBLG	44.23	28.37	15.87	47.00	14.74	15.67
	CHS	60.45	28.55	31.90	51.00	20.15	17.00
	CMD	42.85	24.50	18.50	43.00	21.43	21.50
	CPHMVS	46.80	25.93	20.87	48.00	15.60	16.00
	CSE	29.32	11.65	22.67	37.00	9.77	12.33
	Total Male	311.85	160.02	149.05	283.00	18.34	16.65
Female	CASE	52.43	25.52	26.92	45.00	17.48	15.00
	CBLG	39.30	24.90	14.40	31.00	13.10	10.33
	CHS	54.30	32.97	21.33	57.00	18.10	19.00
	CMD	20.00	10.17	9.83	19.00	6.67	6.33
	CPHMVS	40.30	22.68	17.62	55.00	13.43	18.33
	CSE	19.60	6.55	13.05	19.00	6.53	6.33
		Total Female	225.93	122.78	103.15	226.00	12.55

- Notes: video analysis metrics by college and by gender (n=3 for all colleges for both female and male except for CMD where male n=2).
- “Interaction” refers to each time a question-asker addressed the student with a comment/question.
- Mean total time calculated by dividing total time by number of male/female students
- Mean interactions calculated by dividing total number of questions/comments by number of male/female student,

Interaction type

Overall, the largest proportion of interactions for Male students were questions, while more Female students received comments about their work. There were observable differences between colleges by gender in regards to proportion of interaction type. For example, Male students in CASE, CBLG and CSE had a higher proportion of comments than questions. While Female students in CASE and CHS received more questions than comments. The largest proportion of interactions for Female CSE students were question- comments.

Table 3: Proportion of interactions where question asker asked a question, provided a comment or a question-comment for Male and Female students across colleges

		Proportion of interaction as:		
		Questions	Comments	Question-comments
Male	CASE	0.33	0.35	0.12
	CBLG	0.34	0.55	0.11
	CHS	0.61	0.27	0.16
	CMD	0.35	0.23	0.09
	CPHMVS	0.60	0.35	0.04
	CSE	0.35	0.41	0.24
	Total Male	0.43	0.36	0.12
Female	CASE	0.51	0.38	0.09
	CBLG	0.32	0.55	0.19
	CHS	0.49	0.39	0.11
	CMD	0.32	0.37	0.32
	CPHMVS	0.36	0.56	0.07
	CSE	0.21	0.37	0.42
	Total Female	0.40	0.45	0.15

Notes: video analysis metrics by college and by gender (n=3 for all colleges for both female and male except for CMD where male n=2). Proportion calculated by dividing no of instances (as questions, comment, or question comment) by number of total interactions for that college by gender

Speaking Time

The estimated marginal means for question askers’ Z standardised Total Time speaking demonstrates some observable differences between gender and college (Figure 1). Most colleges had a similar speaking time by question-askers for male and female students with the exception of CMD and CASE. For both CMD and CASE, more questioning time was directed towards Male students than Female students. For differences between colleges, CSE question-askers spoke observably less than other colleges, with a pairwise comparison indicating that CASE question-askers spoke for significantly longer compared to CSE ($H(5) = 12.77, p = .0.026$) (Figure 1).

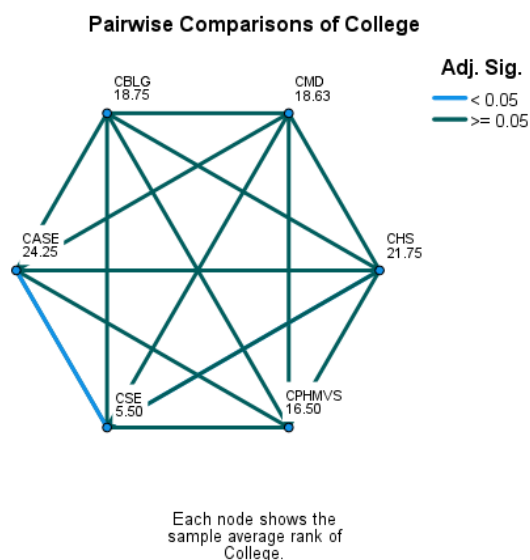
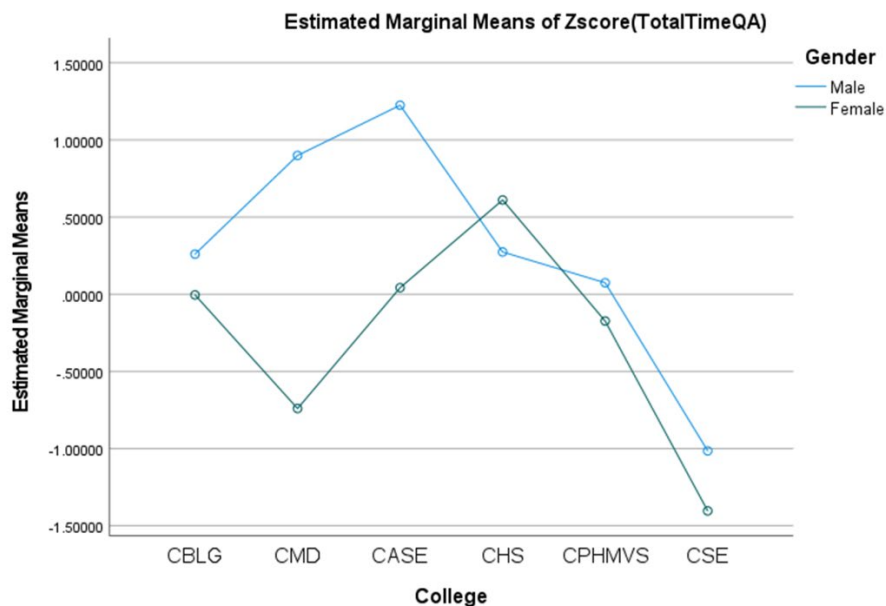


Figure 1: Total Time speaking (z standardised) by Question Askers (Top) broken down by College and gender of presenter. Kruskal-Wallis Pairwise comparison of Colleges for Total Time speaking by the Question Asker (Bottom).

