CHUA WEN XIN B.Sc. (Honours) Statistical Computing and Operations Research

ENHANCING UTAR FRESHMAN MENTAL HEALTH SCREENING: INTEGRATING POWER AUTOMATE TO IMPROVE COUNSELLING SUPPORT

CHUA WEN XIN

BACHELOR OF SCIENCE (HONOURS) STATISTICAL COMPUTING AND OPERATIONS RESEARCH

FACULTY OF SCIENCE

UNIVERSITI TUNKU ABDUL RAHMAN

MAY 2025

ENHANCING UTAR FRESHMAN MENTAL HEALTH SCREENING: INTEGRATING POWER AUTOMATE TO IMPROVE COUNSELLING SUPPORT

By

CHUA WEN XIN

A project report submitted to the Department of Physical and

Mathematical Science

Faculty of Science

Universiti Tunku Abdul Rahman

in partial fulfilment of the requirements for the degree of

Bachelor of Science (Honours)

Statistical Computing and Operations Research

MAY 2025

ABSTRACT

ENHANCING UTAR FRESHMAN MENTAL HEALTH SCREENING: INTEGRATING POWER AUTOMATE TO IMPROVE COUNSELLING SUPPORT

CHUA WEN XIN

Depression may be the second most common global disease burden after HIV/AIDS in the year 2030. Approximately one million people in Malaysia aged 16 years and above suffer from depression. In alignment with ongoing efforts to promote mental health support, UTAR counsellors conducted mental health screening tests for freshmen using the Warwick-Edinburgh Mental Well-being Scale (WEMWBS) during orientation week. The WEMWBS test is designed to measure the mental well-being of individuals. However, the current screening process lacks immediate feedback or actionable suggestions for UTAR students. Additionally, there is uncertainty regarding the construct validity of the WEMWBS questionnaire in capturing the underlying psychological structure. Therefore, this research proposes an automated support system using Microsoft Power Automate to enhance UTAR freshman mental health screening procedures. The workflow may reduce the student's waiting period for support and the counsellors' workload by streamlining the management of student responses. Moreover, a series of multivariate analyses was conducted to gain deeper insight into the underlying structure of the well-being data. Particularly, Principal Component Analysis (PCA), Exploratory Factor Analysis (EFA), Confirmatory Factor Analysis (CFA), and Structural Equation Modelling (SEM) were applied to identify and validate the core dimensions of student mental well-being. A series of reliability analyses and convergent and discriminant validity were tested throughout the analysis. In conclusion, this research successfully achieved all the objectives by developing a working automation system and identifying key factors to student mental well-being. The EFA identified three underlying factors, namely Positive Emotion, Personal Growth, and Social Exploration. Among the three identified factors, Positive Emotion was the most influential. Promoting emotional positivity may most effectively enhance student well-being, offering a practical approach to improving university mental health support.

Keywords: Mental Health Screening, Power Automate, Multivariate Analyses

ACKNOWLEGEMENTS

Firstly, I would like to thank Dr. Lim Huai Tein, my final-year project supervisor. Her helpful guidance was significant in shaping the direction of this project and boosting my confidence in dealing with complex ideas and methods. She gave me the opportunity to explore and think about what I wanted to do for my final year project, and I really appreciate her trust. She had a really busy schedule, but she is still willing to spend her time on a physical discussion biweekly throughout two long trimesters. I enjoyed the time I spent with her. Apart from academics, I appreciate her kindness, understanding, and motivation, especially when I am desperate or overwhelmed. I enjoy the chit-chat moments after our discussion whenever we meet. Her belief in my potential has meant a lot to me, and her guidance has dramatically impacted my studies and personal development. I am thankful for the opportunity to learn from her guidance, and I will never forget the lessons and values I have gained from this experience.

Secondly, I want to express my gratitude to the UTAR Centre for Healthy Minds and Wellbeing and the university's counselling team for their generosity in providing me with the dataset and for sharing crucial insights about UTAR's mental health practices. I am incredibly grateful for their trust in allowing me to work with actual student data, which added depth to the research and motivated me to carry out the project with greater responsibility and care.

Thirdly, I appreciate Universiti Tunku Abdul Rahman (UTAR) for providing the tools and platforms that made this research possible. Microsoft Power Automate was crucial for creating the automated system, and the resources and digital databases from the UTAR library were super helpful for the literature review and analysis. I appreciate how the school is dedicated to giving us the resources to explore, innovate, and succeed in our studies.

Lastly, I want to express my gratitude to all my lecturers, friends, and family. Their trust, encouragement, and emotional support have helped me a lot during my final year. Their support, whether it was through academic advice, encouraging words, or just being there during hard times, really helped me overcome the challenges.

DECLARATION

I hereby declare that the project report is based on my original work except for quotations and citations which have been duly acknowledged. I also declare that it has not been previously or concurrently submitted for any degree at UTAR or other institutions.

CHUA WEN XIN

APPROVAL SHEET

This project report entitled "**ENHANCING UTAR FRESHMAN MENTAL HEALTH SCREENING: INTEGRATING POWER AUTOMATE TO IMPROVE COUNSELLING SUPPORT**" was prepared by CHUA WEN XIN and submitted as partial fulfilment of the requirements for the degree of Bachelor of Science (Honours) Statistical Computing and Operations Research at Universiti Tunku Abdul Rahman.

Approved by:

____

(Dr. Lim Huai Tein) Date: 23 April 2025

Supervisor

Department of Physical and Mathematical Science

Faculty of Science,

Universiti Tunku Abdul Rahman

FACULTY OF SCIENCE

UNIVERSITI TUNKU ABDUL RAHMAN

Date: 23 April 2025

PERMISSION SHEET

It is hereby certified that **CHUA WEN XIN** (ID No: **21ADB02307**) has completed this final year project entitled "**ENHANCING UTAR FRESHMAN MENTAL HEALTH SCREENING: INTEGRATING POWER AUTOMATE TO IMPROVE COUNSELLING SUPPORT**" under supervision of Dr. Lim Huai Tein (Supervisor) from the Department Physical and Mathematical Science.

I hereby give permission to the University to upload the softcopy of my final year project in pdf format into the UTAR Institutional Repository, which may be made accessible to the UTAR community and public.

Yours truly,

____

(CHUA WEN XIN)

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS

Symbols

H_1	Alternative hypothesis
χ^2	Chi-square statistic
R^2	Coefficient of determination
df	Degrees of freedom
ε_i	Error term for observed variable i
R^2	Coefficient of determination
H_0	Null hypothesis
N	Sample size
λ_i	Standardised factor loading for item i
β_i	Standardised regression coefficient
Res_i	Residual of latent variable i
$Var(\varepsilon_i)$	Variance of the error term for observed variable i

LIST OF ABBREVIATIONS

Abbreviations

AHP	Analytic Hierarchy Process
AI	Artificial Intelligence
AIIC	Average Inter-Item Correlation
AIDS	Acquired Immunodeficiency Syndrome
AMOS	Analysis of Moment Structures
ANOVA	Analysis of Variance
AVE	Average Variance Extracted
BMI	Body Mass Index
CFA	Confirmatory Factor Analysis
CFI	Comparative Fit Index
CHMW	Centre for Healthy Minds and Wellbeing
CR	Composite Reliability
CS	Counselling Services
DASS-21	Depression Anxiety Stress Scales (21 items)
EEG	Electroencephalogram
EFA	Exploratory Factor Analysis
GAD-7	Generalized Anxiety Disorder Scale (7 items)
GFI	Goodness of Fit Index
HIV	Human Immunodeficiency Virus
HTML	HyperText Markup Language
HTMT	Heterotrait-Monotrait Ratio

ID	Identification
IPAQ-SF	International Physical Activity Questionnaire – Short Form
KMO	Kaiser-Meyer-Olkin Measure
L	Loading (Factor Loading)
LOOCV	Leave-One-Out Cross-Validation
MDD	Major Depressive Disorder
ML	Maximum Likelihood
MLR	Maximum Likelihood with Robust Standard Errors
NFI	Normed Fit Index
PAF	Principal Axis Factoring
PCA	Principal Component Analysis
PCC	Pearson Correlation Coefficient
PHQ-9	Patient Health Questionnaire (9 items)
RFI	Relative Fit Index
RMSEA	Root Mean Square Error of Approximation
SEM	Structural Equation Modelling
SPSS	Statistical Package for the Social Sciences
SVM	Support Vector Machine
TLI	Tucker-Lewis Index
UKM	Universiti Kebangsaan Malaysia
UTAR	Universiti Tunku Abdul Rahman
WEMWBS	Warwick–Edinburgh Mental Well-being Scale

CHAPTER 1

INTRODUCTION

1.1 Research Background

Mental well-being refers to a person's ability to cope with life's stresses, build positive relationships, have a sense of purpose, and realise their full potential. In addition to preventing mental illness, well-being encompasses positive emotional, psychological, and social functioning and is a crucial component of overall mental health (Ruggeri et al., 2020; Huppert, 2009). Subjective well-being is the same as positive mental health (Ruggeri et al., 2020).

University environments can significantly influence well-being and academic achievement. Saidi et al., (2024) emphasises that supportive environments that enhance mental health awareness and coping strategies are linked to improved academic performance for students. Those who get early mental health support and develop better stress management skills tend to stay more engaged in their studies and achieve higher performance in university. The results indicate that focusing on student well-being is good for emotional health and plays a key role in improving educational success.

In contrast, low levels of mental well-being are often linked to a higher risk of depression. Li, Xia and Zhang (2023) indicate that individuals who report feeling less happy or having lower subjective well-being frequently exhibit more symptoms of depression. Next, the National Health and Morbidity Survey (2022) shows that ¼ of teenagers have faced depression, ⅛ have thought about suicide, and ⅒ have attempted suicide. In the following year, the National Health and Morbidity Survey (2023) reported that one million Malaysians aged 16 and above are currently experiencing depression, a figure that has doubled since 2019. The number of people suffering from depression has increased over the years. Depression is expected to be the second major leading of the global disease burden after HIV/AIDS by 2030 (Hock et al., 2012; Salleh, 2018; Peng et al., 2023; Tamil et al., 2023). This suggests that encouraging mental well-being could serve to prevent depression in university students.

To support this, many universities have adopted mental health screening tools to identify students in need and provide help earlier. For example, Universiti Kebangsaan Malaysia (UKM), a public university in Malaysia, is using the Depression Anxiety Stress Scale-21 (DASS-21) questionnaire to assess the mental health of students (Ismail and Kahwa, 2020). Besides, to investigate the student's mental condition before and after the COVID-19 pandemic, students from Universiti Tunku Abdul Rahman (UTAR) Sungai Long campus are required to complete General Anxiety Disorder–7 (GAD-7), Patient Health Questionnaire-9 (PHQ-9), and International Physical Activity Questionnaire–Short Form (IPAQ-SF) via Google Forms (Mir et al., 2023). Furthermore, in European countries' view, a Slovenia university utilised the Warwick-Edinburgh Mental Well-being Scale (WEMWBS) to measure student mental well-being (Cilar, Pajnkihar and Stiglic, 2020). These mental health screening tools can detect signs of distress and help provide support before things worsen (Jurewicz, 2015; Yulia et al., 2021). Research indicates that when screening is accompanied by adequate support, it can enhance students' mental well-being (Leventhal, Brissette and Leventhal, 2003; Sikorski et al., 2021).

After mental health screening, counsellors play a vital role in analysing the student's mental health condition. In Malaysia, university counselling services are crucial to improve students' quality of life. According to Thuryrajah and Jeyakumar (2017), university counselling interventions significantly impact students’ personal and educational development. However, the analysis from Arifin et al., (2022) indicates that most of the respondents from public universities in Selangor still have a negative attitude toward attending counselling. This resistance is often influenced by cultural norms. In Malaysia, students usually avoid counselling due to the cultural belief of "losing face", which makes seeking help a sign of weakness. This fear can lead to shyness and damage to self-image (Ma, Zhu and Bresnahan, 2021; Loong et al., 2024).

Thus, to make the counselling process more manageable for the student to step out, universities can help by providing private, accessible resources, reducing discomfort, and increasing confidence in seeking help. One way to help with this is by using easy and private screening tools, so students can reflect on how they’re doing without feeling pressured to talk to someone right away. At UTAR Kampar campus, the Warwick-Edinburgh Mental Well-being Scale (WEMWBS) is used as the primary tool to evaluate students’ mental well-being. WEMWBS measures a variety of concepts related to mental well-being, including affective-emotional aspects, cognitive-evaluative dimensions, and psychological functioning. The emphasis is on the positive aspects and aims to promote mental health support (Tennant et al., 2007; Cilar, Pajnkihar and Stiglic, 2020).

