**Paper #20 – 2018**

**Visualisation and Statistical Analysis of Survey Data to Ascertain The Effectiveness of Library Training Efforts on Students’ Research Skills**

Mahima Damani; Manisha Goyal; Mohamed Lukman Bin Mohamed Syed

Singapore Management University

# ABSTRACT

The SMU Libraries Analytics and Research Department strives to develop a data-informed approach for achieving strategic objectives related to library operations and user needs. The library currently aims to develop it’s trainings (content, methodology & availability) catering to solving specific student problems associated with those skills. For this purpose, they have conducted an initial survey for the freshman batch of 2017 to evaluate the difference in their confidence level in various research skills before and after their first semester at SMU. This paper aims to identify factors that relate to and predict student confidence in performing library research tasks and help improve library training initiatives in terms of relevance and effectiveness. We will conduct exploratory analysis to discover trends specifc to schools and problems associated with research skills. Moreover, we will conduct statistical analysis to understand the change in the confidence of students that can be attributed to library trainings.

# 1.0 introduction

## 1.1 BACKGROUND

SMU Libraries offers its services through two physical libraries, as well as virtually through various online services. SMU Libraries is the centre for academic and professional knowledge resources that support the research and learning needs of SMU. The library facilitates knowledge creation and discovery through its electronic search platform and access to a wide array of research. We aim to help the library succeed in this mission of making all resource management exceedingly data-driven, automated and micro-tracked.

The aforementioned survey will be the main focus of our study. It measures metrics like confidence level in using library resources, trainings attended, and modules taken during the semester. We wish to ascertain the effectiveness of the trainings by determining if they have any relationship with the confidence levels of students.

Following are the categories of skills that the library is conducting their research in:

1. **Scoping the topic** and narrowing it down to create a research question
2. **Searching using keywords** and synonyms
3. **Searching** commercial **online databases** such as SPH Newslink, ProQuest etc.
4. **Searching the Internet using Google** or other search engines
5. **Evaluating information** using a set of criteria e.g. currency, purpose
6. **Citing references** using a citation style (e.g. American Psychological Association (APA); Harvard Referencing Style)
7. **Creating a reference list** or bibliography

## 1.2 MOTIVATION

The library currently aims to optimise its student training and workshops based on findings from the survey. Considering the importance of using library resources efficiently, it is central to understand the different trends and patterns students demonstrate in their usage of library resources and how it relates to attributes like modules undertaken, trainings attended and so forth. This will help us to provide the library with specific problems and targeted solutions based on schools and modules to eventually make the research process of an SMU student more effective and efficient.

## 1.3 OBJECTIVEs

The objectives of the project are the following:

1. Business objective: To discover the current confidence level of freshmen across different faculties and indentify trends. Moreover, to explain, with clear visuals, how students have responded to different trainings for each skill at the end of the semester
2. Technical objective: To use data analytics tools and statistical methods to study the data and obtain insights to facilitate the business objective

To achieve our two primary objectives, we will need:

* To understand the data domains
* To understand the library training process
* To identify if there exist any students who experience high or low confidence and its contributing factors
* To create a visual representation of the effectiveness of the trainings conducted during the semester, and provide recommendations.

# 2.0 LITERATURE REVIEW

## 2.1 DIVERGENT STACKED BAR GRAPHS

Visualisations give quick insights from data, allowing targeted analysis. For Likert data, using divergent stacked bar graphs is especially appropriate. It facilitates a visual comparison of respondents’ answers to the survey (Heiberger and Robbins, 2011). For our purpose, divergent stacked bar graphs allow us to see the general levels of confidence for the different research skills. By comparing the graphs using pre and post confidence data, we can understand whether there was an improvement in the responses.

## 2.2 CHI-SQUARED TESTS FOR INDEPENDENCE

To statistically determine whether the change, specifically improvements, in confidence were significant, the chi-squared test for independence was used. The chi-squared tests allow for us to conclude whether the distribution of the categorical variable (confidence) is related to the variable of our groups (training) (Kim, 2017).

There is ongoing debate as to the appropriateness of a chi-squared test on paired data. Some discuss using paired t-test or Wilcoxon test when working with paired data (Derrick and White, 2017). However, these approaches assume an equal spacing between the categories on the Likert scale, which is spurious for the confidence categories in our data. Furthermore, in conducting the chi-squared analysis, we obtain contingency tables that help us see detail in any improvements in confidence.

## 2.3 WORD FREQUENCY ANALYSIS

Text comments from the end of the survey can be analysed to find out respondents’ concerns not captured by the survey. This is done with the underlying assumption words that appear more frequently indicate an issue that students care more about (Stemler, 2001).

# 3.0 METHODOLOGY

Our methodology can be summarised as:

1. We began our analysis by transforming the data that was provided
2. We carried out exploratory data analysis using Tableau 10.0. This is where we did the visualisation analysis using the divergent stacked bar graphs
3. From the initial insights, we sought to statistically prove the relationships that were observed. For this, we used JMP Pro 13 to carry out the chi-squared tests
4. We conducted the text analysis to find out if students had any major issues with the trainings conducted
5. Lastly, we used all the analysis done to give recommendations to the library

# 4.0 DATA

The sponsor conducted two surveys with the freshman batch of 2017. Pre-survey was conducted before the start of the semester (Aug 2017) and post-survey at the end of the semester (Nov 2017). The pre and post survey datasets contain responses of students before and after the first semester on their confidence level in research skills. After having cleaned and compiled the two sheets, the record attributes are as follows:



Table 1. Data Dictionary

# 5.0 DATA CLEANING AND PREPARATION

## 5.1 MISSING DATA

The number of records in both datasets was not the same. While the pre-survey dataset had 1455 records, the post-survey dataset had 414 records. To address this problem, we matched the two datasets on the key attribute, the student email. On combining only 292 records matched. Thus the remaining records with no complementary entry in either dataset were considered missing and were removed.

## 5.2 DATA TRANSFORMATION

Both pre and post survey datasets contain questions that ask the students whether they were trained for a particular research skill. The responses in the pre-survey dataset are categorical, whereas those in the post-survey dataset are numerical. It was necessary for both these columns to be standardised for comparison, thus we changed the pre-survey reponses to numerical format.

## 5.3 DATA PREPARATION FOR EXPLORATORY DATA ANALYSIS

Data transformation was needed for the analysis. We used a 1-5 scale to represent the answers as it facilitated the visualisation process. Response columns were pivoted to create a database format. Then, we created calculated fields:

1. Sum(Answers): For each research skill, for each response (Not At All Confident – Very Confident)
2. Negative Scores: If the score was less than 3 (Not At All Confident and Slightly Confident), it was considered a negative score, and if it was more than 3 (Confident and Very Confident) it was considered a positive score. For scores equal to 3 (Somewhat Confident), we split total responses in two – half would be considered negative while the other half would be considered positive. We assumed a fair split because we cannot be sure about the distribution of confidence within this response
3. Total Negative Scores: Total of the negative scores summed up across each skill category
4. Total Scores: Total of the responses for each skill category
5. Gantt Start: Total Negative Scores divided by Total Scores. This would allow our visualisation to be offset from the y-axis depending on the responses
6. Percent of Total Sizing: Sum of Number of Records divided by Total Scores. This would define the length of each response for each question
7. Gantt Percent: Indicates the Gantt lines to separate the different responses on the bar by colour

## 5.4 DATA PREPARATION FOR CHI-SQUARED TESTS

To conduct the chi-squared tests, we split the data into different skills. Within each skill we filltered out all the students who had been trained before the start of the semester and the ones who had not been trained even at the end of the semester. This would help us understand the true difference in confidence that can be attributed to training conducted during the semester only.

## 5.5 DATA PREPARATION FOR TEXT ANALYSIS

The post survey contains a column with the comments of the resondents. First, we removed all the comments that were either blank, or ‘nil’. After that we used VBA code to transform the comments column into a two column sheet, one column for the unique words, and one with the frequency of each word. This data was used to conduct text analysis on the comments to derive insights.

# 6.0 EXPLORATORY DATA ANALYSIS

## 6.1 percentage of students who received training in library research skills

|  |  |  |
| --- | --- | --- |
| **Skilll** | **Highest Majority** | **Lowest Majority** |
| Citing References | SOSS students (78.05%) | SOL students (54.84%) |
| Creating Reference Lists | SOSS students (68.29%) | 43.33% of SIS students |
| Searching the Internet using Google | SIS freshmen (76.67%) | 58.06% of SOL students |
| Searching using Keywords | SIS freshmen (83.33%) | 57.89% of SOE students |
| Evaluating Information | SIS freshmen (70%) | 38.71% of SOL students |
| Scoping your Topic | SOA freshmen (70%) | 48.39% of SOL students |
| Searching using Library Online Databases | SIS freshmen (83.33%) | 50% of SOE students |

Table 2. Percentage of students trained

The resultant vertical bar graphs are attached in Appendix A.

## 6.2 EFFECT OF TRAINING ON CONFIDENCE LEVELS

Confidence level of SMU students who did not receive any training in conducting overall library research tasks:

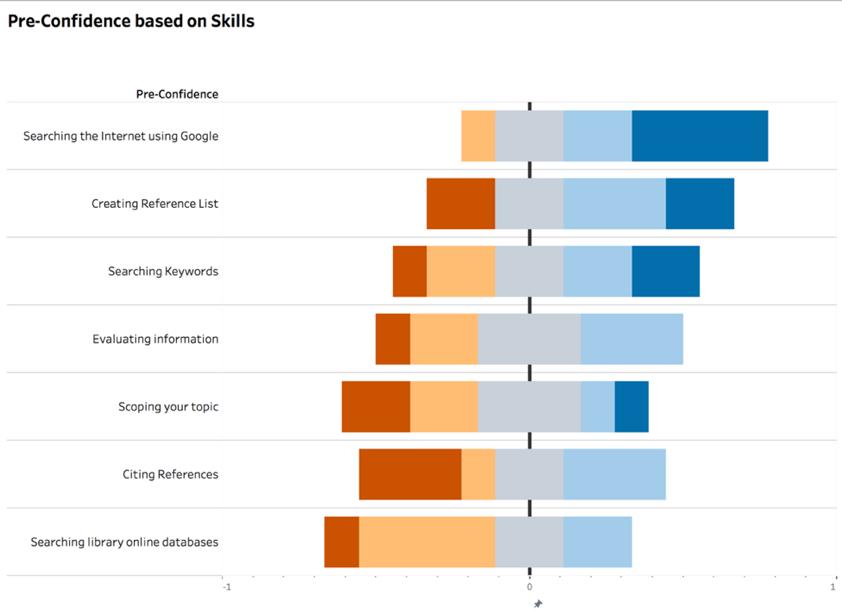


Figure 1. Pre-Confidence of library research tasks without training

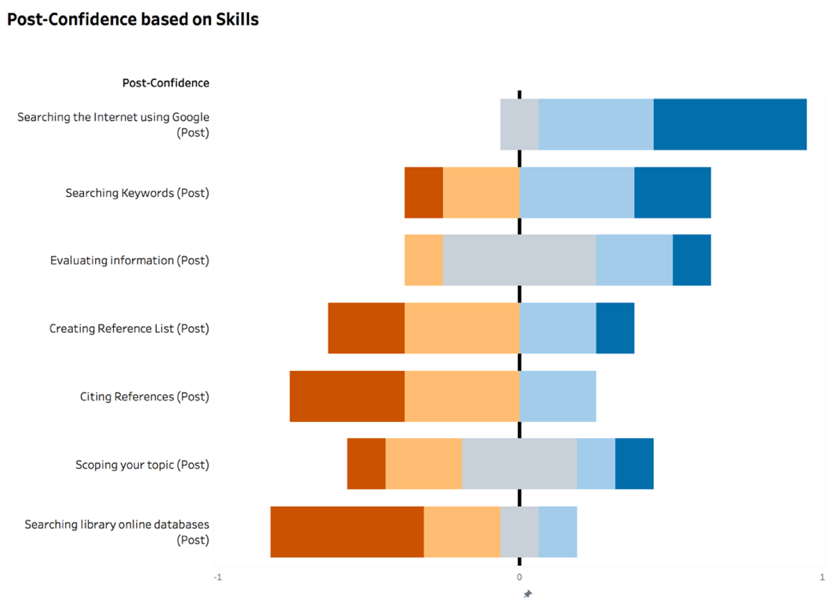


Figure 2. Post-Confidence of library research tasks without training

There is a positive shift in the confidence of students in the skills of ‘Searching the Internet Using Google’, ‘Evaluating Information’, ‘Scoping your Topic’ and ‘Searching Keywords’. The confidence for ‘Citing References’, ‘Creating Reference Lists’ and ‘Searching Library Online Databases’ shows a negative shift.

Confidence level of SMU students who received training in conducting overall library research tasks.

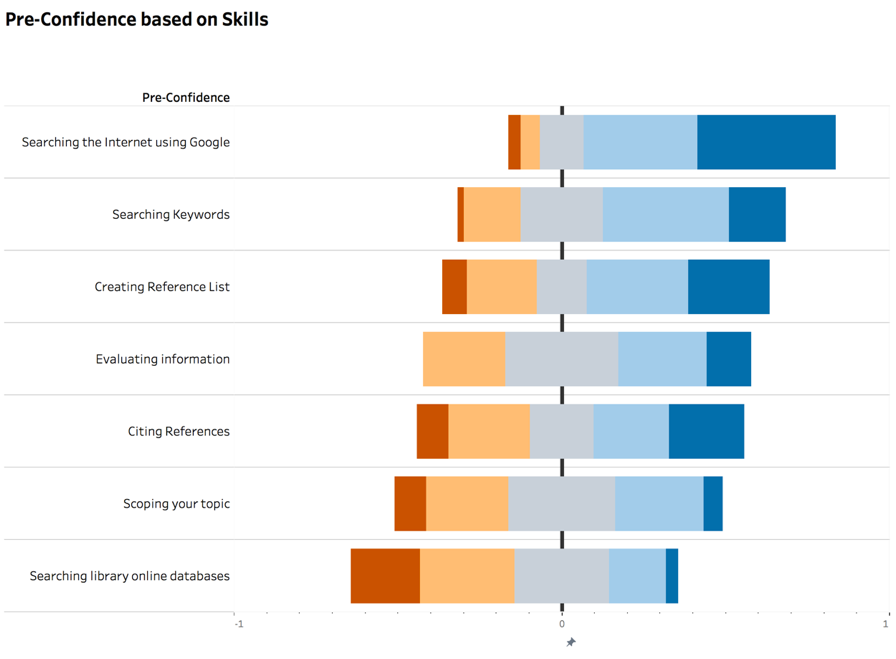


Figure 3. Pre-Confidence of library research tasks with training

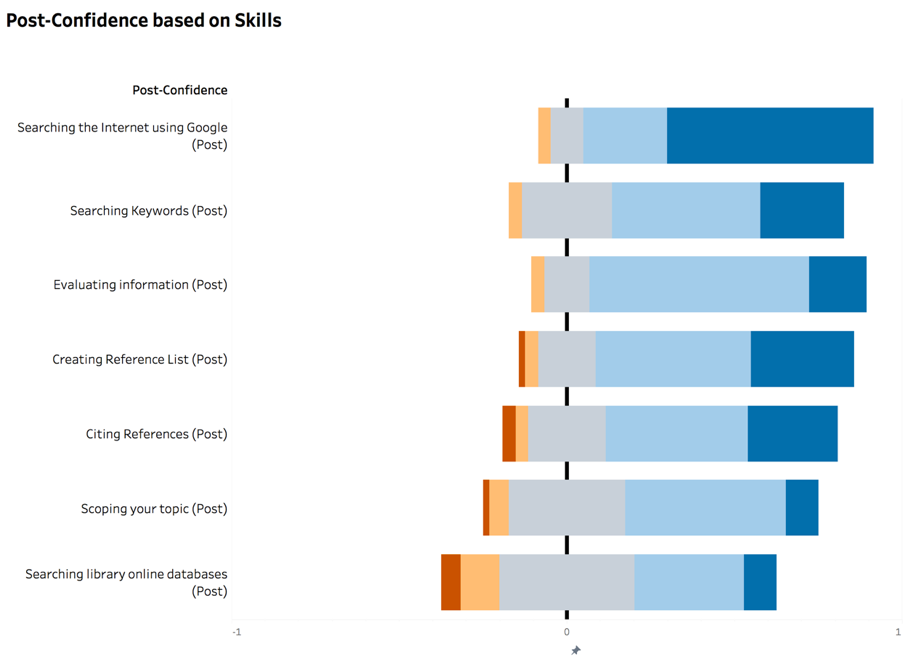


Figure 4. Post-confidence of library research tasks with training

There is a positive shift in the confidence of students in all the skills.