Plotting of the estimated marginal means of the Z scores for Total Time speaking by the student demonstrates some observable differences between gender and college (Figure 2). Overall, across all colleges female students had less speaking time in comparison to male students, though CBLG and CPHMVS were closer in scores. Beyond these observable differences, there were no significant differences for gender or college.

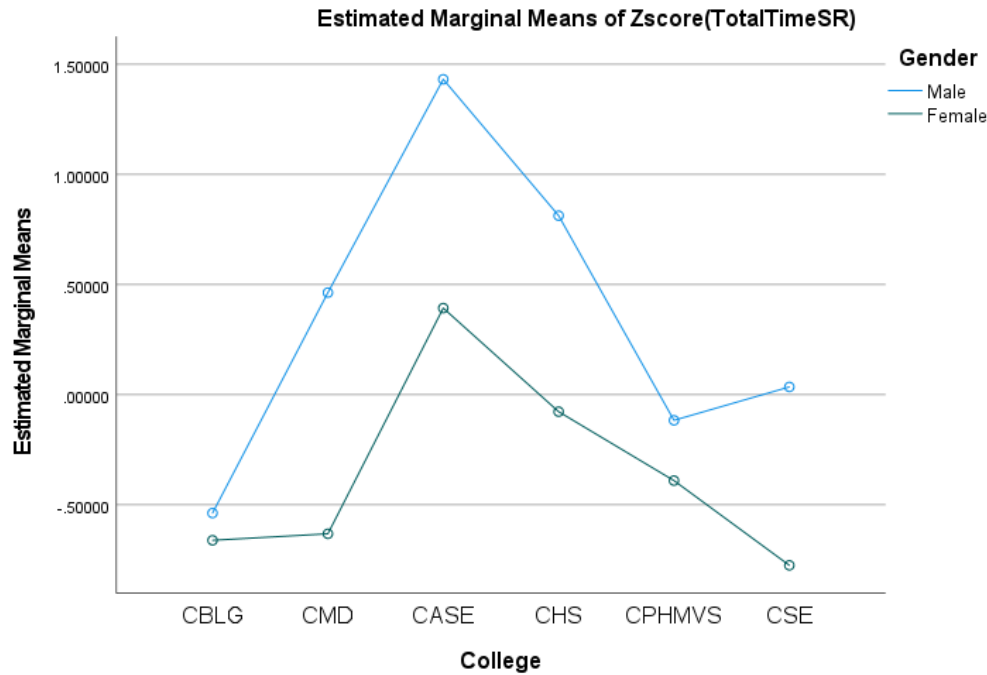


Figure 2: Total Time speaking (z standardised) by Student broken down by College and gender of presenter.

Plotting of Time Talk Ratio (ratio of speaking time between question-askers and students) shows that most colleges have a relatively equal proportion of speaking time between question-askers and students across genders (Figure 3). CSE does present below 1, indicating it is the only college in which students speak for longer than question askers (regardless of the gender of the student). The notable observable difference is in CBLG, whereby there is an inequity in speaking time for female students', with question-askers speaking for noticeably longer than students. The ratio for male students in CBLG is more equitable distribution of speaking time for question askers and students. A Kruskal-Wallis pair-wise comparison indicates a significant difference between CBLG and CSE, with CBLG question-asker having a higher proportion of speaking time than CSE, where CBLG had more speaking time afforded to question-askers ($H(5) = 13.24, p = .021$) (Figure 3).

