國家科學及技術委員會補助專題研究計畫報告

「以多模式方法探討如何提升工科生的潛在性別偏見意識及降低在STEM環境裡的性別偏見 (L02)」

報 告 類 別 ： 成果報告
計 畫 類 別 ： 個別型計畫
計 畫 編 號 ： NSTC 112-2629-H-224-001-SSS
執 行 期 間 ： 112年08月01日至113年07月31日
執 行 單 位 ： 國立雲林科技大學前瞻學士學位學程

計 畫 主 持 人 ： 吳碧瑛

計畫參與人員： 碩士班研究生-兼任助理：賴筱晴
大專生-兼任助理：許愷珈
大專生-兼任助理：徐于婷
大專生-兼任助理：楊嘉文
大專生-兼任助理：余竣詮
大專生-兼任助理：魏震帆

本研究具有政策應用參考價值：■否 □是，建議提供機關
（勾選「是」者，請列舉建議可提供施政參考之業務主管機關）
本研究具影響公共利益之重大發現：□否 □是

中 華 民 國 113 年 10 月 31 日

中 文 摘 要 ： 這項研究旨在透過兩個VIDS影片和引導討論，以協助減低工程系學生潛在的性別偏見。我們對非工程系和工程系學生進行了相似但獨立的研究。二十三名非工程系學生參與了Pre-Study II研究，而二十七名工程系學生參與了Main Study研究。在二十七名工程系學生中，只有二十二名完成了自我反省日記和後測問卷II。研究中使用了三份問卷：前測、後測I和後測II，分別是矛盾性別主義量表（ASI）、男性兩極性別思維問卷（MPGQ）和女性科學家量表（WiSS），這些量表都已被證明具有實證效度。

分析了三份與女性在現代社會角色相關的量表，均由同一組參與者完成，其結果顯示潛在性別偏見有多種面向。儘管在指導後，工程系和男性學生不再支持兩極化的性別思維（如MPGQ結果所示），並且不同意傳統上對女性在科學領域角色的性別歧視信念和刻板印象（如WiSS結果所示），但他們對女性抱有較高的仁慈型性別歧視（BS）態度（如ASI結果所示），進一步支持了東亞社會普遍持有的父權觀念。

本研究對開放式問卷和自我反省日記進行了內容分析。儘管並非所有參與者都能識別VIDS影片中存在的潛在性別偏見，但他們都能認識到這種偏見對個人和社會的負面影響。參與者表示，許多仍然存在的性別偏見是源於廣泛接受的「不同性別，不同期望」觀念。儘管性別平等教育已經在學校實施了一段時間，人們仍會被期望以符合社會規範的方式行事。幾位參與者分享了個人經歷，描述了他們自己、朋友、家人和同學遇到的（潛在）性別偏見。當參與者被問及女性應該如何回應涉及潛在性別偏見的情況時，許多人建議女性應該「做自己」，追求卓越，不要過度擔心潛在的批評。大多數參與者也表示，公司應該被視為採取「無偏見」的立場，並定期組織講座和敏感度訓練，以幫助減少工作場所中的此類偏見。最後，超過三分之二的學生對參與研究表示滿意。他們感謝能夠在包容性頗高的環境中公開、誠實地討論性別偏見，促進對自己態度和行為的自我反思。

儘管研究並未按原計劃進行，但所得到的結果證明了研究團隊的努力，因為他們提供了寶貴的見解，並有助於更深入地理解這個主題。潛在的性別偏見不是人們願意承認的事情，讓他們承認、接受並反思它更是艱鉅的任務。因此，研究團隊希望這項研究能夠提供一些學生對潛在性別偏見這一主題的心理的一瞥，並激發進一步的討論和行動，以促進性別平等和包容性。

中文關鍵詞： 潛在性別偏見、VIDS 影片、STEM、仁慈型性別歧視、性別平等、偏見意識、性別友善、工程系學生

英 文 摘 要 ： This research aimed to use two VIDS videos with guided discussions to help minimize latent gender bias among engineering students. Two separate yet similar instructions were conducted among non-engineering and engineering students. Twenty-three non-engineering students participated in Pre-Study II, while twenty-seven engineering students participated in the Main Study. Of the twenty-seven engineering students, only twenty-two completed their self-reflective journals and the post-

treatment questionnaires II. Three questionnaires were used in tandem: the pre-treatment, post-treatment I, and post-treatment II, and they are the Ambivalent Sexism Inventory (ASI), the Men's Polarized Gender Thinking Questionnaire (MPGQ), and the Women in Science Scale (WiSS), which had all been proven their empirical validity.

The results from three different scales related to women's roles in modern society, completed by the same set of participants, showed that latent gender bias has a variety of dimensions. Even though, after the instruction, engineering and male students do not endorse polarized gender thinking (as seen from the result from MPGQ), and they disagreed with traditional sexist beliefs and stereotypes regarding women's roles in science (as seen from the result from WiSS), they did hold heightened benevolent sexism (BS) attitudes towards women (as seen from the result from ASI), further endorsing the paternalistic views commonly held in East Asian societies. Content analysis was done on the open-ended questionnaires and the self-reflective journals. Even though not all participants were able to identify the latent gender bias that was present in both VIDS videos, all of them were able to recognize the negative impact that such biases can have. Participants stated that much of the gender bias that is still happening was due to the widely accepted notion of "different gender, different expectations." Men and women are expected to behave in ways that conform to societal norms, even though gender equality education has been implemented in schools for some time. Several participants shared personal anecdotes about encountering (latent) gender biases in their own lives and in the lives of their friends, family, and classmates. When participants were asked how women should respond to situations involving latent gender bias, many suggested that women should simply "be themselves," strive for excellence, and not be overly concerned about criticism. Most participants also indicated that companies should be seen as taking a "no bias" position and organize regular talks and sensitivity training to help minimize such biases in the workplace. Finally, over two-thirds of the students were appreciative of the opportunity to openly and honestly discuss gender bias in a supportive environment, facilitating self-reflection on their attitudes and behaviors.

Although the research did not proceed as planned, the results justified the research team's effort as they provided valuable insights of the topic. Having latent gender bias is not something people would admit willingly, and getting them to acknowledge, accept, and reflect upon

it is an even more herculean task. Thus, the research team hoped that this research would provide a small glimpse into students' psyche on latent gender bias and inspire further discussion and action to promote gender equality and inclusivity.

英文關鍵詞： Latent gender bias, VIDS videos, STEM, benevolent sexism, gender equality, bias awareness, gender-friendly, engineering students

附件一

國家科學及技術委員會補助專題研究計畫報告

「以多模式方法探討如何提升工科生的潛在性別偏見意識及降低在STEM環境裡的性別偏見 (L02)」

報告類別：□進度報告
■成果報告：■ 完整報告/□精簡報告
計畫類別：■ 個別型計畫 □整合型計畫
計畫編號：NSTC112－2629－H－224－001－SSS
執行期間： 112 年 08 月 01 日至 113 年 07 月 31 日

執行機構及系所：國立雲林科技大學高階管理碩士學位學程

計畫主持人： 吳碧瑛 助理教授
共同主持人：
計畫參與人員：許愷珈、楊嘉文、余竣詮、徐于婷、魏震帆、賴筱晴

本計畫除繳交成果報告外，另含下列出國報告，共 0 份：
□執行國際合作與移地研究心得報告
□出席國際學術會議心得報告
□出國參訪及考察心得報告

本研究具有政策應用參考價值： ■ 否 □是，建議提供機關________
(勾選「是」者，請列舉建議可提供施政參考之業務主管機關)
本研究具影響公共利益之重大發現：■ 否 □是

中 華 民 國 113 年 10 月 30 日

1. Introduction

Many researchers have delved into the ways and means to eliminate or minimize the latent gender bias within STEM. This study hopes to add to the scholarship by looking into a possible means of changing or influencing the attitude of male engineers toward their female counterparts while at the same time raising the awareness of female engineers of the latent gender biases that they may hold. Using the Tripartite Model of Attitude (Breckler, 1984), this study hopes to convince male engineers by influencing their affect, behavior, and cognition, as shown in Fig. 1. Multimodal pedagogic discourse involves the modes of gesture, gaze, voice, facial expression, and spatial position (Lim, O'Halloran & Podlasov, 2012). This study wants to expose these engineering students to the latent gender bias encountered by women engineers through VIDS videos. VIDS videos have been shown to increase the bias literacy of viewers, which are both the male and female engineering students in this research, by raising the awareness of gender bias, increasing knowledge of gender inequality, and feeling of efficacy at being able to notice bias (Pietri et al., 2017). More importantly, VIDS videos were found to reduce modern sexism, improve attitudes towards women in STEM environments, and engage the emotions of empathy and anger on the part of both men and women (Moss-Racusin et al., 2018). The subsequent guided group discussion that produces consensus reports aims to encourage these engineering students to act or react appropriately to minimize latent gender bias in the workplace. Finally, this study believes that in the end, these engineering students will conclude that women engineers are as capable as male engineers. Thus, the modified research questions for this study, due to the unusable Neurosky Mindwave devices, are updated as follows:

1. Does the treatment (watching videos and having guided group discussions) minimize the latent gender bias of male and female engineering students?
2. Does listening to different viewpoints help provide a more equity-friendly view?
3. Do the effects of the treatment persist 3 weeks later?
4. What can the management learn from the results of this research and what are the steps they can take to minimize latent gender bias in their organizations?

