LIS_Xinwei_Kuang

Xinwei(Wendy) Kuang

Summary

This report analyzes long-term Impacts and Outcomes of Experiential Learning(IEL) in International Service Learning(ISL) and Global Seminar Programs.Office of Regional and International Community Engagement(ORICE) has conducted a piolt survey to UBC alumini who have participated in funded ISL and Global Seminar program and got responses.The objective of this report is to evaluate the effectiveness of the survey and distribute the responses. Comparing with responses from funded and nonfunded students after analyzing both, findings in data for funded students are more solid since it gets a good model fitting. This report will only show analysis for responses from funded students.

Introduction

Both ISL and Global Seminar are course-based placements which provide students the knowledge and skills to engage with international partners. According to researches, IEL experiences have shown great positive influences in student's personal, academic and professional development. However, the research about the long-term impacts and outcomes of IEL programs is very limited. A 15-minute online survey with questions(on a Likert scale) for UBC alumini 1-8 years out from the ISL and/or Global Seminar program. The survey measures the growth of individuals in four dimensions which are personal impact and self-perception,knowledge and skills, education and career trajectory and civic engagement.The two questions to be addressed are assessing the quality of the data and finding distributions of the responses; analyzing how responses to the questions in the survey are correlated with and between dimensions. To achieve this, barplots of responses to all questions will be displayed to show how students response. Besides, a confirmatory factor analysis will be performed to get useful categories of questions within each dimension for improving the questionnaire for future studies.

Data Description

Completed responses to the survey were collected and recorded in Excel.The funded dataset consists of 92 students who were funded for both ISL and Global Seminar programs. There are 95 questions in the dataset and consist of likert ones which have 3-point scale, 5-point scale, or 6-point scale(with Not Applicable) and non-likert ones(open-ended questions), while likert scale questions are what the client is interested in:

Each question refers to a specific variable under one of the following dimensions: 1. Personal Impact and Self-Perception: resiliency, open-mindedness, and self-confidence 2. Knowledge and Skills: course, relevancy, type of knowledge 3. Education and Career Trajectory: goal selection, problem solving, planning and self-appraisal, positive or negative reinforcement, international experience as an employability asset. 4. Civic Engagement: network, local and global awareness, and how individuals view issues around them.

Each question together with any sub-questions is treated as a variable in the dataset since one question assesses a specific variable as given.

Data Cleaning

The responses need to be cleaned and transformed for analysis. All non-likert variables and unused rows were removed; then, observations with above 50% missing values were removed. After that, all likert items were mapped to the same rating scale from 1 for low to 5 for high(Likert scale Standardization) since having checked that non of the responses is 'Not Applicable' for further data analysis. Potential missing data approaches will

be explored to see if a more complete dataset can be created. To deal with the rest nonresponses, the means of the responses were assigned to the nonresponses for each variable(Ronald,Downey & Craig, 1998). The direction of questions was also considered because one of the questions(Q9_2) has rating opposite to others, 1 for high and 5 for low. As a result, their values were switched for this question.