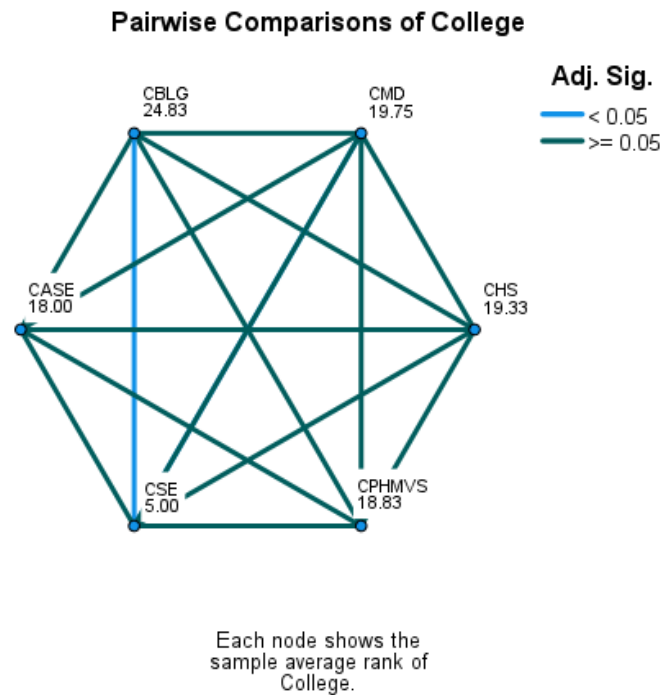
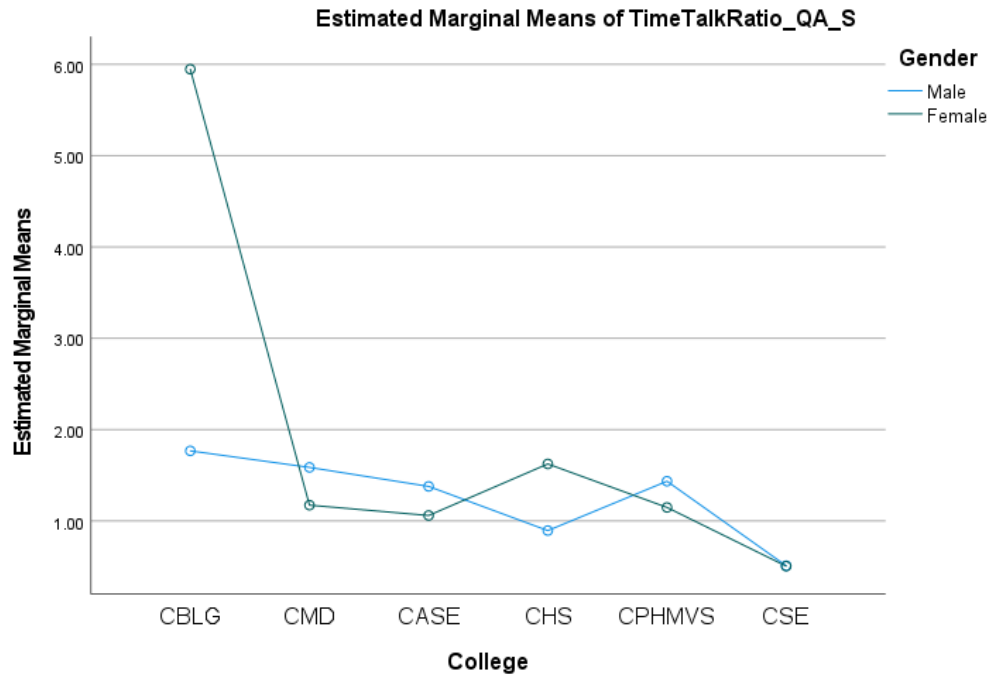


Figure 3: Time Talk Ratio (proportion of speaking time between Question-Askers and Students (Top). Kruskal-Wallis pairwise comparison of time talk ratio between colleges (Bottom).

Gender of Question-Askers

Plotting the Gender of Question-Asker Ratio (-1=Male; +1 Female) shows that there are some observable differences between college and gender of student (Figure 4). CBLG, CHS, CPHMVS report scores close to zero across both student genders, indicating that a relatively even proportion of male and female question-askers question both male and female students. CMD has more males questioning male students and more females questioning female students, whilst CSE leans toward more males questioning female students. In CASE, overall more females than males question students. This effect is noticeably more pronounced for female students than male students. Between group comparisons indicated no significant differences beyond these observable differences. A Spearman’s *rho* correlation indicated a significant relationship between Question-Asker Gender Ratio and frequency of Challenging Validity of Student Claim, with male question-askers more likely to ask these types of questions to students overall ($rho(34) = -.344, p = .039$). When examining male and female students separately, this relationship was no longer significant. However, there is a significant positive relationship of moderate effect size between Question-Asker Gender Ratio and frequency of overall politeness in questioning towards female students ($rho(17) = .541, p = .025$), indicating that female questioners are more likely to be polite to female students

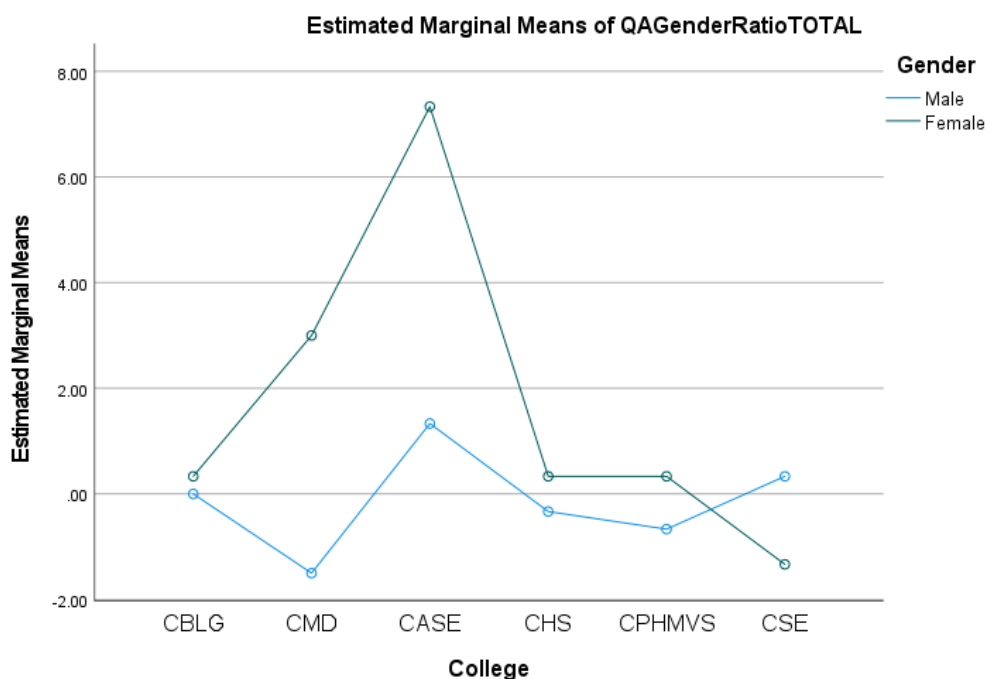


Figure 4: Gender of Question-Asker Ratio (-1=Male; +1 Female) by gender and college