The videos used in this research belonged to VIDS (Video Interventions for Diversity in STEM; obtained from https://academics.skidmore.edu/blogs/vids/ and free to use for research purposes), which consists of two sets of short videos that expose participants to empirical results from published gender bias research. One set of videos illustrated gender bias using narratives (i.e., compelling stories), and the other presented the same bias using expert interviews (i.e., straightforward presentation of facts). VIDS videos have been shown to increase the bias literacy of viewers by raising awareness of bias, increasing knowledge of gender inequality, and increasing the feeling of efficacy in noticing bias (Pietri et al., 2017). More importantly, VIDS videos were found to reduce modern sexism, improve attitudes towards women in STEM environments, and engage the emotions of empathy and anger on the part of both men and women (Moss-Racusin et al., 2018). For this research, the team used two narrative-style videos (videos No. 02 and 04 of the narrative series) that are relatively more relatable to non-graduate students.

2. The Experiments

This project aimed to gain insights into possible latent gender biases that may be present in engineering students. Through watching two videos and having guided group discussions, the students would learn how some or all of these communication modes would not just constitute latent gender bias but perpetuate it. Data collected before and after the video watching and group discussion, as shown in Fig. 2, consisted of questionnaires, group reports, and individual reflective journals. All these data were provided by the participating students willingly and consciously. Fusch and her colleagues (2018) stated that triangulation is how one explores different levels and perspectives of the same phenomenon. Thus, these multiple sources of data constitute methodological triangulation (Bekhet & Zauszniewski, 2012), and methodological triangulation adds depth to the data that is collected (Manganelli et al., 2014).

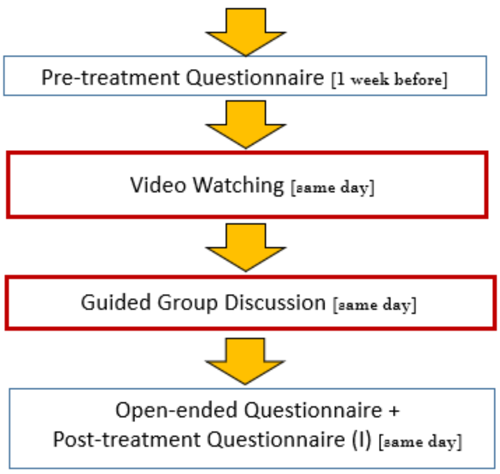

Fig.1 Pre-Study (II)

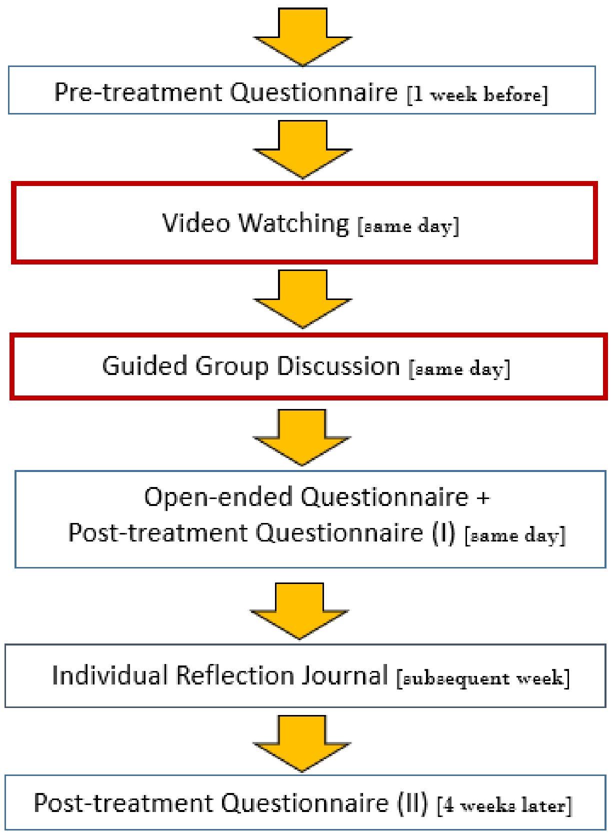

Fig. 2 Main Study

The video-watching sessions were held in a 20-person classroom for video watching and discussion. Unfortunately, the Neurosky Mindwave devices required for this project are inoperable, and with the project budget being limited, there is no extra budget to purchase these devices as each cost slightly less than NTD$ 10K. The Main Study, Pre-Study Sessions (I), and (II) were all held in the exact location to ensure consistency in the neuro signals collected.

The type of participants for this project consisted of non-engineering students and engineering students. For the Pre-Study Session (I), 26 students from the Bachelor's Program in International Management took part, 22 of whom were females. The main objective of holding a Pre-Study Session (I) was to finetune the guided group discussion questions. Thus, the project team only collected the Post-treatment sets of questionnaires. For the Pre-Study Session (II), 23 students from non-engineering programs participated, with 14 female and nine male students. Five were graduate students, while the rest were undergraduates. For the Main Study, there were 27 participants consisting of 24 male and three female engineering students. Among them, eight were graduate students. The Pre-Study Session (II) and the Main Study interventions are shown in Fig. 1 and 2, respectively.

3. Data Analysis

3.1 Quantitative Analysis

For this project, a complete set of pre and post-treatment questionnaires comprised of three verified questionnaires, namely the *Ambivalent Sexism Inventory* (*ASI*; Glick & Fiske, 1996), the *Men's Polarized Gender Thinking Questionnaire* (*MPGQ*; Bergman, Larsman & Love, 2014), and the *Women in Science Scale* (*WiSS*; Owen et. al., 2007). Table 1 illustrates the general demographics of the data gathered from the Pre-Study (II) and the Main Study. The following sections will discuss the result of the analysis.

Table 1: *Information of data collected.*

	Pre-Study (II)		Main Study	
Pre-treatment Questionnaires	**N=23**	Female=14; Male=9.	**N=27**	Female=3; Male=24.
Post-treatment Questionnaires I	**N=23**	Female=14; Male=9.	**N=27**	Female=3; Male=24.
Post-treatment Questionnaires II	-	-	**N=22**	Female=2; Male=20.

3.1.1 Analysis: The Ambivalent Sexism Inventory (ASI)

3.1.1.1 Reliability

The internal consistency of the ASI questionnaire was examined using the SPSS software. The value of Cronbach's α was .824, indicating that the ASI questionnaire was highly reliable.

3.1.1.2 Normality

The normality test was conducted to examine whether the collected data was a normal distribution. According to the result of the Shapiro-Wilk test ($p > .05$), the distribution of the mean scores on the ASI questionnaire in the non-engineering and engineering groups was normal. Additionally, females' and males' mean scores on the ASI questionnaire were distributed normally. Consequently, parametric test was conducted for data analysis.

3.1.1.3 Findings

Table 2 reports that there is no difference between the non-engineering and engineering groups in the pre-mean scores of the ASI questionnaire. After instruction, **the non-engineering group's post-mean score was significantly lower than the engineering group's**.

Table 2

Differences between non-engineering and engineering groups in pre- and post-ASI

					Independent samples *t*-test			
Aspects	Groups	*n*	*M*	*SD*	*df*	*t*	*p*	Effect size (Cohen's *d*)
Pre-ASI	Non-engineering	23	3.22	.57	48	.377	.708	.102
	Engineering	27	3.16	.61				
Post-ASI	Non-engineering	23	3.13	.68	48	-3.676	.001**	1.044
	Engineering	27	3.78	.56				

Note. The effect size (Cohen's *d*) is defined as small ($.2 \leq d < .5$), medium ($.5 \leq d < .8$), and large ($d \geq .8$) (Sullivan & Feinn, 2012). *$p < .05$, **$p < .01$, ***$p = 0.000$.