Although WEMWBS is a helpful tool for understanding students' well-being, there is currently no structured system in place at UTAR that provides students with instant feedback or follow-up after completing the screening. In most cases, the process ends once students submit the form, and only those identified as high risk may be contacted by a counsellor later. This will lead to many students may be left without support, particularly those who are scared to ask for help or are unsure of how they feel. As a result, the opportunity for early intervention may be missed, even though the screening data is already available.

1.2 Problem Statement

In practical, UTAR uses the WEMWBS to assess student well-being, however, the current process lack of immediate feedback after screening. Once the form is submitted, the process becomes manual as students must wait to be contacted by a counsellor, which a counsellor's feedback may merely depend on the availability and case severity. According to Leventhal et al., (2003) and Sikorski et al., (2021), provide efficient response or feedback directly to individuals will enhance their engagement in help-seeking. However, without immediate feedback, students may be less likely to seek help as they are unaware of available resources or do not know how to take the next action. Furthermore, when students do not know where to get counselling resources, their concern about losing face with their peers is one of the significant barriers to their approach to help from counsellors (Ma, Zhu and Bresnahan, 2021;

Loong et al., 2024). Therefore, they are passively or not likely to ask for resources from their peers.

From the perspective of UTAR counsellors, the WEMWBS scores may not clearly indicate which specific aspects of mental well-being to focus on. Although the scores can highlight general areas where students may be struggling, the lack of specificity limits the ability to provide targeted and effective early interventions. Recent research emphasizes the importance of early awareness as counsellors must be able to evaluate students' psychological difficulties clearly to provide timely and practical support. Without this, students may struggle to receive help at the right time (Cong et al., 2024). When counsellors respond proactively and align their strategies with student needs, mental health outcomes improve through a more collaborative and supportive environment (Wahyuni et al., 2024).

1.3 Research Questions

1. How can students' feedback be provided promptly and automatically so that appropriate support can be delivered based on their well-being screening results?
2. What are the key factors that represent the well-being of UTAR freshmen based on their WEMWBS responses?
3. How well does the proposed measurement model accurately fit the data ~~and~~ to reflect student well-being?
4. What is the relationship between the identified underlying well-being factors?

1.4 Research Objectives

The main objective of this research is to develop an automated workflow that provides personalized support in an effective manner. Moreover, the key underlying factors are explored in the WEMWBS for student well-being through multivariate statistical analysis. By combining both automation and data insights, this research seeks to offer practical recommendations for improving mental health services in the university setting. The specific objectives are listed as below:

1. To generate an automated mental well-being system for UTAR freshmen. The system may be used for well-being screening and provide supportive resources.
2. To identify the underlying factors of mental well-being by using Principal Component Analysis (PCA) and Exploratory Factor Analysis (EFA).
3. To validate the measurement model of the WEMWBS using Confirmatory Factor Analysis (CFA).
4. To examine the relationship between the identified underlying well-being factors using Structural Equation Modelling (SEM).

1.5 Limitation of the Research

Although this research explores the potential for providing prompt assistance through automation and the underlying structure of UTAR students' well-being. This research included only freshmen students from UTAR. The results may not be entirely applicable to all education level of university students.

The WEMWBS test is reliable and valid, but it only looks at the positive aspects of mental health such as the condition of optimism, happiness, and social connection. Consequently, other aspects of mental health may be omitted, for example stress, anxiety, and depression.

Besides, the research evaluates the mental well-being of freshmen at a single point in time. Thus, the data collected may limit the comprehension of how their mental well-being evolves throughout their university life and in response to academic challenges.

1.6 Significance of the Research

This research contributes to mental health intervention strategies and technological applications in higher education. Practically, it introduces an automated screening-response workflow using Microsoft Power Automate. It is expected to deliver instant and personalised feedback to students after they complete the WEMWBS questionnaire. This helps bridge the standard gap between screening tests and social support, thus making early intervention more accessible and less dependent on manual follow-up.

Academically, the study applies multivariate statistical techniques such as PCA, EFA, CFA, and SEM to explore the latent structure of student well-being based on the WEMWBS. Pinpointing important underlying factors gives UTAR counsellors a more precise direction for focusing on the interventions. This addresses a known limitation that the total well-being scores often lack exploration, resulting in unclear areas of student struggling.

CHAPTER 2

LITERATURE REVIEW

This chapter reviews key topics connected to the research, such as measuring mental well-being with the Warwick–Edinburgh Mental Well-being Scale (WEMWBS), utilising Microsoft Power Automate and applying multivariate analysis techniques. This chapter explores how Power Automate can help mental health systems by offering quick and automated responses. Since WEMWBS is commonly used to evaluate positive mental well-being and serves as the foundation for the screening tool in this research, this paper discusses multivariate methods like Principal Component Analysis (PCA), Exploratory Factor Analysis (EFA), Confirmatory Factor Analysis (CFA), and Structural Equation Modelling (SEM) to explore the structure and the underlying factors.

2.1 Warwick-Edinburgh Mental Well-being Scale (WEMWBS)

The Warwick–Edinburgh Mental Well-being Scale (WEMWBS) was developed in 2006 by a panel of experts in the United Kingdom. They constructed it on a review of existing literature, discussions in focus groups, and an earlier measure known as Affectometer 2. The main objective was to develop a straightforward, positively framed questionnaire to evaluate mental well-being in the general population (Tennant et al., 2006; Tennant et al., 2007; Stewart-Brown et al., 2011; Stewart-Brown, 2016; Konaszewski et al., 2021).

WEMWBS quantitatively assesses mental well-being through hedonic and eudaimonic well-being. Hedonic well-being emphasises happiness and satisfaction, while eudaimonic well-being encompasses psychological functioning and self-realisation (Tennant et al., 2006; Fung, 2019).

WEMWBS comprises 14 items, each rated on a 5-point Likert scale, resulting in a total score between 14 and 70, where higher scores reflect improved overall mental well-being (Santos et al., 2015). The complete set of WEMWBS questionnaires can be found in Appendix A (Warwick Medical School, 2020).
WEMWBS is freely accessible. However, researchers and practitioners must obtain a free licence before using through an application process (Koushede et al., 2019). The scale has been validated internationally, including in countries like Brazil (Santos et al., 2015), Sri Lanka (Perera et al., 2022), Norway (Smith et al., 2017), and Denmark (Koushede et al., 2019), showing that it is relevant across different cultures.

WEMWBS has been utilised in Malaysia to evaluate mental well-being across different groups. A study at the International Islamic University Malaysia focused on students' resilience and self-stigma, demonstrating strong reliability and suitability for the local context (Thartori and Nordin, 2019). During the COVID-19 Movement Control Order (MCO), WEMWBS was used to assess well-being among couples, showing high internal consistency and effectively tracking changes in emotional well-being (Chua et al., 2021).

The findings show that WEMWBS is a popular and flexible tool for assessing mental well-being in different cultures, including Malaysia. The regular use of this method in school and practical situations shows that it is a good fit for assessing well-being in university settings, setting the stage for more investigation in this research.

2.2 Multivariate Analysis

Multivariate analysis is a statistical technique used to look at several variables simultaneously. This helps researchers find patterns, relationships, and structures in the variables of the datasets. Multivariate analysis differs from univariate and bivariate analyses because it examines multiple variables simultaneously instead of focusing on just one or two (Mwange, Chiseyeng'i, and Matoka, 2023). This makes it helpful for looking into psychological concepts such as mental well-being.

Some common techniques for multivariate analysis are Principal Component Analysis (PCA), which helps reduce dimensions, Exploratory Factor Analysis (EFA) that uncovers underlying factor structures, Confirmatory Factor Analysis (CFA) used to evaluate measurement models, and Structural Equation Modelling (SEM) that analyzes relationships among latent variables.

Multivariate analysis is categorized into dependency techniques and interdependency techniques. Dependency techniques like regression and SEM are used when some variables depend on others. The goal is to explain or predict what happens next. On the other hand, interdependency techniques like PCA and EFA group variables based on shared variance without labelling any as dependent, making them helpful in exploring underlying structures (Mwange, Chiseyeng'i, and Matoka, 2023).

2.3 Power Automate

Robotic Process Automation (RPA) is a new technology that enables people to automate humans performing rules-based tasks or routines easily. These tasks include repetitive actions like data entry, file moving, and other jobs behind the scenes. RPA bots can perform those repetitive tasks faster than humans, and precision can be guaranteed. Thus, it reduces errors and improves productivity when completing these repetitive jobs (Afrin, Roksana and Akram, 2024).

RPA software has become popular in business due to its cost-effectiveness. It included AutomationEdge, UIPath, Microsoft's Power Automate (van der Aalst, Bichler and Heinzl, 2018; Kannan and Chin, 2022). However, this research will only focus on Microsoft Power Automate. Microsoft Power Automate is an online automated solution developed by Microsoft to help users optimise business processes through low-code and Artificial Intelligence (AI). It enables individuals and organisations to create automated workflows to connect various applications and services, including Microsoft Teams, SharePoint, Outlook, Microsoft Forms, Power BI, and many other tools (Microsoft, n.d.). According to Microsoft (2025), Microsoft Power Automate is a new technology published in April 2021. Workflows or Flows are the final outputs in Power Automate. There are two kinds of components in flow which are triggers and actions. Every flow begins with a trigger. After the flow is triggered, it will proceed through one or more actions. Both the triggers and the actions need to connect to other systems that are beyond Power Automate. The type of flow depends on the trigger type, and it consists of three types of flow, which are automated cloud flow, instant cloud flow, and scheduled cloud flow (Ding, 2023).

Microsoft Power Automate is becoming increasingly popular across fields like business and healthcare due to its wide range of automation capabilities.

In the healthcare field, Kannan and Chin (2022) proposed that Power Automate was used to create an automated data backup system for Malaysia vaccination records gathered via Google Forms. After respondents filled out the form, their information was saved in a Google Sheet. Power Automate was set up to pull the responses from the Google Sheet at regular times to ensure the data remains accurate and consistent. After extraction, the flow automatically generated a new Excel file in Microsoft OneDrive, with the latest data as a backup copy. The layout of this Excel file is the same as the original response. This ensures that all the information is easy to find and look over whenever needed. The backup process was set up to run automatically, which made it less likely that data would be lost due to mistakes, form restarts, or sudden system problems. After automation, the data is stored on the MyVAS servers, which are the government's official vaccination record systems. The automation utilized different connectors in Power Automate to effectively connect Google Forms and Sheets with Microsoft's OneDrive storage, eliminating the need for manual downloads or uploads. This implementation shows how Power Automate can work across different platforms and strengthen healthcare data practices by offering a secure, cloud-based backup strategy.

Furthermore, Baggyalakshmi, Dhanya, and Revathi (2024) also show how Power Automate can be used in the human resources (HR) area. Their study suggests a Power Automate solution that makes the employee onboarding process easier and faster by reducing repetitive manual tasks in a business setting. It begins by gathering responses from new employees using Microsoft Forms, which automatically records their information upon submission of the form. After that, Power Automate creates customized onboarding kits that contain documents such as welcome letters, policies, and various HR materials. The documents automatically fill in employee details and can be saved as PDFs or in other formats. The system checks the product or stock quantity, if applicable, and sends real-time email notifications to HR regarding the current or low-stock status. Also, managers and IT support will get notifications regarding what they need to do to prepare for the new hire. Reminders and tasks get set up automatically to help everyone stay on track. The overall structure is set up to integrate with HR systems while ensuring data privacy is maintained. This solution automates tasks such as kit creation, notification, approval, and delivery. This makes things more consistent and helps tackle common onboarding issues like delays, errors, and communication gaps. This workflow highlights how essential low-code tools like Power Automate are becoming in actual HR processes.

Microsoft Power Automate remains relatively new, with limited studies investigating its applications within psychology and other social science disciplines. This indicates that significant discoveries remain, and this research aims to develop the potential of automation to enhance mental health workflows within a university context.

2.4 Principal Component Analysis (PCA)

In this significant data era, efficiently handling large-scale multidimensional data has become more challenging. Principal Component Analysis (PCA) is an important method for handling large amounts of data, especially for extracting and simplifying data. This is a strong statistical tool for reducing the number of dimensions in data. The main objective of PCA is to transform data from a high-dimensional space to a lower-dimensional space and retail as much information as possible. This transformation helps in better understanding the relationships between the new variables and the original variables. In other words, principal component analysis focuses on transforming the original variables into fewer composite variables. The composite variables are also known as principal components. Principle components are the original variables' linear combinations and are uncorrelated (Li and Qin, 2024).