Question Characteristics

Plotting of the estimated marginal means of the frequency of Challenging questions asked presents some observable differences between gender and college (Figure 5). Overall, with the exception of CASE and CHS, female students were challenged observably less than male students. CMD and CBLG present the observably largest differences between male and female students. An independent samples Mann-U Whitney test that compared male and female students across all colleges demonstrates that, though there is not a significant difference

($U=90.00$; $z= -1.89$; $p=.059$), it is approaching significance with a medium effect size $>.3$ [$u/(n1 \times n2)$]. Overall, the finding represents a difference between male and female students in how frequently they are asked challenging questions.

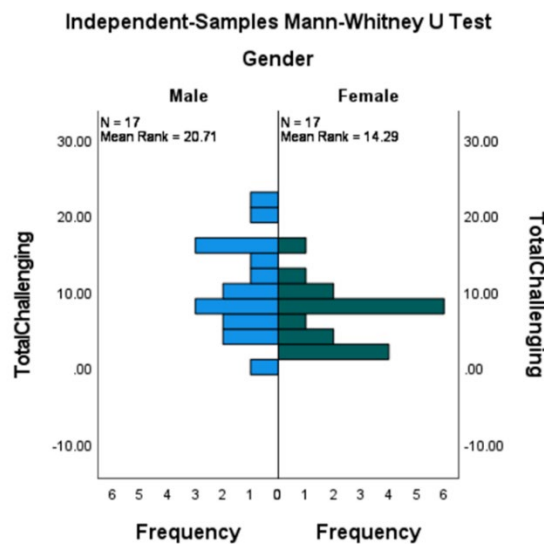
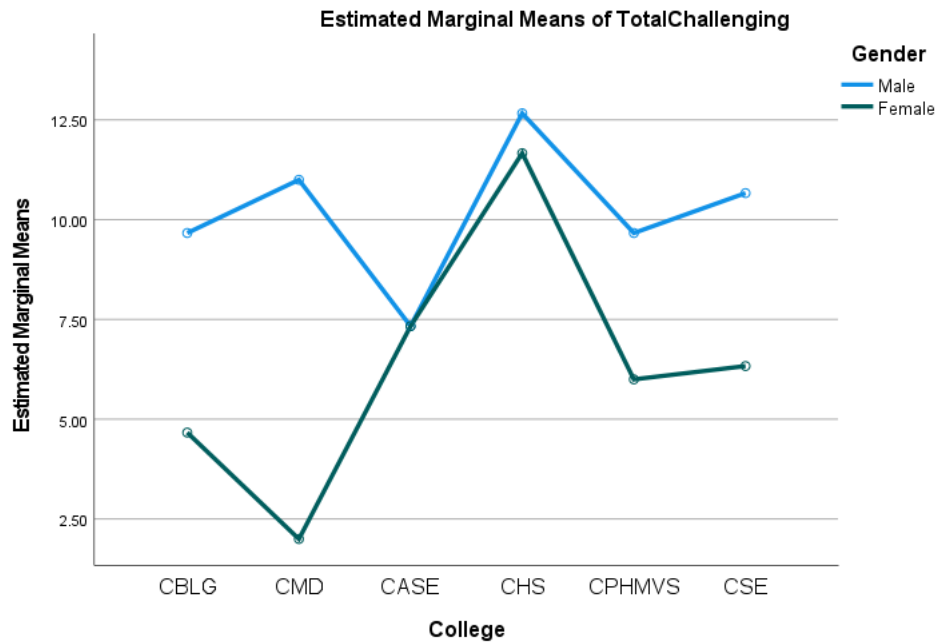
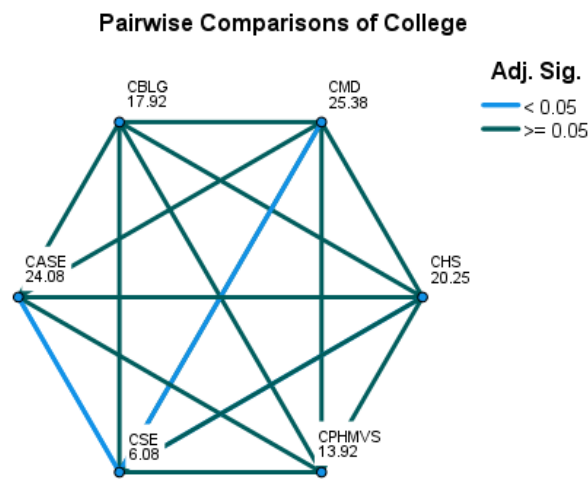
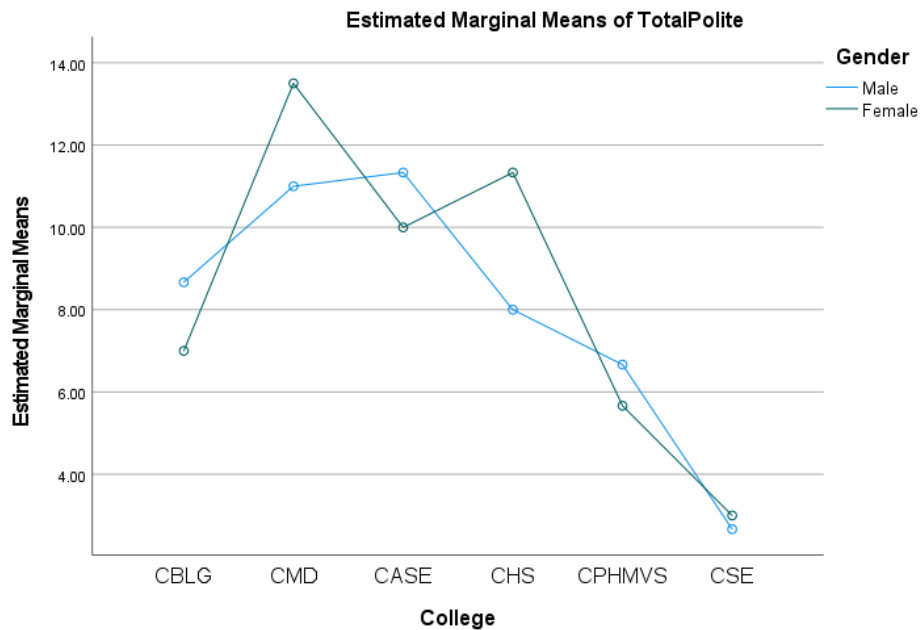