Table 3 reports no significant difference between pre- and post-mean scores of the ASI questionnaire in the non-engineering group. Contrarily, in the engineering group, the post-mean score was significantly higher than the pre-mean score. Similarly, Table 4 reports that no difference between females and males was found in the pre-mean score of the ASI questionnaire. However, **after instruction, males showed a higher mean score on the ASI questionnaire than females**.

Table 3

Differences between pre- and post-ASI among non-engineering and engineering groups

					Paired samples *t*-test			
Groups	Aspects	*n*	*M*	*SD*	*df*	*t*	*p*	Effect size (Cohen's *d*)
Non-engineering	Pre-ASI	23	3.22	.57	22	1.052	.304	.220
	Post-ASI	23	3.13	.68				
Engineering	Pre-ASI	27	3.16	.61	26	-2.855	.008**	.550
	Post-ASI	27	3.78	.56				

Table 4

Differences between females and males in pre- and post-ASI

					Independent samples *t*-test			
Aspects	Gender	*n*	*M*	*SD*	*df*	*t*	*p*	Effect size (Cohen's *d*)
Pre-ASI	Female	17	3.10	.57	48	-.729	.470	.220
	Male	33	3.23	.60				
Post-ASI	Female	17	3.21	.70	48	-2.094	.042*	.607
	Male	33	3.62	.65				

Table 5 reports no significant difference was found between females' pre- and post-mean scores of the ASI questionnaire. In contrast, males' post-mean scores were significantly higher than their pre-mean scores.

One-way repeated measures ANOVA was utilized to examine the differences among testing times of the ASI questionnaire in the engineering group. According to the results in Table 6, the assumption of Mauchly's Test of Sphericity was violated, indicating that the variance of three testing times was heterogeneous (W = .185, p = .000***). Accordingly, the results of the Greenhouse-Geisser test were used, as shown in Table 5. Table 5 reports **significant differences among the three testing times of the ASI questionnaire** (p = .001***).

Specifically, Table 7 shows that the **first post-mean score of the ASI questionnaire was significantly higher than the pre-mean score (*p* = .007**) and the second post-mean score (*p* = .003**)**. There was no difference between the pre-mean score and the second post-mean score.

Table 5

Differences between pre- and post-ASI among females and males

					Paired samples *t*-test			
Groups	Aspects	*n*	*M*	*SD*	*df*	*t*	*p*	Effect size (Cohen's *d*)
Female	Pre-ASI	17	3.10	.57	16	-.457	.654	.114
	Post-ASI	17	3.21	.70				
Male	Pre-ASI	33	3.23	.60	32	-2.390	.023*	.422
	Post-ASI	33	3.62	.65				

Table 6

Differences among testing times of ASI questionnaire in the engineering group

				One-way repeated measures ANOVA			
				Greenhouse-Geisser test			
Testing times	*n*	*M*	*SD*	*df*	*F*	*p*	Effect size (Partial eta)
Pre-ASI	22	3.09	.58	1.102	12.659	.001**	.376
Post1-ASI	22	3.86	.49				
Post2-ASI	22	3.11	.47				

Table 7

Pairwise comparison between testing times of ASI questionnaire

One-way repeated measures ANOVA				
Pairwise comparison (1 vs. 2)	M_1	M_2	Difference score (1-2)	*p*
Pre-ASI vs. Post1-ASI	3.09	3.86	-.767	.007**
Pre-ASI vs. Post2-ASI	3.09	3.11	-.017	1.000
Post1-ASI vs. Post2-ASI	3.86	3.11	.750	.003**

There exist two positively correlated components of sexism, namely Hostile Sexism (HS) and Benevolent Sexism (BS), as stipulated by Glick and Fiske (1996). Thus, the team further analyzed the data collected and found several exciting observations as stated in the following sections.

3.1.1.4 Reliability (for HS and BS)

The Cronbach's α coefficient was .811 for the benevolent sexism (BS) subscale and .813 for the hostile sexism (HS) subscale. The overall Cronbach's α was .824, indicating a high level of reliability for the ASI questionnaire.

3.1.1.5 Findings

Table 8 reported no significant difference between the non-engineering and engineering groups in ambivalent sexism in terms of benevolent sexism and hostile sexism before instruction. However, after instruction, the ambivalent sexism of the engineering group was significantly higher than that of the non-engineering group. Notably, **a significant difference in benevolent sexism (BS) was found between both groups, with the engineering group scoring higher than the non-engineering group**.

Table 8

Differences between the non-engineering and engineering groups in ambivalent sexism

					Independent samples *t*-test			
Aspects	Groups	*n*	*M*	*SD*	*df*	*t*	*p*	Effect size (Cohen's *d*)
Pre-benevolent sexism	Non-engineering	23	3.11	.82	48	.845	.402	.247
	Engineering	27	2.91	.80				
Post-benevolent sexism	Non-engineering	23	2.98	.79	48	-4.147	.000***	1.195
	Engineering	27	3.93	.82				
Pre-hostile sexism	Non-engineering	23	3.34	.77	48	-.343	.733	.100
	Engineering	27	3.41	.62				
Post-hostile sexism	Non-engineering	23	3.28	.88	32.58	-1.643	.110	.482
	Engineering	27	3.62	.47				
Pre-ambivalent sexism	Non-engineering	23	3.22	.57	48	.377	.708	.102
	Engineering	27	3.16	.61				
Post-ambivalent sexism	Non-engineering	23	3.13	.68	48	-3.676	.001**	1.044
	Engineering	27	3.78	.56				

Note. The effect size (Cohen's *d*) is defined as small ($.2 \le d < .5$), medium ($.5 \le d < .8$), and large ($d \ge .8$) (Sullivan & Feinn, 2012). *$p < .05$, **$p < .01$, ***$p = 0.000$.

Table 9 reported the results of the paired samples *t*-test for the non-engineering group. Table 10 shows the results for the engineering group. In the non-engineering group, no significant changes were found in ambivalent sexism, including benevolent and hostile sexism, before and after instruction. Contrarily, **the engineering group showed a significant increase in ambivalent sexism after instruction, with a particular increase in benevolent sexism (BS).**

Table 9

Non-engineering group's ambivalent sexism before and after instruction

		Pre-survey		Post-survey		Paired samples *t*-test			
Aspects	*n*	*M*	*SD*	*M*	*SD*	*df*	*t*	*p*	Effect size (Cohen's *d*)
Benevolent sexism	23	3.11	.82	2.98	.79	22	.984	.336	.205
Hostile sexism	23	3.34	.77	3.28	.88	22	.443	.662	.092
Ambivalent sexism	23	3.22	.57	3.13	.68	22	1.052	.304	.219

Table 10

Engineering group's ambivalent sexism before and after instruction

		Pre-survey		Post-survey		Paired samples *t*-test			
Aspects	*n*	*M*	*SD*	*M*	*SD*	*df*	*t*	*p*	Effect size (Cohen's *d*)
Benevolent sexism	27	2.91	.80	3.93	.82	26	-3.407	.002**	.668
Hostile sexism	27	3.41	.62	3.62	.47	26	-1.057	.300	.207
Ambivalent sexism	27	3.16	.61	3.78	.56	26	-2.855	.008**	.560

Table 11 showed no significant differences between females and males in ambivalent sexism. However, **after instruction**, **males showed a higher level of ambivalent sexism than females, with a larger difference in benevolent sexism (difference score = .52) compared to hostile sexism (difference score = .31).**

Table 11

Differences between females and males in ambivalent sexism

					Independent samples *t*-test			
Aspects	Gender	*n*	*M*	*SD*	*df*	*t*	*p*	Effect size (Cohen's *d*)
Pre-benevolent sexism	Female	17	2.95	.76	48	-.344	.732	.100
	Male	33	3.03	.84				
Post-benevolent sexism	Female	17	3.15	.98	48	-1.945	.058	.564
	Male	33	3.67	.86				
Pre-hostile sexism	Female	17	3.26	.67	48	-.841	.404	.263
	Male	33	3.44	.70				
Post-hostile sexism	Female	17	3.26	.74	48	-1.492	.142	.439
	Male	33	3.57	.67				
Pre-ambivalent sexism	Female	17	3.10	.57	48	-.729	.470	.222
	Male	33	3.23	.60				
Post-ambivalent sexism	Female	17	3.21	.70	48	-2.094	.042*	.607
	Male	33	3.62	.65				

Table 12 presents the results of the paired samples *t*-test for females and Table 6 reports the results for males. As shown in Table 5, no significant differences were observed among females in ambivalent sexism, including benevolent and hostile sexism after instruction. Contrarily, there was **a significant difference among males in ambivalent sexism before and after instruction**. **Specifically, males' benevolent sexism increased significantly after instruction**.