The same process was repeated for each of the skills (Appendix B). Following are the insights discovered:

|  |  |  |
| --- | --- | --- |
| **Skill** | **Change in confidence without training** | **Change in confidence with training** |
| Citing References | There is a positive shift in the confidence of SOSS, LKCSB and SOE students. The confidence of SOL, SIS, and SOA students shows a negative shift. | There is a positive shift in the confidence of students of all the schools. |
| Creating Reference Lists | There is a positive shift in the confidence of SOL, SIS, LKCSB and SOA students. The confidence of SOSS, SOE, and SOA students shows a negative shift. | There is a positive shift in the confidence of SOA, SOSS, LKCSB, SOE and SIS students. The confidence of SOL students shows a negative shift. |
| Searching the Internet using Google | There is a positive shift in the confidence of students of all the schools. | There is a positive shift in the confidence of students of all the schools. |
| Searching using Keywords | There is a positive shift in the confidence of SOL, SIS, SOSS, SOA, and SOE students. The confidence of LKCSB students shows a negative shift. | There is a positive shift in the confidence of students of all the schools. |
| Evaluating Information | There is a positive shift in the confidence of SIS, SOA, SOSS, LKCSB and SOE students. The confidence of SOL students shows a negative shift. | There is a positive shift in the confidence of students of all the schools. |
| Scoping your Topic | There is a positive shift in the confidence of students of all the schools. | There is a positive shift in the confidence of students of all the schools. |
| Searching using Library Online Databases | There is a positive shift in the confidence of SIS, SOSS, SOA, and SOE students.The confidence for LKCSB and SOL students shows a negative shift. | There is a positive shift in the confidence of students of all the schools. |

Table 3. Comparison of confidence levels with and without training

# 7. CHI SQUARE TEST

To establish whether the observed improvements were statistically significant, we conducted chi-squared goodness of fit tests. For each research skill category, pre levels of confidence were measured against postlevels. The sample chosen to conduct the analysis was of students who had gone through training during the semester to allow us to conclude if the trainings had any real impact on confidence levels.

H0: Pre-confidence and post-confidence are independent.

H1: H0 is false.

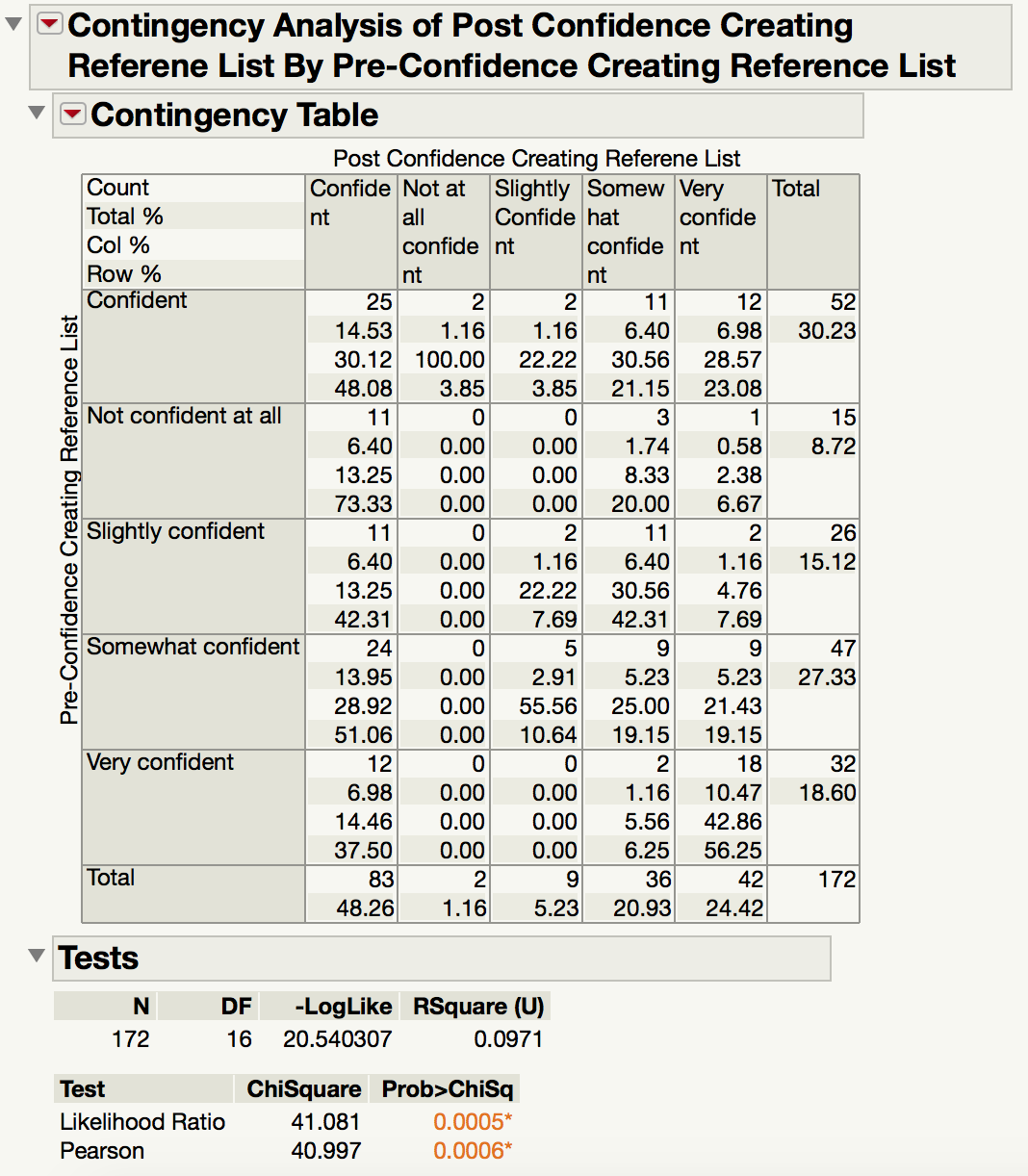


Figure 5. Change in confindence of students in creating reference lists

Since the p-value for our test is <0.05, we have statistically significant evidence that H0 is false, i.e. pre-confidence and post-confidence for Creating Reference Lists are not independent.

We furthur analysed the contingency table to see the effectiveness of the training for Creating Reference Lists. Notably, all respondents who reported “Not Confident At All” in the pre-survey experienced an improvement in their confidence, with 73.33% becoming “Confident” at the time of the post-survey.

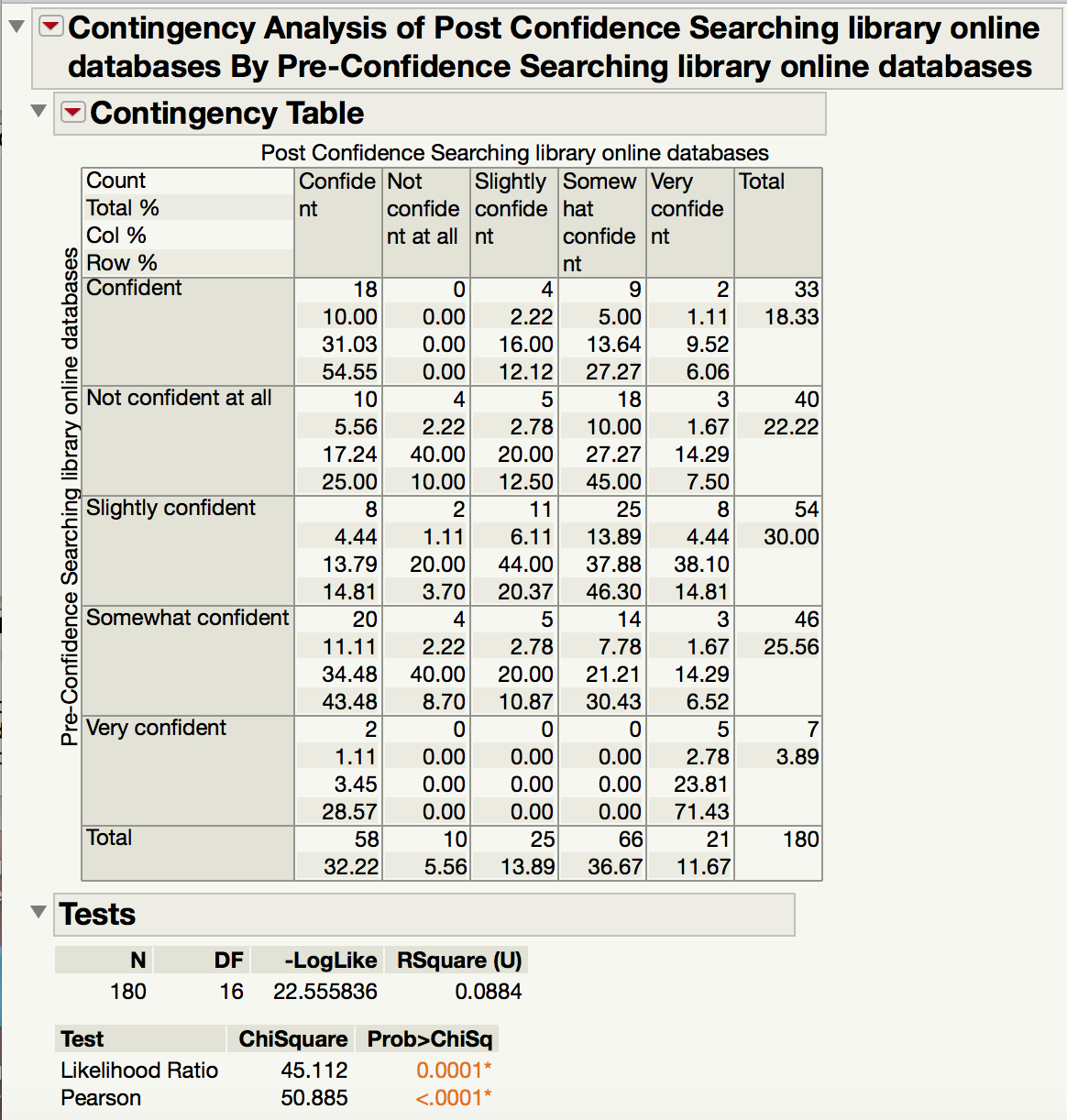


Figure 6. Change in confindence of students in Searching Library Online Databases

With a statistically significant test, we conclude that pre-confidence and post-confidence for Searching Library Online Databases are not independent. Results seem generally positive. However, 10% of students who reported “Not Confident At All” did not experience any improvement at the end of the semester. Further, 12.5% became “Slightly Confident” and almost half, 45%, reported “Somewhat Confident”. This indicates that the training for Searching Library Online Databases might not be very effective.

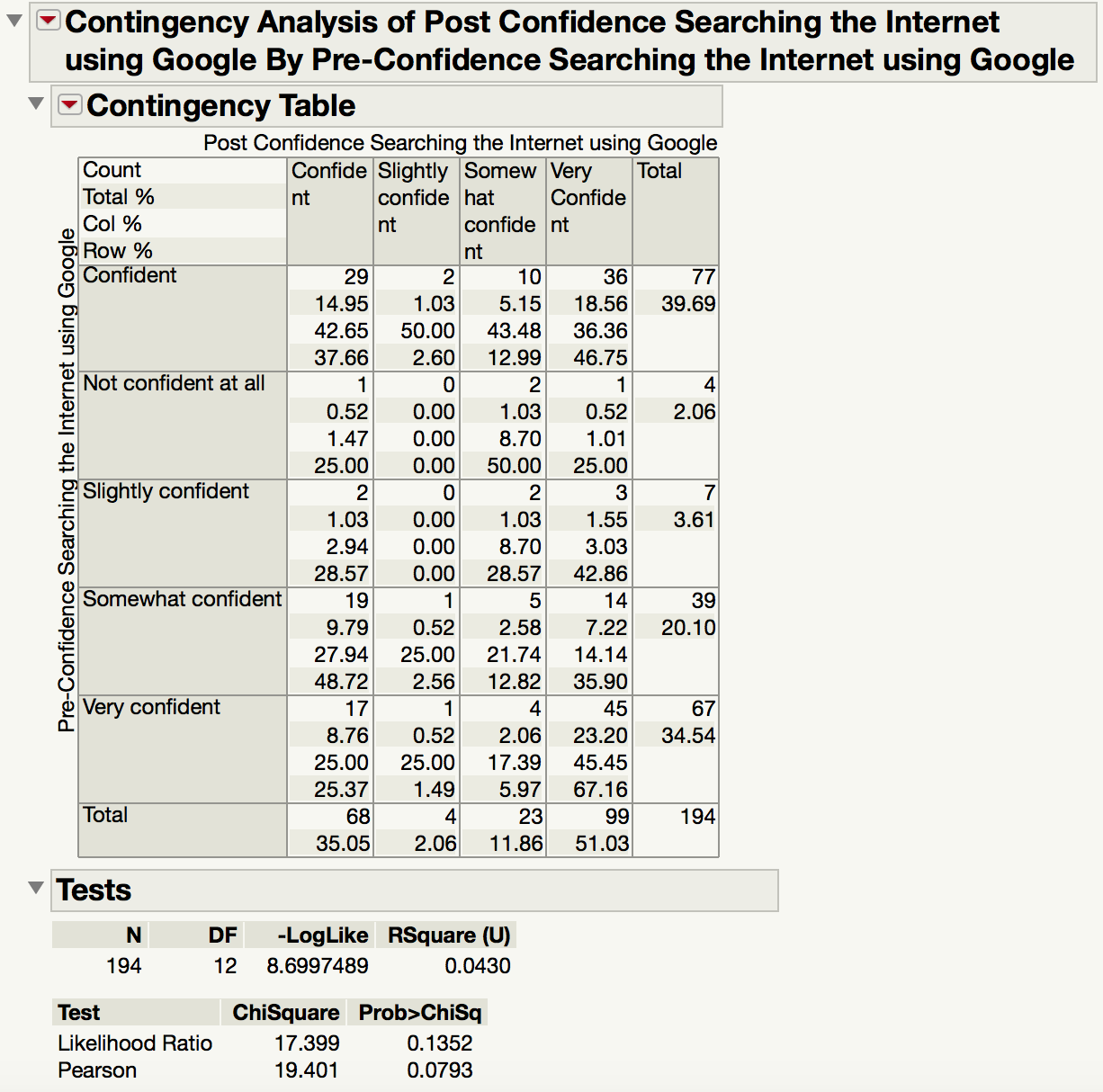


Figure 7. Change in confindence of students in Searching the Internet using Google

The Pearson statistic for the test on Searching the Internet Using Google was not significant at the 5% level of significance. Thus, we cannot reject H0 and conclude that there is no evidence of a relationship between pre and post levels of confidence for this skill. This implies that training might not have been effective.

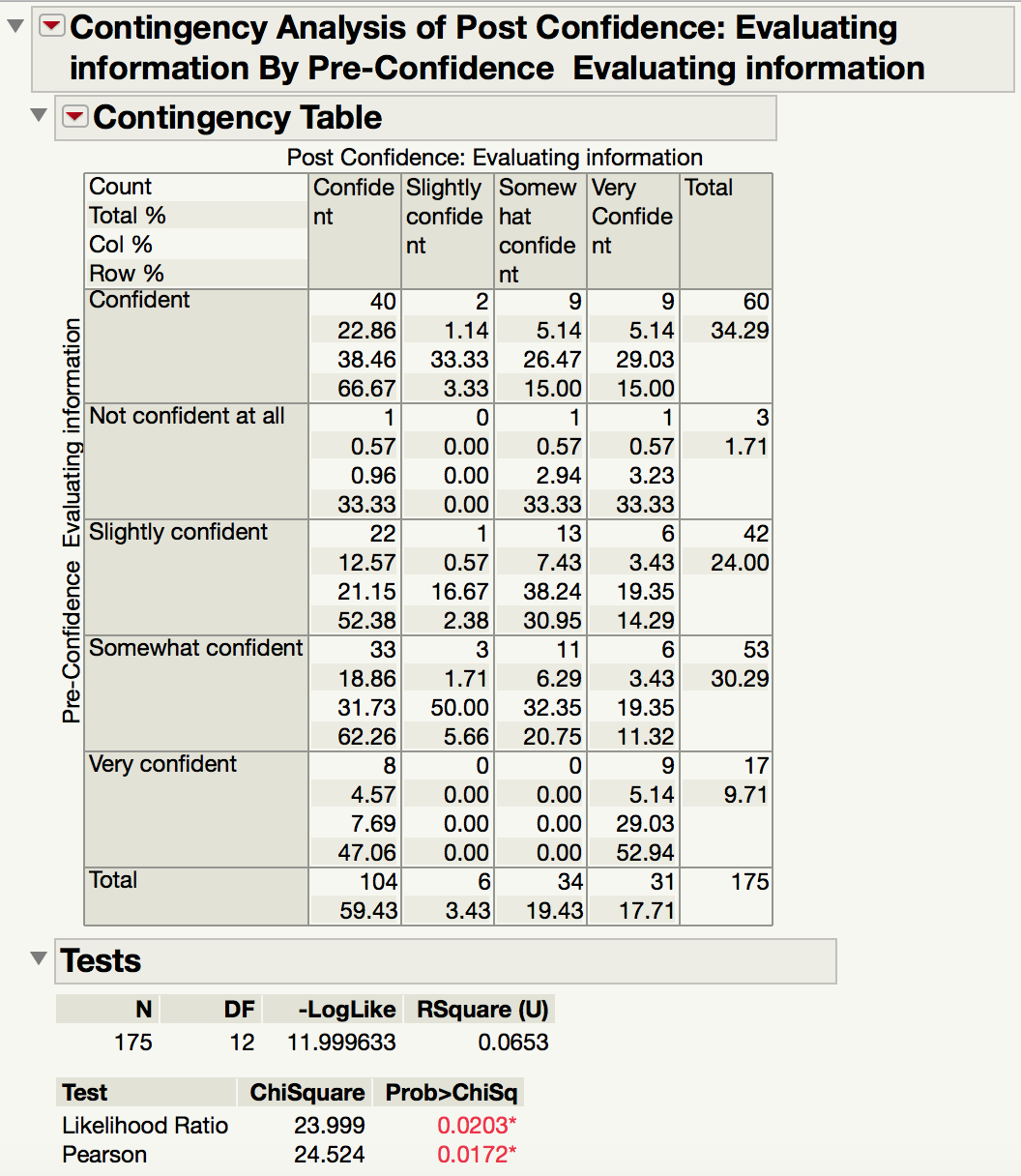


Figure 8. Change in confindence of students in Evaluating Information Lists

With a statictically significant test, we conclude that pre-confidence and post-confidence for Evaluating Information are not independent. Almost half of respondents who reported “Very Confident” experienced a loss in confidence at the end of semester.  On the other hand, 62.26% of students who were “Somewhat Confident” became “Confident”. Training can be concluded to have positive effects, but there is room for improvement.