Figure 5: estimated marginal means of the frequency of Challenging questions asked by Question-Askers (Top). Mann-U Whitney by gender (Bottom)

Plotting of the estimated marginal means of the frequency of Polite question/comments asked by question-askers presents some observable differences between gender and college (Figure 6). Though relatively similar within colleges, CMD and CHS represented the largest observable gap

with a higher frequency of Polite questions/comments towards female students than male students. Between colleges, CSE is observably different to the other colleges with overall the lowest frequency of Polite questions/comments towards students. A Kruskal-Wallis pairwise comparison indicates significant differences between CASE and CSE, and CMD and CSE, with CASE and CMD both asking polite questions/comments to students more frequently than CSE ($H(5) = 14.40, p = .013$) (Figure 6).



Each node shows the sample average rank of College.

Figure 6: Estimated marginal means of the frequency of Polite question/comments asked by Question-Askers by gender and college (Top). Kruskal-Wallis pair-wise comparison by college (Bottom)

Quantitative form ratings

Of the 35 forms, 3 received inadequate ratings in the seminar evaluation section (Table4):

Table 4: Summary of CoC forms which received inadequate ratings, total number of forms was 35

Gender	College	Inadequate rating	Candidature confirmed
M	CASE	<ul style="list-style-type: none"> • Clear statement and justification of the aims of the research • Understanding of theoretical / historical context, • Hypotheses / research questions, • Overall comprehensiveness Total=4	Confirmed with progress support
F	CBLG	<ul style="list-style-type: none"> • Method selected for study • Understanding of methods Total=2	Candidate confirmed
F	CMD	<ul style="list-style-type: none"> • Understanding of methods • Overall comprehensiveness Total=2	Confirmed with progress support

Generative AI

Investigation throughout this project found that AI is not recommended for the qualitative assessment of CoC Q&A. There are ethical issues associated with the use of generative AI to perform qualitative analysis. Qualitative analysis involves interpreting and reflecting on themes found in data. The predictive algorithms behind generative AI models do not have the ability to ‘make meaning’ out of the language being analysed instead it may simulate an approximation of the process (Webster 2025). Generative AI also has the tendency to replicate societal inequities in its results (Ukanwa 2024). This is an important consideration, especially as this project involves investigating equity. Therefore AI should be used with caution in projects such as these.

However, this project did find that Generative AI could be used as an accompaniment to quantitative analysis with further refinement. For example, by extracting and identifying timestamps from CoC video transcripts.

Quantitative ratings on CoC forms could not completely be successfully extracted by Generative AI tools. The two AI models tested (Copilot and Claude), only identified a third of the inadequate markers correctly (Table 5).

Table 5: Summary comparison of data extraction from CoC forms performed manually versus AI tools (CoPilot, Claude)

Method	Number inadequate identified
Manual data extraction	3
Copilot	1
Claude	1

Fairly extensive preparation of files was required in order to make this process efficient (Figure 7). CoC forms are digitally signed and often have student proposals attached, this means that

uploading documents into AI models quickly reaches the page read limit either necessitating multiple rounds of uploads or editing signed PDFs in adobe acrobat to remove pages. Using both Claude and Copilot, we also ran into issues with number of files being able to be uploaded, users are restricted to 3 uploads at a time for a maximum of 20 files per conversation (CoPilot) or reach the daily upload limit (Claude) with our sample size. Copilot also performed worse at being able to identify ticked/checked boxes on the forms (with 3 forms being unable to be read by the algorithm). Overall, this took more time than extracting the information from the forms manually.



Figure 7: AI data workflow

Discussion

The purpose of this project was to identify whether inequities were experienced between male and female PhD students during their Confirmation of Candidature (CoC) in how they were evaluated and questioned during Question and Answer (Q&A) time. It is acknowledged that the findings are preliminary and should be interpreted with caution due to sample size. The project set out to identify whether this was an area that warranted further investigation given the lack of research on gender differences in experiences during candidature milestones of a PhD. This area of research was considered important to investigate due to the PhD representing the foundation of a person's research career and as such where expectations on performance are formed. In all, the current project has identified differences in experiences across college and gender for students undertaking their Confirmation of Candidature.