Table 12

Differences in ambivalent sexism before and after instruction among females

		Pre-survey		Post-survey		Paired samples *t*-test			
Aspects	*n*	*M*	*SD*	*M*	*SD*	*df*	*t*	*p*	Effect size (Cohen's *d*)
Benevolent sexism	17	2.95	.76	3.15	.98	16	-.712	.487	.173
Hostile sexism	17	3.26	.67	3.26	.74	16	.000	1.00	.000
Ambivalent sexism	17	3.10	.57	3.21	.70	16	-.457	.654	.111

Table 13

Differences in ambivalent sexism before and after instruction among males

		Pre-survey		Post-survey		Paired samples *t*-test			
Aspects	*n*	*M*	*SD*	*M*	*SD*	*df*	*t*	*p*	Effect size (Cohen's *d*)
Benevolent sexism	33	3.03	.84	3.67	.86	32	-2.646	.013*	.461
Hostile sexism	33	3.44	.70	3.57	.67	32	-.896	.377	.156
Ambivalent sexism	33	3.23	.60	3.62	.65	32	-2.390	.023*	.416

One-way repeated measures ANOVA was utilized to examine the differences across the engineering group's three testing times of the ASI questionnaire. According to the results, the assumption of the Mauchly's Test of Sphericity was violated, indicating that the variance of three testing times was heterogeneous ($p < .001$***). Accordingly, the results of the Greenhouse-Geisser test were used, as shown in Table 14. Table 14 reports **a significant difference in ambivalent sexism across the three testing times, particularly in the aspect of benevolent sexism**.

Table 14

Differences in ambivalent sexism across three testing times in the engineering group

					One-way repeated measures ANOVA			
					Greenhouse-Geisser test			
Aspects	Testing times	*n*	*M*	*SD*	*df*	*F*	*p*	Effect size (Partial eta)
Benevolent sexism	Pre	22	2.82	.80	1.088	12.418	.001**	.372
	Post1	22	4.03	.81				
	Post2	22	2.93	.76				
Hostile sexism	Pre	22	3.37	.59	1.311	3.853	.05	.155
	Post1	22	3.69	.36				
	Post2	22	3.29	.44				
Ambivalent sexism	Pre	22	3.09	.58	1.102	12.659	.001**	.376
	Post1	22	3.86	.49				
	Post2	22	3.11	.47				

Specifically, Table 15 shows the results of the pairwise comparisons between testing times in ambivalent sexism, including benevolent and hostile sexism. **Regarding ambivalent sexism, the students' first post-mean score was significantly higher than the pre-mean score (*p* = .003**) and the second post-mean score (*p* = .007**). Similarly, regarding benevolent sexism, the first post-mean score was significantly higher than the pre-mean score (p = .004**) and the second post-mean score (p = .008**).**

Table 15

Pairwise comparison between testing times of the ASI questionnaire

	One-way repeated measures ANOVA				
Aspects	Pairwise comparison (1 vs. 2)	M_1	M_2	Difference score (1-2)	p
Benevolent sexism	Pre vs. Post1	2.82	4.03	-1.215	.004**
	Pre vs. Post2	2.82	2.93	-.116	.469
	Post1 vs. Post2	4.03	2.93	1.099	.008**
Hostile sexism	Pre vs. Post1	3.37	3.69	-.318	.346
	Pre vs. Post2	3.37	3.29	.083	1.000
	Post1 vs. Post2	3.69	3.29	.401	.051
Ambivalent sexism	Pre vs. Post1	3.09	3.86	-.767	.007**
	Pre vs. Post2	3.09	3.11	-.017	1.000
	Post1 vs. Post2	3.86	3.11	.750	.003**

3.1.2 The Men's Polarized Gender Thinking Questionnaire (MPGQ)

3.1.2.1 Reliability

The internal consistency of the MPGQ questionnaire was examined using the SPSS software. The value of Cronbach's α was .828, indicating that the MPGQ questionnaire was highly reliable.

3.1.2.2 Normality

The normality test examined whether the collected data was a normal distribution. According to the result of the Shapiro-Wilk test ($p > .05$), the distribution of the scores on the MPGQ questionnaire in the non-engineering and engineering groups was normal. Additionally, females' and males' scores on the MPGQ questionnaire were distributed normally except for males' post-mean scores. According to the central limit theorem (Kwak & Kim, 2017), when the sample size is more than 30 (i.e., 33 males in this study), the sampling distribution can be assumed to be normal. Therefore, the parametric tests were utilized to analyze the data.

3.1.2.3 Findings

Table 16 reports that no statistically significant difference between the pre-mean score of the MPGQ questionnaire was found. However**, after instruction, the mean score of the engineering group was significantly lower than that of the non-engineering group, indicating that instruction with group discussion could help reduce the students' gender discrimination**.

Table 16

Differences between non-engineering and engineering groups in pre- and post-MPGQ

					Independent samples *t*-test			
Aspects	Groups	*n*	*M*	*SD*	*df*	*t*	*p*	Effect size (Cohen's *d*)
Pre-MPGQ	Non-engineering	23	3.89	.64	35.13	.977	.335	.283
	Engineering	27	3.74	.39				
Post-MPGQ	Non-engineering	23	3.93	.80	30.97	3.302	.002**	.965
	Engineering	27	3.32	.40				

Table 17 presents no significant difference between pre- and post-mean scores of the MPGQ questionnaire among the non-engineering students. In contrast, **the engineering students' mean score significantly reduced after instruction**.

Table 17

Differences between pre- and post-MPGQ among non-engineering and engineering groups

					Paired samples *t*-test			
Groups	Aspects	*n*	*M*	*SD*	*df*	*t*	*p*	Effect size (Cohen's *d*)
Non-engineering	Pre-MPGQ	23	3.89	.64	22	-.256	.800	.053
	Post-MPGQ	23	3.93	.80				
Engineering	Pre-MPGQ	27	3.74	.39	26	2.901	.007**	.558
	Post-MPGQ	27	3.32	.39				

Table 18

Differences between females and males in pre- and post-MPGQ

					Independent samples *t*-test			
Aspects	Gender	*n*	*M*	*SD*	*df*	*t*	*p*	Effect size (Cohen's *d*)
Pre-MPGQ	Female	17	3.84	.51	48	.368	.715	.096
	Male	33	3.79	.53				
Post-MPGQ	Female	17	3.84	.58	48	1.848	.071	.560
	Male	33	3.48	.70				

Table 19

Differences between pre- and post-MPGQ among females and males

					Paired samples *t*-test			
Groups	Aspects	*n*	*M*	*SD*	*df*	*t*	*p*	Effect size (Cohen's *d*)
Female	Pre-MPGQ	17	3.84	.51	16	.017	.987	.004
	Post-MPGQ	17	3.84	.58				
Male	Pre-MPGQ	33	3.79	.53	32	2.174	.037*	.378
	Post-MPGQ	33	3.48	.70				

Table 18 reports no significant difference in the pre-mean score of the MPGQ questionnaire between females and males. The same result was also found in the post-mean score. Similarly, Table 19 reports no statistically significant difference between females' pre- and post-mean scores of the MPGQ questionnaire was evident. However, **males' mean score showed a significant reduction after instruction, indicating that instruction could help reduce males' gender discrimination**.

Table 20

Differences among testing times of the MPGQ questionnaire in the engineering group

				One-way repeated measures ANOVA			
				Greenhouse-Geisser test			
Testing times	*n*	*M*	*SD*	*df*	*F*	*p*	Effect size (Partial eta)
Pre-MPGQ	22	3.69	.37	1.398	5.821	.014*	.217
Post1- MPGQ	22	3.35	.37				
Post2- MPGQ	22	3.76	.44				

Table 21

Pairwise comparison between testing times of MPGQ questionnaire

One-way repeated measures ANOVA				
Pairwise comparison (1 vs. 2)	M_1	M_2	Difference score (1-2)	*p*
Pre- MPGQ vs. Post1- MPGQ	3.69	3.35	.344	.094
Pre- MPGQ vs. Post2- MPGQ	3.69	3.76	-.065	1.000
Post1- MPGQ vs. Post2-MPGQ	3.35	3.76	-.409	.035*

One-way repeated measures ANOVA was utilized to examine the differences across testing times of the MPGQ questionnaire in the engineering group. As the assumption of Mauchly's Test of Sphericity was violated ($W = .185$, $p = .000$***), the results of the Greenhouse-Geisser test were used. As shown in Table 20, there were significant differences across the three testing times of the MPGQ questionnaire ($p = .014$*). More specifically, Table 21 reports no significant differences between the pre- and the first post-mean score ($p = .094$) and between pre- and the second post-mean score ($p = 1.000$). However, **the first post-mean score was significantly lower than the second post-mean score ($p = .035$*)**.