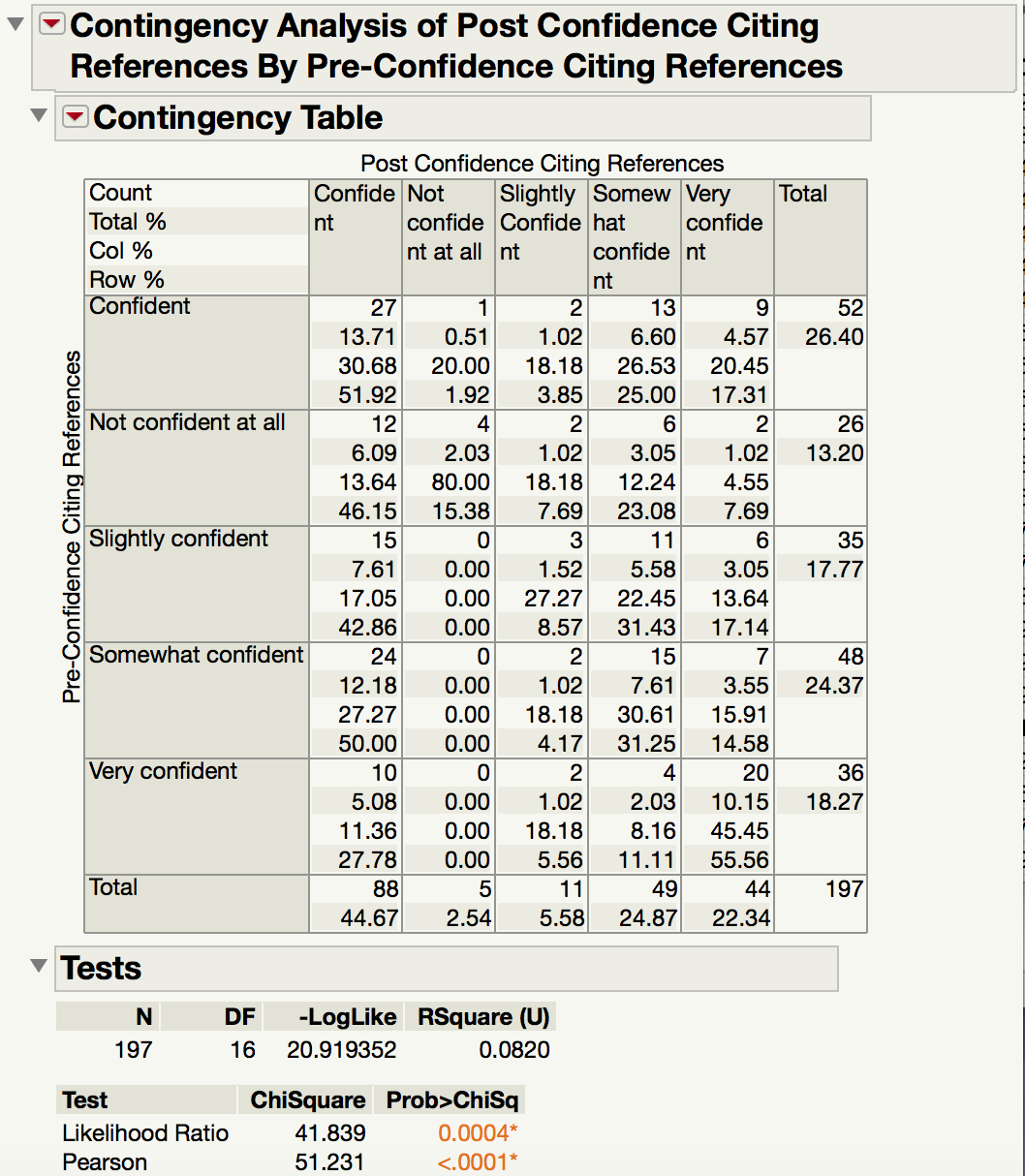


Figure 9. Change in confindence of students in Citing References

With a statistically significant test, we conclude that pre-confidence and post-confidence for Citing References are not independent. Notably, 46.15% of the respondents who reported “Not Confident At All” and 42.86% of the students who reported “Somewhat Confident” experienced an improvement and moved up to “Confident”. On the flipside, we see that 25% of the students who were “Confident” become “Somewhat Confident”.

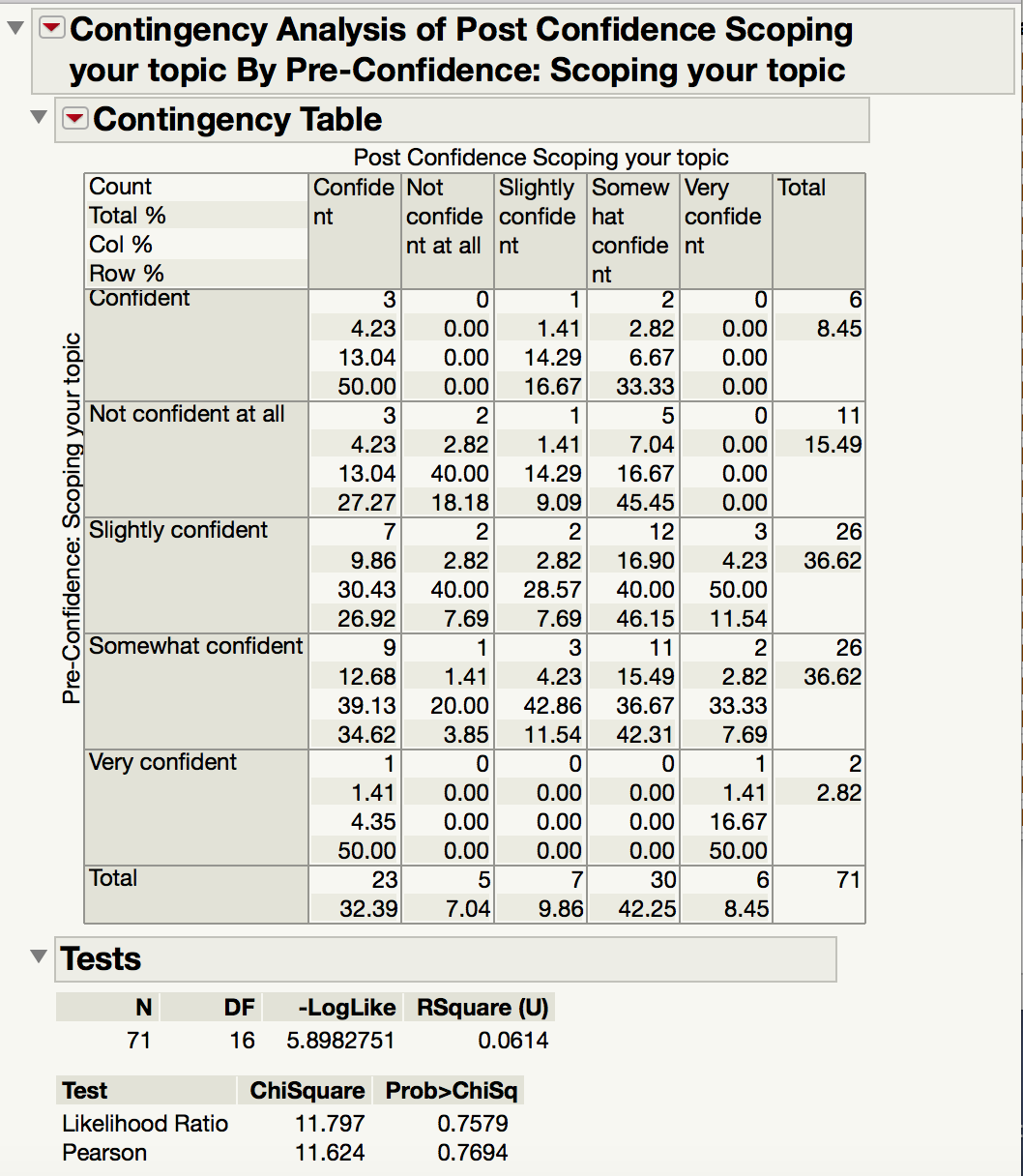


Figure 10. Change in confindence of students in Scoping youy Topic Pre and Post Survey

The Pearson statistic for the test done on Scoping your Topic was not significant at the 5% level of significance. Thus, we cannot reject H0 and conclude that there is no evidence of a relationship between pre and post levels of confidence for this skill. This implies that training might not have been effective.

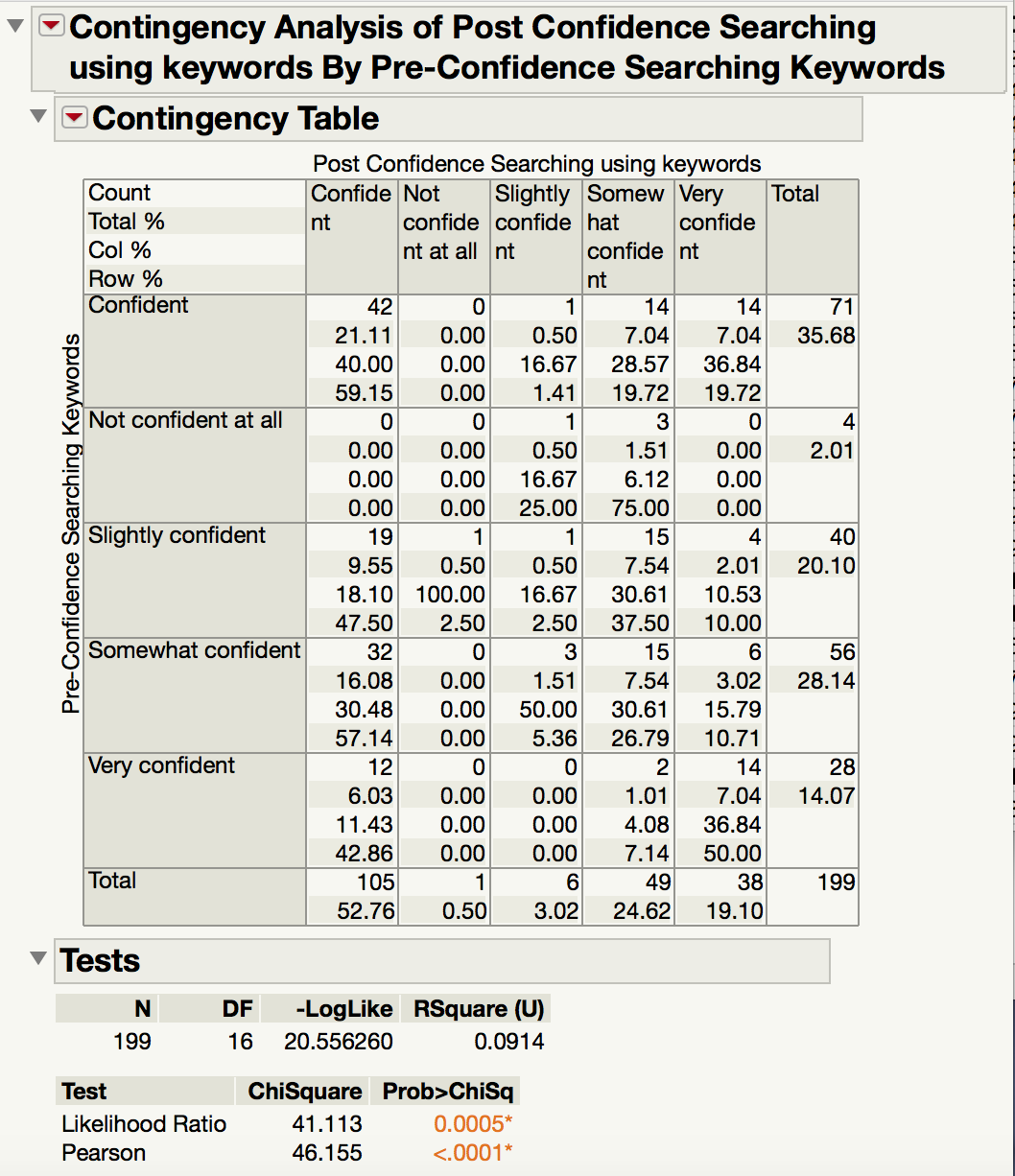


Figure 11. Change in confindence of students in Searching using Keywords Pre and Post Survey

With a statistically significant test, we conclude that pre-confidence and post-confidence for Searching using Keywords are not independent. Notably, only 59.15% of the students who responded to be “Confident” remained “Confident”. The remaining students went both ways into “Somewhat Confident” and “Very Confident” which shows a clear divide in the effectiveness of the training. While 47.5% of the students who responded to be “Slightly Confident” and 57.14% of the student who responded to be “Somewhat Confident” experienced improvement and moved up to “Confident”, 42.86% of the “Very Confident” students moved down to only “Confident”.

# 8. TEXT ANALYSIS

We had a total of 53 comments to analyse. We used tableau with our prepared data to create a word frequency chart. Only words with a frequency greater than 1 were included for analysis. After the first round of analysis, we found that there were several stop words, and other irrelevant nouns. We re-analysed the words after removing all such words. Finally, the following chart was created. From this, we found that the most common words were positive adjectvies - ‘conducive’, ‘helpful’, and ‘useful’.

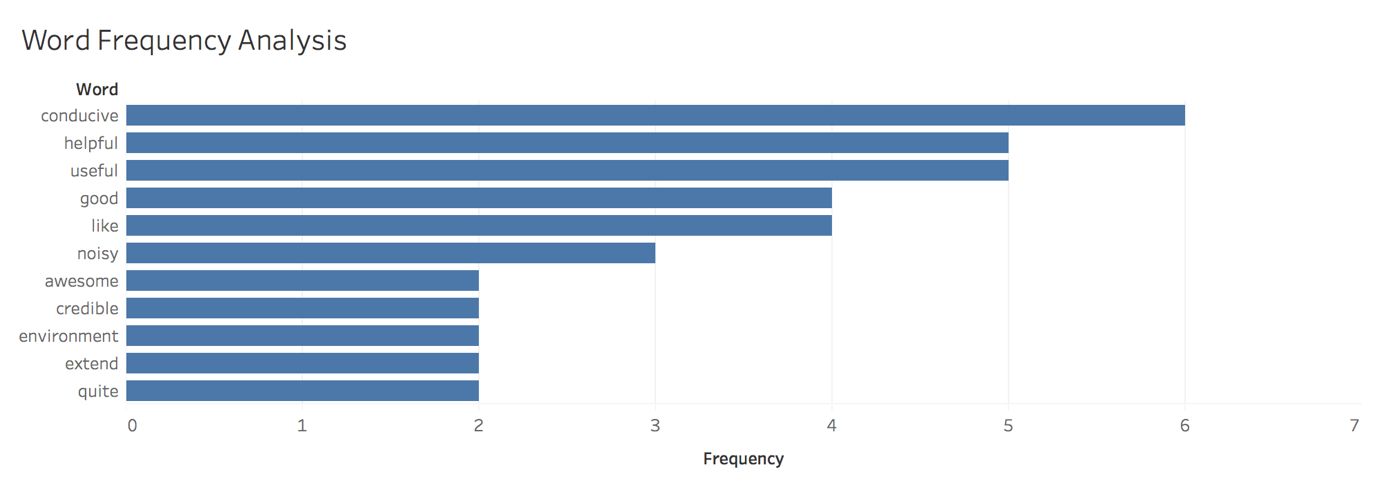


Figure 12. Word frequency analysis of post-survey comments

As the number of unique words that are representative of how students feel about the library were very few, furthur analysis could not be performed.

# 9. RECOMMENDATIONS

Based on the exploratory data analys, we see that only for skills like citing references and searching library online databases, there is a negative shift in the confidence of students who did not receive training during the semester. Whereas, there is a an overall positive shift in the confidence of students in applying all the skills after receiving training. This shows that the trainings conducted by the library are more relevant for skills that require the use of library resources like the online databases. Moreover, they were also more helpful for skills that students might not have been taught in high school, as there are different standards at university, like citing of references.