Overall, there were observable differences in the duration of the Q&A time for male and female students across colleges. That is, there was considerably less Q&A time in female CoC's, including less question-asker time (except CHS), less student response time, less interactions, and less time spent per interaction on average. Though these were not significant differences, the consistency with which these differences are presented across the indicators suggest that these differences warrant further exploration. These findings may not indicate a negative experience for female students given they are being questioned or challenged less than male students and may instead represent female students as more prepared and competent for their CoC. However, these findings could also imply that female students have already internalised the normative expectation that women must outperform men in order to be seen as competent, further evidencing the potential experience of a "chilly climate" (Palmgren et al., 2013). Indeed, when the gender of the question-asker is taken into account, there is consideration here for who is asking the questions and the types of questions being asked. Male question-askers were more likely to ask both male and female students challenging questions about the validity of a claim the student made during their presentation. Male question-askers were also less likely than female question-askers to frame questions politely towards female students. Moreover, female students had a higher proportion of comments and question-comments as opposed to questions from

question-askers than male students. These findings suggest that, across the colleges, there is a risk of female students experiencing a “chilly climate”, whereby they feel less supported within their academic careers (Biggs et al., 2018; Greene et al., 2010).

When considering differences between and within the colleges, there were some notable differences. For overall question-asker time, there was significantly more time spent asking questions in CASE than there is in CSE. One of the more informative indicators of quality of discussion in the Q&A session is the ratio of time spent talking between the question-askers and the students. Most colleges had relatively equitable time distribution between question-askers and students. CSE was the only college that had higher proportion of time for student responses than question-askers. Of particular note was the observable and considerably higher proportion of time used by question-askers in CBLG for female students, with an overall significant difference between CBLG and CSE for ratio of time spent talking between question-askers and students. This observable and significant difference both demonstrate that in CBLG, the question time is being dominated by the question-askers, with little time being afforded to students and potentially compounded by gender.

What these findings regarding speaking time suggest is that there appears to be great variation in the quality of robust discussion occurring between colleges. The purpose of the CoC is for students to provide sound rationale for the purpose and design of their research, with question time providing opportunities to clarify or further justify decision-making. With this in mind, it would be expected that there should be equitable proportion of speaking time between question-askers and students, or, ideally, students having a higher proportion of speaking time. This premise is consistent with past research that describes the impact of power imbalance on experiences of HDR students (Jones & Blass, 2019). Therefore, the implications of the current findings are that CASE perhaps demonstrates the most robust discussion given both the length of time on average discussing the student’s work as well as the equitable speaking time between question-askers and students. CSE, though having the least “robust” discussion time (given the significantly shorter time), perhaps provides an empowering environment to students given the higher proportion of speaking time afforded to students. However, given CSE was significantly less likely to politely frame questions or comments, the quality and supportiveness of this environment needs to be explored further as research has shown that politeness tendencies can be gendered (Jarvis et al., 2022). Finally, with CBLG having significantly higher proportionality of speaking time for question-askers suggests a risk of a discussion time that is not empowering students. Moreover, given CBLG female students had observably less proportion of speaking time, and asked fewer challenging questions, than male students suggest a risk of establishing unfavourable norms and expectations for female students consistent with past research description of a “chilly climate”. For example, the perceived silencing of women can lead to higher intentions of exiting academic careers (Biggs et al., 2018).

Overall, these results suggest that gender differences exist at the university level as well as within the colleges, and that there are also overall differences in experiences for both male and female students between the colleges. Therefore, more research is recommended to explore these differences.

Limitations

First and foremost, the sample size of the current study is an acknowledged limitation. However, given the mixed methods analysis and caution taken in the interpretation, this impact has been as minimised as possible. Though the deductive content analysis of the Q&A time was

considered most appropriate for the parameters and purpose of the current study, future research should consider an inductive thematic analysis of the transcripts to identify themes in the quality of questions being addressed to male and female students so that the experiences of students can be understood more fully. In addition, practically, there were some issues with audio quality while analysing video, and as such conveners of CoC's should make concerted efforts to ensure improved recording quality (e.g. microphone use for questions) as this would enable similar or extended review of CoC presentations in the future.

Future work

While this project quantitatively and qualitatively analysed CoC from an external, historical perspective, future work should undertake a phenomenological (lived experience) approach to gauge in-depth understanding of the experience of CoC questioning and progress monitoring from the student perspective and from the perspective of established academics/panel members. For example, exploring if students felt they gained valuable feedback on their planned research, if they felt the experience met their expectations, and if they felt they were adequately prepared for the experience. Other research on student experience of progress reporting has shown that female and male students interpret and experience these milestones differently (Mewburn et al., 2024). As such, there is some evidence to support an examination of gender inequity from a student perspective. An in-depth exploration of this nature would also enable JCU to identify areas where support could be provided (e.g. professional development in public speaking) or action taken if students feel questioning is unfair. Additionally, it would be beneficial to explore how academic staff members approach CoC presentations, as other research highlights that academic staff members 'set the tone' for progression milestones having a large impact on the student experience (McCulloch et al., 2025; Jones & Blass, 2019). This would enable JCU to ensure that standards set by JCU policy, student expectations and staff member participation in candidature progression milestones align.

As the current project only examined gender equity, we also recommend that other forms of diversity and inclusion and how these interact be considered in the future. In particular, the impact of domestic versus international visa status and students who have English as a Second Language (ESL) should be explored as several instances were observed of advisors answering questions on ESL student behalf.

Conclusions and recommendations

The aim of the current study was to identify whether further research is needed to determine if inequity of experiences exist at the post-graduate level for male and female students. The conclusion of the authors of this report is that further research is required given the sufficient evidence of differences identified for male and female students both within and between colleges. A series of short-term and long-term recommendations have been developed based on these findings.