3.1.3 The Women in Science Scale (WiSS)

3.1.3.1 Reliability

The internal consistency of the WiSS questionnaire was measured using the SPSS software. The result showed that the WiSS questionnaire was highly reliable (Cronbach's $\alpha = .942$).

3.1.3.2 Normality

The normality test was utilized to assess whether the collected data followed a normal distribution. The results of the Shapiro-Wilk test ($p < .05$) presented that the scores on the WiSS questionnaire were non-normally distributed in the non-engineering and engineering groups. Similarly, the mean scores in both female and male groups showed non-normal distributions. Consequently, non-parametric tests were applied to enhance the robustness of the statistical analysis (Orcan, 2020; Zimmerman & Zumbo, 1993).

3.1.3.3 Findings

Table 21 reports no significant difference between the non-engineering and engineering groups in the mean score on the WiSS questionnaire before instruction. However, **after instruction, the engineering group's mean score on the WiSS questionnaire was significantly lower than that of the non-engineering group**.

Table 21

Differences between the non-engineering and engineering groups in the pre- and post-WiSS

					Mann-Whitney U test			
Aspects	Groups	*n*	*M*	*SD*	*U*	*Z*	*p*	Effect size (η^2)
Pre-WiSS	Non-engineering	23	5.02	1.04	292.50	-.351	.725	.003
	Engineering	27	5.07	.69				
Post-WiSS	Non-engineering	23	5.39	.59	.000	-6.051	.000***	.747
	Engineering	27	1.79	.74				

Table 22 reports the results of the Wilcoxon signed-rank test. As shown in Table 22, the **non-engineering group's mean score on the WiSS questionnaire significantly increased after instruction. Contrarily, the engineering group showed a significant reduction after instruction**.

Table 22

Differences between the pre- and post-WiSS among the non-engineering and engineering groups

		Pre-survey		Post-survey		Wilcoxon signed-rank test		
Groups	*n*	*M*	*SD*	*M*	*SD*	*Z*	*p*	Effect size (Pearson's *r*)
Non-engineering group	23	5.02	1.04	5.39	.59	-3.026	.002**	.631
Engineering group	27	5.07	.69	1.79	.74	-4.541	.000***	.874

Table 23 shows no significant differences between females and males in the pre-mean scores on the WiSS questionnaire. However, after instruction, a significant difference was found, with females scoring significantly higher than males.

Table 23

Differences between females and males in pre- and post-WiSS

					Mann-Whitney U test			
Aspects	Gender	*n*	*M*	*SD*	*U*	*Z*	*p*	Effect size (η^2)
Pre-WiSS	Females	17	5.10	1.08	235.00	-.934	.350	.018
	Males	33	5.02	.74				
Post-WiSS	Females	17	4.68	1.72	140.00	-2.881	.004**	.169
	Males	33	2.80	1.73				

Table 24

Differences between pre- and post-WiSS among females and males

		Pre-survey		Post-survey		Wilcoxon signed-rank test		
Gender	*n*	*M*	*SD*	*M*	*SD*	*Z*	*p*	Effect size (Pearson's *r*)
Females	17	5.10	1.08	4.68	1.72	-.934	.350	.227
Males	33	5.02	.74	2.80	1.73	-4.289	.000***	.747

Table 24 reports no significant differences before and after instruction among females. In contrast, **males significantly reduced the mean scores on the WiSS questionnaire after instruction**. Meanwhile, Friedman's test was utilized to examine the differences across the testing times of the WiSS questionnaire in the engineering group. The results of Friedman's test are shown in Table 25, reporting **a significant difference across the three testing times of the WiSS questionnaire (*p* = .000***)**. Lastly, Table 26 reports that the **pre-mean score was significantly higher than the first post-mean score (*p* = .000***). The first post-mean score was significantly lower than the second post-mean score (*p* = .000***).** No significant difference was found between the pre- and the second post-mean score (*p* = .660).

Table 25

Differences among testing times of the WiSS questionnaire in the engineering group

				Friedman's test		
Testing times	*n*	*M*	*SD*	*df*	x^2	*p*
Pre-WiSS	24	5.07	.65	2	34.587	.000***
Post1-WiSS	24	1.78	.70			
Post2-WiSS	24	4.74	1.25			

Table 26

Pairwise comparison between testing times of the WiSS questionnaire

One-way repeated measures ANOVA				
Pairwise comparison (1 vs. 2)	M_1	M_2	Difference score (1-2)	*p*
Pre-WiSS vs. Post1-WiSS	5.07	1.78	3.29	.000***
Pre-WiSS vs. Post2-WiSS	5.07	4.74	.33	.660
Post1-WiSS vs. Post2-WiSS	1.78	4.74	-2.96	.000***

3.2 Qualitative Analysis

After watching the two narrative VIDS videos, the Pre-Study II and Main Study participants engaged in a 30-minute discussion with their group members. The guided discussion was based on Bandura's Social Learning Theory and consisted of the following questions:

a. *Attention* – "What was wrong in the VIDS videos?" and "Were you aware that latent gender biases had occurred?"
b. *Retention* – "Have you seen or encountered similar situations in real life?" and "How would you have handled those situations now?"
c. *Reproduction* – "How would you handle such situations in the future?" and "Will you speak out if you or your colleague face this bias?"
d. *Motivation* – "What ways you may aid in minimizing latent gender bias in your future workplace?"

Participants for Pre-Study II were then given ample time to write their answers to the questions listed in Table 27. Participants were encouraged to write as much (or as little) as they wanted. The project lead gave the participants as much time as they required, and the participants could leave once they had finished answering the

Participants for the Main Study were also given a list of guided questions to write their self-reflective journals. These sets of questions for both the open-ended questionnaire (OEQ) and self-reflective journals (SRJ) were as follows. All questions were in Traditional Chinese. All the participants answered these questions in Chinese, too. For this report, these questions are translated into English.

Questions posed in the open-ended questionnaire:

OEQ1-1. For video 01, what are your thoughts on what happened to Chen?

OEQ1-2. For video 01, what are your thoughts on what happened to Kee-Youn?

OEQ1-3. For video 01, do you think what you saw is common? Do you think it is reasonable?

OEQ1-4. Do you think gender biases have happened in this scenario (video 01)?

OEQ1-5. What are your thoughts on what happened in this video?

OEQ2-1. For video 02, what are your thoughts on what happened to Sarah?

OEQ2-2. For video 02, what are your thoughts on what happened to Kevin?

OEQ2-3. For video 02, do you think what you saw is common? Do you think it is reasonable?

OEQ2-4. Do you think gender biases have happened in this scenario (video 02)?

OEQ2-5. What are your thoughts on what happened in this video?

OEQ3. What other thoughts or opinions do you have about this entire experiment?

Questions used for the self-reflective journal:

SRJ1-1. Do you think gender biases have happened in this scenario (video 01)? Please elaborate.

SRJ1-2. For video 01, do you think what you saw is common? Do you think it is reasonable? Have you encountered similar scenarios? Have your friends or family encounter similar scenarios?

SRJ1-3. If you were Kee-Youn in video 01, how would you support Chen?

SRJ1-4. If you were Chen in video 01, how would you want Kee-Youn to support you?

SRJ2-1. Do you think gender biases have happened in this scenario (video 02)? Please elaborate.

SRJ2-2. For video 02, do you think what you saw is common? Do you think it is reasonable? Have you encountered similar scenarios? Have your friends or family encounter similar scenarios?

SRJ3. What other thoughts or opinions do you have about this entire experiment?

Table 27: *Information of data collected* (qualitative)

	Pre-Study (II)		Main Study	
Open-ended Questionnaire	**N=23**	F=14; M=9.	**N=27**	F=3; M=24.
Self-reflective Journal	-	-	**N=22**	F=2; M=20.

Content analysis was conducted to answer some of the research questions posed above. As the research questions provided a set of pre-existing ideas for analysis, three coders analyzed and coded the content from open-ended questionnaires from Pre-Study II and Main Study and self-reflective journals from Main Study. Two of the coders are research assistants, while the third coder is the lead researcher. One of the main reasons for engaging research assistants as coders is that the participants and the coders are from the same generation, and thus, the coders will be able to sense the nuance of the answers the participants gave. Participants were free to write as much or as little as they wanted for the open-ended questionnaires and self-reflective journals. Often, they did not state outright what they thought but instead wrote in vague terms. Also, the research did not check the relevancy of their answers. Thus, some questions garnered more responses while others less. Any disagreement with what was written by the participants was resolved through discussion between the three coders.

