Gender Differences in the Use of Storytelling, Analogy, and Metaphor to Promote Empathy among Science Graduate Students

Lan Luo[1] and Qingqing Xing[2,*]

[1] Society Hub, The Hong Kong University of Science and Technology (Guangzhou), Guangzhou, China
[2] College of Education Sciences, The Hong Kong University of Science and Technology (Guangzhou), Guangzhou, China
Email: lluo476@connect.hkust-gz.edu.cn (L.L.); claireqqxing@hkust-gz.edu.cn (Q.X.)
*Corresponding author

***Abstract*—Although extensive research has emphasized gender differences in empathy, the results have been mixed, and there is limited understanding of how these differences are reflected in students' empathetic awareness within their research contexts. To fill in the gap, we apply narrative pedagogy using storytelling, analogies, and metaphors in a design thinking course for 236 science master students at China's first interdisciplinary research university, and examine its effect on their empathy awareness in multi-modal presentation works through sentiment discourse analysis. Using Natural Language Process method with a Naive Bayes Classifier, we then conducted a discourse analysis of the 3 sequential multi-modal tasks and found that male students tended to use more words in the 3-minute video report than female students. Different from the previous research, there was hardly any marginal difference detected between male and female students in terms of emotional awareness. Both female and male students showed similar levels of emotional expression before empathy training, and both genders showed significant improvement after training. These results provide new insights into gender differences in empathy awareness among research students and highlight the effectiveness of innovative pedagogical approaches to promote empathy.**

***Keywords*—gender difference, storytelling, analogy, metaphor, science communication, Natural Language Process (NLP)**

I. Introduction

Science communication ability has been increasingly recognized in the literature as one of the most important soft skills for research students [1–3], and teaching in this area is now practiced at research-oriented universities worldwide [4]. However, little attention has been paid to how science communication skills differ between female and male graduate students. This ignorance is not only evident in the literature, but also in pedagogical observation: we have rarely seen practitioners focus on gender-specific communication characteristics and educators seldom take care to address gender differences in soft skill building. A gender-inclusive study of the development on the soft skills courses for research students is therefore in dire need. Our starting point is empathy building through classroom training and project work among science and engineering graduate students at a newly established interdisciplinary university in China.

Research on the relationship between science communication and gender offers findings ranging from direct behavioral measurements [5–8] to complex neurobiological [9, 10] and sociocultural analysis [11]. There is evidence that there are differences in the capacity for empathy between men and women. Stereotypes typically portray women as more caring and empathetic and men as less emotional and cognitive and research evidence support this [12]. In addition, research has shown that women in lead author positions in psychology publications are less likely to use generic language compared to men [6]. Researchers have also developed Scientific Empathy Index (SEI) to measure these traits of scientific empathy [13]. The SEI measures students' sensitivity, situational interest, scientific imagination, empathetic concern, and empathetic understanding of others, reflecting both cognitive and affective dimensions of their engagement in scientific activities. However, research on promoting empathy among science graduate students is sparse both in theory and in practice.

Design thinking and scientific empathy are closely linked in educational contexts, particularly in how they both foster a deeper understanding and engagement with subjects. Design Thinking, as an interdisciplinary course, not only focuses on the mindset and skills required for creative problem solving, but is also based on innovation through human-centered design principles. Design thinking courses have attracted the attention of both researchers and educational practitioners with the aim of equipping students with an innovative and problem-solving mindset and preparing them for various industries, the design thinking course has been shown to be beneficial for students' innovative learning from various

Manuscript received March 11, 2024; revised June 3, 2024; accepted June 11, 2024; published October 18, 2024.

disciplinary perspectives, such as eliciting changes in brain activation [14], improving product, process, and organizational innovation [15], and fostering social entrepreneurship [16]. Rarely have we seen literature on gender differences in the impact of design thinking courses.

Training research students with design thinking skills can be an effective way to foster and cultivate empathy [17]. Practices in healthcare and aging studies have explored the relationship between empathy and narratives in building a shared understanding among stakeholders [18]. However, there is little research conducted in a specific course development context using longitudinal data to measure academic empathy. At the same time, demonstration on the impact of empathy teaching is hardly observed. Our research aims to address this gap by collecting and analyzing empirical data from a group of science and engineering students taking a design thinking course over three months and examining the gender differences in their writing works on empathy awareness.