We can also see that there is an overall positive shift in the confidence of students of most schools for all skills after receiving training. However, students from the school of law are an exception for some of the skills, such as creating reference lists, as their confidence showed a negative shift after they received training. This is perhaps because the nature of the curriculum for the School of Law demands more in-depth research skills as compared to other schools, as the profession requires constant consulting to global publications and journals (SMU Law, 2018). The library training being catered to all the schools is unable to match up to the needs of the students of the law school, leading to a drop in their confidence levels despite training. The library should look into having separate training for the School of Law students, specifically catered to their research needs.

Particularly, as noted in the chi-squared test, results for Scoping Your Topic and Searching the Internet Using Google were insignificant. This implies that there is no evidence that the trainings that were provided had any impact on these skills. Thus we recommend that trainings for these skills should be paid specific attention for improvement. Based on our visualisation analysis, Searching the Internet using Google was already a skill that students were generally very strong in at the pre-survey. Thus, the library should focus its efforts on Scoping Your Topic only to make the best use of limited resources.

Particular skills like Searching the Library Online Databases and Searching using Keywords showed significance in the chi-squared test. However, several students showed a negative trend in confidence at the end of the semester. This counterintuitive trend could have been because of the ineffectiveness in the current trainings leading to the students losing confidence. We recommend the library change their training format and content for the above mentioned skills.

Looking into the future, we recommend the library to continue tracking the confidence levels. With more data over time, analysis may be modelled to more precise measurements. One important thing noted is that more specific data could be collected to allow for more granular analysis. For instance, each research skill could be further broken down into component tasks or skills and students’ corresponding confidence can be tracked. With this level of specificity, the library can better understand where to target improvements in their training programmes. Furthermore, surveys can be conducted at the end of each training, where the training session is evaluated based on the specific tasks and skills that were targeted.

# 10. LIMITATIONS AND FUTURE IMPLICATIONS

## 10.1 RESPONSE RATE

We had a relatively small sample of responses for both the surveys. The analysis can be greatly improved, if more students participated in both surveys. With more data points, the results will be more robust and convincing. For example, the library could follow the example of the end of term course evaluation survey that rewards students with e-dollars for their participation.

## 10.2 TRAINING DATA

It was challenging to give recommendations on improving research skills. Respondents only answered about whether or not they attended trainings, but no further information was asked of them about trainings. Moreover, our analysis showed comments that were given were inconsequential to library research skills or training. The chi-squared test analysis gave some insight into the effectiveness of trainings, but more information is needed for studying specific areas of weakness for library trainings.

# 11. CONCLUSION

We sought to provide the library with analysis that would help them provide better training for the research skills. The results of our visualisation and statistical chi-squared analysis are promising, and indicate that the trainings have had a general positive impact on students’ confidence levels. For more specific targetting on where the library can improve their trainings, further data collection should be done by the library, focussing on obtaining more detailed inputs from respondents.

# 12. ACKNOWLEDGEMENT

We want to acknowledge the invaluable input from our faculty supervisors, Professor Kam Tin Seong and Professor Meenakshi Gopalakrishnan and thank them for suppporting us throughout the project. We would also like to thank our sponsors Mr. Aaron Chee Hsien and Mr. Wee Lian William Koh for giving us the trust to work with their data and flexibility to explore solutions.

# 13. CONTACT INFORMATION

Mohamed Lukman Bin Mohamed Syed: [mohamedlms.2014@economics.smu.edu.sg](mailto:mohamedlms.2014@economics.smu.edu.sg)

Mahima Damani: [mahimad.2014@business.smu.edu.sg](mailto:mahimad.2014@business.smu.edu.sg)

Manisha Goyal:  [manishag.2014@sis.smu.edu.sg](mailto:manishag.2014@sis.smu.edu.sg)

# 14. REFERENCES

Derrick, B. and White, P. (2017) *Comparing two samples from an individual Likert question*. International Journal of Mathematics and Statistics, 18 (3). ISSN 0974-7117 <http://eprints.uwe.ac.uk/30814>

Heiberger M., Robbins, N B. (2011). *Plotting Likert and Other Rating Scales.* Proceedings of the 2011 Joint Statistical Meeting

Kim HY. (May 2017).  *Statistical notes for clinical researchers: Chi-squared test and Fisher's exact test*.   Restor Dent Endod. 42(2):152-155.   [https://doi.org/10.5395/rde.2017.42.2.152](https://doi.org/10.5395/rde.2017.42.2.152" \t "_blank)

Stemler, Steve. (2001). *An overview of content analysis*. Practical Assessment, Research & Evaluation, 7(17). <http://PAREonline.net/getvn.asp?v=7&n=17>