The majority of the participants were aware that there was latent gender bias happening in both videos, and the percentage of female participants who were aware of the biases was higher (76% of female participants vs 57% of male participants). Those who noticed the latent gender biases in the experimental videos were also outraged that such biases still exist and are pretty standard. They said, in various ways but meant the same thing, the basis of such gender bias is due to the widely accepted notion of ‘different gender, different expectations.’ Men and women are expected to behave in ways that are expected of them. Men are supposed to be aggressive and take the leadership role. Women are supposed to be demure and not be technologically savvy. Men aren’t supposed to take care of their sick children. Women aren’t supposed to be good at driving or fixing things. Once such role incongruity occurs, people might not accept it and, thus, could judge men and women harshly and unfavorably. Here are three excepts from participants of Pre-Study II (Student SP001 and SP019) and Main Study (Student M014)

我出生在一個比較傳統的家庭，上面剛好有一個哥哥，而爺爺奶奶有極度嚴重的重男輕女，對哥哥寵上天卻對我不聞不問，但好在有爸媽加倍的疼愛讓我也很快樂的長大了，但爸爸一直認為男人就要養一個家，所以都會要求哥哥要跟他一起學換燈泡等一些雜事，但卻從不要我學；且我的成績較哥哥優秀，但爸爸一直覺得女生只要找一個好老公嫁就好，不必那麼努力，對此這個研究真的讓我很有感觸。[SP001]

性別偏見，在生活中算是一個常態發生，而且可以說在我們前淺意識中影響我們的立場和偏見，或許在和我們生活成長時收到的男生要做什麼，女生要做什麼的刻板印象有關，在某些領域上，我們可能會對於性別產生預設立場，像是看到男護士會覺得粗魯。[M014]

而日常生活中，我們總是在無意識地做出「性別」的判斷，就如現在要搬東西時，我們一般都會說「請男生去幫忙」，但或許有力氣大的女生也可以去搬，又或者說當要檢查報表等細膩的工作時，通常則是女生，但有些男生的心思細膩跟女生不相上下，那他應該也能做才對；而我們總是常常對於男生跟女生有不同的期待，男生就力氣大、直接、不會讀書、愛玩等，女生愛漂亮、柔弱，但人們總是認為他們不同。[SP019]

The majority of participants (78%) said that even though the state of gender equality in Taiwan has improved tremendously due to the education and promotion of such equity and equality for the past decades, much still needs to be done as gender biases and discrimination have become implicit and less noticeable. As written by Student M0024, they were glad to have the opportunity to reflect upon themselves. Student SP012 stated that they believed that the state of gender equality seemed to stagnate as there are few well-known esteemed female figures to champion this cause.

從大概國小開始其實就有開始聽到再宣導男女平等了，但台灣的社會文化、家庭教育等都依然還停留在舊時代，所以就算學校宣導效果依然有限，而這

次參加這次的實驗其實也是增加對自己的了解，挖掘自己的潛意識，了解自己是否有性別刻板印象但不自覺，藉由這個機會來反思自己、了解自己。而經由這次的實驗我對我自己也有了一些新的發現，我一直以為我算很平等的不太會有偏見，但實驗中跟旁邊的同學討論完後其實有一些地方還是有刻板印象但我卻不自覺，經過此次之後我會將我自己不曾發現但不好的地方修正，並時時反省自身，避免同樣的情況在一次發生。[M024]

透過這 2 個影片，其實還是很明顯感受到性別偏見的存在，大家可能很努力在翻轉這樣的想法，但我認為這當中最大的瓶頸是很多人並不覺得女生的權益受到不公平的對待，而是已經習慣生活在這樣的環境中，同時也缺乏更多有社會地位或權力的女性發聲，因此才會導致女權好像推行很多年，卻有點停滯不前的感覺，也許可以透過憲法強制要求大家，但如何讓社會真正認可男女平等的想法是一個很困難的課題。[SP012]

A few of them shared their own experiences or the incidents that had happened to their friends, classmates, and family members. One of the participants, who is a male, shared the experience of one of his female friends. This female friend is a tech wiz and would love to work with pieces of machinery at her job. Still, she was denied the chance, time and time again, because her bosses did not think that she was competent enough to handle “complicated” machinery and placed her in an administrative position. In the end, she had to quit because she felt being stifled due to these ingrained gender biases. Another male participant shared his experience of how other family members simply assumed he would have better driving skills than his sister because he is a man. Men are supposed to be better drivers than women. In this case, he admitted honestly that his sister was a far better driver than he was. There were several similar incidents that the participants shared that showed that latent gender biases still exist in society, even though these are less obvious than outright discrimination.

A few participants, such as Students M006 and M012, did not view the treatment received by the female protagonist in video 01 as the consequence of (latent) gender bias. They viewed the differences in treatment by the faculty member as the result of the female protagonist's behavior and oral presentation skills and not due to her gender. To these participants, the female protagonist came across as “aggressive” in her second presentation compared to her first. It was this contrast that led to the unfavorable feedback by the faculty member, and not due to her gender. In video 02, the latent gender bias suffered by the female protagonist was much more apparent than in video 01. Thus, more participants were able to point out that the male protagonist was biased against the female protagonists due to his stereotypical views.

沒有，只是個人問題，女主太緊張，然後資料不全，不夠好。[M006]

其實我在第一次觀看影片時並沒有感受出有性別偏見的情況，在我個人眼裡我認為比起性別上的差異更偏向於男演講者和女演講者熟練度問題，再他們第一次報告時台教授也是有刁難它們，兩次報告差異只差在是男性還女性刀刃主演講者而已，我認為今天演講講者角色性別互換男生報告的差也一樣會被刁難。[M012]

When faced with the latent gender bias scenarios, much of the advice from male or female participants was for the women to “be themselves” (做自己), do their best, and not worry too much about what others say. This seemingly positive advice would perhaps reflect the naivete of participants as they have yet to work in an organization. Lastly, about one-third of the participants voiced their appreciation for helping them to be aware of latent gender bias through this experiment. Many of them said that even though awareness promotion has been going on for quite some time, the cultural inertia and traditional views still favor men over women. Having more opportunities to learn about such latent gender bias would help them avoid behaving in a biased manner.

Many participants, such as Students M008 and M010, choose not to speak out or speak up immediately when they or their friends, colleagues, or family members face such latent gender biases. Instead, they chose to address such incidents indirectly, such as privately encouraging those being treated unfairly or talking to their HR separately after the incident. Many, too, said that they want their companies to have fair and transparent policies on handling any such gender bias allegation once it is reported. Those who wish to report such discriminatory behaviors should be protected, and the company should safeguard their anonymity so that they are not targeted for speaking out. All investigations should be done professionally and transparently, and results should only be known to relevant parties. Companies should be seen as taking a “no gender bias” stand. They, including Student M016, also suggested that companies should have regular talks and sensitivity training so that employees are aware of and do not discriminate against the opposite gender, albeit unknowingly.

我認為若是只有我一個人的話我不會選擇發聲，畢竟槍打出頭鳥，但若是我或是我熟識的人遭到類似的狀況，我會選擇默默地幫助他，像是提供他管道去檢舉，或是幫他檢舉，抑或是提供他心靈上的支持等等，但我不會選擇直接與犯人對撞的方式。 [M008]

如果在職場上遇到了性別偏見的狀況，我認我我不會直接跳出來為當事人發聲，因為我會擔心這樣會不會影響到我自己的工作，如果得罪了上司可能自己也會不好過。如果要幫助當事人我可能會選擇私下鼓勵他或者是當一個聽眾聆聽他的情緒，盡我所能的讓他的心情不要那麼不好。[M010]

定期舉辦性別平等和多樣性培訓，提高員工的性別意識，並鼓勵員工參與討論和學習，又或者可以建立性別平等的工作環境，制定和執行反性別歧視的政策，確保所有員工了解並遵守。建立透明的舉報和處理機制，讓員工放心舉報性別偏見。[M016]

At the end of the experiment, most participants commented that the experiment and the videos had made them more aware of some of the still ongoing latent gender biases that society might have, as written by Students M005, M015, and M017. Many said the instructions, including the video watching and guided discussion, helped them understand and realize the seriousness of such latent gender biases. Having candid discussions without judgment and prejudice, even among people they do not know, was beneficial in showing them the viewpoints of others.

