Project Report

Tim, Xiaofan, Yanwen, Jingning

11/20/2019

Introduction

Client is investigating how foreign language teachers feel about and utilize methods from the Teacher Effectiveness for Language Learning (TELL), and seeking advice about how to improving the survey.

Our purpose for our client in this project: 1. A lot of people don't answer the survey because it's long. Can we reduce the number of questions? 2. Is the survey currently answering the research questions?
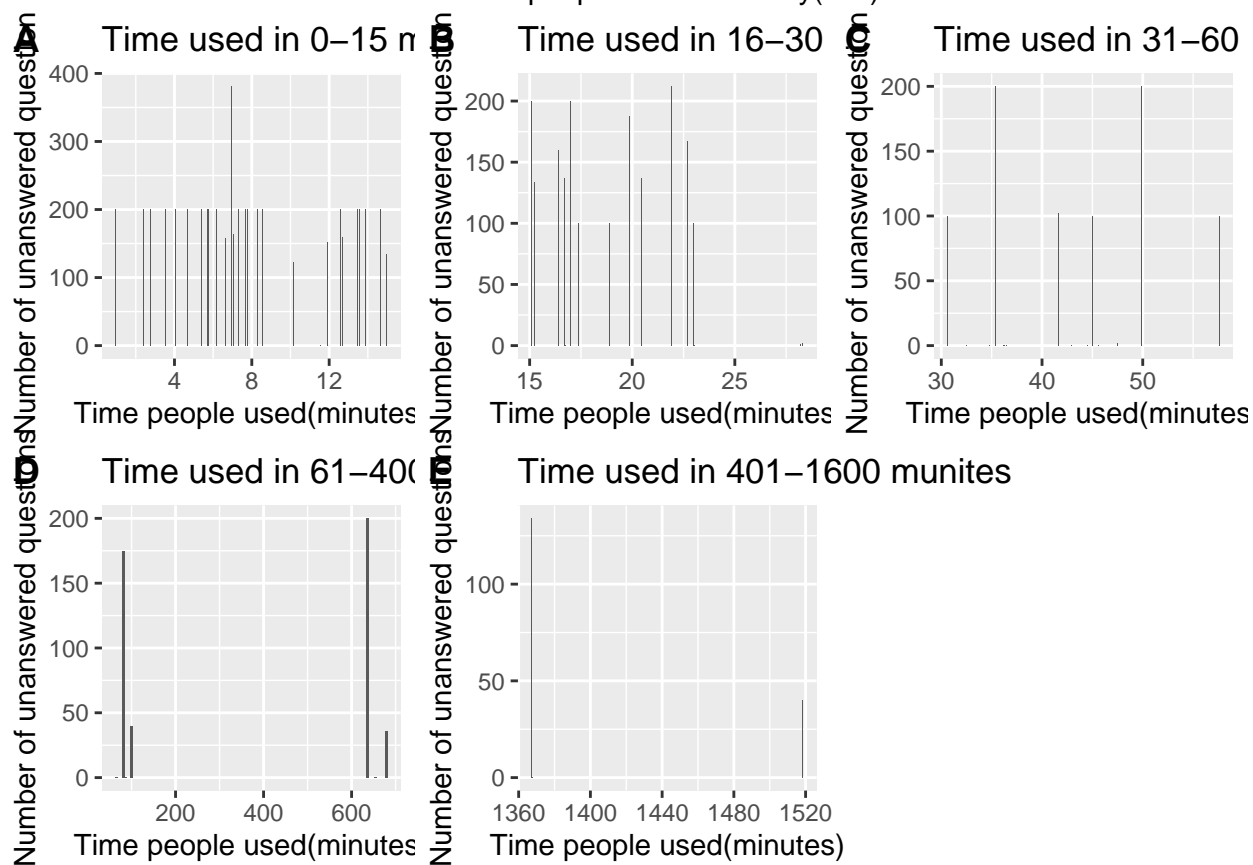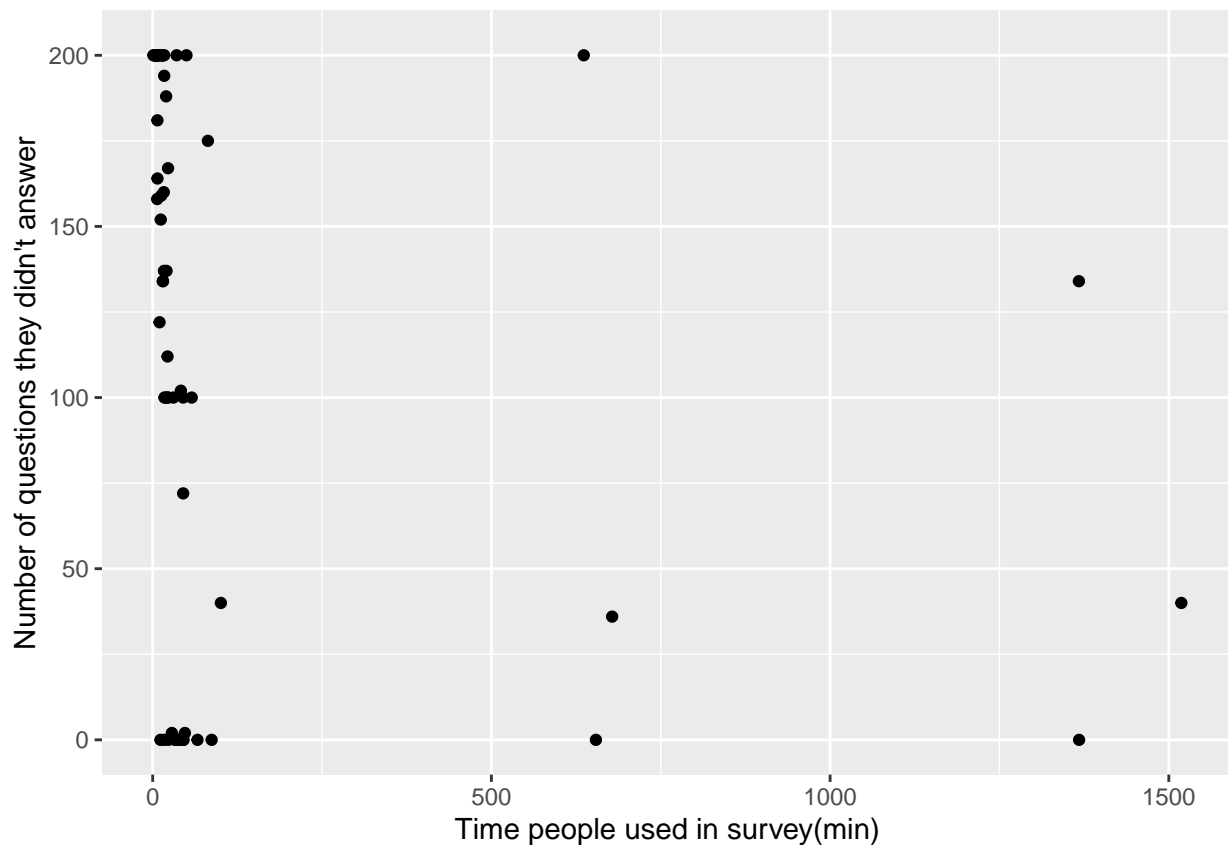
EDA & Conerns