Law.smu.edu.sg. (2018). [online] Available at: https://law.smu.edu.sg/research

# APPENDIX

## APPENDIX A: percentage of students ho received training in library research skills

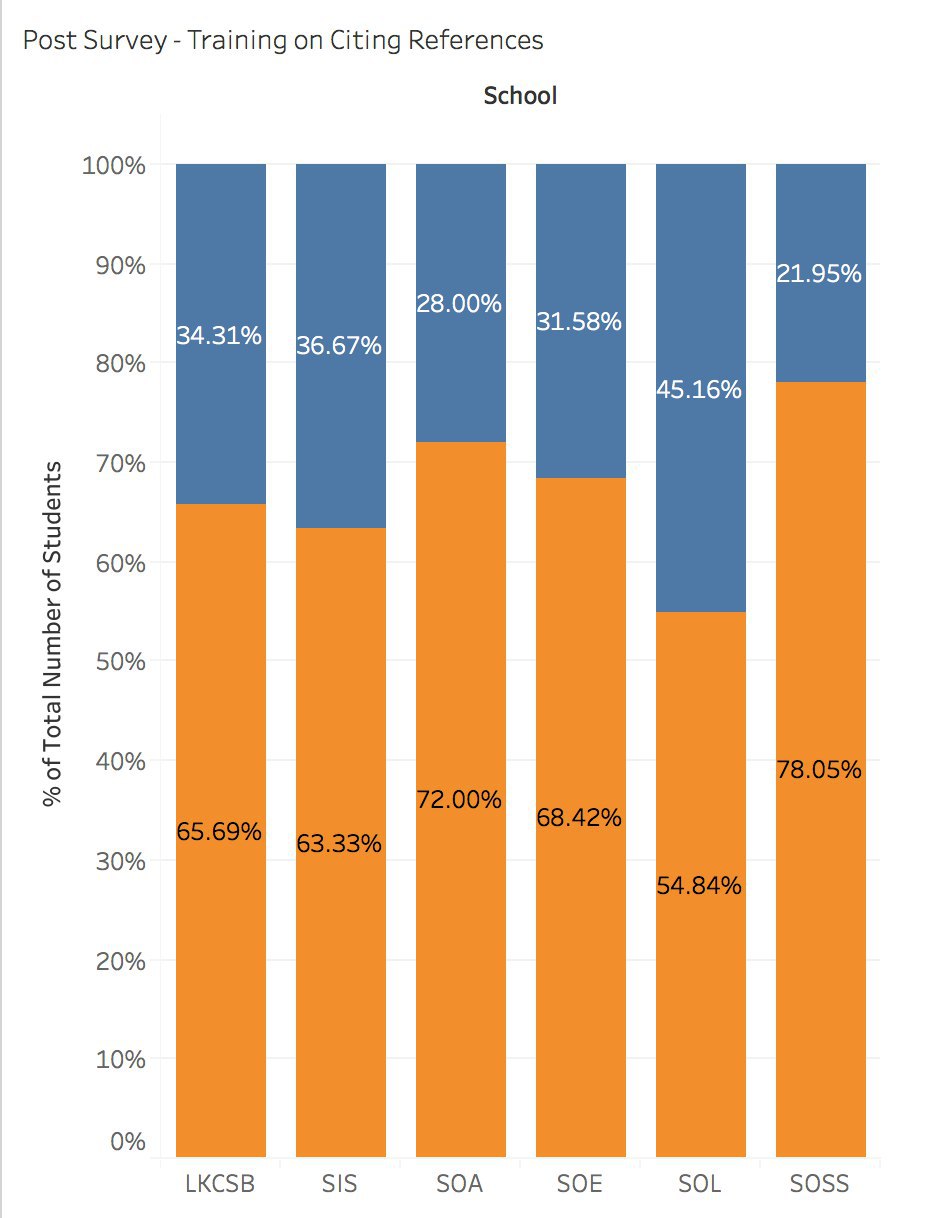


Figure 13. Training for citing references

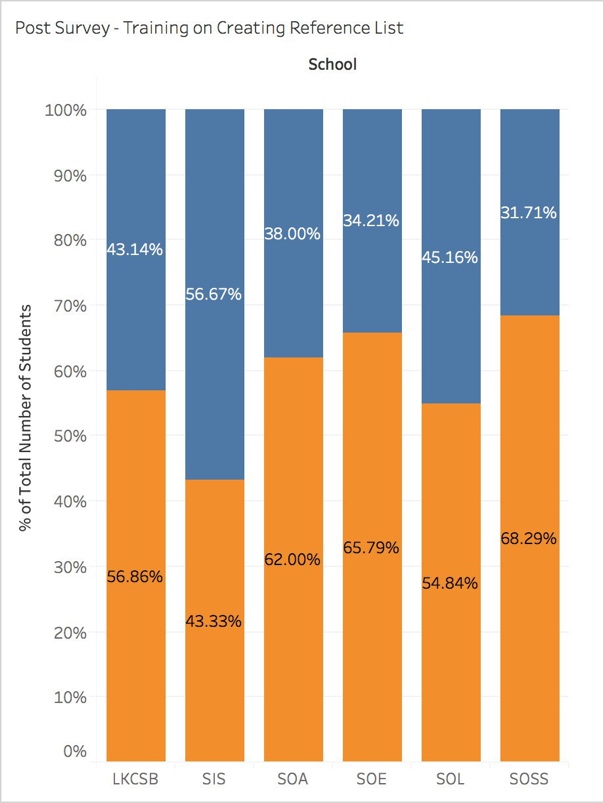


Figure 14. Training for creating reference lists

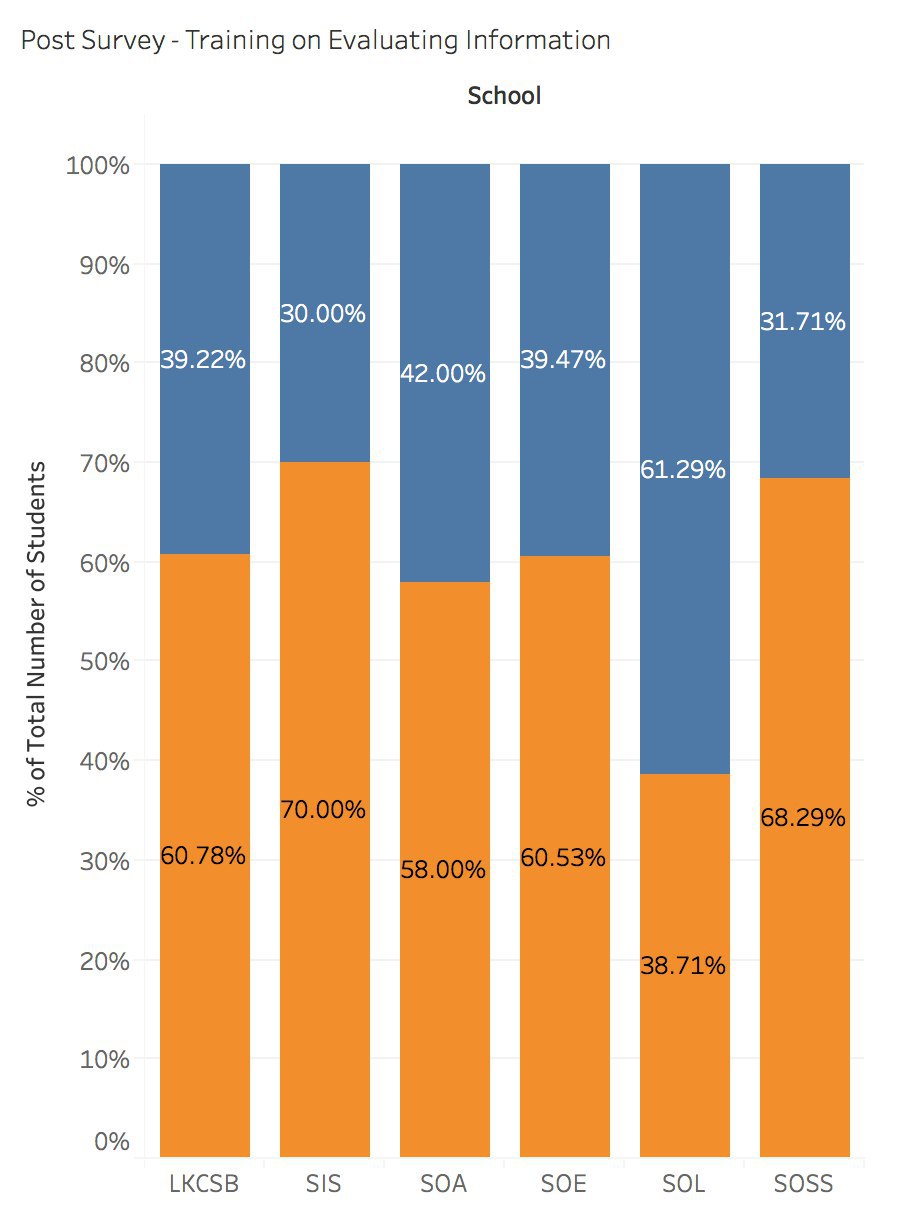


Figure 15. Training for evaluating information

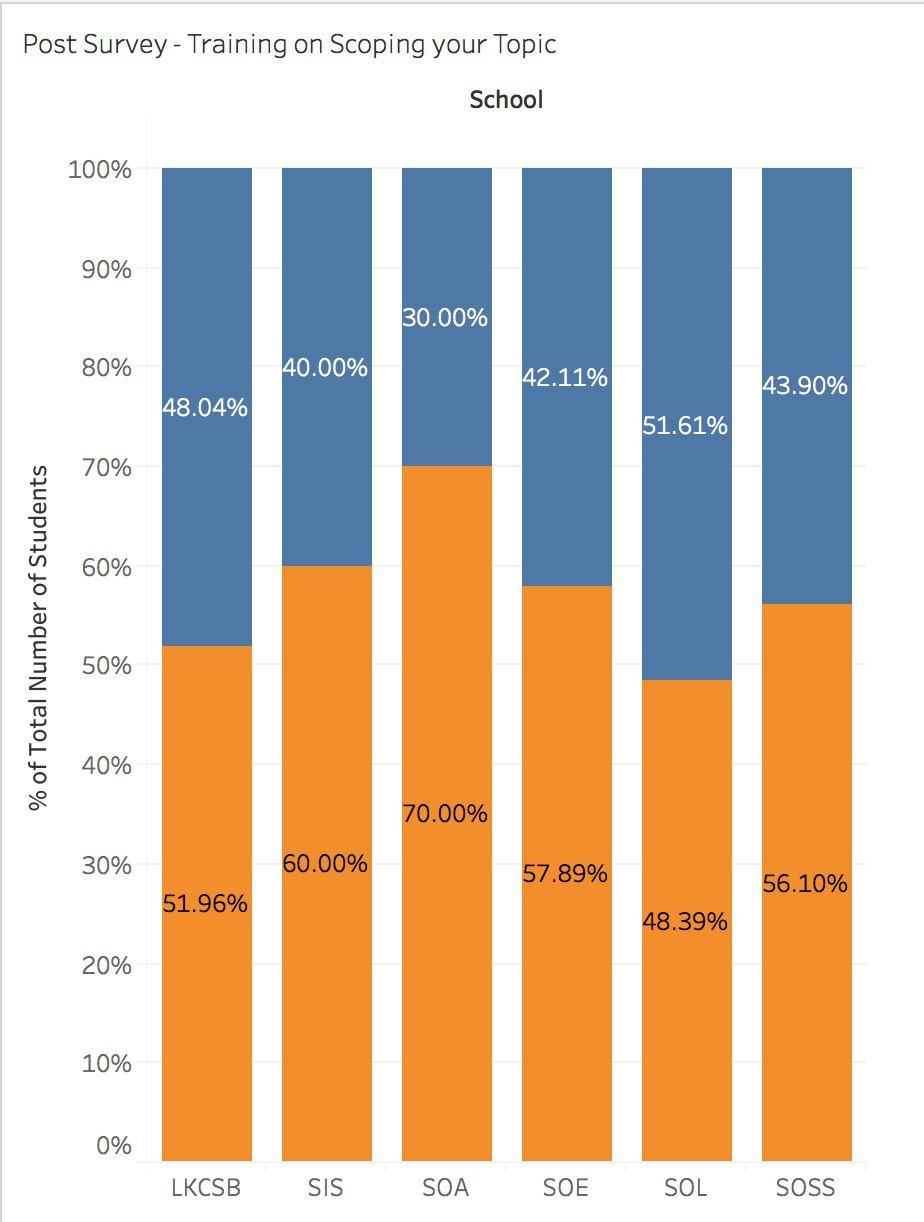


Figure 16. Training for scoping your topic

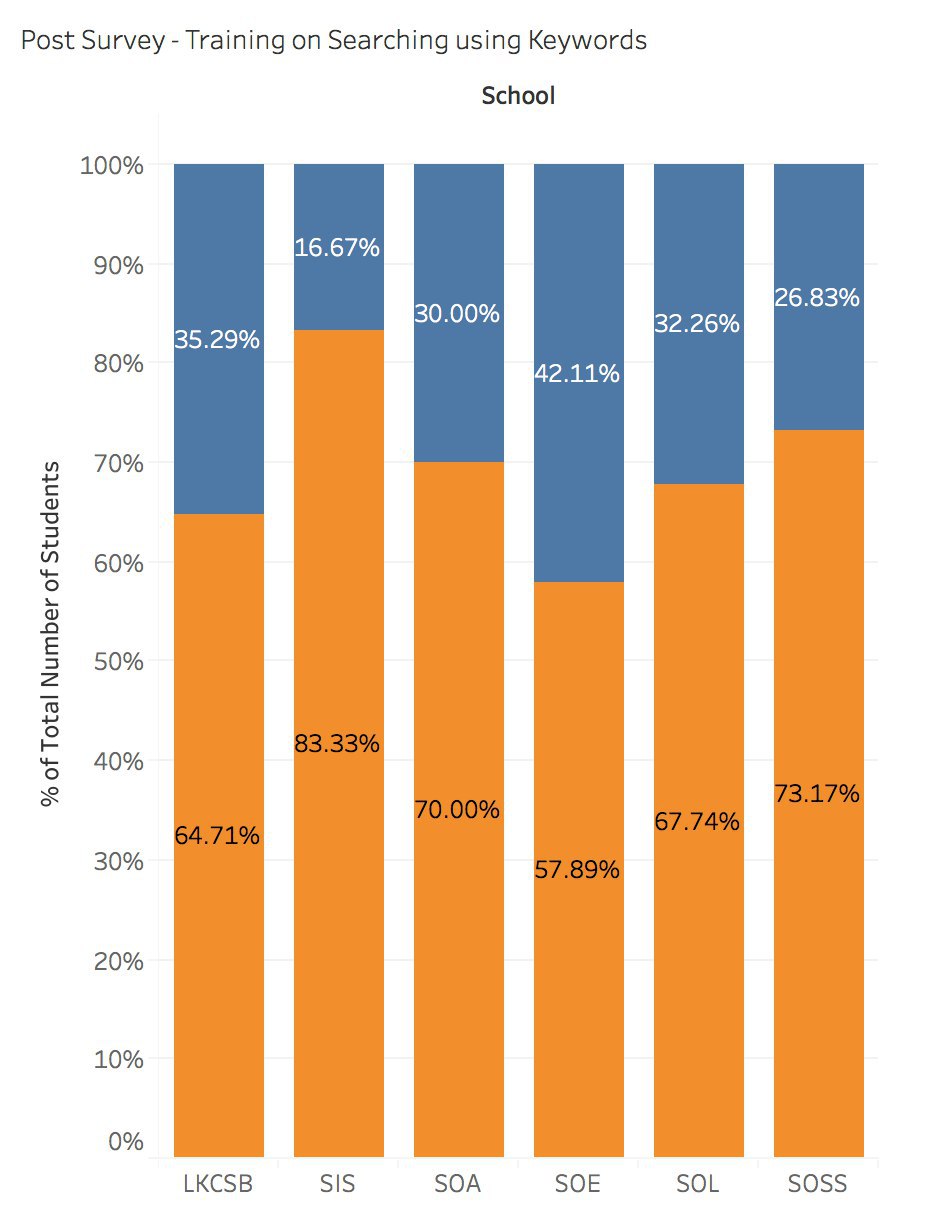


Figure 17. Training for searching using keywords

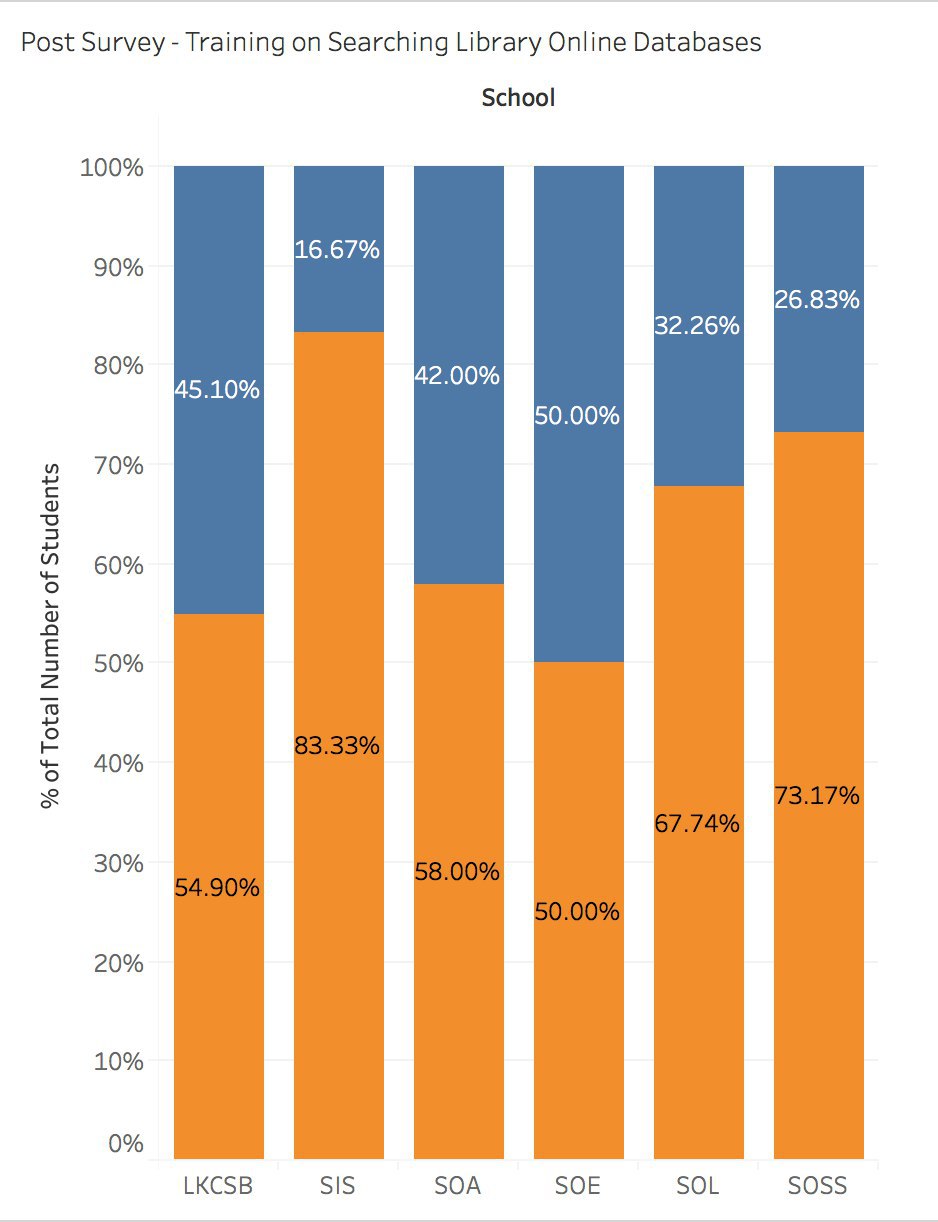


Figure 18. Training for searching library online databases

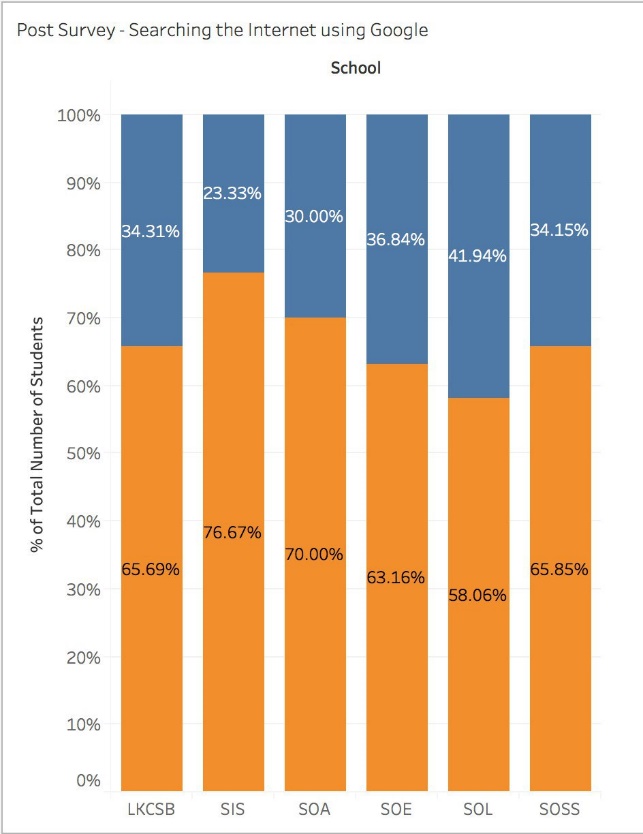


Figure 19. Training for searching the internet using google

## APPENDIX B: EFFECT OF TRAINING ON CONFIDENCE LEVELS

Confidence level of SMU students who did not receive any training in citing references

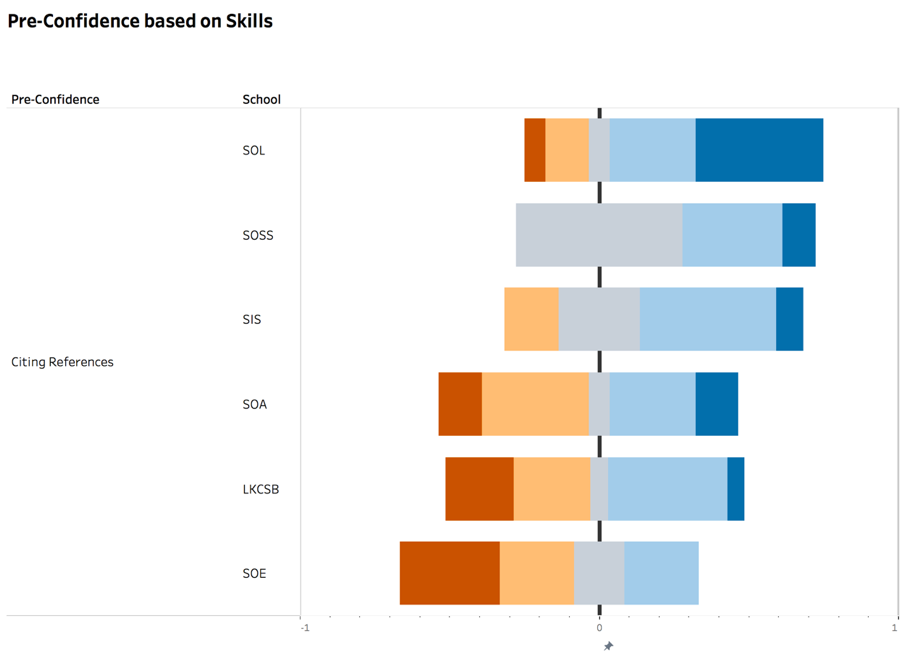


Figure 20. Pre-Confidence of citing references without training

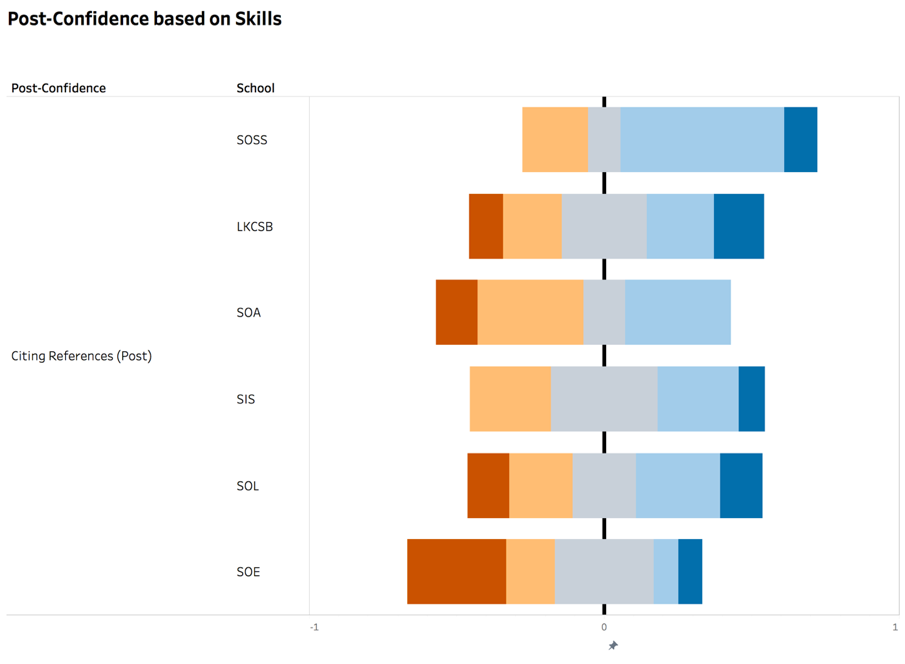


Figure 21. Post-Confidence of citing references without training

Confidence level of SMU students received training in citing references

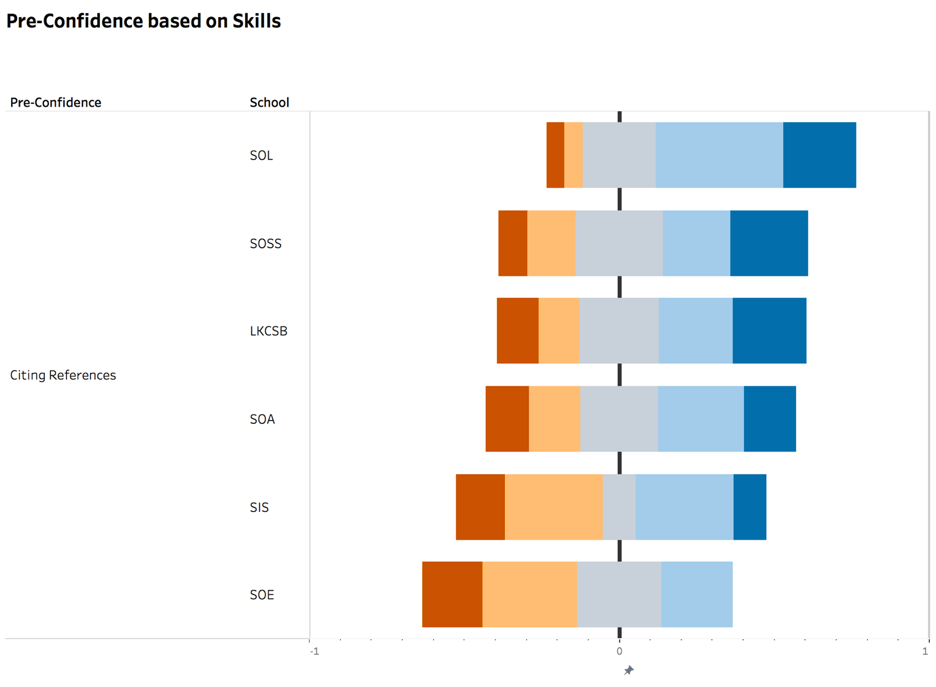


Figure 22. Pre-Confidence of citing references with training

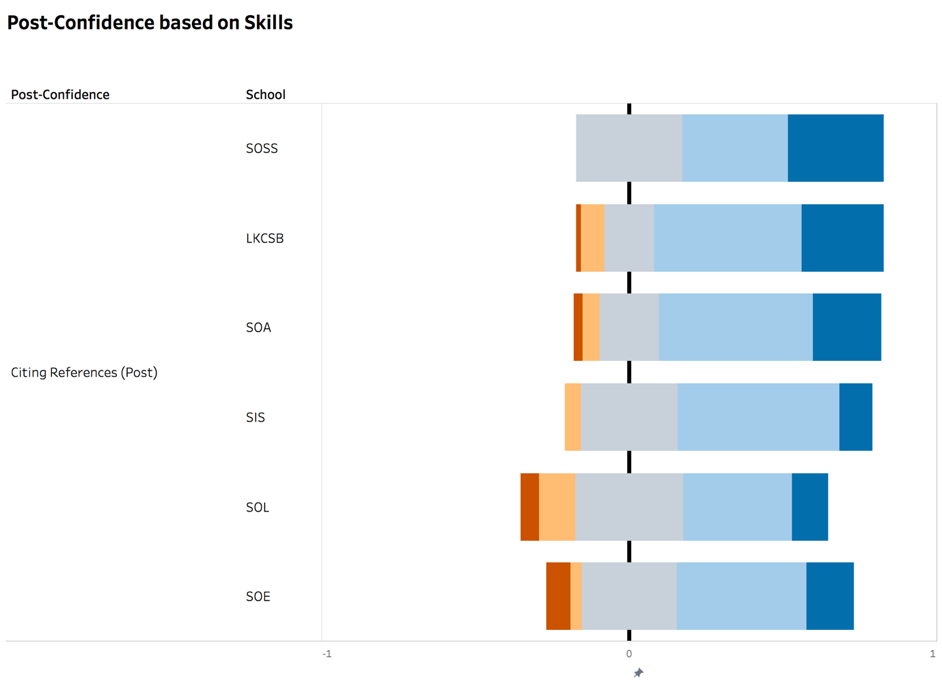


Figure 23. Post-Confidence of citing references with training

Confidence level of SMU students who did not receive any training in creating references lists