性別歧視問題不管是天生產生的還是社會構建形成的，都需要人們去體會以及了解才可以降低性別歧視的發生率。參與這次實驗讓我更加深刻地意識到性別歧視的存在和嚴重性。我深信，只有通過積極的行動和教育，我們才能夠減少甚至消除性別歧視，讓每個人都能夠在一個公平和尊重的環境中自由發展。[M005]

在這次的實驗中讓我了解到性別偏見是社會中一個根深蒂固的問題，它不僅影響著個人的發展和自由，也導致了社會的不公平和分裂。　在現今的社會裡，我們依然可以看到許多不同形式的性別歧視和刻板印象存在於各個層面。在職場上，男性往往被認為更適合擔任領導職位，而女性則常常受到性別歧視和職場欺凌。在教育領域，依然存在著對於男女學生不同的期望和要求，這種偏見影響著學生的學習和成就。也阻礙了整個社會的發展和進步。要解決性別偏見問題，我們需要從教育、制度和行動等多個層面入手。[M015]

性別偏見確實是一個亟要解決的潛在問題，但是因表現的方式實在太多元且無形，很難去定義甚麼較偏見，甚麼叫做沒偏見，什麼又是平等對待。這種偏見不僅限於男女之間的差異，同性間也可能存在著細微的歧視和偏見。透過討論的方式聽見不同人的發言也可以知道大家的想法，說不定也能從他人的想法裡面獲得不同感悟或是發現原來自己以前的想法是錯的，別人並不這麼認為，這種來自他人的觀點和感悟是寶貴的，它挑戰了我們的舊有思維模式，也促使我們學會從不同角度來看待問題。[M017]

4. Conclusions

Data analysis showed a significant increase in the Ambivalent Sexism Index (ASI) among the engineering group after treatment, especially among male participants compared to female participants, as indicated in Tables 3, 4, and 5. Looking deeper, it was found that the engineering group showed a significant increase in benevolent sexism (BS) (Table 10) compared to the non-engineering group, with a similar significant increase among males than females (Table 11). Male participants in the engineering group showed a significantly higher sense of BS after the instruction. This could be explained by the fact that these participants were reminded of their gender equality education, which made them want to protect and support the weaker gender. However, this increase in BS did not last, as seen in Table 15, where the first mean score for BS was significantly higher than the pre-mean (p=.004**) and second post-mean scores (p=.008**). This meant that male participants felt a heightened sense of benevolent sexism (BS) immediately after the guided discussion. Still, this sense of BS fell back to their default baseline when they did their second post-questionnaires four weeks later.

The result from the Ambivalent Sexism Index can be collaborated with the data analysis results from the Men's Polarized Gender Thinking Questionnaire (MPGQ) and the Women in Science Scale (WiSS). Analyzing the data for MPGQ, after the instruction to view the videos and conduct guided discussions, a) the mean score of the engineering group was significantly lower than that of the non-engineering group, as shown in Table 16; b) the engineering students' mean scores reduced significantly, as indicated in Table 17; and c) the male participants' mean scores showed a significant reduction, as indicated in Table 19. Similarly, for the WiSS, a) the mean score of the engineering group was significantly lower than that of the non-engineering group, as shown in Table 21; b) the engineering students' mean scores reduced significantly, as indicated in Table 22, and c) the male participants' mean scores showed a significant reduction, as indicated in Table 23, after the instruction to view the videos and conduct guided discussions. Taken together, this means that this type of instruction, namely video-watching and guided discussion, can reduce the gender biases of these participants.

Nonetheless, the effect of this instruction was not long-lived. After four weeks, the Post2 mean scores for ASI, MPGQ, WiSS, although they did not return to their pre-instruction mean scores (Pre-), were either slightly higher or lower. For ASI, the mean scores of Benevolent Sexism (BS), Hostile Sexism (HS), and Ambivalent Sexism (AS) for [Pre, Post1, and Post2] were [2.82, 4.03, and 2.93], [3.37, 3.69, and 3.29], and [3.09, 3.86, and 3.11] respectively for the engineering group. The engineering group's MPGQ mean scores for Pre, Post1, and Post2 were 3.69, 3.35, and 3.76, respectively. For WiSS, the mean scores for Pre, Post1, and Post2 were 5.07, 1.78, and 4.74 for the same group. A summary of results between various comparison settings can be seen in Tables 28, 29, and 30 for ASI, MPGQ, and WiSS.

Table 28. Instructional Effects using ASI

Immediately after the instruction	Effectiveness
The engineering group was **MORE** biased than the non-engineering group.	Not Effective
The engineering group was **MORE** biased than before instruction.	Not Effective
Male participants were **MORE** biased than female participants.	Not Effective
Male participants were **MORE** biased than before instruction.	Not Effective
The engineering group was **MORE** biased compared to 4 weeks later.	Not Effective
The engineering group's **BS was higher** than that of the non-engineering group.	Not Effective
The engineering group's **BS was higher** than that before the instruction.	Not Effective
Male participants' **BS was higher** than female participants.	Not Effective

Table 29. Instructional Effects using MPGQ

Immediately after the instruction	Effectiveness
The engineering group was **LESS** biased than the non-engineering group.	Effective
The engineering group was **LESS** biased than before the instruction.	Effective
Male participants were **LESS** biased than before the instruction.	Effective
The engineering group was **MORE** biased compared to 4 weeks later.	Not Effective

Table 30. Instructional Effects using WiSS

Immediately after the instruction	Effectiveness
The engineering group was **LESS** biased than the non-engineering group.	Effective
The engineering group was **LESS** biased than before the instruction.	Effective
Female participants were **MORE** biased than male participants.	Not Effective
Male participants were **LESS** biased than before the instruction.	Effective
The engineering group was **MORE** biased compared to 4 weeks later.	Not Effective

This research might have failed at its objectives, with so many mishaps plaguing it from the beginning. Yet, there were several silver linings to all these, even when it lost one of its data triangulation points. The results from three different scales, completed by the same set of participants, showed that latent gender bias has a variety of dimensions. Even if one measurement indicated latent gender bias among the participants, another tool might show otherwise. Even though, after the instruction, engineering and male students do not endorse polarized gender thinking (as seen from the result from MPGQ), and they disagreed with traditional sexist beliefs and stereotypes regarding women's roles in science (as seen from the result from WiSS), they did hold heightened benevolent sexism (BS) attitudes towards women (as seen from the result from ASI), further endorsing the paternalistic views commonly held in

East Asian societies that are still somewhat heavily influenced by Confucianism. However, the team could not determine the cause-effect relationship due to the limited sample size.

5. Limitations and Future Improvement

On top of the deteriorated conditions of the Neurosky Mindwave devices that render them useless, the main issue for this research project is the limited number of participants. As advised by the IRB board, participation in this research has to be completely voluntary, and students shouldn't be pressured into participating by any means. Thus, participants were recruited only through posters around the campus and word-of-mouth communication among students. No promotion for this event was done through "official means" or encouragement from other faculty members. Furthermore, even after doubling the remuneration of participation, from the initial NTD $400 for the Main Study to NTD $800, the response remained lukewarm at best. Another issue with the data collection was that there was no leverage to enforce how much (or little) the students could write for their open-ended questionnaires and self-reflective journals. Lastly, many of the self-reflection journals received from engineering students contain shorter paragraphs and fewer words written than those of non-engineering students.

Future improvements to similar research on this topic should include a more extended participant recruitment period, minimal length requirement for self-reflective journals, and using AI to analyze the contents from open-ended questionnaires and self-reflective journals.

Reference:

Bekhet, A. K., & Zauszniewski, J. A. (2012). Methodological triangulation: An approach to understanding data. *Nurse Researcher*, 20, 40–43.

Bergman, B., Larsman, P., & Löve, J. (2014). Psychometric evaluation of the "Men's Polarized Gender Thinking Questionnaire (MPGQ)". *Gender in Management: An International Journal*, *29*(4), 194-209.

Breckler, S. J. (1984). Empirical validation of affect, behavior, and cognition as distinct components of attitude. *Journal of personality and social psychology*, *47*(6), 1191.

Fusch, P., Fusch, G. E., & Ness, L. R. (2018). Denzin's paradigm shift: Revisiting triangulation in qualitative research. *Journal of Social Change*, 10(1), 2.