Our research aims to investigate the gender difference in soft skills development by observing a graduate design thinking course. Despite the fact that students feel AIGC tools facilitate science communication [19], effective verbal communication and the expression of thoughts and ideas with layman audiences remain important for graduate students. However, the "deficit model" of science communication is still prevalent, and the ability to communicate science to laypeople is not yet a focus of teaching or learning in most research-oriented universities. Therefore, our research aims to shed light on the promotion of research students' soft skills in today's digital age.

We want to investigate the following questions:

First, how do narrative techniques such as storytelling, analogy, and metaphor as pedagogical devices influence the development of empathy in female and male students in a design thinking course?

Second, how effective are storytelling, analogy, and metaphor in promoting empathy skills in male and female students in a design thinking course in comparison?

II. Institutional Setting and Research Background

Empathy, as the first step in Design Thinking, plays an important role in the design thinking process to search for rich stories and discover what people truly need [20]. One of the intended learning outcomes of the design thinking course at this interdisciplinary university aims to improve students' empathy through academic communication. The course in 2022–2023 Fall semester engaged 236 MPhil students over a 3-month period, teaching them the five iterative steps of design thinking: Empathizing with Users, Defining the Problem, Developing Solutions, Prototyping Options, and Testing Results. A diverse team of lecturers, including language tutors focusing on effective research communication, supported these students. Rather than traditional lectures on science communication, the course used project design scenarios that challenged students to design their own projects. This approach is intended to foster their ability to communicate empathically with non-professional audiences but also enhances their capacity to handle interdisciplinary project topics, catering to their diverse academic backgrounds.

We developed the idea of "Empathic Narratology for Science Communication" for the first time and put it into teaching practice. Based on our need in the interdisciplinary context, with prior knowledge on the narrative intelligence and scientific empathy, we introduced an empathy-based communication model using storytelling, analogies, and metaphors in project-based learning process. Storytelling, in which the research question or gap is presented as "the tension" or "the conflict" to elicit empathetic understanding from the audience, fosters a shared space between the sender and receiver of the information; analogy facilitates the explanation of the scientific mechanism. Complex scientific concepts are easier to understand when comparing an unfamiliar mechanism with a familiar one. Metaphors are used to stimulate creativity and visualize abstract and vague ideas. For example, when students are trying to name a new algorithm designed by themselves, we encourage them to use metaphorical expressions.

Students were then asked to finish three distinctive tasks in accordance with the Design Thinking steps. The first, Opportunity Finding", is about exploring research questions. Students should identify existing research gaps and opportunities that they are interested in. This task involves collecting data, reviewing scientific literature, and working with industry professionals to discover unmet needs and new areas for innovation. The second task is called "Empathizing to understand the problem". It follows on from the first task to select key users and stakeholders in a structured process. Students created an interview schedule identifying the number of interviews and the questions with different groups to gather information. They were then required to observe one or two interviewees and record key behaviors and interactions. In addition, they must reflect on their observations to identify areas for deeper understanding and create empathy maps from the interview data. The third task, "Storytelling, Analogy and Metaphor to present your research with 1 A4 page of paper", was about using narratives for science communication: students develop compelling narratives about their projects by using rhetorical devices to communicate complex scientific concepts to a non-professional audience.

The introduction of this "Empathy-Based SAM model" echoes the interdisciplinary nature of the university, we have found that simpler language is the first choice for better communication between these young researchers when clear understanding is needed. This is especially true in the practice of academic presentations, and some of the work has been tried out at the university's Three Minute Thesis Competition and set the foundation for a modularized blended course.

III. RESEARCH METHOD

We use the Natural Language Process (NLP) with a Naive Bayes Classifier to conduct the sentiment analysis of student works, and the Kolmogorov-Smirnov test and Mann-Whitney U-test to analyze the gender differences in the sentiment scores.

Fig. 1 shows the process of NLP training of the classifier, analyzing the sentiment of students' assignments, and determining whether scores are related to gender. The model is trained with an extensive corpus categorized into positive and negative sentiments and undergoes Inverse Document Frequency (IDF) pre-processing to determine word meaning and convert the sentences into feature sets for training. Sentiment analysis then processes the students' assignments and scores each line to calculate a cumulative sentiment score, which provides insights into the emotional tone of the assignments through sophisticated text analysis. The final step in Section III.*D* is to compare these aggregated sentiment scores between the genders and visualize them in a histogram to identify any differences in sentiment between the male and female groups.