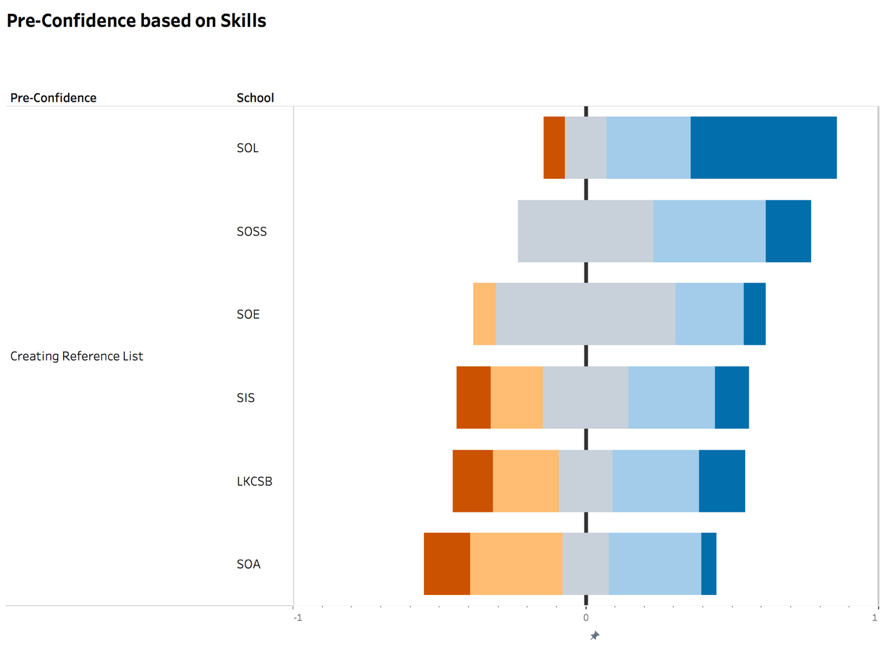


Figure 24. Pre-Confidence of creating reference lists without training

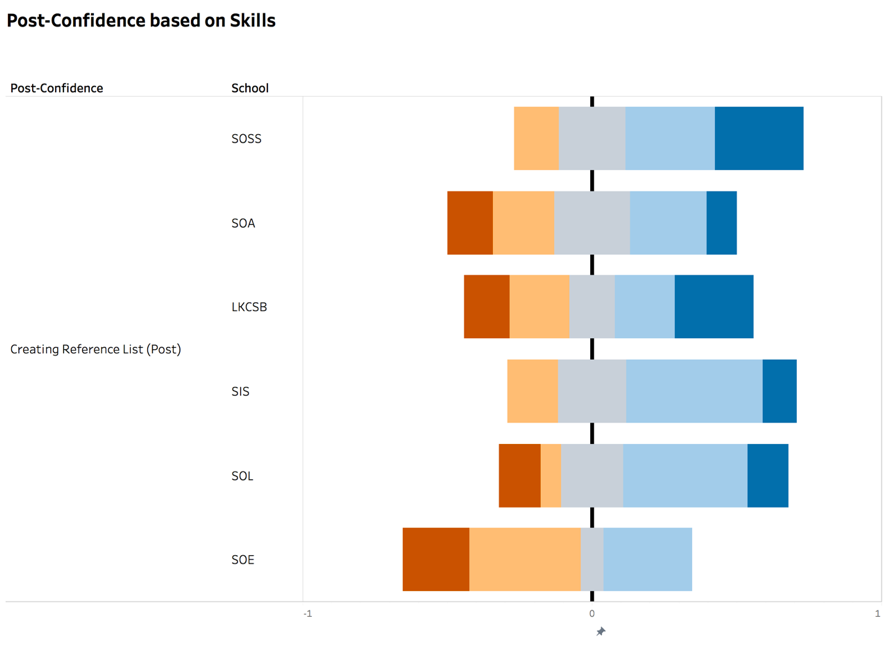


Figure 25. Post-Confidence of creating reference lists without training

Confidence level of SMU students who received training in creating references lists

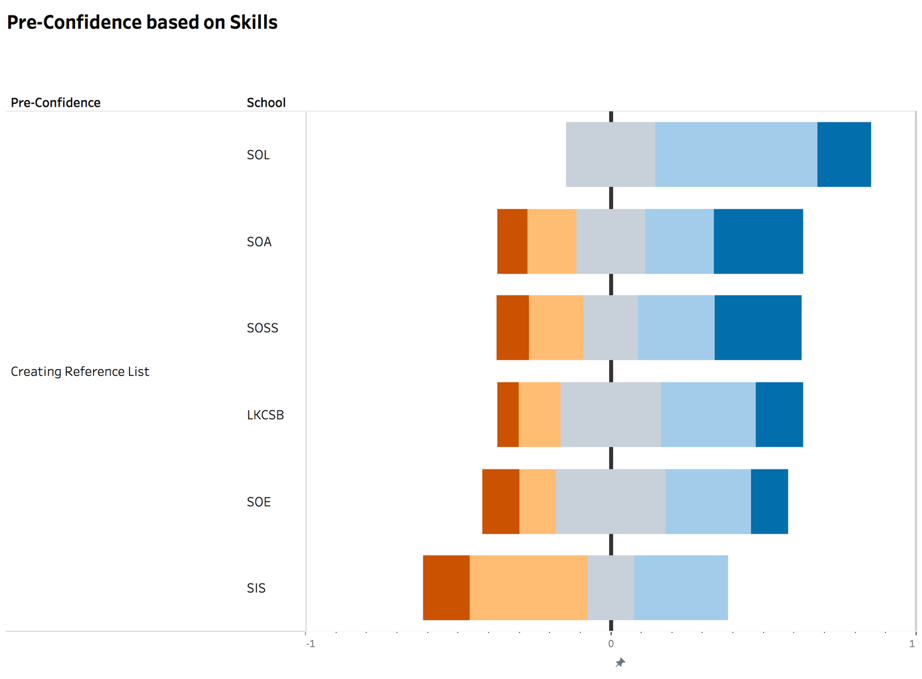


Figure 26. Pre-Confidence of creating reference lists with training

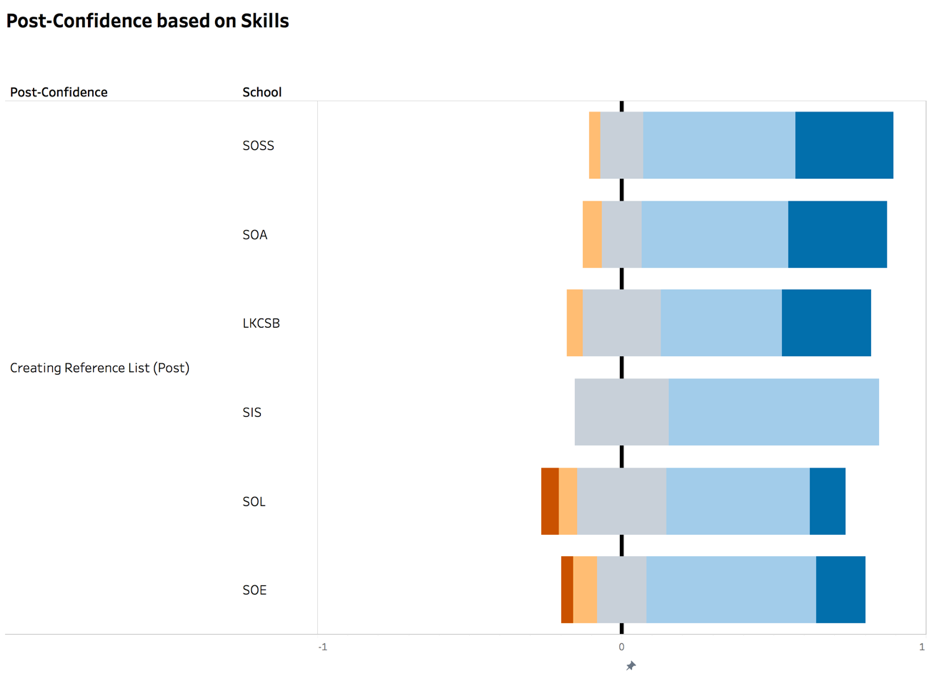


Figure 27. Post-Confidence of creating reference lists with training

Confidence level of SMU students who did not receive any training in evaluating information

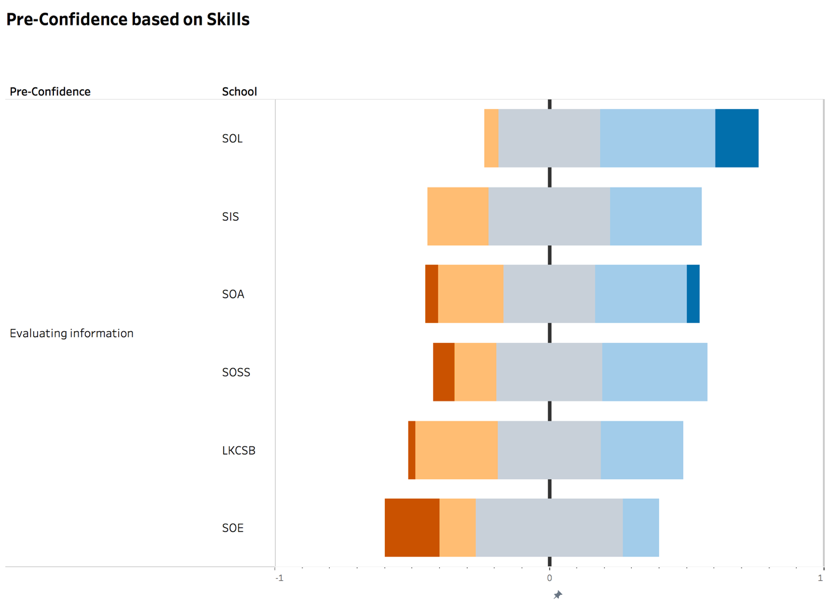


Figure 28. Pre-Confidence of evaluating information without training

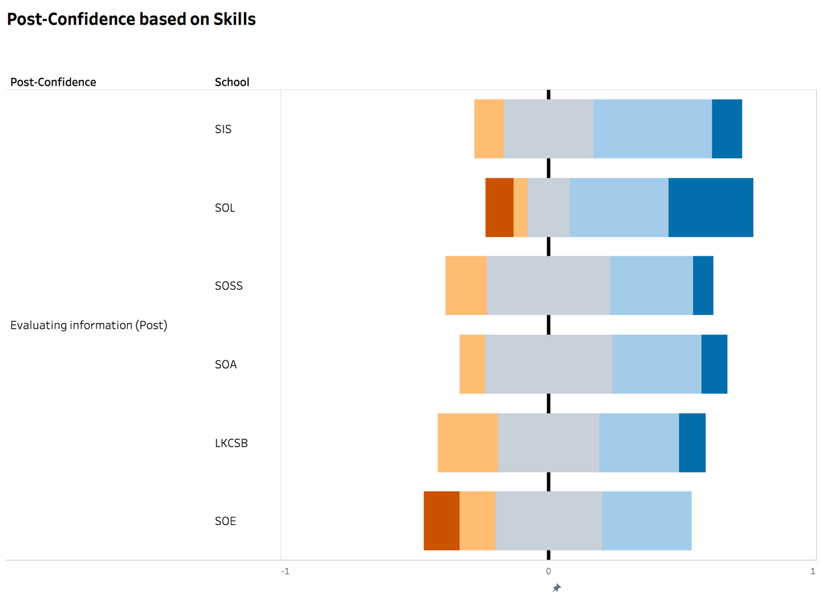


Figure 29. Post-Confidence of evaluating information without training

Confidence level of SMU students received training in evaluating information

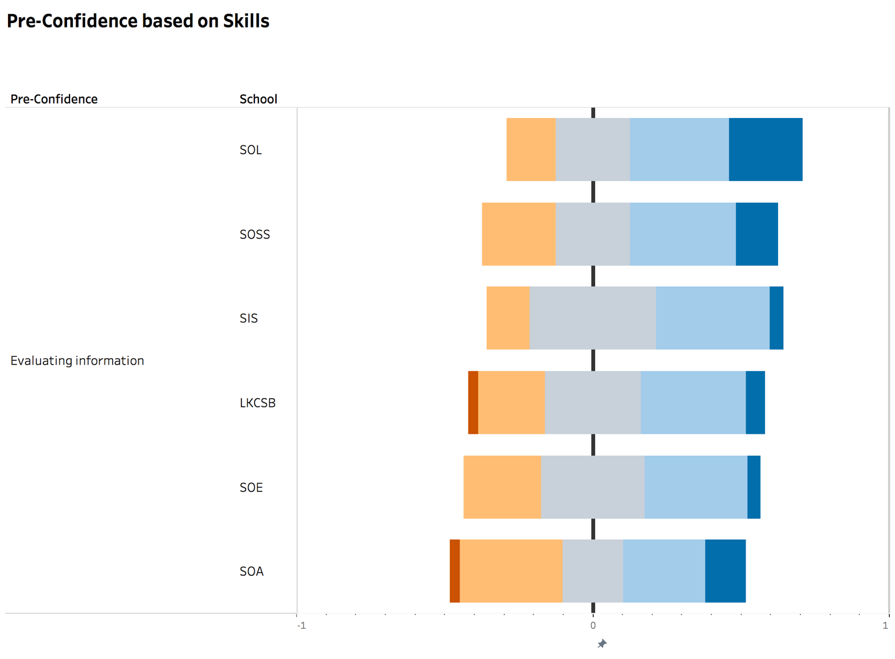


Figure 30. Pre-Confidence of evaluating information with training

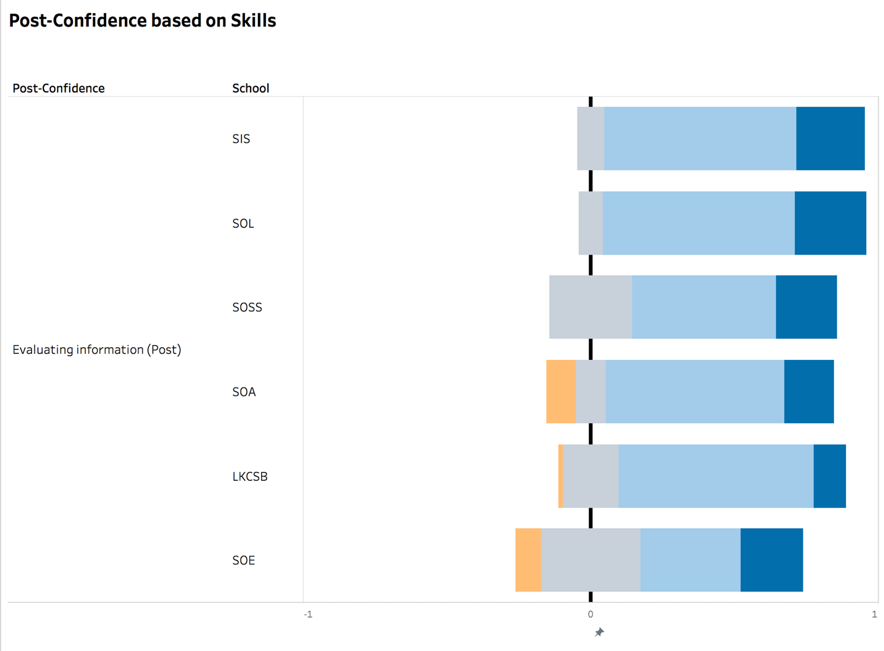


Figure 31. Post-Confidence of evaluating information with training

Confidence level of SMU students who did not receive any training in scoping your topic