In the Fernández-Cardero et al., (2024) study, PCA was utilised to identify the primary factors contributing to overweight and obesity. The PCA reduced 26 observed variables into seven principal components, which explained 81.5% of the total variance in the dataset. This reduction highlights the efficiency of PCA in simplifying complex datasets while retaining most of the information. The first principal component was the Caloric and lipid factor, which explained 25.9% variability. This component was mainly related to nutrition, emphasising the function of lipids and energy in obesity. The second principal component was the adiposity factor at 16.9%, and the third Cardiometabolic risk factor explained 13.2% of the variability. The results indicate that animal-derived, high-fat diets are associated with obesity, whereas plant-based protein sources (legumes, nuts, seeds) are linked to better health outcomes.

In the financial market, Shah et al., (2021) used PCA to create a cryptocurrency index that adapts dynamically to market movements. Over 911 days, the researchers examined daily market value statistics for 1200 cryptocurrencies. This showed that the dataset consisted of 1200 variables, each corresponding to a different cryptocurrency. The study analysed historical market data and found that PC1 alone was sufficiently representative of overall market trends, capturing over 94% of the cryptocurrency price variance. The PCA-based index was proven reliable as a financial analysis tool using a three-factor pricing model that included market, size, and momentum factors. The study showed that PCA can find hidden market patterns, lower market noise, and make adaptive financial indices that work better than standard market capitalisation-weighted indexes.

In the psychological field, Merkulova et al., (2023) utilised PCA in distinguishing between healthy individuals and those with major depressive disorder (MDD) by analysing data from psychological questionnaires and EEG spectral characteristics. The psychological questionnaire initially consisted of 22 variables. Through the application of PCA, these were reduced into 10 principal components. PC1 reflects emotional stability and negative affect tendencies, making it a key factor of MDD status. The top six principal components accounted for over 95% of the total variance, confirming PCA validity in extracting relevant psychological features. Simultaneously, PCA was also applied to EEG spectral data. The PC1 shows delta and theta power in frontal and central regions as depression indicators, and the PC2 capture the alpha power variability for emotional regulation differences. By combining the PCA-reduced data from the questionnaires and EEG features, the study significantly enhanced the classification accuracy of MDD by up to 85% by using Support Vector Machines (SVM).

To increase the reliability of Principal Component Analysis (PCA), cross-validation is essential. According to Wilimitis and Walsh (2023) and Lumumba et al., (2024), cross-validation is a statistical modelling and machine learning technique that focusses on evaluating how the outcomes of a statistical study will compute to an independent dataset. This strategy is used to prevent overfitting, a problem in which the model over-captures noise instead of the true underlying pattern in the data. By splitting the data into training and testing sets, cross-validation helps ensure that the models perform well on the unseen data, thus achieving a balance between bias and variance, leading to more robust model validation. Lumumba et al., (2024) also mentioned that there consists of several types of cross-validation, including Leave-One-Out Cross-Validation (LOOCV), k-folds Cross-Validation, and Repeated k-folds Cross-Validation. This study underscores that, k-folds Cross-Validation and Repeated k-folds Cross-Validation are generally efficient.

2.5 Exploratory Factor Analysis (EFA)

Exploratory factor analysis (EFA) is a multivariate statistical method for identifying the factors that explain the variation in participants' responses to survey instruments, such as Likert-type scale questionnaires and tests. EFA investigates the relationships among a set of variables without prior expectations or assumptions about their structure. This method is generally applied in cases when the underlying factors influencing responses are not predetermined (Phakiti, 2018). Phakiti (2018) claims that two main reasons support the use of EFA in studies. First, EFA simplifies a dataset by reducing a large number of interrelated variables into fewer, more manageable factors, which allows researchers to analyze data more efficiently. Items strongly influenced by the same latent factor can be combined to create composite scores representing broader constructs. Second, Brown (2015) and Phakiti (2018) stated that EFA can help to obtain evidence that the theoretical constructs of interest are true in both convergent and discriminant validity. When it comes to validity, convergent validity is how closely items that measure the same construct are related, where the items that are highly linked will be those that are influenced by the same underlying factor. In contrast, discriminant validity is how closely items that measure different constructs are related. Results from EFA can be used as part of the development and validation of instruments.

In Appendix B, a figure illustrates the conceptual differences between EFA and PCA, which were discussed in the previous section (Phakiti, 2018). EFA uses arrows to indicate the latent factor, which is assumed to cause observed responses and focuses on finding latent structures that explain shared variance among observed variables. Conversely, PCA treats components as linear combinations of observed variables without assuming an underlying causal factor. It transforms observed variables into new uncorrelated components for data reduction without modelling latent constructs or measurement errors (Phakiti, 2018).

According to Omura et al., (2022), EFA is one of the most common ways to use statistics in psychology research. By referring to Appendix C, psychologists use a questionnaire that lists behaviours like drinking alcohol, eating breakfast, smoking, being active, and using illegal drugs. The questionnaire asks respondents things like how often they drink alcohol, eat breakfast, or smoke. Once the data from questionnaires has been collected, EFA groups these variables and finds the number of latent factors that cause people to act in specific ways. In this example, EFA divides five behaviours into two groups, and based on the factor loading, two latent factors have been extracted. By looking at the actions in each group, psychologists deduced that latent factor 1 is health-risking and latent factor 2 is health-promoting mindsets.

To further highlight the use of EFA in psychological research, another example is provided by Brooks et al., (2022), who applied EFA to examine the underlying structure of depressive symptoms. According to Brooks et al., (2022), Exploratory Factor Analysis (EFA) was used to examine the underlying factor structure of the Center for Epidemiological Studies Depression Scale (CES-D) among 150 American Indian women aged 18 to 50. After testing models with one to five factors, the researchers found that the four-factor answer best suited the data, accounting for 61.1% of the variance. The four latent factors identified were conceptually labelled as distracted thoughts, sadness, mixed affect, and interpersonal conflict (refer to Appendix D).

To better understand consumer decision-making in Thailand's last-mile delivery services, Rattanakijsuntorn (2023) showed how Exploratory Factor Analysis (EFA) and Analytic Hierarchy Process (AHP) can be combined. EFA was used to find and group the relevant factors that affect customer preferences, which led to the extraction of ten important factors. They were divided into two groups, where the first group was Primary Performance, which included price, quality, time, closeness, safety, and visibility whereas the other group was Secondary Performance, which included innovation, environment, approachability, and extra services. Furthermore, AHP was used to rank these elements following EFA, showing that price, quality, and time were the most important. The combined usage of EFA and AHP to offer both structure and prioritising in decision-making analysis is underlined in this paper.

2.6 Confirmatory Factor Analysis (CFA)

After performing the Exploratory Factor Analysis (EFA), it is important to validate the factor structure using Confirmatory Factor Analysis (CFA). Confirmatory Factor Analysis (CFA) is a process that examines constructs or measurement models to evaluate how well the items represent their underlying constructs. It helps to assess both the reliability and construct validity of the measurement model. The CFA results perform the factor loading and fitness indexed for each construct item and its residual. This method provides the simultaneous correlations between the constructs, as they were linked together during the “covariance” procedure (Nizar et al., 2019; Haron@shafiee, Abd Halim and Ismail, 2023). Therefore, the analysis starts by linking the constructs and identifying their factor loadings and correlations (Haron@shafiee, Abd Halim and Ismail, 2023). The measurement model consisted of the relationship between factors and items in the questionnaire. Appendix E shows an example of a measurement model with two factors (Sarmento and Costa, 2019).

Although both CFA and EFA are closely related, they serve different purposes. EFA is used to explore the underlying structure of a set of items without a predefined model, while CFA is employed to confirm whether the proposed structure fits the data and supports the theoretical assumptions. Thus, CFA requires specification of which variables load on each factor and the specific number of factors (Sarmento and Costa, 2019).

According to Sarmento and Costa (2019), CFA is mostly used in social research. In educational tourism, CFA was used to confirm a measurement model that includes six key constructs, which include tour operators, event management, local communities, educational institutions, investment, and tourism organisations. The researcher collected data for the study from 384 respondents, all of whom are tourists visiting the Edu-tourism destinations in Terengganu. It was emphasised how important the roles of educational institutions and local communities are. This shows that strong relationships between education providers and community stakeholders can greatly assist the growth of educational tourism. The CFA findings showed that these constructs capture the important areas needed for sustainable growth in educational tourism, helping investors make more focused strategic decisions (Haron@shafiee, Abd Halim and Ismail, 2023).

Riana and Syamsudin (2024) studies focus on parents' dominant factors in selecting an ideal kindergarten for their kids by using CFA. This study collected data from 154 parents using a structured questionnaire that had 22 items. The CFA results confirmed five dominant factors which included, school excellence, social development, teacher professionalism, education costs, and language development, together explaining 60.9% of the total variance. Among these, good social development was found as the leading factor, with a factor loading of 0.777, where the least dominant factor among these five factors is educational cost. The validated CFA model showed that choosing a school for kids involves many factors, considering both academic and non-academic aspects. As a result, schools can use these findings to focus on enhancing the curriculum and school discipline, making sure they meet what parents expect.

2.7 Structural Equation Modelling (SEM)

Structural Equation Modelling (SEM) is a statistical method used to examine the relationships between observed and latent variables (Meydan and Şen, 2011; Civelek, 2018). Observed variables can be interpreted as the measured variables in the data collection process, like the questionnaire items, where latent variables cannot be directly measured and are measured by connecting them to the observed variables (Civelek, 2018). Latent variables should be represented by more than one observed variable as they represent abstract concepts (Civelek, 2018). To examine the accuracy of the conceptual model, SEM consists of a measurement model and a structural model. The measurement model needs to be tested before proceeding to the structural model. The measurement model measures how well the observed variables represent the latent variables. The measurement model mainly involves CFA and the construct validity of scales. Thus, the structural model will not be tested if the measurement model fit indices are low (Dursun and Kocagöz, 2010; Civelek, 2018).

According to Morrison, Morrison and McCutcheon (2017), SEM is similar to a few statistical techniques like analysis of variance (ANOVA) and multiple regression, as SEM often assumes linear relationships. The difference between SEM and other general linear models is that SEM can use multiple measures to represent the construct variables and address the issue of measurement-specific error. In contrast, the other general linear models cannot account for measurement error, and their constructs are only represented by one measure.
SEM is commonly used in psychological and educational research because it allows researchers to simultaneously test theoretical models that include several latent constructs and their relationships. By combining measurement and structural models, SEM helps researchers validate their tools and examine how latent psychological factors interact (Civelek, 2018).

SEM is often used in psychology and education, but researchers have found it helpful in healthcare and business. A study looked into how diabetes and depression are connected by using SEM to investigate how depression could affect diabetes outcomes through self-care behaviours. This study looked at depression as a hidden factor that was assessed through various symptoms, including a lack of interest in activities, feelings of sadness or hopelessness, and difficulties with sleep. SEM was helpful because these symptoms could not be quantified easily with a single question. The researchers also examined how well patients got their medication and how often they checked their blood glucose and haemoglobin A1C levels. Their model indicated that depression did not directly affect diabetes control, but it did have an indirect effect by diminishing self-care behaviours. This study showed how SEM allows researchers to understand the indirect links between mental and physical health (Ronaldson et al., 2020).

In a different scenario, SEM was utilized in a business study that explored how internal management practices influence the stability of manufacturing companies. The researchers created a model with three hidden variables: operating effectiveness, risk management, and company stability. Several variables were used to measure these concepts, such as on-time reporting, resource efficiency, following operational processes, internal control systems, risk assessment efforts, revenue growth, cost control, and sustainability. Since these concepts are complicated and not easily quantifiable, SEM effectively integrated the observed data into significant latent variables. The results indicated that having solid risk management and efficient internal operations played a significant role in keeping the business stable, highlighting the importance of SEM in examining theoretical models within organizations (Vo and Nguyen, 2023).

Next, Austin et al., (2020) used SEM in healthcare to explore whether survivorship care plans (SCPs) effectively enhanced health outcomes for cancer survivors. The researchers examined a structural model that featured SCP receipt as an observed variable and various latent variables like patient-centred communication (PCC), health self-efficacy, health behaviours, and physical health. Different indicators were used to measure these latent variables: communication quality, confidence in managing health, physical activity, nutrition, tobacco use, BMI, and the number of chronic conditions. The SCP did not directly impact physical health, but the SEM model showed some significant indirect effects. Improving communication through SCPs helped boost self-efficacy, resulting in healthier behaviours and, in the end, better physical health. This finding would not have been obvious with more straightforward methods, but SEM allowed us to follow the whole path of influence. These business and healthcare examples show how SEM lets researchers try complicated models and find both direct and indirect links between important variables in the real world.