Short-term Recommendations:

1. Improve recording quality for CoCs
 - a. Ensure reliable audio (e.g., microphone use for questions) so future reviews are feasible and accurate.
2. More Explicit Chair Responsibilities
 - a. Prior to CoC seminar, convene panel for briefing on purpose of a CoC (constructive critique), role expectations, and empowering students.

- b. Moderating question-time for inequities in speaking time for students as well as for adversarial rather than constructive questioning.
3. Similar to SFT pulse survey frameworks, develop a post-CoC pulse survey for students to track inequities in experiences going forwards.
 - a. Survey can address expectation alignment, perceived preparedness, perceived support, perceived quality of experience.

Long-term Recommendations:

2. Expansion of research program
 - a. Conduct an inductive thematic analysis of the Q&A transcripts to better inform quality of experience by gender.
 - b. Conduct an in-depth qualitative study using a phenomenological framework to investigate the experiences of students and panel members.
 - c. Extend the investigation beyond gender and investigate intersectionality of experiences by minority groups including international students and where English is not the first language.
 - d. Increase the sample size of the current study through extracting more recordings from the GRS. Bring on an AI researcher to develop a framework for analysing the recordings and therefore providing the basis for an auditing tool.

References

- Asr FT, Mazraeh M, Lopes A, Gautam V, Gonzales J, Rao P, et al. (2021) The Gender Gap Tracker: Using Natural Language Processing to measure gender bias in media. *PLoS ONE* 16(1): e0245533. <https://doi.org/10.1371/journal.pone.0245533>
- Babcock, L., Recalde, M. P., Vesterlund, L., & Weingart, L. (2017). Gender differences in accepting and receiving requests for tasks with low promotability. *The American Economic Review*, 107(3), 714–747. <https://doi.org/10.1257/aer.20141734>
- Bendels, M. H. K., Müller, R., Brueggmann, D., & Groneberg, D. A. (2018). Gender disparities in high-quality research revealed by Nature Index journals. *PloS One*, 13(1), e0189136–e0189136. <https://doi.org/10.1371/journal.pone.0189136>
- Biggs, J., Hawley, P. H., & Biernat, M. (2018). The academic conference as a chilly climate for women: Effects of gender representation on experiences of sexism, coping responses, and career intentions. *Sex Roles*, 78(5-6), 394-408. doi:<https://doi.org/10.1007/s11199-017-0800-9>
- Bol, T. (2023). Gender inequality in cum laude distinctions for PhD students. *Scientific Reports*, 13(1), 20267–20267. <https://doi.org/10.1038/s41598-023-46375-7>
- Feldon, D. F., Peugh, J., Maher, M. A., Roksa, J., & Tofel-Grehl, C. (2017). Time-to-Credit Gender Inequities of First-Year PhD Students in the Biological Sciences. *CBE Life Sciences Education*, 16(1), ar4-. <https://doi.org/10.1187/cbe.16-08-0237>
- Glorieux, A., Spruyt, B., Minnen, J., & van Tienoven, T. P. (2025). Calling it quits: a longitudinal study of factors associated with dropout among doctoral students. *Studies in Continuing Education*, 47(1), 155–173. <https://doi.org/10.1080/0158037X.2024.2314694>
- Greene, J., Stockard, J., Lewis, P., and Richmond G. (2010) Is the Academic Climate Chilly? The Views of Women Academic Chemists. *Journal of Chemical Education* 87(4), 381-385 DOI: 10.1021/ed800042z
- Hall, R. M., & Sandler, B. R. (1982). The Classroom Climate: A Chilly One for Women?. <https://eric.ed.gov/?id=ED215628>
- Hamilton, L., Elliott, D., Quick, A., Smith, S., & Choplin, V. (2023). Exploring the Use of AI in Qualitative Analysis: A Comparative Study of Guaranteed Income Data. *International Journal of Qualitative Methods*, 22. <https://doi.org/10.1177/16094069231201504> (Original work published 2023)
- Hamilton, M; Williams, A; Baird, M (2022) Gender Equity and Inclusion by Design: A Toolkit for the Australian University Sector. Universities Australia Women: Canberra. <https://universitiesaustralia.edu.au/wp-content/uploads/2023/02/UA-2023-004-Gender-Equity-Toolkit-web-FA.pdf>
- Holman, L., Stuart-Fox, D., & Hauser, C. E. (2018). The gender gap in science: How long until women are equally represented? *PLoS Biology*, 16(4), e2004956-. <https://doi.org/10.1371/journal.pbio/2004956>
- Harris, R., Mate-Sanchez-Val, M., & Ruiz Marín, M. (2024). Gender disparities in promotions and exiting in UK Russell Group universities. *Applied Economics*, 1–17. <https://doi.org/10.1080/00036846.2024.2361384>

Janz, T. A., & Pyke, S. W. (2000). A Scale to Assess Student Perceptions of Academic Climates. *Canadian Journal of Higher Education (1975)*, 30(1), 89–122. <https://doi.org/10.47678/cjhe.v30i1.183347>

Jarvis, S. N., Ebersole, C. R., Nguyen, C. Q., Zhu, M., & Kray, L. J. (2022). Stepping Up to the Mic: Gender Gaps in Participation in Live Question-and-Answer Sessions at Academic Conferences. *Psychological Science*, 33(11), 1882–1893. <https://doi.org/10.1177/09567976221094036>