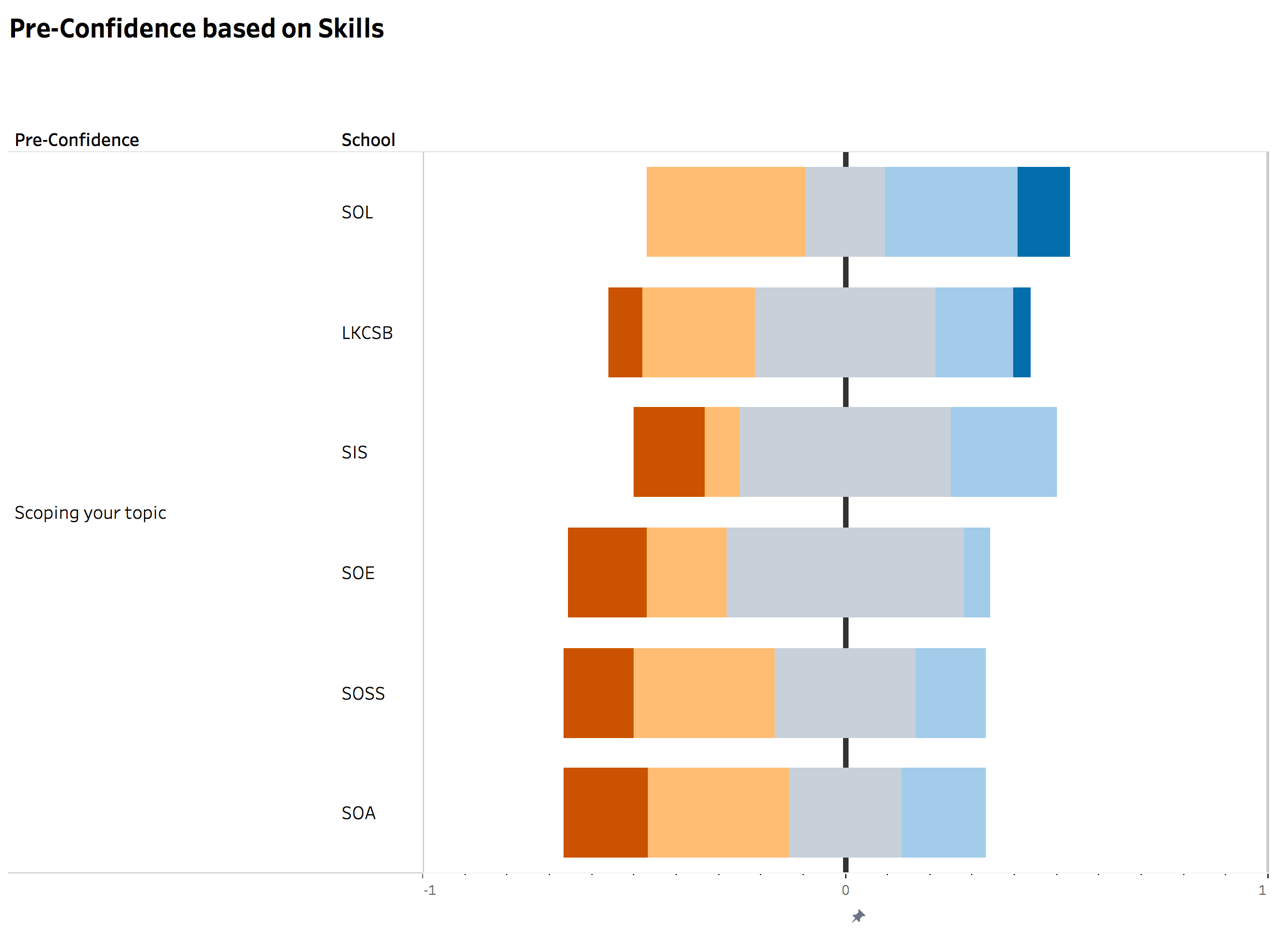


Figure 32. Pre-Confidence of scoping your topic without training

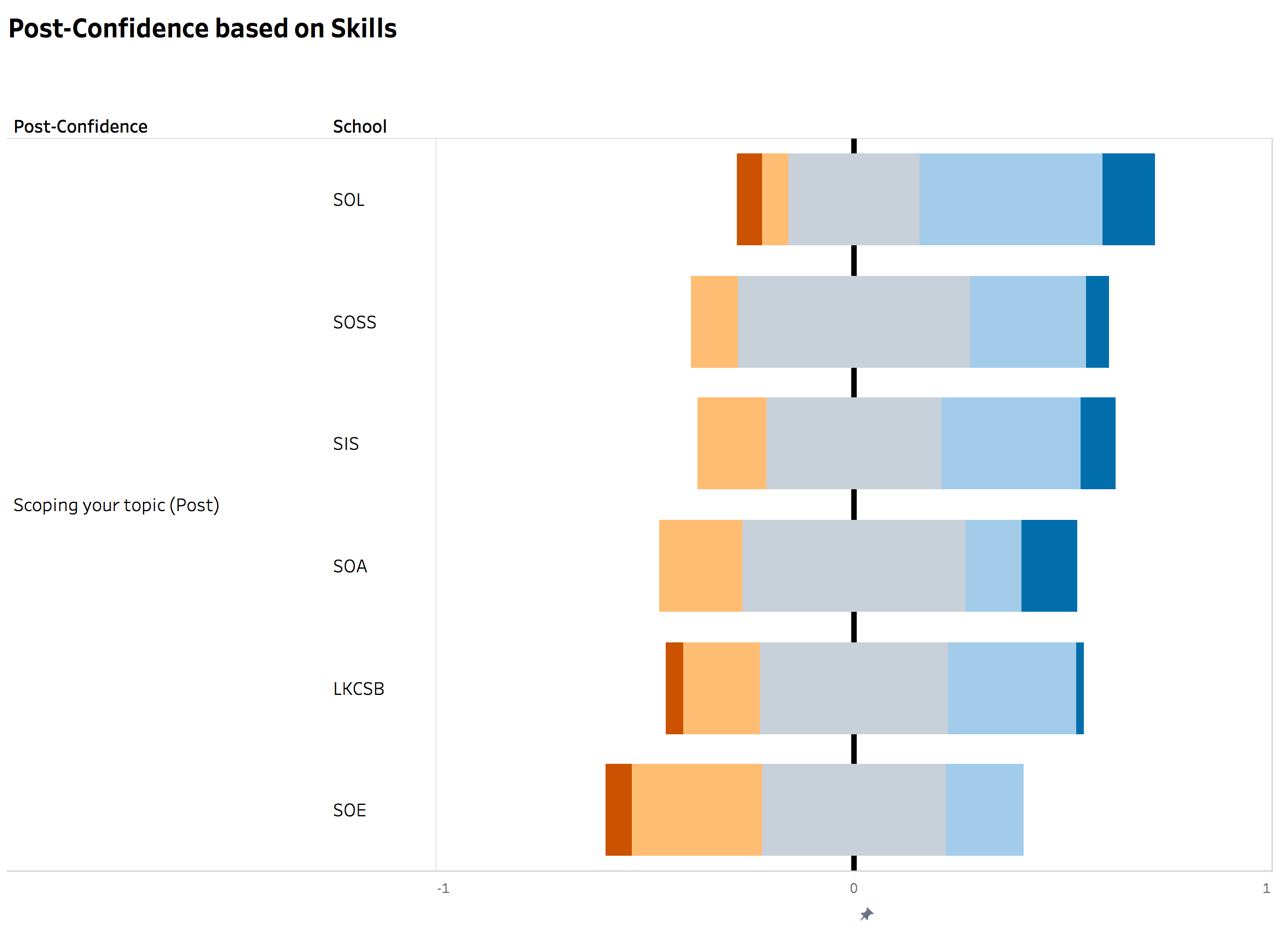


Figure 33. Post-Confidence of scoping your topic without training

Confidence level of SMU students received training in scoping your topic

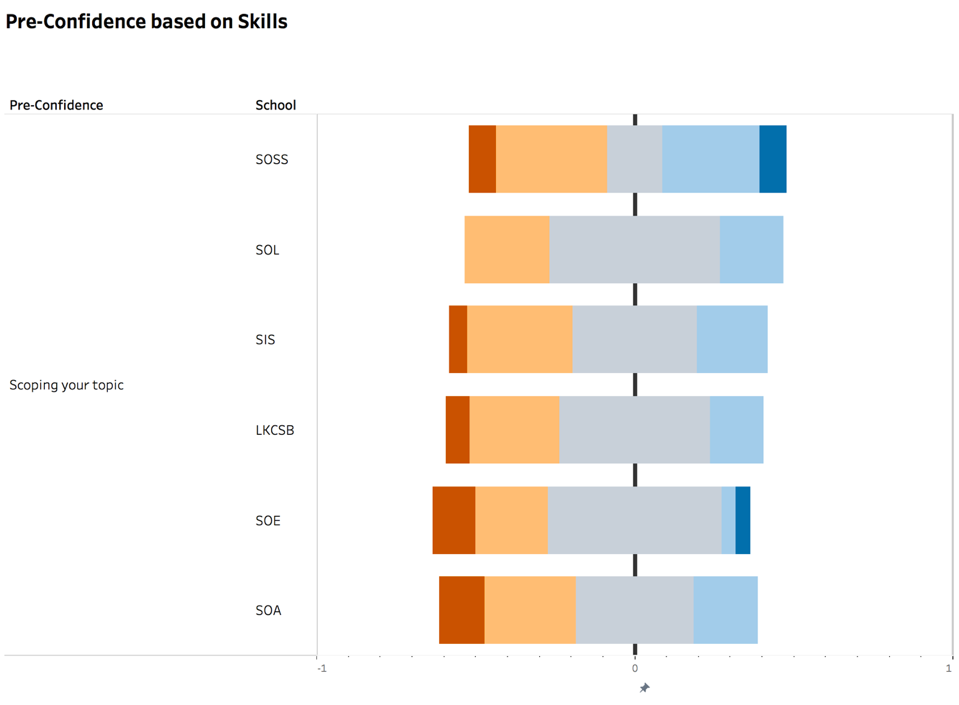


Figure 34. Pre-Confidence of scoping your topic with training

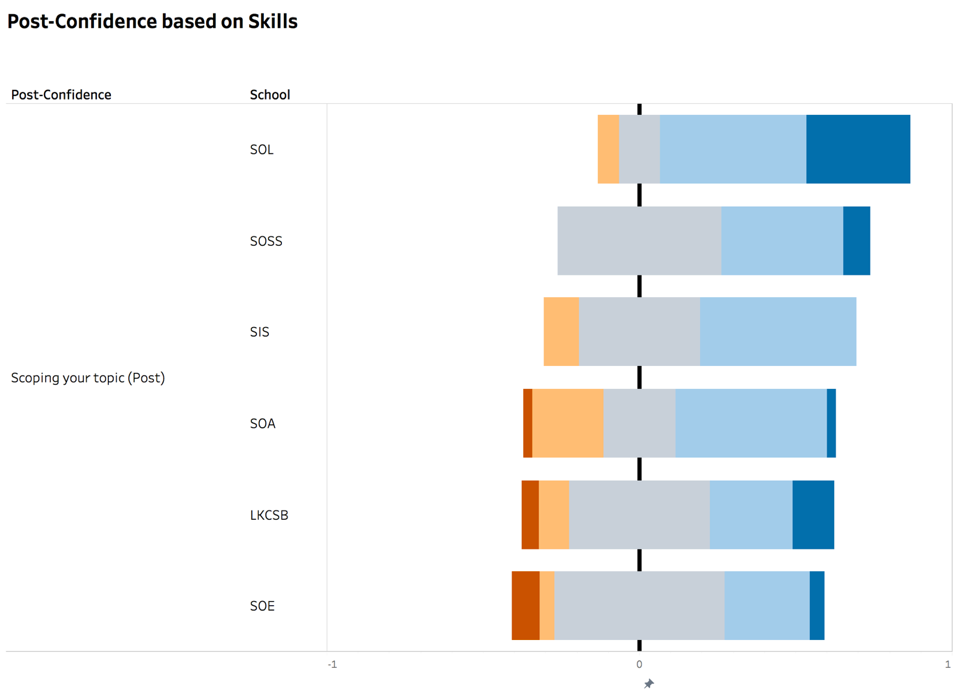


Figure 35. Post-Confidence of scoping your topic with training

Confidence level of SMU students who did not receive any training in seaching keywords

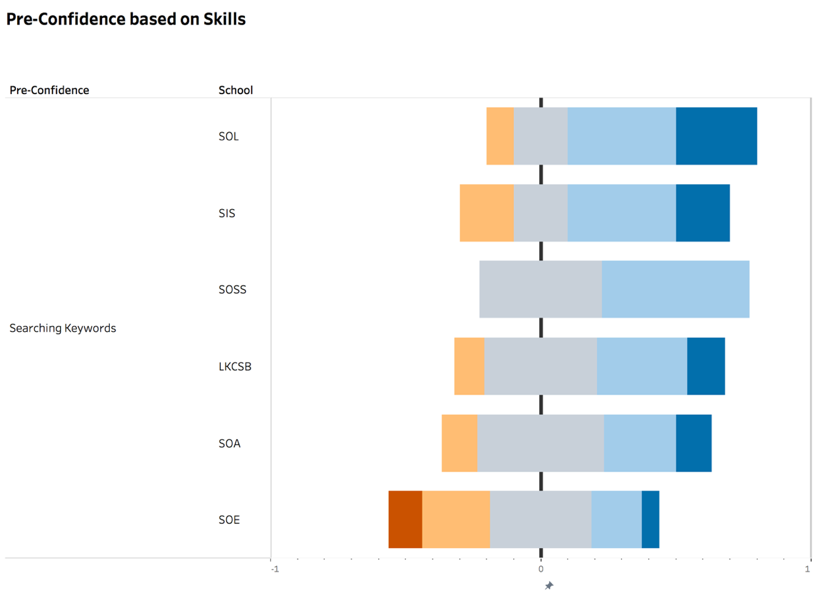


Figure 36. Pre-Confidence of searching keywords without training

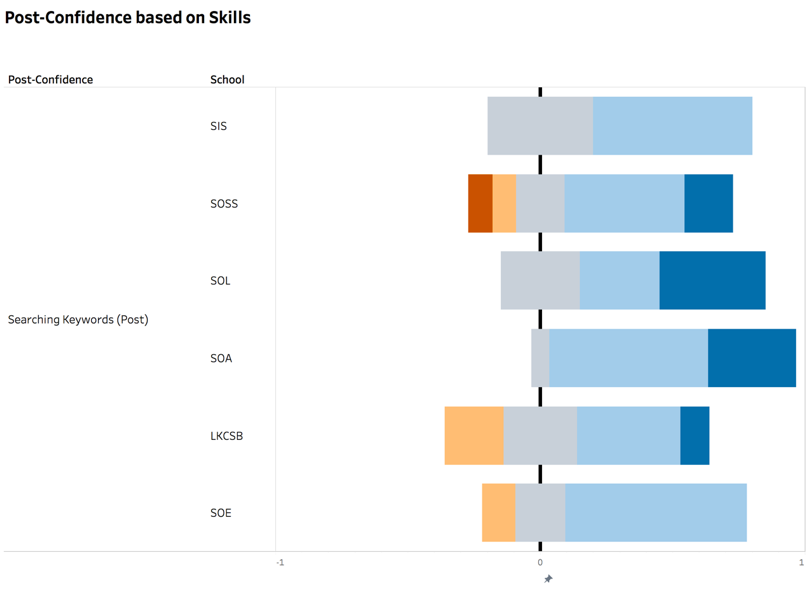


Figure 37. Post-Confidence of searching keywords without training

Confidence level of SMU students received training in seaching keywords

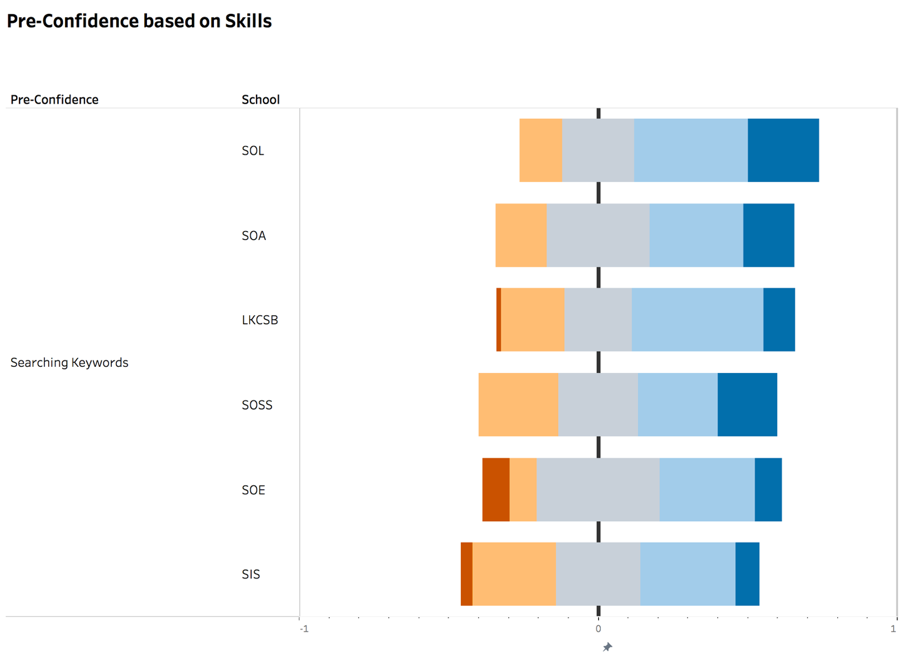


Figure 38. Pre-Confidence of searching keywords with training

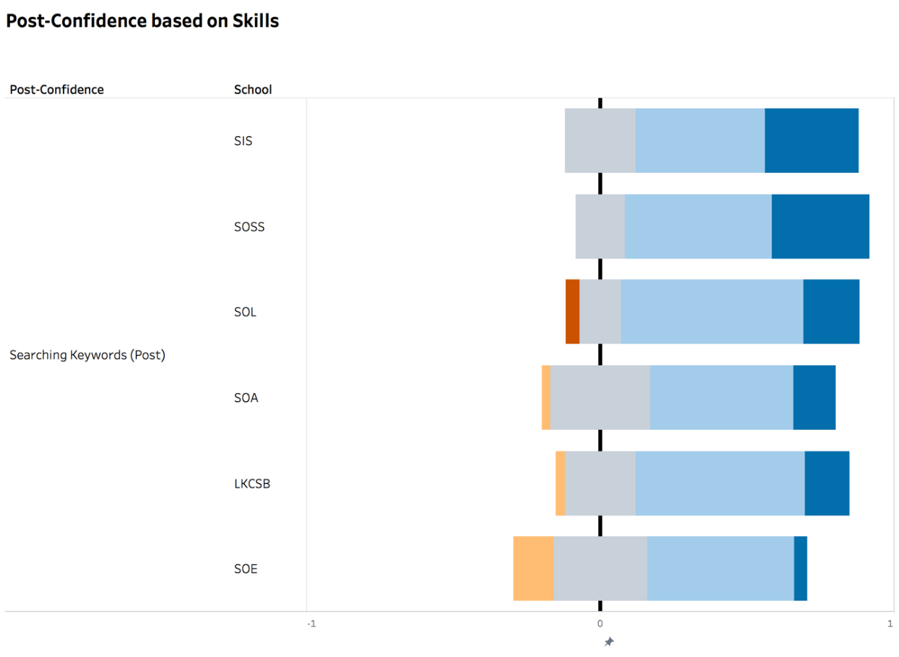


Figure 39. Post-Confidence of searching keywords with training

Confidence level of SMU students who did not receive any training in seaching library online databases