Table 1 shows 1-10 observations with 10 variables of cleaned data

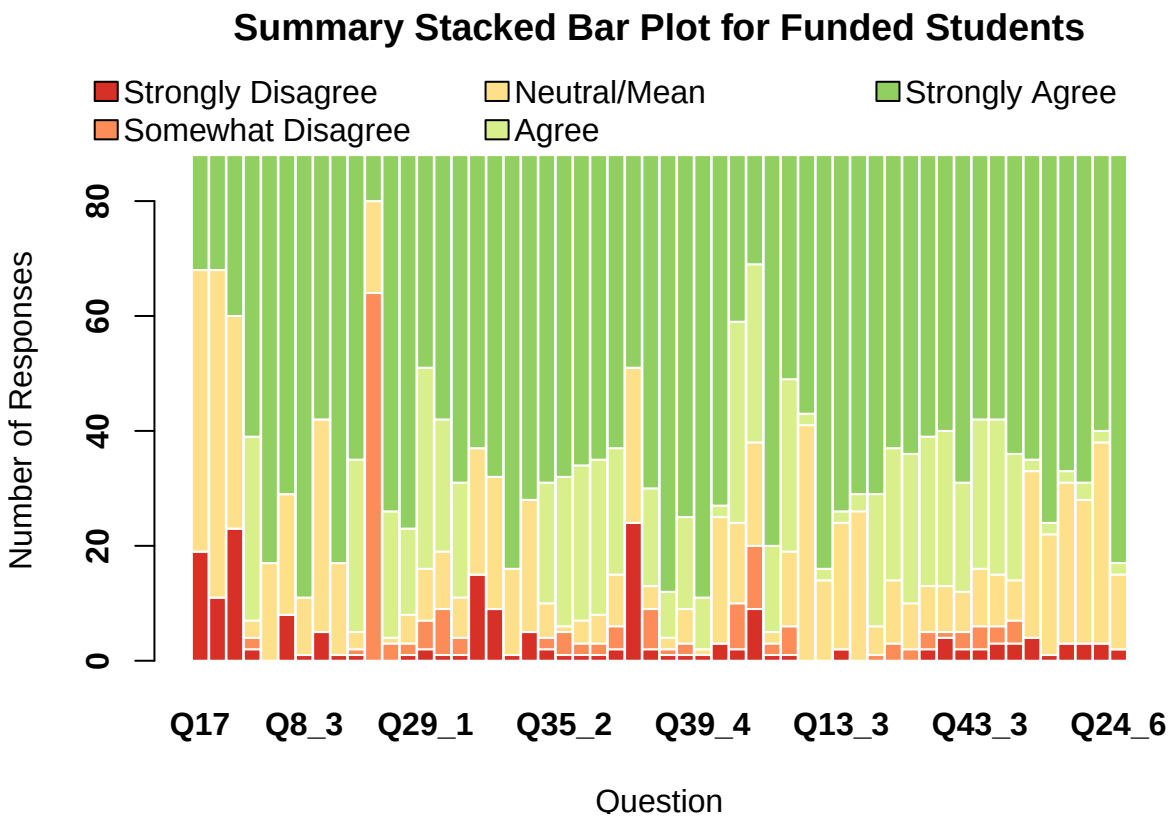
Table 1: parts of Cleaned Data

Q17	Q22	Q20	Q7_1	Q8_1	Q8_2	Q8_3	Q8_4	Q8_5	Q9_1
3	3.4	5.0	4	5	5	5	5	5	5
5	5.0	5.0	5	5	5	5	5	5	5
5	3.0	3.0	5	5	5	5	3	5	5
3	3.4	3.0	4	5	1	5	5	5	5
1	3.4	3.1	5	5	1	3	3	5	5
3	3.4	5.0	4	5	5	5	5	5	5
5	3.0	3.1	5	5	1	5	3	5	4
3	3.4	5.0	1	3	1	3	1	3	3
3	5.0	1.0	4	3	5	5	3	5	4
3	3.0	3.0	5	5	3	5	5	5	5

Data distribution

Stacked bar plot was used to show the distribution of data. All questions are included.

```
barplot(counts_bind_funded, col=brew_col_funded, border="white", space=0.04, font.axis=2,
        names.arg = questions_funded, xlab = "Question", ylab = "Number of Responses")
par(xpd=TRUE)
par(cex=1)
legend("top", fill=brew_col_funded, inset=c(0,-0.2),
       c("Strongly Disagree", "Somewhat Disagree", "Neutral/Mean", "Agree","Strongly Agree"),
       x.intersp = 0.2,y.intersp = 1, text.width = 20, ncol = 3, bty='n')
title(main = "Summary Stacked Bar Plot for Funded Students", line = 3)
```



We used processed data to plot the graph. It's clear to see that the majority of responses are Strongly agree.

Methods

Confirmatory Factor Analysis(CFA)

The confirmatory factor analysis is a method which helps to find out which variables should be under which dimension by setting up fixed number of dimensions and building up model at first. We used the processed data and aggregated responses of each question as a single variable by taking the means if one has subquestions because they measures the same dimensional variables in the dimensions. In this way, we reduced the number of variables, making it less than the number of observations in order to conduct this method. Since the data given by the client specifies which dimension each question is measuring, we analyzed the extent to which the factor loadings for these questions match up with what the client defined.

Combine sub-questions

```
fQ8 <- (funded$Q8_1 + funded$Q8_2 + funded$Q8_3 + funded$Q8_4 + funded$Q8_5)/5
```

For each question with sub-questions, it was combined and a new dataset was got.

```
combined_f <- cbind(fQ17,fQ20,fQ22,fQ7,fQ8,fQ9,fQ28,fQ29,fQ34,fQ35,fQ37,fQ39,fQ10,
                    fQ11,fQ13,fQ24,fQ43,fQ45)
```