Data Structure

We are provided the data in an excel file with 6 spreadsheets including one sheet of notes, one sheet of personal information and 4 sheets of Teacher Effectiveness for Language Learning (TELL) framework survey questions. The dataset of personal information contains questions regarding respondents' teaching language and education background. The dataset of Teacher Effectiveness for Language Learning (TELL) framework survey contains around 200 questions asking about respondents' attitudes of contribution and confidence towards each practice in the framework. There are 4 domains of Teacher Effectiveness for Language Learning (TELL) survey questions: planning, learning experience, learning tools, and performance & feedback. Each domain contains several subdomains and each subdomain contains a different number of questions. For this project, we focus on reducing the number of questions in the dataset of Teacher Effectiveness for Language Learning (TELL) framework survey.

EDA

We conduct a basic Exploratory Data Analysis (EDA) for this project. Firstly, We focus on the time for respondents to complete this survey.

Number of questions they didn't answer

Time people used in survey(min)

A Time used in 0−15 m

Number of unanswered questions

Time people used(minutes)

B Time used in 16−30

Number of unanswered questions

Time people used(minutes)

C Time used in 31−60

Number of unanswered questions

Time people used(minutes)

D Time used in 61−400

Number of unanswered questions

Time people used(minutes)

E Time used in 401−1600 munites

Number of unanswered questions

Time people used(minutes)

Data Cleaning

Concerns

We come up with several concerns after the initial EDA. Firstly, the observations we can use in the analysis are very limited since there are many N/A in the dataset. Secondly, some respondents seem like choosing the same answer through the whole survey and if we identify these answers as non-valid, then our sample size would become even smaller. With this limited sample size, the accuracy and reference of results from our subsequent analysis could be affected.

Method we used

We will use Confirmatory Factor Analysis (CFA) to reduce the survey questions number. CFA is a special form of factor analysis and mostly used in social science research. It is used to check whether measures of a construct are consistent with a researcher's understanding of the nature of that construct.
Here we will use CFA to see if there are survey questions equivalent to each other so we can reduce those repeated questions. We will analysis each subdomain separately and will only consider the problems regarding confidence or not. Within each subdomain, there will be several questions and our null hypothesis is that all survey questions are identical to each other. Then our alternative hypothesis is that the questions are not all equal.
We will focus on the p value result we have from CFA and we will take a p value larger than 0.05 to reject our null hypothesis. When we are not able to reject our null hypothesis, we will look at our factor loading to check the correlations between questions. Then we will fit new model by dropping question with lowest factor loading and see if we will reject our null hypothesis now. We will keep doing this until we have a subdomain with an acceptable p value, which gives us a set of survey questions are not identical to each other.

Analysis

Conclusion / Discussion

Appendix