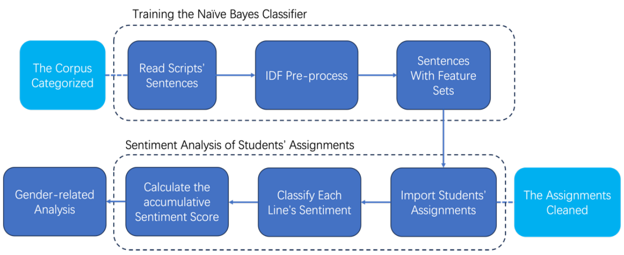


Fig. 1. The NLP algorithm process.

A. Data Collection and Methodology Validation

We processed our data with two parts of the work. In the first part, the data to train Naïve Bayes classifier for sentiment analysis is from https://www.cinafilm.com/ with a self-developed crawler. The corpus is categorized into positive and negative categories. In total, there are 104,523 negative words and 105,670 positive words to train the three assignments of the students. The second part of the data is the text from the students' tasks, which contains project report videos, group project reports, and one-page text-image reports. To extract text from various sources, different strategies are applied. For videos, speech-to-text technology converts spoken words into written text, accommodating audio quality and technical jargon. Text-based reports are processed through document parsing techniques, where software tools extract readable text from formats like PDFs or Word documents. This multi-faceted approach ensures comprehensive text extraction from diverse sources, necessitating subsequent review and refinement to ensure accuracy and context preservation.

The dataset is chosen from the movie review website for the following reasons. First, design thinking itself is an interdisciplinary approach that combines knowledge and skills from different fields such as art, science, and technology. Using the sentiment analysis method of movie reviews, the techniques of literary and art criticism can be introduced into the study of design thinking, which is in line with the interdisciplinary nature of design thinking. Secondly, film criticism is the emotional expression of the critic based on the plot of the movie. It is thus suitable for analyzing students' emotions when using analogies, metaphors, and narratives. Third, the multimodal assignments in this course are not only written expressions, but also visual art, dynamic demonstrations, and other expressions in videos. Sentiment analysis based on movie reviews can help to analyze the expressions of emotions and attitudes in these different modes.

B. Training the Naïve Bayes Classifier

The classifier is trained by reading and processing sentences from the two film review scripts, one containing positive sentences and the other containing negative sentences. It calculates the frequency and relevance of sentiment words by using techniques such as Inverse Document Frequency (IDF) to weigh the importance of each term in the context of the document [21]. The words are subjected to IDF preprocessing before the sentences are converted into feature sets for training the classifier. The classifier is then used to analyze the sentiment in the text files located in the local directory and then to classify the sentiment in each line as positive or negative of students' assignments.

Naive Bayes classifiers offer several advantages over traditional corpus-based methods in Natural Language Processing (NLP) tasks. From the simplicity and efficiency perspective, naive Bayes models are known for their simplicity, allowing for quick training and rapid predictions on new data. This makes them particularly useful for handling large datasets. Next, due to our small datasets in the test training model, Naive Bayes can perform well even with limited data. This is due to its application of Bayes' theorem and the assumption of independence among features, which simplifies the learning process.

Naive Bayes provides not just classification outcomes but also probabilistic estimates for belonging to different classes, which is useful for decision-making processes where weighing uncertainty is crucial. In our study, use Naive Bayes to train the sentence "The cinematography is pretty great in this movie" and has a 0.69 probability of being positive. And the sentence "The cinematography is not much pretty great in this movie" has a 0.79 probability of being negative.

C. Sentiment Analysis of Students' Assessments

The Naive Bayes Classifier, after being trained on a diverse corpus, processes the text from the students' tasks by breaking it down into tokens. Each word or sentence is analyzed for sentiment polarity based on the training. The classifier then categorizes these sentiments as positive and negative. By summarizing these sentiments across the entire task, it calculates a cumulative sentiment score that reflects the overall emotional mood. This score can provide insight into the student's engagement and understanding, and potentially highlight areas where emotional support or academic interventions could be beneficial. This detailed analysis allows educators to tailor their feedback and support to create a supportive learning environment.