CHAPTER 3

METHODOLOGY

This chapter explains the research design, the process of collecting data, the tools that were used, and the analysis techniques that were applied in this research. The process starts by detailing how responses from students were gathered using the Warwick–Edinburgh Mental Well-being Scale (WEMWBS) via Microsoft Forms. The chapter explains how Microsoft Power Automate is used to automate the process of scoring and delivering feedback. In conclusion, it discusses the multivariate analysis methods that were applied to explore the underlying structure of the WEMWBS. This includes Principal Component Analysis (PCA), Exploratory Factor Analysis (EFA), Confirmatory Factor Analysis (CFA), and Structural Equation Modelling (SEM).

3.1 Data Collection

3.1.1 Data Source

The dataset used in this research was provided by UTAR's Centre for Healthy Mind and Well-being (CHMW). It contains personal information and well-being conditions and was collected through a psychological questionnaire initiated by UTAR.

3.1.2 Target Population

The respondents were first-year undergraduates and foundation students from UTAR. Data were collected from respondents enrolled in seven semesters, specifically from June 2022 to June 2024. A total of 2627 voluntary responses were gathered, representing a wide range of academic faculties and student intakes at UTAR.

3.1.3 Procedure and Data Collection Method

The data collection took place during the orientation sessions for each new student intake. During orientation, UTAR counsellors invited students to participate voluntarily by completing an online questionnaire via Microsoft Forms and responses will automatically store in Excel.

3.1.4 Instruments and Measures

The questionnaire consisted of two primary sections:

Section A: Personal Information

This section contains basic information about the respondent including name, gender, email address, telephone number, faculty of study, current year of study, and the agreement of privacy consent on whether to allow the CHMW to contact.

Section B: Warwick-Edinburgh Mental Well-Being Scale (WEMWBS)

The second section utilized the Warwick-Edinburgh Mental Well-being Scale (WEMWBS), comprising 14 items that are scored on a Likert scale ranging from 1 to 5. According to the UTAR's Centre for Healthy Minds and Wellbeing (CHMW), the total scores from the responses were divided into four groups: 'Very Low' for scores between 14 and 40, 'Below Average' for scores between 41 and 44, 'Average' for scores between 45 and 59, and 'Above Average' for scores between 60 and 70.

3.2 Power Automate

3.2.1 Power Automate Flow Design

In UTAR's current practice, freshmen will complete the Warwick-Edinburgh Mental Well-being Scale (WEMWBS) through Microsoft Forms during the orientation. After answering all 14 questions, students are required to manually add up their scores to find their total out of 70. The form then requires students to choose their calculated score range before filling in their particulars. However, after submitting the form, students do not receive confirmation or personalised feedback about their results. The process ends here unless the students decide to get more help. At the same time, the UTAR counsellors will review the responses gathered and determine whether to contact students via email, Microsoft Teams, or phone. This manual method takes a lot of time and doesn't ensure every student gets the support they need on time.

This research created an automated workflows by Microsoft Power Automate to fill these gaps. After students submit the form, their scores are automatically calculated and sorted into their well-being groups. They will receive immediate personalised feedback through Microsoft Teams. This feedback included the scores, gentle suggestions based on their well-being category, and an invitation to check out helpful resources for more support. Besides, students who are willing to communicate further can choose to join the UTAR counselling resource group in Microsoft Teams. The process is automated, consistent, and timely, reducing counsellors' manual workload and ensuring that every student gets immediate support that fits their needs.

Below is a simplified flow diagram of the whole automation process.

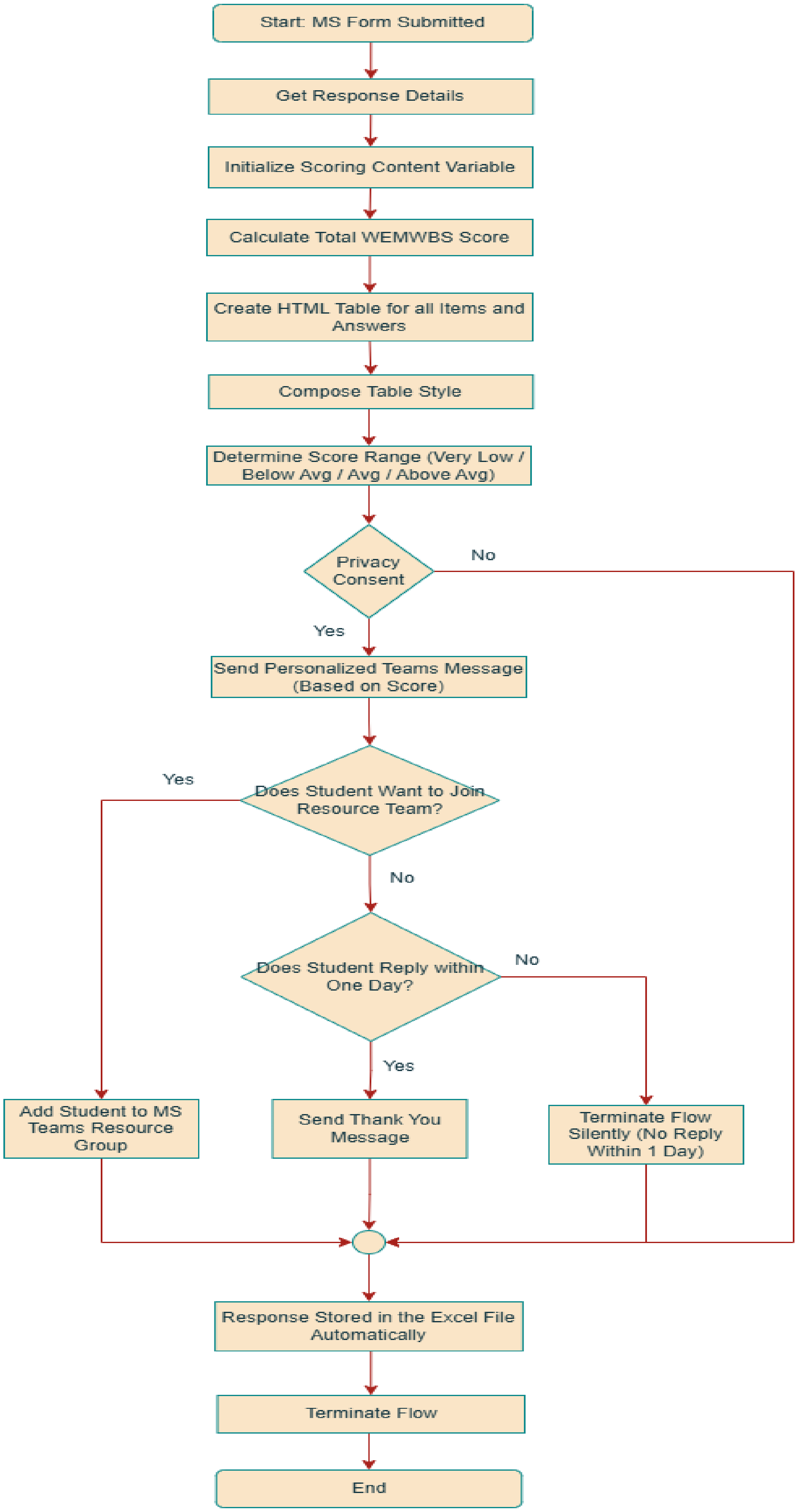


Figure 3.1: Flowchart of the Automated Mental Well-being Support Workflow

3.2.2 Workflow Implementation

In figure 3.2, an automated cloud flow type has been created to ensure the workflow runs automatically whenever a student submits the WEMWBS form.

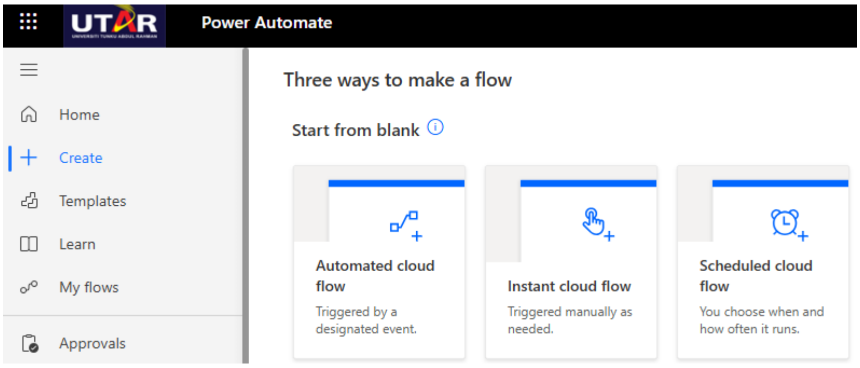


Figure 3.2: Power Automate Flow Options

In Figure 3.3, the workflow begins with a trigger called ‘When a new response is submitted’. Next, the action ‘Get response details’ is selected to extract the respondent's information. Under the Form ID, the Microsoft Forms used for the WEMWBS test are selected. The Response ID refers to the unique identifier for each form submission, and this enables the information to be extracted smoothly and correctly by pointing to the correct Response ID.

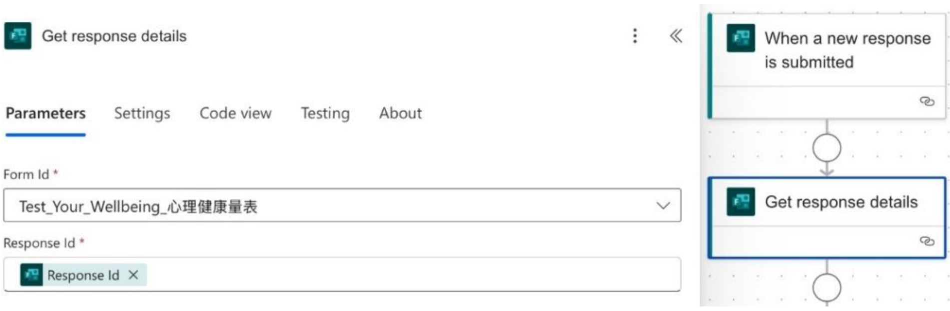


Figure 3.3: Trigger and Data Extraction Setup in Power Automate

In Figure 3.4, the questionnaire’s submission time is extracted and converted to a time unit that is easier to view. Besides, selecting the respondent's email and including it in the ‘Get user profile’ action will enable the system to extract the respondent’s information, such as the display name and given name, that will be used later on.

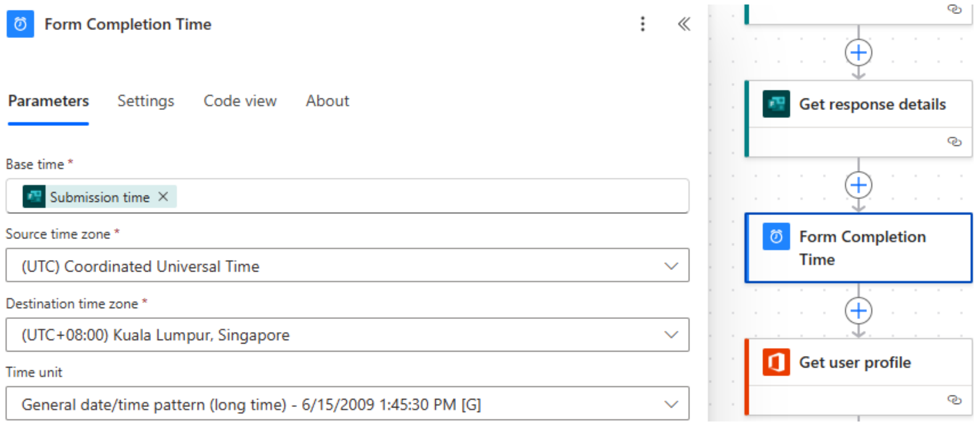


Figure 3.4: Form Completion Time Conversion in Power Automate

In Figure 3.5, the total score for the 14 questions of WEMWBS has been summed up. Next, parameters consisting of all the items in WEMWBS and the respondent's respective score are stored in each set of arrays, starting and closing with a curly bracket. This action is used to create an HTML Table with two columns representing the ‘Question’ and ‘Score’, and 14 rows of items. Furthermore, the following action consists of the table styling, which includes the font size and colour, and the alignment of the items in the table.