Glick, P., & Fiske, S. T. (1996). The Ambivalent Sexism Inventory: Differentiating hostile and benevolent sexism. Journal of Personality and Social Psychology, 70(3), 491-512. https://doi.org/10.1037/0022-3514.70.3.491

Kwak, S. G., & Kim, J. H. (2017). Central limit theorem: the cornerstone of modern statistics. Korean journal of anesthesiology, 70(2), 144-156.

https://doi.org/10.4097/kjae.2017.70.2.144

Lim, F. V., O'Halloran, K. L., & Podlasov, A. (2012). Spatial pedagogy: Mapping meanings in the use of classroom space. *Cambridge Journal of Education*, 42(2), 235-251.

Manganelli, J., Threatt, A., Brooks, J., Healy, S., Merino, J., Yanik, P., & Green, K. (2014). Confirming, classifying, and prioritizing needed over-the-bed table improvements via methodological triangulation. *Health Environments Research & Design Journal*, 8, 94–114.

Moss-Racusin, C. A., Pietri, E. S., Hennes, E. P., Dovidio, J. F., Brescoll, V. L., Roussos, G., & Handelsman, J. (2018). Reducing STEM gender bias with VIDS (video interventions for diversity in STEM). *Journal of Experimental Psychology: Applied*, *24*(2), 236.

Orcan, F. (2020). Parametric or non-parametric: Skewness to test normality for mean comparison. *International Journal of Assessment Tools in Education*, *7*(2), 255-265. https://doi.org/10.21449/ijate.656077

Owen, S. V., Toepperwein, M. A., Pruski, L. A., Blalock, C. L., Liu, Y., Marshall, C. E., & Lichtenstein, M. J. (2007). Psychometric reevaluation of the Women in Science Scale (WiSS). Journal of Research in Science Teaching, 44(10), 1461-1478. https://doi.org/10.1002/tea.20187

Pietri, E. S., Moss-Racusin, C. A., Dovidio, J. F., Guha, D., Roussos, G., Brescoll, V. L., & Handelsman, J. (2017). Using video to increase gender bias literacy toward women in science. *Psychology of Women Quarterly*, *41*(2), 175-196.

Sullivan, G. M., & Feinn, R. (2012). Using effect size—or why the P value is not enough. Journal of Graduate Medical Education, 4(3), 279–282. https://doi.org/10.4300/JGME-D-12-00156.1

Zimmerman, D.W., & Zumbo, B. D. (1993). The relative power of parametric and nonparametric statistical methods. In G. Keren & C. Lewis (Eds.). *A handbook for data analysis in the behavioral sciences: Methodological issues* (pp. 481-517). New York, NY: Lawrence Erlbaum Associates.

附件三

國家科學及技術委員會補助專題研究計畫成果彙整表

<table>
<tr><td colspan="3">計畫主持人：吳碧瑛</td><td colspan="3">計畫編號：NSTC112－2629－H－224－001－SSS</td></tr>
<tr><td colspan="6">計畫名稱:「以多模式方法探討如何提升工科生的潛在性別偏見意識及降低在 STEM 環境裡的性別偏見(L02)」</td></tr>
<tr><td colspan="3">成果項目</td><td>量化</td><td>單位</td><td>質化
（說明：各成果項目請附佐證資料或細項說明，如期刊名稱、年份、卷期、起訖頁數、證號...等）</td></tr>
<tr><td rowspan="6">國內</td><td rowspan="6">學術性論文</td><td>期刊論文</td><td>1</td><td rowspan="2">篇</td><td>投稿中</td></tr>
<tr><td>研討會論文</td><td>0</td><td></td></tr>
<tr><td>專書</td><td>0</td><td>本</td><td>請附專書資訊。</td></tr>
<tr><td>專書論文</td><td>0</td><td>章</td><td>請附專書論文資訊。</td></tr>
<tr><td>技術報告</td><td>0</td><td>篇</td><td></td></tr>
<tr><td>其他</td><td>0</td><td>篇</td><td></td></tr>
<tr><td rowspan="6">國外</td><td rowspan="6">學術性論文</td><td>期刊論文</td><td>0</td><td rowspan="2">篇</td><td>請附期刊資訊。</td></tr>
<tr><td>研討會論文</td><td>0</td><td></td></tr>
<tr><td>專書</td><td>0</td><td>本</td><td>請附專書資訊。</td></tr>
<tr><td>專書論文</td><td>0</td><td>章</td><td>請附專書論文資訊。</td></tr>
<tr><td>技術報告</td><td>0</td><td>篇</td><td></td></tr>
<tr><td>其他</td><td>0</td><td>篇</td><td></td></tr>
<tr><td rowspan="10">參與計畫人力</td><td rowspan="5">本國籍</td><td>大專生</td><td>5</td><td rowspan="10">人次</td><td rowspan="10">請填寫依「國家科學及技術委員會補助專題研究計畫研究人力約用注意事項」所實際約用專任、兼任人員。</td></tr>
<tr><td>碩士生</td><td>1</td></tr>
<tr><td>博士生</td><td>0</td></tr>
<tr><td>專任人員(博士級)</td><td>0</td></tr>
<tr><td>專任人員(非博士級)</td><td>0</td></tr>
<tr><td rowspan="5">非本國籍</td><td>大專生</td><td>0</td></tr>
<tr><td>碩士生</td><td>0</td></tr>
<tr><td>博士生</td><td>0</td></tr>
<tr><td>專任人員(博士級)</td><td>0</td></tr>
<tr><td>專任人員(非博士級)</td><td>0</td></tr>
<tr><td colspan="3">其他成果
(無法以量化表達之成果如辦理學術活動、獲得獎項、重要國際合作、研究成果國際影響力及其他協助產業技術發展之具體效益事項等，請以文字敘述填列。)</td><td colspan="3">有助於提升學生(理工及非理工生)的潛在性別歧視的認知及意識。 大部分參與的學生都認為性別意識已相當普及，故沒有必要進行類似的教學，但參與此項研究後,他們發現自己的認知和意識仍不足，此研究中所使用的短片讓他們看到了潛在的性別歧視,進而激發他們的反思。</td></tr>
</table>

附件八

國家科學及技術委員會補助專題研究計畫

「研究中的性別考量」報告表

計畫編號：NSTC112－2629－H－224－001－SSS

研究人員姓名： **吳碧瑛**

任職機關系所： **國立雲林科技大學高階管理碩士學位學程**　職稱： **助理教授(專案)**

計畫名稱：**以多模式方法探討如何提升工科生的潛在性別偏見意識及降低在 STEM 環境裡的性別偏見 (L02)**

說明：本年度專題研究計畫涉及「人體試驗」或「人體研究」，請於計畫進度報告/成果報告時一併繳交「研究中的性別考量」報告表。

<table>
<tr><th>項次</th><th>項目</th><th>說明</th></tr>
<tr><td>1</td><td>是否有<u>記錄</u>已招募/納入之研究參與者或人體檢體樣本數之生理性別比例？</td><td>■ 有，比例如下：
Male: __33　（66%）
Female: _17___（34%）

□ 有，請參考進度報告/成果報告第__頁。
□ 無，本計畫採單一性別研究設計，理由：
______________________**（結束填答）**
□ 無，本計畫規劃不記錄性別，理由：
______________________**（結束填答）**</td></tr>
<tr><td>2</td><td>是否有依生理性別<u>分組</u>報告結果？
（例：Clayton & Tannenbaum, 2016, JAMA）
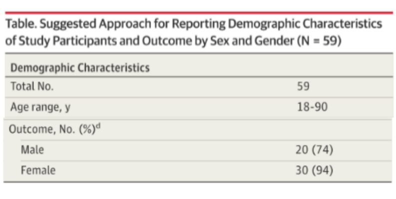
Table. Suggested Approach for Reporting Demographic Characteristics of Study Participants and Outcome by Sex and Gender (N = 59)<table><tr><th>Demographic Characteristics</th><th></th></tr><tr><td>Total No.</td><td>59</td></tr><tr><td>Age range, y</td><td>18-90</td></tr><tr><td>Outcome, No. (%)[d]</td><td></td></tr><tr><td>Male</td><td>20 (74)</td></tr><tr><td>Female</td><td>30 (94)</td></tr></table></td><td>□ 有，研究結果已發表，請參考文獻：
____
■ 有，研究結果未發表，請參考進度報告/成果報告第 5-18 頁。

□ 無，研究進行中，尚無結果。(結束填答)
□ 無，本計畫規劃不依性別分組報告結果，理由：
______________________**（結束填答）**</td></tr>
</table>

3	是否有對生理性別進行分析（例如差異分析、相關與迴歸分析等）？	□ 有，研究結果已發表，請參考文獻： ______________________________ ■ 有，研究結果未發表，請參考進度報告/成果報告第 5-18 頁。 □ 無，本計畫未規劃對生理性別進行分析。