D. Gender-Related Analysis

We then conducted gender-related analysis with the sample test as shown in Fig. 2. We calculated the emotional expressions in assignment 1 and compared the difference between genders and adjusted our method. To analyze the relationship between gender and sentiment scores, our first consideration was to apply the histograms and quantile-quantile plots to obtain a visual assessment of the normality of the "sum" values [22] and Kolmogorov-Smirnov test to calculate the exact value of the normality of the "sum". If the histograms and quantile-quantile plots show the two datasets are normal distribution, then will use the independent samples t-test with the *p*-value results to determine if they are related. When the *p*-value is less than 0.05, there is a significant difference between the two samples, and vice versa. However, if the distribution of "sum" values within each gender group is not normal, we decided to use the Mann-Whitney U-test instead of the t-test. The Mann-Whitney U-test is the non-parametric equivalent of the t-test for independent samples, which could compare the differences between the genders in their expressions of emotion.

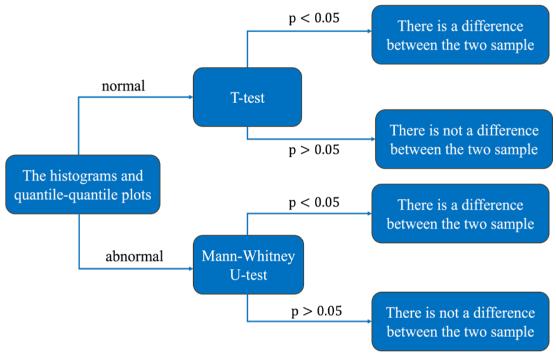


Fig. 2. The method to determine whether two variables are related.

IV. Results and Discussion

The histograms (a), (b), and (c) in Fig. 3, overlaid with a density plot, compare the distribution of sentiment analysis scores for 204 valid assignments between two gender groups, including 127 males and 77 females. The "sum" on the x-axis represents the cumulative sentiment score, and the "density" on the y-axis refers to the probability of a sentiment score occurring within a given range in the data set. Overlaying the two histograms allows a visual comparison of the distribution of sentiment by gender.

In Assignment 1, the sentiment scores for males and females both center around 0, which indicates an overall neutral mood, with women tending slightly toward the negative. The sentiment analysis for Assignment 2 shows a more positive outlook, with female scores peaking around 10 and male scores showing a broad positive sentiment; the mean and median scores for both genders are positively skewed. The multimodal sentiment distribution of Assignment 3 indicates diverse emotional expression, with mean scores for both genders being slightly positive. Overall, each assignment has unique emotional patterns that reflect different emotional intensity and breadth across genders.

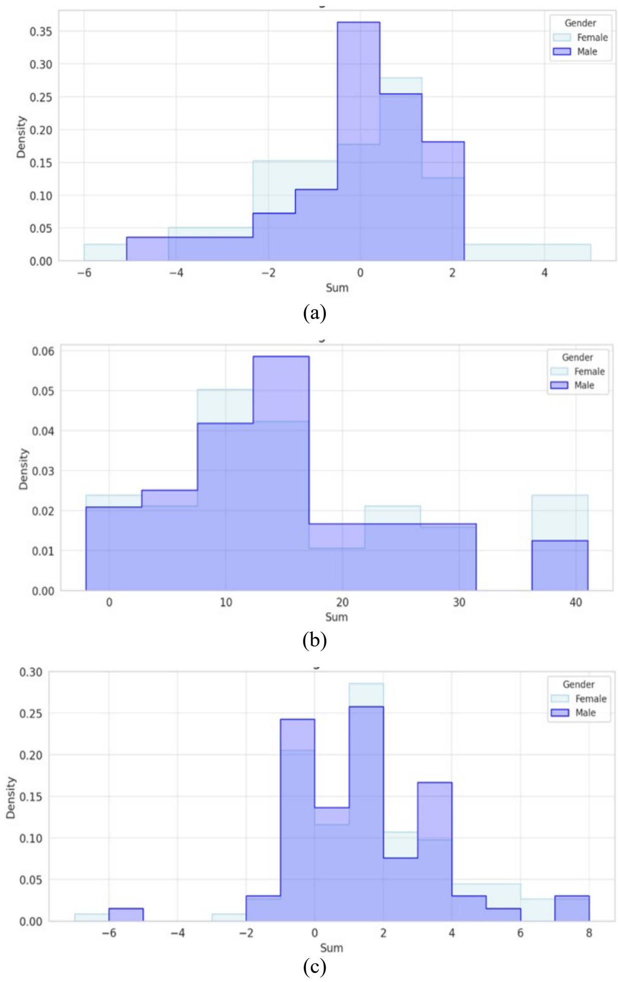
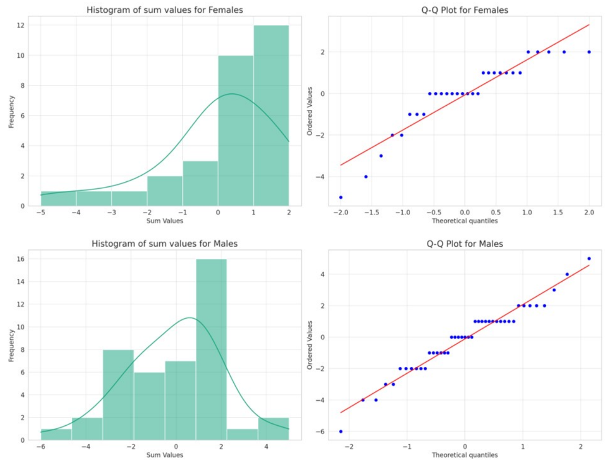