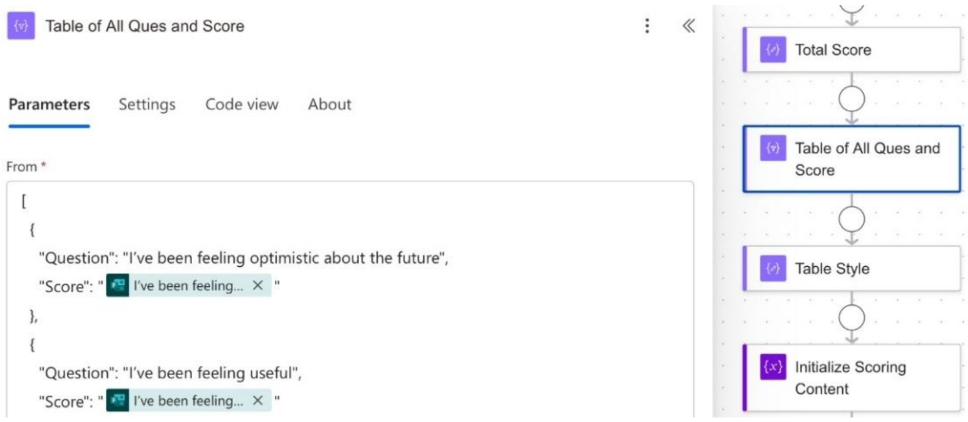


Figure 3.5: Generating Total Score and Table of Questions and Scores

In Figure 3.6, a string variable is initialised to store the scoring content based on the student's score. As mentioned previously, UTAR counsellors divided the well-being conditions into four groups. Next, in Figure 3.7, a conditional loop checks whether the total score falls under which category. After that, the flow executes only one action of the score variable. For example, in Figure 3.8, the content of the range of 14-40 scores is shown. This score variable will update the initialised 'Initialise Scoring Content' variable, which means the variable now stores the content for the suggestion to the student with a 14-40 score.

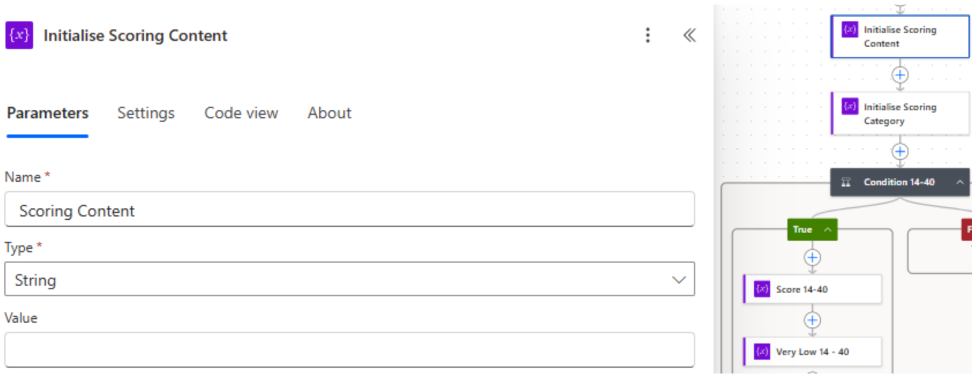


Figure 3.6: Initialising Scoring Variable

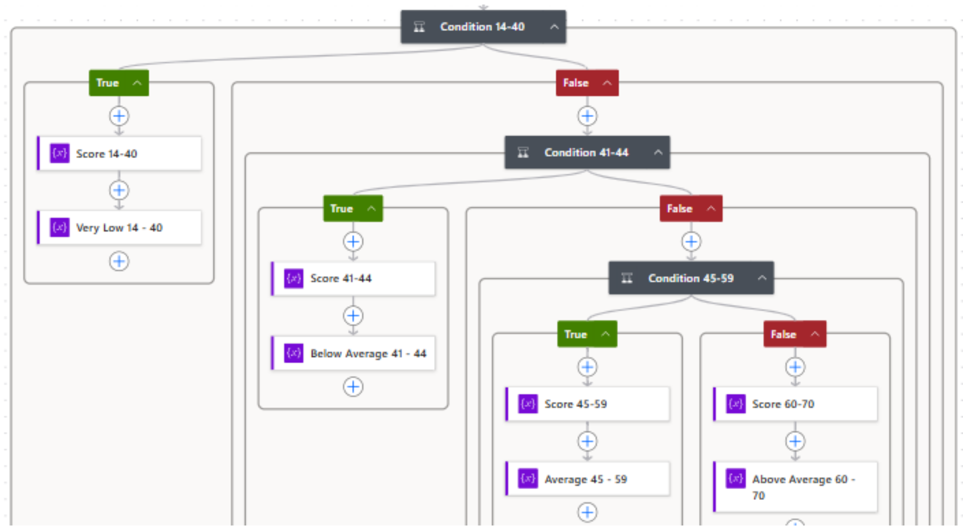


Figure 3.7: Conditional Branching Based on WEMWBS Score

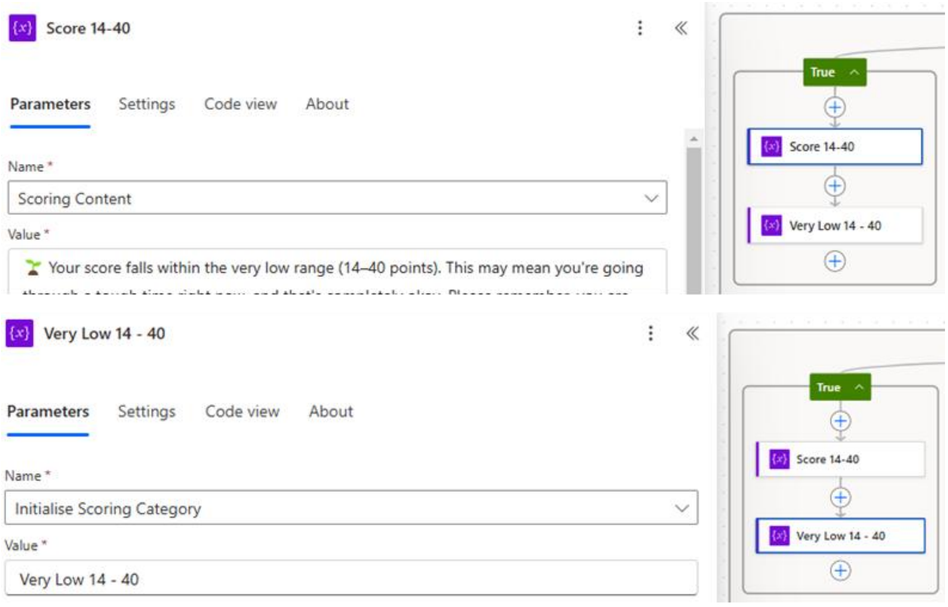


Figure 3.8: Assigning Scoring Content and Category to Variables

The scoring content of all categories refers to the article named User Guide to the Warwick-Edinburgh Mental Wellbeing Scales (2024). The respective contents are shown in Appendix F.

After identifying the scoring category, the respondents' privacy agreement is also checked. If the respondent is unwilling to be contacted by the university, then the flow will be terminated, and no further action will be taken. Before that, their information will automatically be stored in the Excel File, as shown in Figure 3.9. The Excel File will consist of nine columns, including particulars, total score, and privacy consent on not being contacted. In this case, the respondent's display name will be used for file storage. This is to ensure the name saved is the same as the name that appears in Microsoft Teams. This will make the counsellor's workload easier if they get any further action.

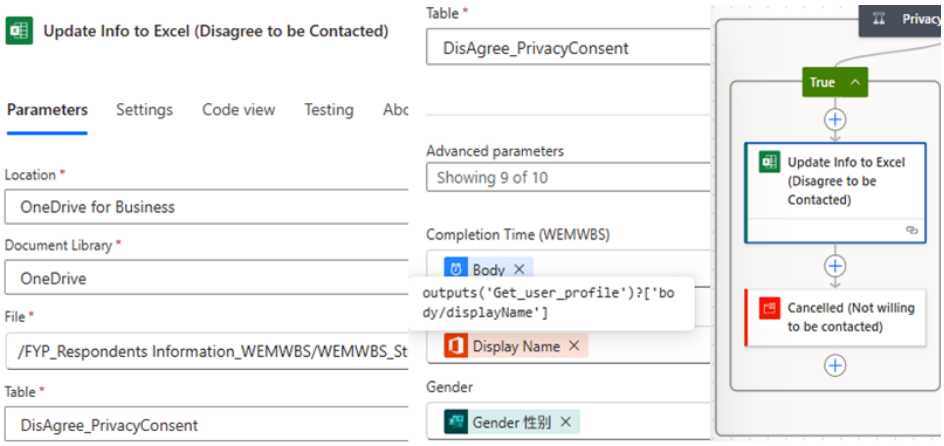


Figure 3.9: Storing Respondent Information in Excel (Disagree to Be Contacted)

In Figure 3.10, a message to be sent to the respondent is composed. Flow Bot identifies the message as system-generated and sends it on behalf of the automated flow. The message consists of an HTML table with styling, the total score of the student's WEMWBS test, the respective scoring content depending on the well-being category group, and a link to access UTAR counselling booking appointments, as shown in Figure 3.11.

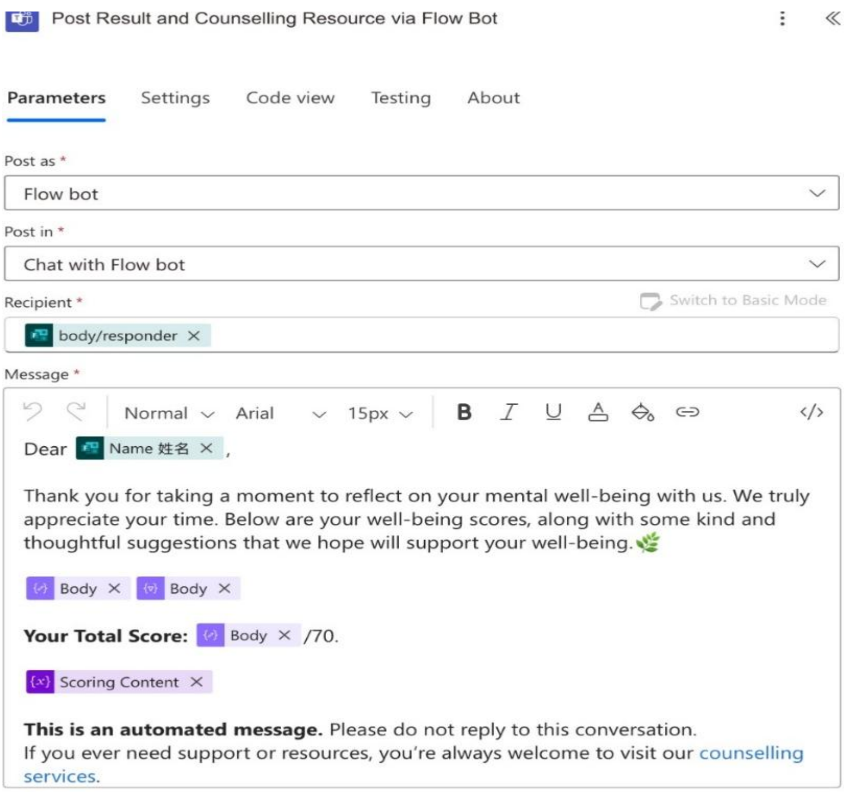


Figure 3.10: Sending Automated Message to Respondent via Flow Bot

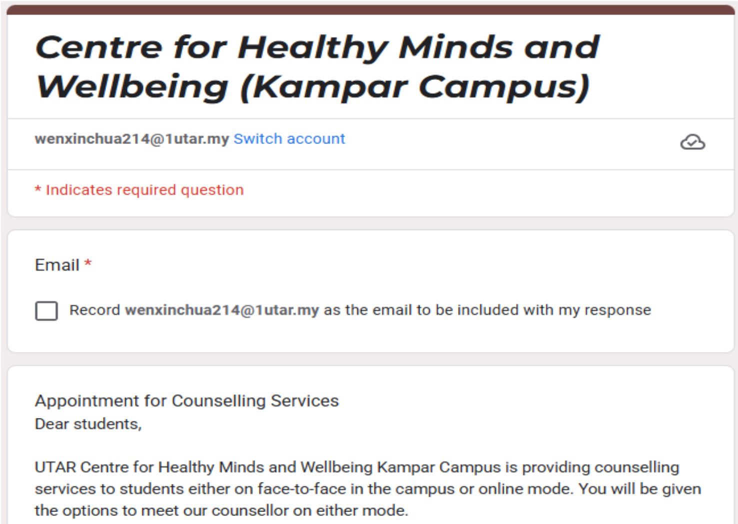


Figure 3.11: Current UTAR Counselling Appointment Form

In Figure 3.12, another message is sent to the respondent. It invites the respondent to join the UTAR Counselling Resource Group in Microsoft Teams. The message includes two clickable response options, which are 'Yes' or 'No'. The respondents can respond to the invitation by clicking the options they wish. It consists of three cases of the respondent's reactions:

1. Selected 'Yes' Option
2. Selected 'No' Option
3. No response to the action

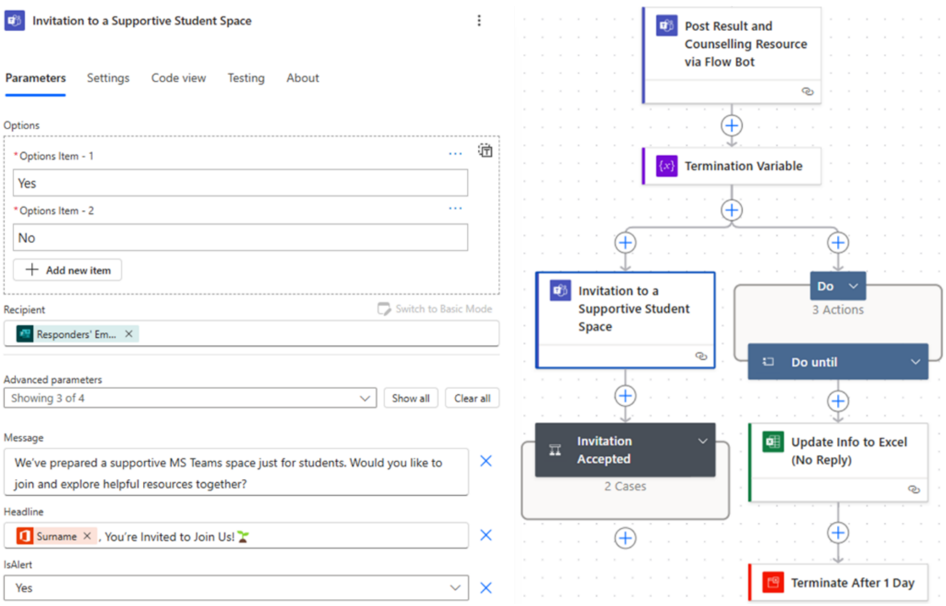


Figure 3.12: Invitation Message to Join the Counselling Resource Group

If the respondents are willing to be added to the resource group, they will be added to the UTAR Counselling Resource Room automatically and immediately, as shown in Figure 3.13. Next, a notification will be sent out to the respondents via Flow Bot to notify them. Besides, their information and willingness to be added will be recorded in the Excel file before the flow ends.

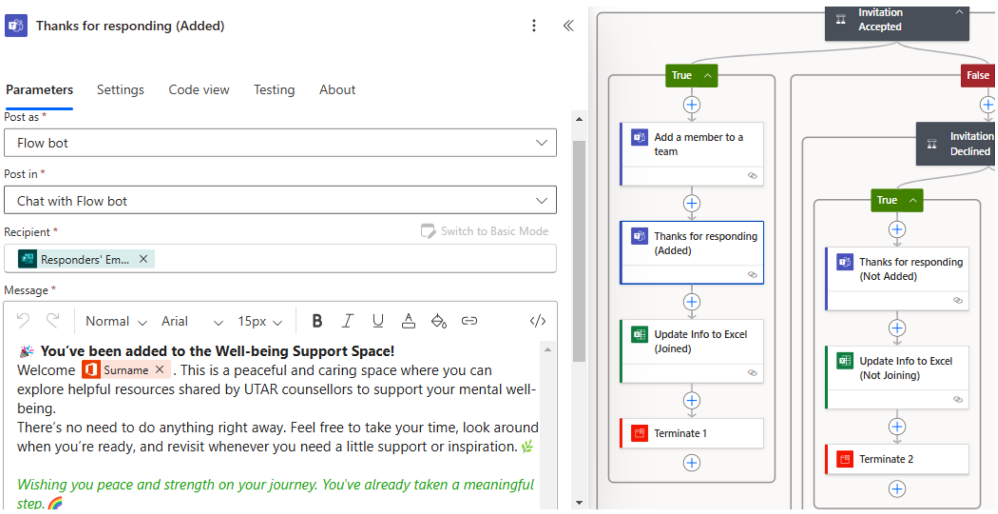


Figure 3.13: Notifying Respondent After Joining the Support Group

In the second situation, some respondents might be added to the Resource Teams. After the respondent selects the ‘No’ Option, an automated message is sent, as shown in Figure 3.14. The message consists of a hyperlink to access the UTAR counselling appointment booking, the same link shown in Figure 3.2.10. The flow will be terminated as it ran successfully after updating the respondent information in the Excel file.

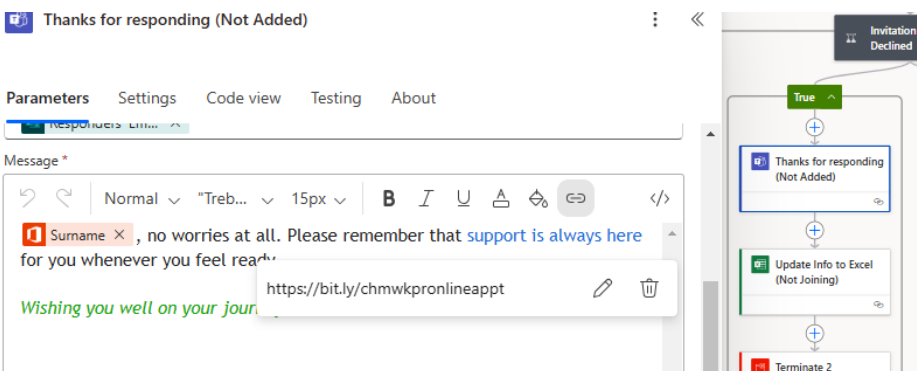


Figure 3.14: Confirmation Message for Students Not Joining the Group

In the last case, the respondent did not react to either option for about 24 hours. A reminder message will prompt the respondent and notify that the flow will be terminated the next day, as shown in Figure 3.15. The flow will be terminated after 24 hours, and the flow status will be successfully run.

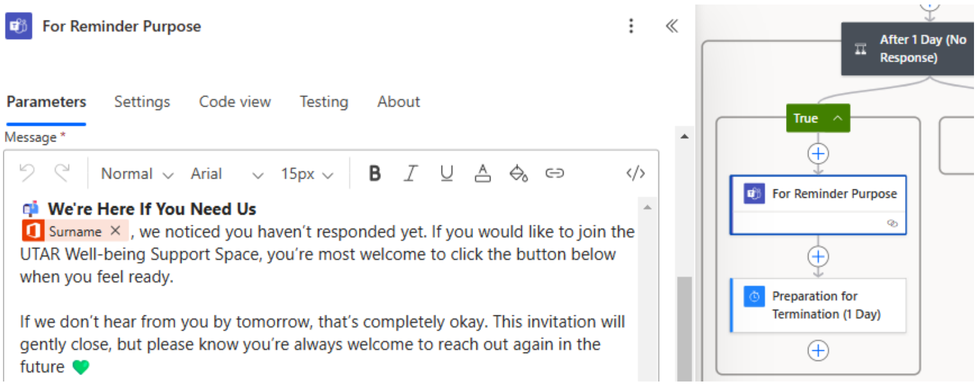


Figure 3.15: Reminder Message for Non-Responding Students

The reminder message is customised. It is not a built-in action in Power Automate. In Figure 3.16, the concept of the reminder message starts with a do-until loop, where the loop consists of an increment variable, which means that once the flow executes the do-until loop, the termination variable will be increased by one unit. Carrying on this concept to the conditional loop, the loop will only flow under the true condition when the termination variable is equal to 48, which means that they do until the loop is executed 48 times. The reminder message will only show up after the do-until loop runs 48 times. After sending out the reminder message, the flow will be delayed and paused for 1 day, waiting for the respondent to react. After that, the conditional loop will exit, leading to the delay action of 30 minutes. This is why 48 times is chosen as both loop conditions. Since the reminder message will only be sent out after 24 hours, and the do-until loops check the status every 30 minutes, the loops need to be executed 48 times to accumulate for 24 hours. This indicates that the increment variable will increase 48 times, leading to a total of 48. The do-until loops will exit only if the termination variable equals 48, leading to the flow termination.

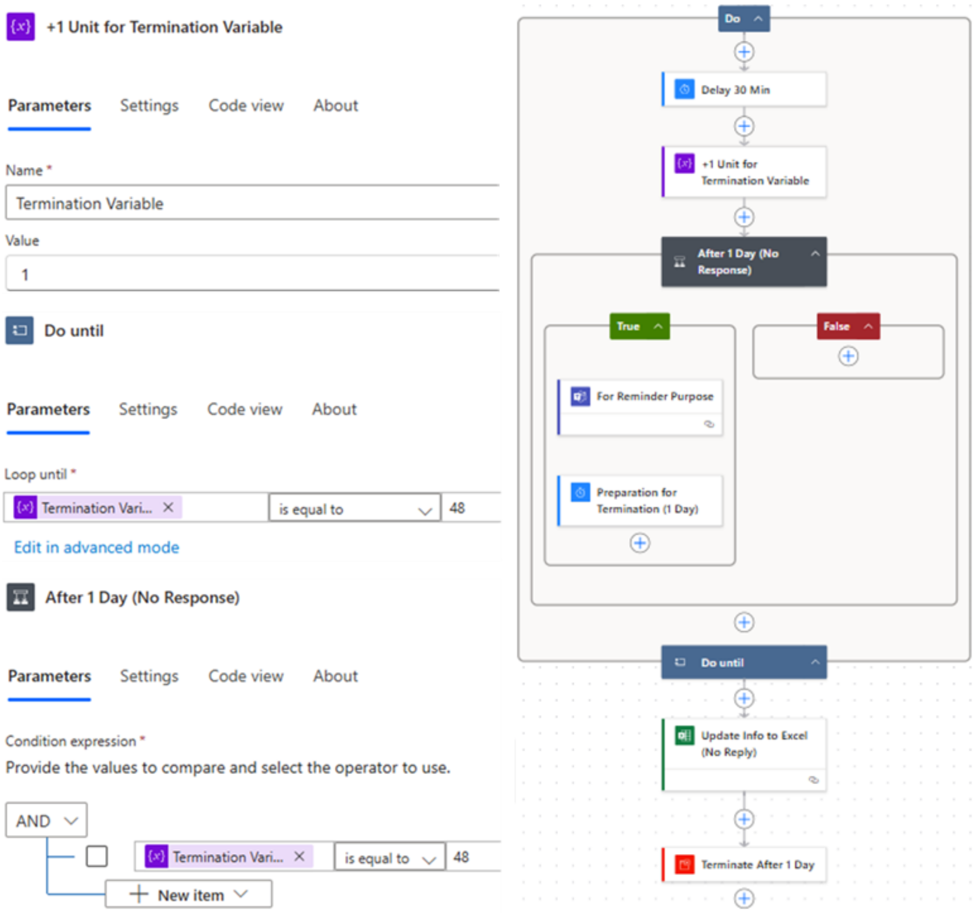


Figure 3.16: Reminder and Termination Loop for Non-Respondents

In conclusion, this section covered the complete design and workflow of the automated response system created with Microsoft Power Automate. By automating the process, the system enhances the overall experience for the students while also lightening the manual workload for UTAR counsellors.

3.3 Principal Component Analysis (PCA)

3.3.1 Data Preparation for PCA

After data collection, Principal Component Analysis (PCA) was used to find the Warwick-Edinburgh Mental Well-being Scale (WEMWBS) dataset's underlying dimensions or components. Firstly, the 14 questions from

WEMWBS were standardized. According to Brereton (2025), standardization before PCA prevents variables with greater magnitudes from degrading the underlying patterns by assuring equal contribution from each variable to the analysis. Practically, standardization involves converting the raw scores into Z-scores, where the mean is zero and the standard deviation is one for each item.

3.3.2 Suitability of Data for PCA

Before conducting Principal Component Analysis (PCA), it's really important to check if the dataset is suitable for factor analysis. This part looks at the determinant score, the Kaiser-Meyer-Olkin (KMO) test, and Bartlett's test of sphericity to check if the dataset is suitable for PCA. According to Shrestha (2021), the determinant score is an important test for detecting singularity or multicollinearity. If the determinant score of the correlation matrix exceeds 0.00001, it indicates that multicollinearity is not present. In contrast, if the determinant value is below 0.00001, it is crucial to identify pairings of variables with a correlation coefficient greater than 0.8 and exclude them from the research. Shrestha (2021) explained that a lower score might indicate high inter-correlations among groups of three or more questions or statements, which could decrease the item exclusion level until this criterion is satisfied.