Jones, A. and Blass, E. (2019). The Impact of Institutional Power on Higher Degree Research Supervision: Implications for the Quality of Doctoral Outcomes. *Universal Journal of Educational Research*, 7(7), 1485 - 1494. DOI: 10.13189/ujer.2019.070702

Lee, J. J., & McCabe, J. M. (2020). Who Speaks and Who Listens: Revisiting the Chilly Climate in College Classrooms. *Gender & Society*, 35(1), 32-60. <https://doi.org/10.1177/0891243220977141> (Original work published 2021)

Llorens, A., Tzovara, A., Bellier, L., Bhaya-Grossman, I., Bidet-Caulet, A., Chang, W. K., Cross, Z. R., Dominguez-Faus, R., Flinker, A., Fonken, Y., Gorenstein, M. A., Holdgraf, C., Hoy, C. W., Ivanova, M. V., Jimenez, R. T., Jun, S., Kam, J. W. Y., Kidd, C., Marcelle, E., ... Dronkers, N. F. (2021). Gender bias in academia: A lifetime problem that needs solutions. *Neuron (Cambridge, Mass.)*, 109(13), 2047–2074. <https://doi.org/10.1016/j.neuron.2021.06.002>

McCulloch, A., Loeser, C., & Bageas, R. (2025). Does the viva matter? PhD student experiences of the oral examination and its contribution to examination outcome and researcher development. *Assessment & Evaluation in Higher Education*, 1–16. <https://doi.org/10.1080/02602938.2025.2566652>

Mewburn, I., Cuthbert, D., & Tokareva, E. (2014). Experiencing the progress report: an analysis of gender and administration in doctoral candidature. *Journal of Higher Education Policy and Management*, 36(2), 155–171. <https://doi.org/10.1080/1360080X.2013.861054>

Palmgren PJ, Chandratilake M, Nilsson GH, Laksov KB. Is there a chilly climate? An educational environmental mixed method study in a chiropractic training institution. *J Chiropr Educ*. 2013 Spring;27(1):11-20. doi: 10.7899/JCE-12-015. PMID: 23518905; PMCID: PMC3604959.

Salem, V., McDonagh, J., Avis, E., Eng, P. C., Smith, S., & Murphy, K. G. (2021). Scientific medical conferences can be easily modified to improve female inclusion: a prospective study. *The Lancet. Diabetes & Endocrinology*, 9(9), 556–559. [https://doi.org/10.1016/S2213-8587\(21\)00177-7](https://doi.org/10.1016/S2213-8587(21)00177-7)

Sandler, B. R. (1996). The chilly classroom climate: A guide to improve the education of women. <https://eric.ed.gov/?id=ED396984>

Smith, C., Townley, C., Huppatz, K. & Bansel, P. (2021). Vice-Chancellor’s GENDER EQUALITY FUND Final Report 2021, Equity, policy and practice: disruptions to candidature and barriers to career progression for women HDR candidates. https://www.westernsydney.edu.au/_data/assets/pdf_file/0006/1948722/Equity_Policy_and_Practice_Final_Report_-_Caroline_Smith.pdf

Smith, J. L., Lewis, K. L., Hawthorne, L., & Hodges, S. D. (2013). When Trying Hard Isn’t Natural: Women’s Belonging With and Motivation for Male-Dominated STEM Fields As a Function of Effort Expenditure Concerns. *Personality & Social Psychology Bulletin*, 39(2), 131–143. <https://doi.org/10.1177/0146167212468332>

Ukanwa, K. (2024). Algorithmic bias: Social science research integration through the 3-D Dependable AI Framework. *Current Opinion in Psychology*, 58, Article 101836. <https://doi.org/10.1016/j.copsyc.2024.101836>

Walton, G. M., Logel, C., Peach, J. M., Spencer, S. J., & Zanna, M. P. (2015). Two brief interventions to mitigate a “chilly climate” transform women’s experience, relationships, and achievement in engineering. *Journal of Educational Psychology*, 107(2), 468-85.

Wang, Y. (2022). Using Machine Learning and Natural Language Processing to Analyze Library Chat Reference Transcripts. *Information Technology and Libraries*, 41(3), 1–10. <https://doi.org/10.6017/ital.v41i3.14967>

Webster, C. S. (2025). Natural and artificial intelligence – the psychotechnical agenda of the 21st century. *Journal of Psychology and AI*, 1(1). <https://doi.org/10.1080/29974100.2025.2491445>

Wollast, R., Boudrenghien, G., Van der Linden, N., Galand, B., Roland, N., Devos, C., De Clercq, M., Klein, O., Azzi, A., & Frenay, M. (2018). Who Are the Doctoral Students Who Drop Out? Factors Associated with the Rate of Doctoral Degree Completion in Universities. *International Journal of Higher Education*, 7(4), 143-. <https://doi.org/10.5430/ijhe.v7n4p143>

Witteman, H. O., Hendricks, M., Straus, S., & Tannenbaum, C. (2019). Are gender gaps due to evaluations of the applicant or the science? A natural experiment at a national funding agency. *The Lancet (British Edition)*, 393(10171), 531–540. [https://doi.org/10.1016/S0140-6736\(18\)32611-4](https://doi.org/10.1016/S0140-6736(18)32611-4)

Zhang, H., Wu, C., Xie, J., Lyu, Y., Cai, J., & Carroll, J. M. (2025). Harnessing the power of AI in qualitative research: Exploring, using and redesigning ChatGPT. *Computers in Human Behavior: Artificial Humans*, 4, 100144-. <https://doi.org/10.1016/j.chbah.2025.100144>