Fig. 3. The sum-gender values in (a) Assignment 1, (b) Assignment 2, (c) Assignment 3.

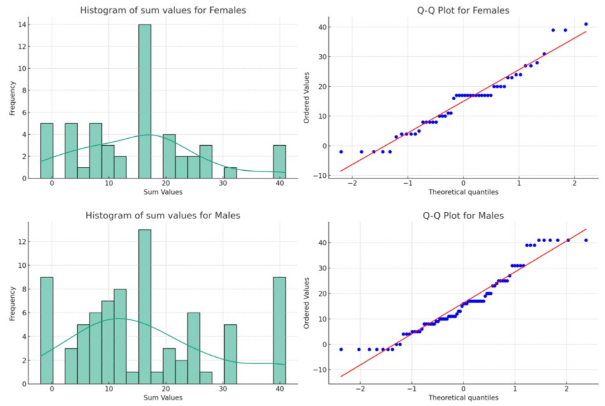
We use quantile-quantile plots (hereafter referred to as "Q-Q plots") to compare the quantiles of the "sum" scores with the quantiles of a normal distribution. The Q-Q plot provides a visual assessment of the normality of the "sum" values for each gender group. The histograms for both females and males show the distribution of 'sum' values with a kernel density estimate superimposed to estimate the probability density function. A linear pattern in the points suggests that the data may follow a normal distribution.

These visual methods are supplemented by statistical tests for normality, such as the Kolmogorov-Smirnov test, to accurately determine whether the "sum" values for each gender follow a normal distribution. If the data deviates significantly from the normal distribution, non-parametric tests such as the Mann-Whitney U-test are recommended for further analysis. The Mann-Whitney U-test was performed to examine the relationship between gender and "sum" values. The result of the test yielded a U-statistic of 626.0 and a *p*-value of approximately 0.833.

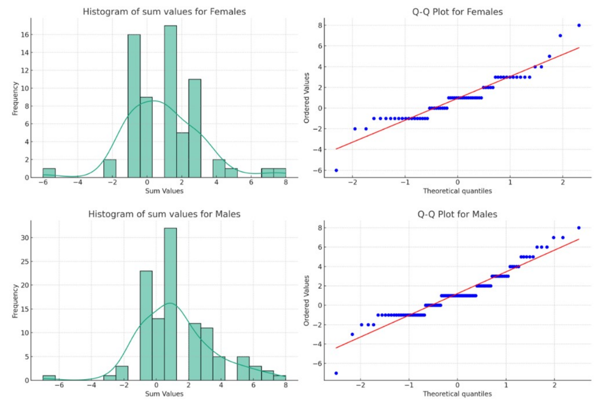
The *p*-value is significantly greater than the typical alpha level of 0.05, indicating that there is no statistically significant difference in the distribution of "sum" scores between female and male groups. (See Fig. 4)



(a)



(b)



(c)

Fig. 4. The normality of "sum" values in (a) Assignment 1, (b) Assignment 2, (c) Assignment 3.

The visual and statistical assessments provide insight into the normality of the sum score distribution for both genders in Assignment 2. The Kolmogorov-Smirnov test was performed for each gender group, comparing their sum values to a normal distribution with the same mean and standard deviation as the observed data. For females, the K-S statistic is approximately 0.136 with a *p*-value of about 0.290. For males, the K-S statistic is approximately 0.134 with a *p*-value of about 0.105. The histograms and Q-Q plots suggest that neither group perfectly follows a normal distribution; however, they do not show significant deviation from normality. The Q-Q plots do not display a perfectly straight line which would indicate normality, but they are not highly deviated either. The Kolmogorov-Smirnov test results show *p*-values higher than the typical alpha level of 0.05, meaning that we do not reject the null hypothesis of normality for either group based on this test.