Besides, Average Inter-Item Correlation (AIIC) is also a significant test to conduct before factor analysis. As stated by Stefana et al., (2025), AIIC evaluates the internal consistency and suitability of the dataset. In detail, AIIC evaluates the degree to which a single item's score is related to the scores of all other items in a scale and examines how consistently the items represent the same content area (Cohen et al., 2013; Raykov & Marcoulides, 2011). According to Briggs and Cheek (1986), the AIIC is required to be between 0.15 and 0.50. This range helps maintain a good balance between having some commonality among items and avoiding too much overlap.

Furthermore, the data's appropriateness for factor analysis was determined using Bartlett's Test of Sphericity and the Kaiser-Meyer-Olkin (KMO) Measure of Sample Adequacy. Bartlett's Test investigates whether the correlation matrix of the variables significantly differs from an identity matrix. In contrast, the KMO Test is intended to measure the suitability of data for factor analysis. In other words, it tests the adequacy of the sample size (Shrestha, N. 2021). According to Napitupulu, Kadar, and Jati (2017), a KMO index greater than 0.50 indicates that the data are suitable for factor analysis, with values approaching 1 suggesting a higher adequacy for performing factor analysis. They also stated that Bartlett's Test is considered statistically significant if the p-value is less than 0.05, thus indicating that the correlation matrix is adequate for factor analysis.

3.3.3 Implementation of PCA

After confirming the data adequacy, PCA was conducted to determine the minimum number of components that can best represent the interrelationships among the WEMWBS. Kaiser’s Criterion (Eigenvalue Sciterion) and Scree Test were used to assist in deciding the number of components to retain.

Eigenvalue is a ratio of the common variance and the specific variance explained by a particular components extracted. An eigenvalue of a factor denotes the proportion of total variance explained by that factor where the factors with eigenvalues greater than one should be retained because they explain a significant portion of the total variance (Pallant, 2010; Guttman, 1954; Kaiser, 1970; Verma, 2013; Shrestha, 2021).

Scree test is used to determine how many factors should be taken out before the amount of unique variance dominates the common variance structure. In general, a scree plot is used to determine the number of factors needed. A scree plot graph comprises the magnitude of the eigenvalues as the y-axis, where the number of factors is the x-axis. The eigenvalues are plotted as dots and connected with a graph line. Factor extraction should be stopped when there is an 'elbow' or plot levelling (Hair et al., 1998; Cattell, 1966; Cattel, 1973; Shrestha, 2021).

However, according to Weismer et al., (2021), eigenvalues below 1 usually mean that the statistical support is lower, but components with eigenvalues slightly below 1 can still be kept if they have useful information for interpretation. To improve the overall interpretability of their research, Weismer et al., (2021) specifically chose to keep a component with an eigenvalue below 1 (0.406), since it distinguished particular profiles within their data. The decision makes it clear that Kaiser's eigenvalue criterion (eigenvalue ≥ 1) does not have to be strictly followed if interpretive clarity and meaningfulness are improved. In addition, Shaharudin and Ahmad (2017) explained that if there are too few components, the observations are not well-represented, but if more cumulative percentages are taken or more components are retained, the result will be poor as it will inflate the importance of noise. Furthermore, Wasdahl

(2024) also highlighted that a 62% cumulative variance explained could be adequate for social science research. While some variance remains unexplained due to the failure to capture these factors, it indicates other considerations that might affect variance in the data.

3.3.4 Validation for PCA

In this research, k-folds cross-validation is used to determine the accuracy and reliability of PCA. The number of folds chosen for this research is k=10, as Wong and Yeh (2020) suggest that performing 10-folds cross-validation is better than 2 or 5-folds. According to Lumumba et al., (2024), in conducting a 10-folds cross-validation, the dataset was partitioned into ten equal subsets. One-fold served as the testing set for each iteration, while the remaining nice folds were utilized to train the model. This process was repeated ten times, ensuring each fold was used exactly once for validation. Model performance was assessed by averaging results from all iterations, providing a reliable estimation of model accuracy and reducing variability. A graphical explanation can be found in Appendix G.

3.4 Exploratory Factor Analysis (EFA)

3.4.1 Introduction to EFA

After extracting principal components using Principal Component Analysis (PCA), Exploratory Factor Analysis (EFA) was conducted to investigate the Warwick-Edinburgh Mental Well-being Scale (WEMWBS) data to identify the latent factors.

While PCA focuses on dimensionality reduction by summarising the variance, EFA is used to identify the latent variables hidden within the dataset collected. EFA can clarify how these items are grouped into meaningful and interpretable latent factors. Therefore, EFA allows the UTAR counsellor to understand the student's mental well-being better and thus focus on that particular area for intervention and support.

3.4.2 Data Preparation for EFA

The dataset needs to be standardised during the PCA process. This ensures that each of the 14 WEMWBS items has equal weight. Therefore, this standardised data was directly reused for the EFA analysis.

According to Victor-Edema (2023), applying multivariate statistics to a real-world issue always involves a combination of univariate and multivariate analyses. This ensures that insights into the relationship between variables are obtained. Normality implies whether a variable follows a pattern like the normal distribution. Multivariate normality examines whether multivariate data follow the properties of a multivariate normal distribution. Computing multivariate normality is important for EFA and the following sessions, as it will affect the factor extraction method. For a dataset to fulfil the criteria of multivariate normality, each variable must follow a normal distribution (Victor-Edema, 2023). In the research, histograms, skewness, and kurtosis were tested for univariate normality.

According to Hatem et al., (2022), to conclude the normality of a designated dataset using a histogram, the dataset is explored to check if it follows into the standard distribution shape. Normality is met if the bars of the histogram form a symmetric bell. Besides, according to Haşıloğlu and Haşıloğlu-Cifciiler (2023), to test the normality of a Likert-type questionnaire, the coefficient of skewness and kurtosis needs to be within the range of (-1,+1) to be accepted for normality. However, Victor-Edema (2023) noted that obtaining univariate normality alone does not prove that the dataset follows a multivariate normal distribution. Thus, Mardia's multivariate skewness and kurtosis test were examined. If the p-value is less than 0.05, the null hypothesis of multivariate normality is rejected, stating that the dataset does not meet the assumption of multivariate normality.

3.4.3 Factor Extraction and Selection

This section explains the factor extraction and rotation method and the number of factors to retain from the 14-item WEMWBS dataset.

To uncover the underlying latent factor structure on 14 items of the WEMBWS dataset, an oblique rotation method, Promax, was applied. According to Sürücü, Yikilmaz and Maşlakçı (2024), Orthogonal and oblique are the two primary types of rotation methods. The main difference is whether the correlations of the factor are allowed. Oblique rotations help to explain the correlations between factors and tend to yield more realistic outcomes compared to orthogonal rotations. If the relationship between factors is unknown, oblique rotation should be performed, aligning with the current research's objective. In the research, Promax was chosen because it has a fast computation speed and works well with large sample sizes (Sürücü, Yikilmaz and Maşlakçı, 2024).

According to Brown (2015) and Kyriazos (2023), the extraction process in the EFA aims to figure out the minimum number of common factors, reproducing correlations among observed variables acceptably. Selecting a suitable extraction method is very important since it will affect the entire research analysis. The extraction method in EFA will affect the results of confirmatory factor analysis (CFA) and Structural Equation Modelling (SEM), which will be discussed in detail in the following sections.

When selecting an extraction method, sample size and normality must be considered. For large datasets (sample size greater than 100), Principal Axis Factoring (PAF) and Maximum Likelihood (ML) are suitable (Brown 2015; Kyriazos, 2023). For normality, PAF does not require distribution assumptions, whereas ML require multivariate normality (Brown 2015; Kyriazos, 2023). However, ML also accept slight normality violations for a large dataset (Bollen, 1989; Raykov and Widaman, 1995; Kyriazos, 2023).

WEMBWS datasets show a large sample size, but the multivariate normality is violated. The research uses the Maximum Likelihood (ML) rotation method rather than the Principal Axis Factoring (PAF). Explanation will be based on Table 3.1 from Kyriazos (2023) and summarised by Watkins (2021).

Table 3.1: Common Estimation Methods in EFA, CFA, and SEM

Estimator	Class	Used in	Description
MINRES[a]	LS	EFA	Minimum Residual
PA[b]	LS	EFA	Principal Factor Solution (aka PAF)
WLS	LS	EFA/CFA/SEM	Weighted Least Squares (aka ADF)
GLS	LS	EFA/CFA/SEM	Generalized Least Squares
ML	ML	EFA/CFA/SEM	Maximum Likelihood
DWLS	LS	CFA/SEM	Diagonally weighted least squares
ULS	LS	CFA/SEM	Unweighted Least Squares
DLS	LS	CFA	Distributionally-weighted Least Squares
PML	ML	CFA/SEM	Pairwise Maximum Likelihood
MLM	ML	CFA/SEM	ML with robust standard errors and SB χ^2
MLR	ML	CFA/SEM	ML with robust standard errors and Yuan-Bentler χ^2

Table 3.1 indicates that PAF only supports the use of ML in EFA, where ML can be used in EFA, CFA, and SEM and fulfils the research needs. However, as mentioned earlier, ML requires a multivariate normality assumption, which the WEMBWS dataset violates. To be further discussed in the following session, Maximum Likelihood with robust (MLR) is an extraction method for large datasets that can be used even when the multivariate normality is violated. However, it is suitable for CFA and SEM but not EFA.

Kyriazos (2023) and Skinner (1989) mentioned that MLR uses a pseudo-ML method, which allows it to provide identical parameter estimates to ML. Therefore, in the research, Maximum Likelihood (ML) will be used for EFA, and Maximum Likelihood with robustness (MLR) will be used in CFA and SEM. This decision is supported by prior research, which integrates two extraction methods in the same datasets, and they also used ML for EFA and MLR for CFA and SEM (Caputo et al., 2023). Besides, Roesel et al., (2024) findings are more closely represented by the WEMWBS dataset, which their findings used a large dataset (*n*=1125) and examined 13 Likert Scale questionnaires related to healthcare. They used ML for EFA and MLR for CFA and SEM, where, in the end, they got an informative output.

Besides, the number of factors designated for extraction was determined based on previous PCA findings, which discovered an appropriate number of underlying dimensions via eigenvalue estimation and scree plot analysis (Phakiti, 2018). According to Sürücü, Yikilmaz and Maşlakçı (2024), it is generally accepted that at least three variables are required to produce a factor from the observable variables in EFA.

3.4.4 Computation of Factor Loadings

Phakiti (2018) explains that in Exploratory Factor Analysis (EFA), factor loadings indicate the correlation coefficients linking each observed variable to the latent factors that have been identified. In this research, 14 items of the Warwick-Edinburgh Mental Well-being Scale (WEMWBS) have been analysed, which helps estimate factor loadings based on the extracted factor solution. The loading values were important for understanding how each questionnaire item relates to a specific aspect of mental well-being.

Higher factor loadings show that the relevant factors are better explained by the items (Yong and Pearce, 2013; Sürücü, YIKILMAZ and Maşlakçı, 2024). Fabrigar and Wegener (2012), Field (2013) and Phakiti (2018) stated that factor loadings above 0.30 are usually seen as significant for interpretation in EFA.

Items with loadings below this value can be seen as weak indicators of any factor and are often considered for exclusion in further analysis (Phakiti, 2018). In this research, the factors that exhibited the highest loading over this threshold were assigned to the items. The loading matrix helped organise the WEMWBS items into three groups identified earlier based on the PCA results.

3.4.5 Communalities

According to Samuels (2017), communalities represent the variance in each observed variable accounted for by the extracted factors during EFA. High communalities indicate that a variable significantly differs from the common factors, making it an ideal retention choice.

According to Haider et al., (2022), all items with suitable communalities were kept in the research, which shows that the reliability of their questionnaire was strong. To discuss the acceptable value, Karimian and Chahartangi (2024) pointed out that a communality value above 0.5 is usually suggested when looking at the acceptable value, and a communality value between 0.3 and 0.4 is also acceptable. However, items should be removed if communalities drop below 0.2 (Child, 2006; Samuels, 2017). Nevertheless, Braman and Azzam (2023) pointed out that it is important not to lightly remove questions from a validated questionnaire, particularly without a solid theoretical reason or if it could compromise the measured construct. Therefore, all the factors will be retained in the research since WEMWBS is a validated questionnaire.