Define model and fit CFA

```
F.model <-'
edu =~ fQ17 + fQ20 + fQ22
per =~ fQ7 + fQ8 + fQ9
kno =~ fQ28 + fQ29 + fQ34 + fQ35 + fQ37 + fQ39
civ =~ fQ10 + fQ11 + fQ13 + fQ24 + fQ43 + fQ45'

fitf <- cfa(F.model, data=combined_f, std.lv=TRUE, missing="fiml")
```

Assumptions

- The sample size is large enough.
  - Although the sample size 92 are not greater than 200, we can use another rule: `N/p >= 10` (Nicholas et al., 2011).
  - `N/p >= 10` could be satisfied as p is the number of factors (4 in our case), and N is the sample size. N/p is greater than 10 in all the cases.
  - Verify sample size by doing analysis on the power of the model.
- The observations should be independent
- Multivariate Normality
- The latent variable is not correlated with the measurement error.
- The latent variables are dependent on each other
- Linearity.

Results

From the summary of fitted CFA model in the Appendix, we observe that the root mean square residul(RMSR=0.085) suggests that the model provides a good fit according to Ali and her team(2011). And, Root Mean Square Error of Approximation (RMSEA) value is 0.096, which is between 0.08 and 0.1, which is considered to be marginal, as noted by Fabrigar and other team members(1999), which means the analysis needs to improve for the future.

Summary of Factor Loadings

```
parameterEstimates(fitf, standardized=TRUE) %>%
  filter(op == "=~") %>%
  select('Latent Factor'=lhs, Indicator=rhs, B=est, SE=se, Z=z, 'p-value'=pvalue,
         Beta=std.all) %>%
  kable(digits = 3,caption="Factor Loadings",format='pandoc')
```

Table 2: Factor Loadings

Latent Factor	Indicator	B	SE	Z	p-value	Beta
edu	fQ17	0.712	0.211	3.383	0.001	0.535
edu	fQ20	0.474	0.225	2.104	0.035	0.312
edu	fQ22	0.272	0.170	1.600	0.110	0.236
per	fQ7	0.306	0.100	3.068	0.002	0.361
per	fQ8	0.574	0.099	5.802	0.000	0.883
per	fQ9	0.079	0.068	1.161	0.246	0.142
kno	fQ28	0.467	0.058	8.027	0.000	0.753

Latent Factor	Indicator	B	SE	Z	p-value	Beta
kno	fQ29	0.534	0.073	7.302	0.000	0.707
kno	fQ34	0.547	0.086	6.338	0.000	0.631
kno	fQ35	0.499	0.061	8.136	0.000	0.760
kno	fQ37	0.648	0.176	3.681	0.000	0.396
kno	fQ39	0.530	0.051	10.315	0.000	0.887
civ	fQ10	0.311	0.134	2.323	0.020	0.293
civ	fQ11	0.427	0.094	4.543	0.000	0.570
civ	fQ13	0.315	0.069	4.562	0.000	0.491
civ	fQ24	0.418	0.081	5.157	0.000	0.562
civ	fQ43	0.520	0.083	6.292	0.000	0.675
civ	fQ45	0.316	0.071	4.447	0.000	0.500

From Table 2, we can see that questions are recategorized into 4 dimensions according to factor loadings: Q17,20,22 belong to Education and Career Trajectory; Q7,8,9 belong to Personal Impact and Self-Perception; Q28,Q29,Q34,Q35,Q37,Q39 belong to Knowledge and Skills; Q39,10,11,13,24,43,Q45 belong to Civic Engagement. Comparing to the original data, Q43 and Q24 are removed from Education and Career Trajectory and moved to Civic Engagement; questions for Personal Impact and Self-Perception and Knowledge and SKills keeps the same.

Most of the factor loadings in the funded model fall in range 0.4 to 0.6. It is suggested that factor loadings in the range 0.30 to 0.40 as the minimal requirement, loadings +/-0.5 as practically significant, and loadings +/-0.7 as significant(Black, Anderson & Tatham, 2006). This implies that most of our factor loadings meet the requirements. However two indicators have much lower loadings compared to the others and they are Q9 with loading 0.079 and Q22 with loading 0.272. This suggests that these two indicators have limited influence in their respective dimensions.