The Mann-Whitney U test yields a U statistic of approximately 1911 and a *p*-value of about 0.758. The *p*-value is well above the conventional alpha level of 0.05, indicating that there is no statistically significant difference in the 'sum' values distribution between genders based on this test. Therefore, we do not have evidence to suggest that 'sum' and 'gender' are correlated.

Here is the result of Assignment 3. The Kolmogorov-Smirnov test compares the 'sum' values to a normal distribution with the same mean and standard deviation. For females, the K-S statistic is approximately 0.171, with a *p*-value of about 0.038. For males, the K-S statistic is approximately 0.187 with a *p*-value of about 0.00064. The histograms suggest that the distributions of 'sum' values for both genders have deviations from a normal distribution, as indicated by the shape of the KDE curves. The Q-Q plots for both genders also show some deviation from the expected straight line for a normal distribution, particularly in the tails. The K-S test results for both genders yield *p*-values below the standard alpha level of 0.05, suggesting that the distributions of 'sum' values for both females and males are not normally distributed.

The Mann-Whitney U test for the new dataset yields a U statistic of approximately 3464 with a *p*-value of about 0.478. This *p*-value is above the conventional alpha level of 0.05, suggesting that there is no statistically significant difference in the distribution of 'sum' values between the female and male groups. Therefore, the test does not provide evidence of a correlation between 'sum' and 'gender' for the given data.

In Assignment 3, the exercise focused on empathy was completed collaboratively in small groups. The analysis below represents the composition of these groups, detailing the number of female and male participants within each group, as well as the sum values that potentially reflect a measure of empathy or related metrics observed or calculated during the exercise. The bar chart in Fig. 5 has been created based on the groups specified. It shows the count of females (in pink) and males (in blue) for each group. The total 'sum' values are also indicated above the corresponding bars for each group.

The histograms for women generally show a normal distribution of sum scores in the Assignment 1 analysis, but this normality is less evident in the second and third analyses, suggesting a change in emotional expression over time or across different assignments. For men, the distribution in the first analysis is relatively normal but with a positive skew, while subsequent analyses show a multimodal distribution, indicating variability in emotional expression.

The Q-Q plots for women consistently show a deviation from normality, particularly at the tails of the distribution in all three analyses, indicating the presence of outliers or extreme sentimental values. The Q-Q charts for men show a similar pattern of deviations from normality, particularly in the Assignment 2 analysis, which may indicate more pronounced expressions of emotion below the extreme values. In the context of empathy, these patterns might suggest that male students express a wide range of feelings, which could indicate different levels of engagement and empathic responses to the projects, especially when the tasks involve topics that elicit strong emotional responses.

The Mann-Whitney U test was performed to check if there is a statistically significant difference between the 'sum' values multiplied by the count of females and males across the groups. The test statistic is approximately 21.0 and the *p*-value is approximately 0.710. A *p*-value higher than the common alpha level of 0.05 suggests that there is no statistically significant difference between the 'sum' values for the different gender counts within the groups. This implies that there is no evidence of a correlation between 'sum' values and gender quantity based on the groups provided.

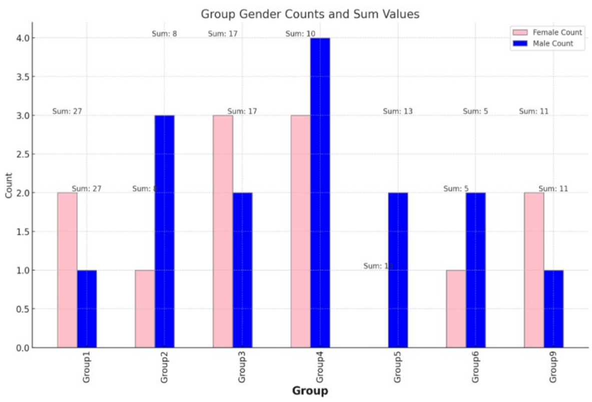


Fig. 5. Group gender counts and sum values.

V. Conclusion

The study examines the effects of storytelling, analogy, and metaphor as pedagogical tools on the development of empathy in male and female students in a design thinking course. It examines their effectiveness in improving empathy skills by analyzing students' tasks using sentiment analysis, employing a Naive Bayes Classifier and natural language processing techniques. A diverse and emotionally varied dataset of movie reviews is used to train the sentiment analysis model, reflecting the range of emotions in students' assignments and aiming to accurately classify and evaluate the nuanced expressions of empathy in instructional texts.