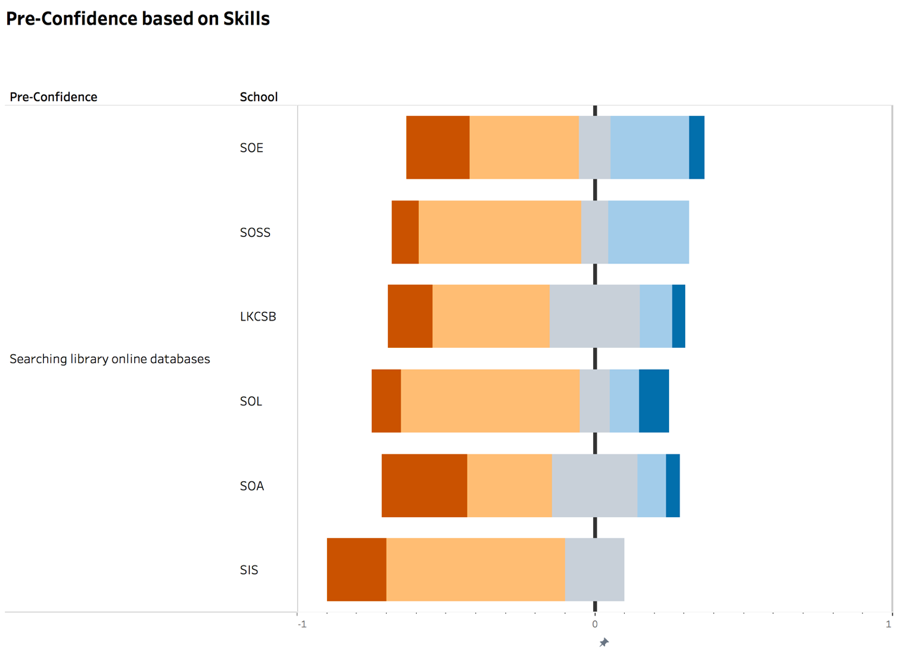


Figure 40. Pre-Confidence of searching library online databases without training

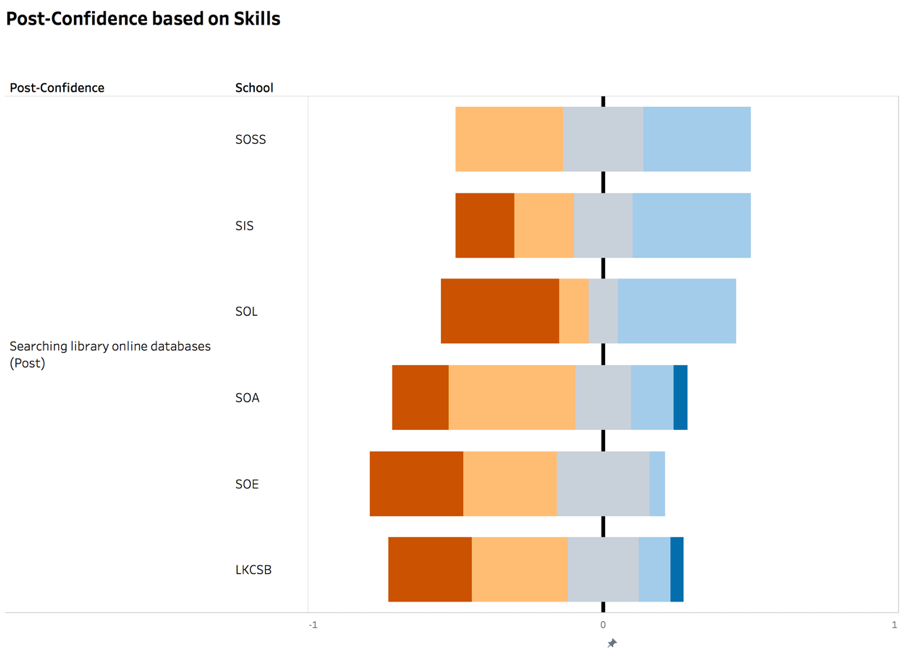


Figure 41. Post-Confidence of searching library online databases without training

Confidence level of SMU students who received training in seaching library online databases

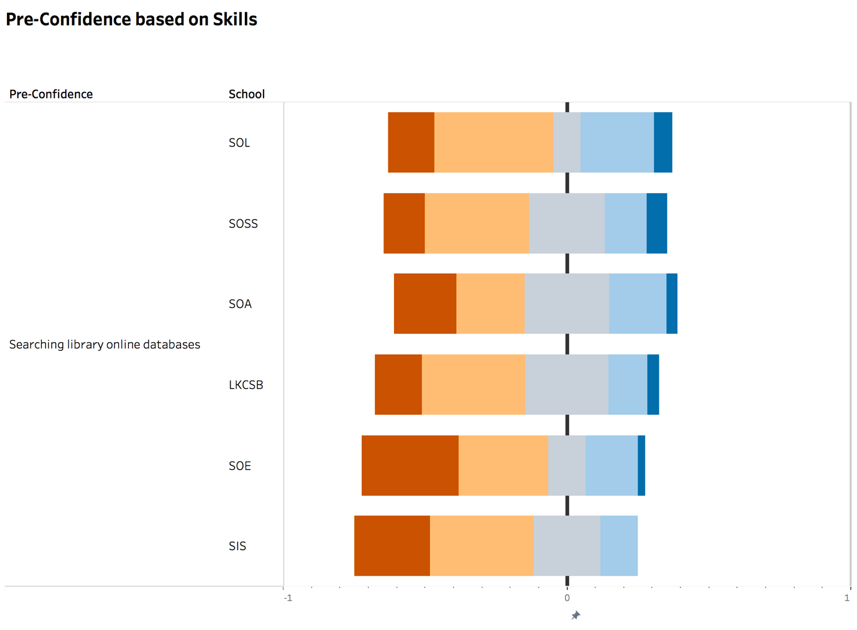


Figure 42. Pre-Confidence of searching library online databases with training

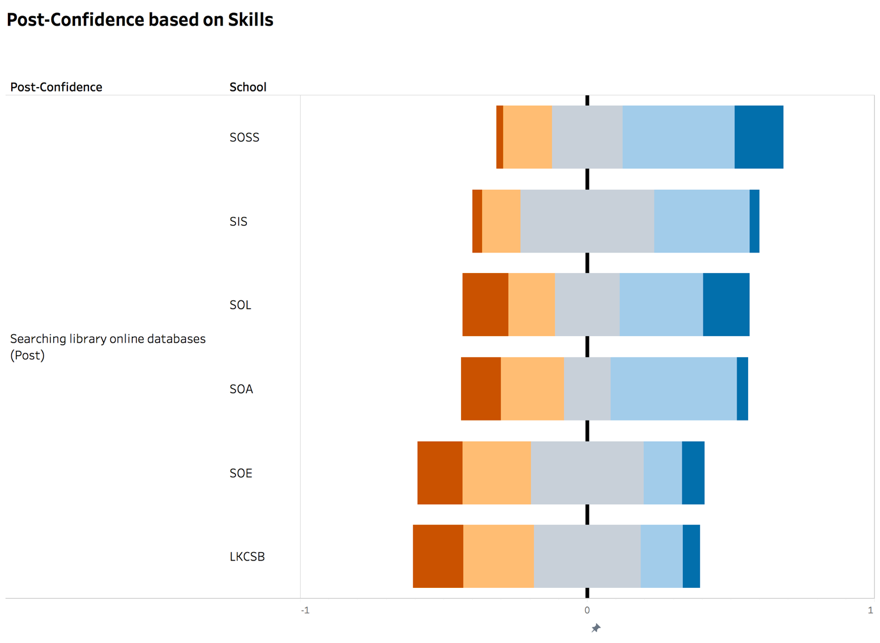


Figure 43. Post-Confidence of searching library online databases with training

Confidence level of SMU students who did not receive any training in seaching the internet using google

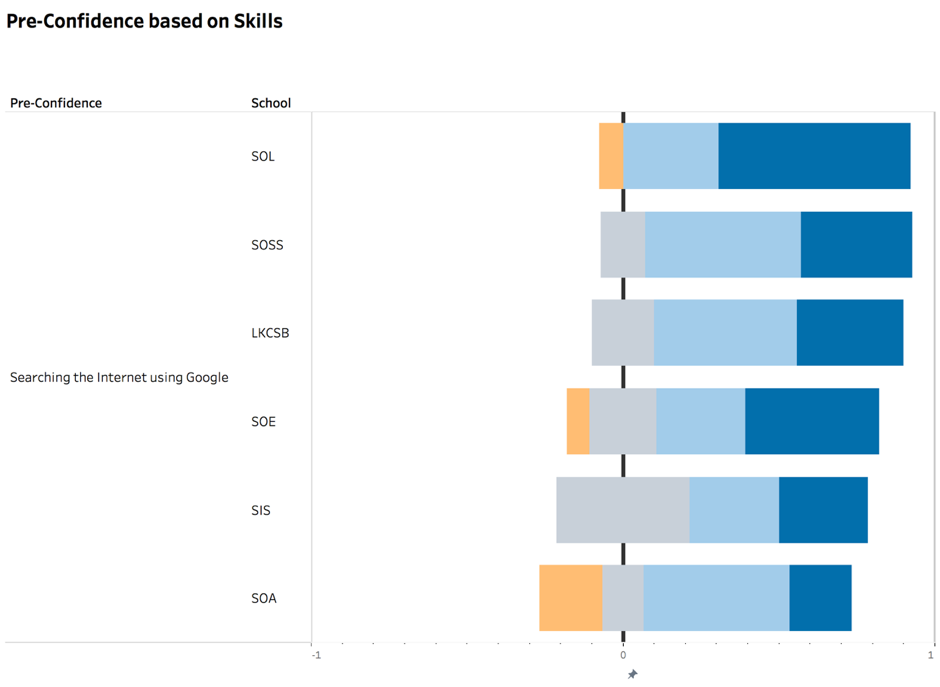


Figure 44. Pre-Confidence of searching the internet using google without training

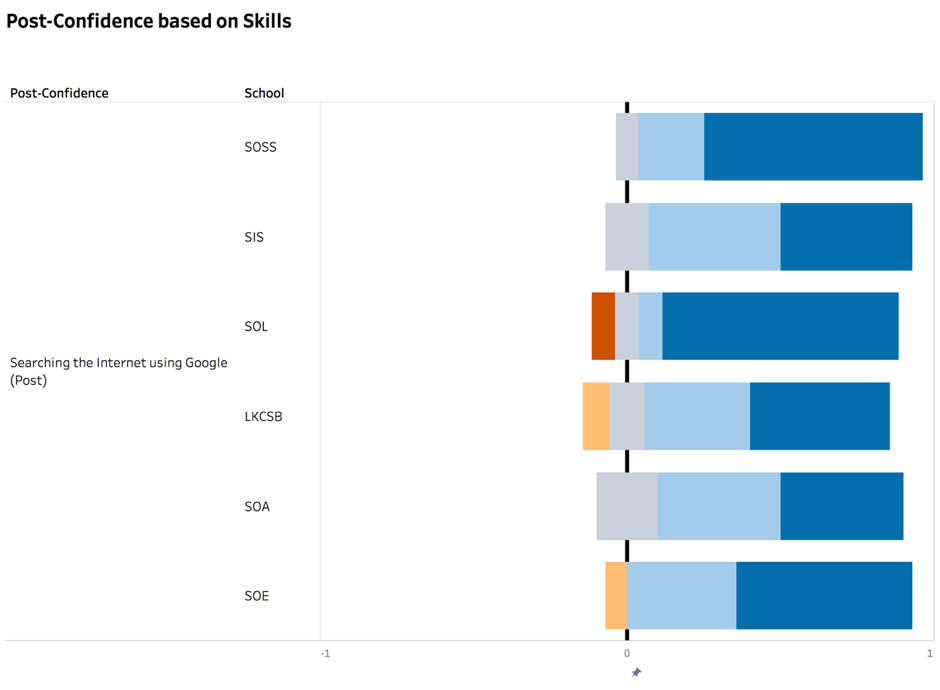


Figure 45. Post-Confidence of searching the internet using google without training

Confidence level of SMU students who received training in seaching the internet using google

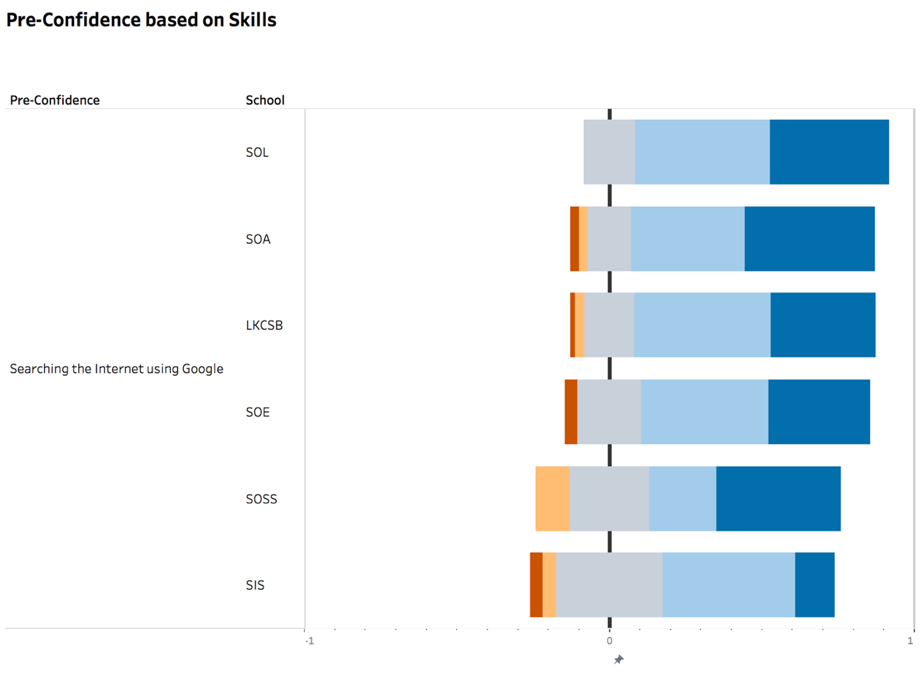


Figure 46. Pre-Confidence of searching the internet using google with training

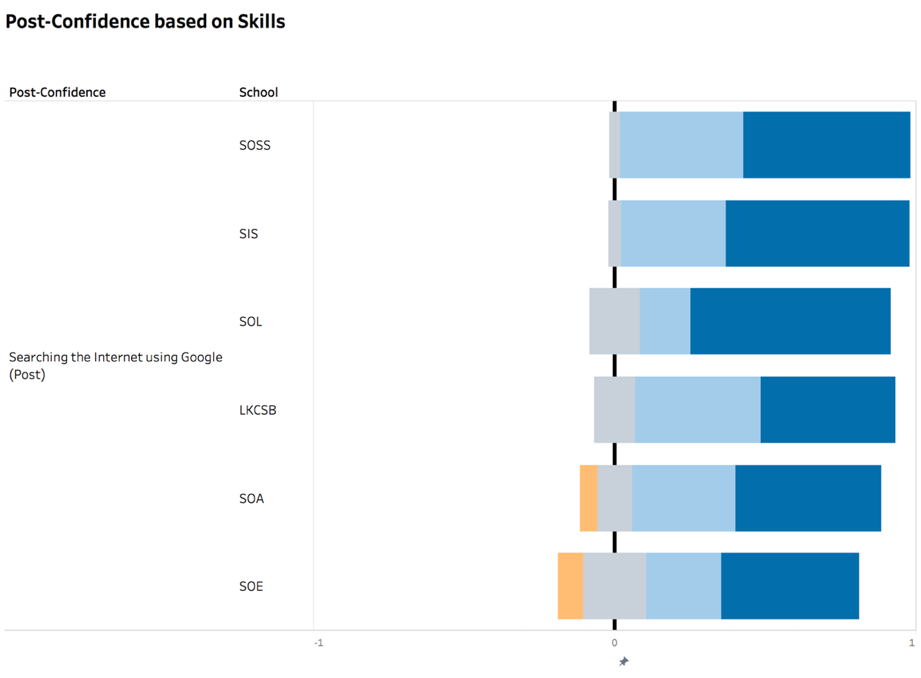


Figure 47. Post-Confidence of searching the internet using google with training