Factor Correlations

```
parameterEstimates(fitf, standardized=TRUE) %>%
  filter(op == "~~",
         lhs %in% c("edu", "per", "kno", "civ"),
         !is.na(pvalue)) %>%
  mutate(stars = ifelse(pvalue < .001, "***",
                        ifelse(pvalue < .01, "**",
                               ifelse(pvalue < .05, "*", "")))) %>%
  select('Factor 1'=lhs,
         'Factor 2'=rhs,
         Correlation=est,
         sig=stars) %>%
  kable(digits = 3, format="pandoc", caption="Latent Factor Correlations")
```

Table 3: Latent Factor Correlations

Factor 1	Factor 2	Correlation	sig
edu	per	0.211	
edu	kno	0.657	***
edu	civ	0.764	***
per	kno	0.713	***
per	civ	0.727	***
kno	civ	0.924	***

From Table 3, it is oberved that most factors are highly positive correlated with each other since they have relatively high significance. For our survey, it measures the growth of an individual in a dimension, so the correlation means that individuals who showed growth in one dimension also showed growth in other dimensions.

Conclusions

Through data distribution, it shows that most of students have achieve growth in all dimensions. From the analysis of CFA, the survey can be improved by moving Q43 and Q24 from Education and Career Trajectory to Civic engagement. And, Q9 and Q22 could be considered to remove from the survey since they have relatively low factor loadings. To improve the RMSR and RMSEA of the fitted model, we will focus on data cleaning since too many missing values with replacement of the means may lead to inprecise and biased results.

References

1. Likert Scale Standardization: http://www-01.ibm.com/support/docview.wss?uid=swg21482329
2. Nicholas D. Myers, Soyeon Ahn, Ying Jin (2011). Sample Size and Power Estimates for a Confirmatory Factor Analytic Model in Exercise and Sport: A Monte Carlo Approach. Retrieved from Mplus website: https://www.statmodel.com/download/Myers,%20Ahn,%20&%20Jin,%202011.pdf
3. Ronald G. Downey & Craig V. King (1998) Missing Data in Likert Ratings: A Comparison of Replacement Methods, The Journal of General Psychology, 125:2, 175-191, DOI: 10.1080/00221309809595542
4. Fabrigar L. R., MacCallum R. C., Wegener D. T., Strahan E. J. Evaluating the use of exploratory factor analysis in psychological research. Psychological Methods. 1999;4(3):272–299. doi: 10.1037/1082-989X.4.3.272.
5. Hair Black Babin Anderson & Tatham, (2006) Multivariate Data Analysis, 6th Edition