We calculated sentiment distributions in the discourse of project report across different genders, and we didn't find sentiments which are more prevalent in one gender's writing than the other's. Though the sentiment analysis shows that the general sentiment toward research project is becoming more positive over time with the empathy training, our analysis revealed no significant gender differences in empathy training, challenging existing stereotypes that suggest otherwise. The overall sentiment score reached the maximum in Assignment 3, and the wide range of positive sentiment scores among males suggests an openness or responsiveness to the research project, which we found to be interesting. This raises questions about further teaching practice on observing any positive skew that indicates that male students could be more engaging with the projects in a constructive manner.

Our assumption was later confirmed by the active participation of our male students in two projects led by female students and dealing with female issues. One project, initiated by a female student after the Design Thinking course, is about coding through interactive storytelling, focusing on female personalities in STEM at the host university, and 4 male students are involved in the design and promotion of the project. Another project, which is about a vending machine to dispense sanitary towels for free, is also carried out by a female leader and 2 male students, which is seen as a sign of the empathetic engagement of male students.

To further substantiate this, qualitative examples from the assignments could be analyzed, looking for instances where male students show encouragement, recognition of effort, or alignment with the values and objectives outlined in female-led initiatives. These examples would illustrate the quantitative findings and provide a narrative around the male students' empathetic approach towards their female peers' projects.

Limitation of this study includes the small sample size and the choice of training data of film review. First, the 236 project reports with only 3 assignments analyzed might be too small to draw generalized conclusions for research students. In addition, they may not capture the diverse range of sentiments typically found in a larger corpus of project reports. This can also lead to a bias in the sentiment analysis results, as the model's training may not be sufficiently comprehensive to understand the nuances in a wider range of texts. Second, training the model on film reviews might not be entirely appropriate for analyzing project reports. The language, tone, and expression style in film reviews can significantly differ from that of project reports. Film reviews are often more narrative, subjective, and emotive, whereas project reports tend to be more formal, objective, and technical. This disparity can lead to inaccuracies in sentiment analysis when the model is applied to project reports.

Despite the limitations, this research provides valuable insights into specific contexts or types of project reports. It can serve as a preliminary study, paving the way for more extensive research in the future. Future studies may consider using a conceptual lens of semiotic cultural psychology that focuses on the complexity and ambivalence that characterizes the transition experience of a female graduate student before choosing the research topic, and intends to demonstrate how personal background, academic culture, pedagogical practice at the university influences various possible developmental trajectories for women. By linking sociological and psychological theory and research, this paper emphasizes the importance of the sociocultural context in which female students' identity development occurs. The importance of developing cultural resources that support women in managing their learning, work, and social tasks, including related transitions, could be further emphasized.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

Qingqing Xing conducted the research and wrote the paper with Lan Luo; Lan Luo analyzed the data and wrote the data-related part; both authors had approved the final version.

FUNDING

This research has been funded as a Key Project of Guangzhou Education Science Planning, with grant number 202315952. In addition, it has been funded by Project of Research on HKUST(GZ) Practices 2023.

ACKNOWLEDGMENT

The authors wish to thank all the students and lecturers involved in the design thinking course in 2022–2023 Fall.

REFERENCES

[1] A. Gibert, W. C. Tozer, and M. Westoby, "Teamwork, soft skills, and research training," *Scientific Life*, vol. 32, no. 2, pp. 81–84, Feb. 2017.

[2] N. Volkova, N. Zinukova, K. Vlasenko, and T. Korobeinikova, "Soft skills, their development and mastering among post graduate students," in *Proc. SHS Web Conf.*, 2020, vol. 75, art. no. 04002. doi: https://doi.org/10.1051/shsconf/20207504002

[3] N. Ten, M. Didkovskaia, M. Kurushkin, and V. Limonova, "Merging science and soft skills for master students," in *Proc. ICERI2021*, 2021, pp. 2278–2285.

[4] A. Flectcher, *Storythinking: The New Science of Narrative Intelligence*, 1st ed., Columbia University Press, 2023.

[5] S. Baron-Cohen, "The essential difference: The male and female brain," *Phi Kappa Phi Forum*, vol. 85, no. 1, p. 23, 2005.

[6] J. M. DeJesus, V. Umscheid, and S. A. Gelman, "When gender matters in scientific communication: The role of generic language," *Sex Roles*, vol. 85, no. 7, Nov. 2021. doi: 10.1007/s11199-021-01240-7

[7] J. Kolev, Y. Fuentes-Medel, and F. Murray, "Gender differences in scientific communication and their impact on grant funding decisions," in *AEA Papers and Proceedings*, Nashville, TN, USA: American Economic Association, 2020, vol. 110, pp. 245–249.

[8] M. V. Mestre, *et al.*, "Bidirectional relations among empathy‐related traits, prosocial moral reasoning, and prosocial behaviors," *Social Development (Oxford, England)*, vol. 28, no. 3, pp. 514–528, 2019. https://doi.org/10.1111/sode.12366

[9] B. Derntl, *et al.*, "Multidimensional assessment of empathic abilities: neural correlates and gender differences," *Psychoneuroendocrinology*, vol. 35, no. 1, pp. 67–82, 2010. https://doi.org/10.1016/j.psyneuen.2009.10.006

[10] M. Schulte-Rüther, *et al.*, "Gender differences in brain networks supporting empathy," *NeuroImage (Orlando, Fla.)*, vol. 42, no. 1, pp. 393–403, 2008. https://doi.org/10.1016/j.neuroimage.2008.04.180

[11] A. H. Eagly and M. Crowley, "Gender and helping behavior: A meta-analytic review of the social psychological literature," *Psychological Bulletin*, vol. 100, no. 3, p. 283, 1986.

[12] G. A. Paganini, K. M. Summers, L. T. Brinke, and E. P. Lloyd, "Women exaggerate, men downplay: Gendered endorsement of emotional dramatization stereotypes contributes to gender bias in pain expectations," *J. Exp. Soc. Psychol.*, vol. 109, 104520, 2023. https://doi.org/10.1016/j.jesp.2023.104520

[13] H. Yang, D. Anderson, and S.-J. Kang, "An instrument for measuring scientific empathy in students' disciplinary engagement: the scientific empathy index," *Front. Educ.*, vol. 8, 1254436, 2023. doi: 10.3389/feduc.2023.1254436

[14] M. Saggar, E.-M. Quintin, N. T. Bott, *et al.*, "Changes in brain activation associated with spontaneous improvisation and figural creativity after design-thinking-based training: A longitudinal fMRI study," *Cerebral Cortex*, vol. 27, no. 7, pp. 3542–3552, Jul. 2017. doi: 10.1093/cercor/bhw171

[15] M. Eradatifam, S. Heydarabadi, and A. Shahbazi, "The impact of design thinking on innovation," *IEEE/OSA Journal of Display Technology*, vol. 1, pp. 49–60, 2020.

[16] C. Docherty, "Perspectives on design thinking for social innovation," *The Design Journal*, vol. 20, pp. 719–724, Nov. 2017. doi: 10.1080/14606925.2017.1372005

[17] P. Stoev and M. Stoeva, "Building empathy in students by developing cyber-physical projects through design thinking," *Journal of Physics Conference Series*, vol. 2701, no. 1, 2024. doi: 10.1088/1742-6596/2701/1/012041

[18] S. Burgess, M. Rogers, and D. Jefferies, "Empathy and narrative – A discussion of contemporary stories from education and health," *Int. J. Elem. Educ.*, vol. 14, pp. 631–646, 2022. doi: 10.26822/iejee

[19] Q. Xing, C. Zheng, N. Zhu, and D. Yip, "AI-Generated content for academic visualization and communication in maker education," in *Proc. 2023 3rd International Conference on Educational Technology (ICET)*, Xi'an, China, 2023, pp. 52–56. doi: 10.1109/ICET59358.2023.10424137

[20] A. Hashim, S. Aris, and Y. F. Chan, "Promoting empathy using design thinking in project-based learning and as a classroom culture," *Asian Journal of University Education*, vol. 15, p. 14, 2019. doi: 10.24191/ajue.v15i3.7817

[21] D. Rahmawati and M. L. Khodra, "Word2vec semantic representation in multilabel classification for Indonesian news article," in *Proc. 2016 International Conference on Advanced Informatics: Concepts, Theory and Application (ICAICTA)*, Aug. 2016, pp. 1–6.

[22] A. Alobaid and O. Corcho, "Linear approximation of the quantile-quantile plot for semantic labelling of numeric columns in tabular data," *Expert Systems with Applications*, vol. 238, art. no. 122152, 2024.