Appendix

Summary of fitted model

```
summary(fitf, fit.measures=TRUE, standardized=TRUE)
```

```
## lavaan 0.6-3 ended normally after 95 iterations
##
##   Optimization method                           NLMINB
##   Number of free parameters                         60
##
##   Number of observations                            88
##   Number of missing patterns                         1
##
##   Estimator                                         ML
##   Model Fit Test Statistic                     232.960
##   Degrees of freedom                               129
##   P-value (Chi-square)                           0.000
##
## Model test baseline model:
##
##   Minimum Function Test Statistic              657.886
##   Degrees of freedom                               153
##   P-value                                        0.000
```

```
##
## User model versus baseline model:
##
##   Comparative Fit Index (CFI)                    0.794
##   Tucker-Lewis Index (TLI)                       0.756
##
## Loglikelihood and Information Criteria:
##
##   Loglikelihood user model (H0)              -1738.477
##   Loglikelihood unrestricted model (H1)      -1621.997
##
##   Number of free parameters                         60
##   Akaike (AIC)                                3596.955
##   Bayesian (BIC)                              3745.595
##   Sample-size adjusted Bayesian (BIC)         3556.260
##
## Root Mean Square Error of Approximation:
##
##   RMSEA                                          0.096
##   90 Percent Confidence Interval          0.076  0.115
##   P-value RMSEA <= 0.05                          0.000
##
## Standardized Root Mean Square Residual:
##
##   SRMR                                           0.085
##
## Parameter Estimates:
##
##   Information                                 Observed
##   Observed information based on                Hessian
##   Standard Errors                             Standard
##
## Latent Variables:
##                   Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
##   edu =~
##     fQ17             0.712    0.211    3.383    0.001    0.712    0.535
##     fQ20             0.474    0.225    2.104    0.035    0.474    0.312
##     fQ22             0.272    0.170    1.600    0.110    0.272    0.236
##   per =~
##     fQ7              0.306    0.100    3.068    0.002    0.306    0.361
##     fQ8              0.574    0.099    5.802    0.000    0.574    0.883
##     fQ9              0.079    0.068    1.161    0.246    0.079    0.142
##   kno =~
##     fQ28             0.467    0.058    8.027    0.000    0.467    0.753
##     fQ29             0.534    0.073    7.302    0.000    0.534    0.707
##     fQ34             0.547    0.086    6.338    0.000    0.547    0.631
##     fQ35             0.499    0.061    8.136    0.000    0.499    0.760
##     fQ37             0.648    0.176    3.681    0.000    0.648    0.396
##     fQ39             0.530    0.051   10.315    0.000    0.530    0.887
##   civ =~
##     fQ10             0.311    0.134    2.323    0.020    0.311    0.293
##     fQ11             0.427    0.094    4.543    0.000    0.427    0.570
##     fQ13             0.315    0.069    4.562    0.000    0.315    0.491
##     fQ24             0.418    0.081    5.157    0.000    0.418    0.562
```

```
##     fQ43              0.520    0.083    6.292    0.000    0.520    0.675
##     fQ45              0.316    0.071    4.447    0.000    0.316    0.500
##
## Covariances:
##                   Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
##   edu ~~
##     per              0.211    0.223    0.946    0.344    0.211    0.211
##     kno              0.657    0.197    3.334    0.001    0.657    0.657
##     civ              0.764    0.220    3.471    0.001    0.764    0.764
##   per ~~
##     kno              0.713    0.110    6.456    0.000    0.713    0.713
##     civ              0.727    0.140    5.190    0.000    0.727    0.727
##   kno ~~
##     civ              0.924    0.071   13.088    0.000    0.924    0.924
##
## Intercepts:
##                   Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
##    .fQ17             3.023    0.142   21.300    0.000    3.023    2.271
##    .fQ20             3.120    0.162   19.283    0.000    3.120    2.056
##    .fQ22             3.395    0.123   27.629    0.000    3.395    2.945
##    .fQ7              4.409    0.090   48.778    0.000    4.409    5.200
##    .fQ8              4.405    0.069   63.620    0.000    4.405    6.782
##    .fQ9              3.561    0.059   60.439    0.000    3.561    6.443
##    .fQ28             4.614    0.066   69.768    0.000    4.614    7.437
##    .fQ29             4.269    0.081   53.018    0.000    4.269    5.652
##    .fQ34             4.188    0.092   45.306    0.000    4.188    4.830
##    .fQ35             4.456    0.070   63.644    0.000    4.456    6.784
##    .fQ37             3.299    0.175   18.888    0.000    3.299    2.013
##    .fQ39             4.647    0.064   72.955    0.000    4.647    7.777
##    .fQ10             4.348    0.113   38.348    0.000    4.348    4.088
##    .fQ11             4.061    0.080   50.867    0.000    4.061    5.422
##    .fQ13             4.376    0.068   63.914    0.000    4.376    6.813
##    .fQ24             4.289    0.079   54.049    0.000    4.289    5.762
##    .fQ43             4.331    0.082   52.671    0.000    4.331    5.615
##    .fQ45             4.461    0.067   66.310    0.000    4.461    7.069
##     edu              0.000                              0.000    0.000
##     per              0.000                              0.000    0.000
##     kno              0.000                              0.000    0.000
##     civ              0.000                              0.000    0.000
##
## Variances:
##                   Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
##    .fQ17             1.265    0.302    4.187    0.000    1.265    0.714
##    .fQ20             2.080    0.348    5.970    0.000    2.080    0.903
##    .fQ22             1.255    0.200    6.276    0.000    1.255    0.944
##    .fQ7              0.625    0.099    6.333    0.000    0.625    0.870
##    .fQ8              0.093    0.096    0.968    0.333    0.093    0.220
##    .fQ9              0.299    0.045    6.587    0.000    0.299    0.980
##    .fQ28             0.167    0.029    5.714    0.000    0.167    0.433
##    .fQ29             0.286    0.049    5.832    0.000    0.286    0.501
##    .fQ34             0.452    0.073    6.167    0.000    0.452    0.602
##    .fQ35             0.182    0.032    5.668    0.000    0.182    0.422
##    .fQ37             2.264    0.348    6.500    0.000    2.264    0.844
##    .fQ39             0.076    0.018    4.130    0.000    0.076    0.212
```

```
##    .fQ10             1.034    0.163    6.353    0.000    1.034    0.914
##    .fQ11             0.379    0.076    4.966    0.000    0.379    0.675
##    .fQ13             0.313    0.050    6.274    0.000    0.313    0.759
##    .fQ24             0.379    0.065    5.876    0.000    0.379    0.684
##    .fQ43             0.324    0.064    5.038    0.000    0.324    0.545
##    .fQ45             0.299    0.050    6.022    0.000    0.299    0.750
##     edu              1.000                                1.000    1.000
##     per              1.000                                1.000    1.000
##     kno              1.000                                1.000    1.000
##     civ              1.000                                1.000    1.000
```