3.4.6 Grouping of Questionnaire Items by Factor

The questionnaire items were grouped based on their strongest factor loadings. For every item in the questionnaire, the absolute highest factor loading was used to decide its assignment to the relevant factor group. The dynamic factor groupings led to clusters of items, with each one clearly representing specific underlying concepts measured by the WEMWBS scale. This approach made sure that every item fit into one factor, which helped with understanding and checking reliability later.

3.4.7 Reliability Analysis After EFA

After completing the Exploratory Factor Analysis (EFA), the internal consistency of each factor grouping was evaluated using Cronbach's Alpha, McDonald's Omega, and Composite Reliability (CR).

Cronbach's Alpha was used as a key measure of internal consistency. This method helps to determine how closely related the items are within the same factors (Shrestha, 2021).

To support this, McDonald's Omega was calculated as well. According to Orçan (2023), Omega gives a better estimate of reliability compared to Alpha, particularly when the items have unequal loadings. The article points out that Omega is often chosen when the assumptions of Alpha are not fully met, which can enhance the accuracy of reliability results. According to Orçan (2023), Omega offers a better estimate of reliability compared to Alpha, particularly when the items have unequal loadings. Based on Youngstrom et al., (2017) and

Stefana et al., (2025), alpha and omega values usually fall between 0.00 and 1.0. A threshold of 0.70 is considered adequate, 0.80 is good, and 0.90 is excellent.

Finally, Composite Reliability (CR) was calculated to further confirm the internal consistency. According to Shrestha (2021), a CR value of 0.6 or higher is seen as acceptable. Adding CR together with Alpha and Omega provides a clearer and more comprehensive understanding of the reliability of the factor structure.

3.5 Confirmatory Factor Analysis (CFA)

3.5.1 CFA Model Specification

The CFA model was developed based on the factor structure from the EFA findings. Specifically, items from the WEMWBS questionnaire were grouped into three distinct categories, each representing different dimensions of mental well-being experienced by UTAR first-year students.

Table 3.2: Grouping of WEMWBS Items by Factor

Factor	Question
Positive Emotion	Q1, Q2, Q3, Q8, Q10, Q12, Q14
Personal Growth	Q6, Q7, Q11
Social Exploration	Q4, Q5, Q9, Q13

3.5.2 Estimation Method for CFA

As mentioned in the previous section, the dataset for the current research did not achieve the multivariate normality assumptions. The CFA was conducted using the Maximum Likelihood Robust (MLR) estimator. The MLR method is a robust version of the traditional Maximum Likelihood (ML) estimation, as it can work with non-multivariate normal data. MLR modifies the standard errors and chi-square statistics, which helps draw reliable conclusions even when the distributions are abnormal (Kyriazos, 2018). However, MLR also requires large samples, which can be fulfilled in this research. Therefore, using MLR was a good option to make sure the CFA results were accurate and reliable based on the dataset of this research.

3.5.3 Model Fit Evaluation

A significant aspect of CFA is its focus on a hypothesis-driven method (Brown, 2015; Alavi et al., 2020). The researchers start by creating a hypothesis about the model structure, which is expressed as particular factors that underlie a set of items. The analysis examines how well the proposed factor structure explains the covariance among the items. In addition to looking at the covariance that the model captures, it is crucial to assess how well the model fits in confirmatory factor analysis (Hooper, Coughlan, and Mullen, 2008; Alavi et al., 2020). This helps to understand how closely the model matches the actual and observed data. According to Alavi et al., (2020), the model fit indices are typically divided into global and local fit indices. Global fit indices examine how well the model matches the observed data structure. Local fit indices focus on specific aspects of the model, like factor correlations and inter-item residual covariances, which can offer ideas for improving the model.

Global fit indices can be further divided into absolute fit indices, incremental fit indices and parsimony fit indices. Absolute fit indices evaluate how closely the suggested model aligns with the actual data without making comparisons to other models. They look at how well the proposed model matches the actual data they collected (Hooper et al., 2008; Kline, 2005; Alavi et al., 2020). The measures under global fit indices are Chi-square, Goodness-of-Fit Index (GFI), Root Mean Square Error of Approximation (RMSEA) and much more (Jöreskog and Sörbom, 1989; Steiger, 2007; Alavi et al., 2020).

Next, Alavi et al., (2020) stated that the incremental, or comparative fit indices look at how the proposed model compares to a baseline model that assumes no relationships between the variables. The baseline model is the hypothesis of no meaningful relationships between variables. It illustrates how much better the hypothesised model fits than assuming all variables are unrelated. The measures under incremental fit indices are Comparative Fit Index (CFI), Normed Fit Index (NFI) and Tucker-Lewis Index (TLI) (Bentler, 1990; Bentler and Bonett, 1980; Alavi et al., 2020).

Parsimony fit indices consider model complexity by adjusting for the number of estimated parameters. This helps prevent overfitting by adding a penalty for overly complex models (Alavi et al., 2020). The measures under parsimony fit indices are the Chi-Square Test of Independence (Mohtar et al., 2024).

To determine how effectively the hypothesised model fits the WEMWBS questionnaire, several model fit indices were evaluated.

The Goodness of Fit Index (GFI) indicates how well the model we proposed aligns with the data when compared to a null model, which assumes there's no relationship between the variables. It functions similarly to R^2 in regression, indicating the extent to which the model can account for the variance or covariance in the data. A value of greater than 0.9 is considered satisfied for the model (Alavi et al., 2020).

$$GFI = 1 - \frac{\chi^2}{\chi^2{}_{null\,model}} \tag{1}$$

Root Mean Square Error of Approximation (RMSEA) is a measure that aims to adjust for the tendency of chi-square statistics to reject models when dealing with large samples. RMSEA focuses on the differences between the proposed model, which uses carefully selected parameter estimates, and the population covariance matrix, helping avoid sample size problems (Maccallum, Browne, and Sugawara, 1996; Sarmento and Costa, 2019).

$$RMSEA = \sqrt{\max\left(\frac{\chi^2{}_{proposed\,model} - df_{proposed\,model}}{df_{proposed\,model} \times (N-1)}, 0\right)} \tag{2}$$

where N represent the sample size and df denotes degrees of freedom. RMSEA gives a one sided test with the following hypothesis:

H_0: The RMSEA equals 0.05 (close-fitting model)

H_1: The RMSEA is higher than 0.05

Thirdly, the comparative fit index (CFI) looks at how well the model fits by checking the differences between the data and the proposed model. It also takes into account the sample size problems that come with the chi-squared test and the normed fit index (Portela, 2012; Sarmento and Costa, 2019).

$$CFI = 1 - \frac{\max\,[\chi^2{}_{proposed\,model} - df_{proposed\,model}, 0]}{\max\,[\chi^2{}_{null\,model} - df_{null\,model}, 0]} \tag{3}$$

Next, the Tucker-Lewis index (TLI) is also a model fit index. The Tucker-Lewis index (TLI) is also called the non-normed fit index (NNFI). It combines a parsimony measure and a comparative index between the proposed and null models (Portela, 2012; Sarmento and Costa, 2019).

$$TLI = \frac{\frac{\chi^2_{null\,model}}{df_{null\,model}} - \frac{\chi^2_{proposed\,model}}{df_{proposed\,model}}}{\frac{\chi^2_{null\,model}}{df_{null\,model}} - 1} \quad (4)$$

Normed Fit Index (NFI) is also one of the model fit indices. Normal Fit Index (NFI) analyses the difference between the chi-squared value of the null model and the chi-squared value of the proposed model (Portela, 2012; Sarmento and Costa, 2019).

$$NFI = 1 - \frac{\chi^2\ (proposed\ model)}{\chi^2\ (null\ model)} \quad (5)$$

Lastly, Relative fit indices (RFI), often referred to as "incremental fit indices" or "comparative fit indices," are important concepts in statistical analysis. It compares and measures the chi-square for the proposed model to a null model. This null model typically includes a scenario where all the variables are uncorrelated, leading to a significantly large chi-square value, which suggests a poor fit. It is considered very good if the RFI is nearest to 1 (Portela, 2012; Sarmento and Costa, 2019).

$$RFI = 1 - \frac{\chi^2_{proposed\,model} / \, df_{proposed\,model}}{\chi^2_{null\,model} / \, df_{null\,model}} \quad (6)$$

Table 3.3 refers to the reference values for adjustment indices as mentioned above (Sarmento and Costa, 2019).

Table 3.3: Reference Thresholds for Model Fit Indices

Fit Indices	Statistic	Very Good	Good	Suffering	Bad
Absolute Fit Indices	GFI	Need to be greater than 0.9			
	RMSEA	≤ 0.05	$[0.05, 0.08]$	$[0.08, 0.10]$	> 0.10
Increment Fit Indices	CFI	≥ 0.95	$[0.90, 0.95]$	$[0.80, 0.90]$	< 0.80
	TLI	≥ 0.95	$[0.90, 0.95]$	$[0.80, 0.90]$	< 0.80
	NFI	≥ 0.95	$[0.90, 0.95]$	$[0.80, 0.90]$	< 0.80
	RFI	The better if closer to 1			

3.5.4 Convergent Validity and Discriminant Validity

To confirm and understand how the observed variables represent the constructs in Confirmatory Factor Analysis (CFA), it is essential to check the scale's reliability and validity. The previous sections tested reliability tests like Cronbach's Alpha, McDonald's Omega, and Composite Reliability, and those results indicate that the factors have strong internal consistency. For validity, it aims to determine whether the scale measures or represents the construct that the researcher wants to assess. Different types of validity methods change based on the research objectives. The primary techniques include convergent and discriminant validity (Sarmento and Costa, 2019).

Convergent validity looks at how well different measures that are supposed to assess the same thing relate to each other, or how they connect to other areas where there should be a positive or negative relationship based on theory (Sarmento and Costa, 2019). This form of validity is usually evaluated through Construct Reliability (CR) and Average Variance Extracted (AVE). These measures show how consistently the items represent a construct by reflecting the average proportion of variance explained (Sarmento and Costa, 2019).

Due to the computation of CR and AVE will include the standardized factor loadings, thus, it will be defined in this section. Standardised factor loadings are assessed for how well they represent the underlying construct. A loading value over 0.4 is usually seen as acceptable for interpretation, but it is better to have higher values for stronger relationships (Stevens, 2002; Cheung et al., 2023). Most of the time, a threshold of 0.5 or higher is suggested, and 0.7 or more is considered ideal (Hair et al., 2009; Cheung et al., 2023).

According to Grewal et al., (2004) and Cheung et al., (2023), Construct Reliability or Composite Reliability (CR) is a common metric used to assess the internal consistency of a latent construct in Confirmatory Factor Analysis (CFA). In contrast to Cronbach's alpha, which treats all items as equally important, CR takes into account the actual standardised factor loadings of each item. This makes it more accurate, particularly when the loadings of the items differ.

The CR formula is shown below:

$$CR = \frac{(\sum \lambda_i)^2}{(\sum \lambda_i)^2 + \sum(1-{\lambda_i}^2)} \tag{7}$$

where λ_i refers to the completed standardized factor loading of the item i. This formula makes sure that when reliability is measured, both the strength of the item-factor relationships and the error variances are taken into account. A CR value of 0.70 or higher is generally seen as acceptable, meaning that the majority of the variance is explained by the construct instead of measurement error. This shows that CR is a strong and preferred measure of reliability in CFA, as it provides a clearer understanding of how consistently the items represent the intended latent factor (Grewal et al., 2004; Cheung et al., 2023).

According to Cheung et al., (2023), Average Variance Extracted (AVE) shows the amount of variance in the observed variables that the latent construct captures compared to the variance resulting from measurement error. This gives a good idea of how much the items in a factor have in common in terms of variance. The AVE is determined by squaring the standardised factor loadings of each indicator that corresponds to a specific factor.

For factor X, AVE is defined as shown below:

$$AVE(X) = \frac{\sum_{i=1}^{p} \lambda_i{}^2}{\sum_{i=1}^{p} \lambda_i{}^2 + \sum_{i=1}^{p} Var(\varepsilon_i)} = \frac{1}{p}\left(\sum_{i=1}^{p} \lambda_i{}^2\right), \quad (8)$$

where:

- $\lambda_i{}^2$represents the standardized factor loading of the item i,
- $Var(\varepsilon_i)$ is the error variance of item i,
- p is the number of items loading onto the factor X.

In practice, this formula is usually simplified to the average of the squared standardised loadings, particularly when the measurement model is standardised (Cheung et al., 2023). An AVE value of 0.50 or higher is seen as acceptable, meaning that the construct accounts for at least 50% of the variance in its indicators (Yu et al., 2021; Cheung et al., 2023). If the values are below this threshold, it indicates that a larger portion of the variance is attributed to error rather than the actual construct, which raises some concerns regarding convergent validity (Fornell and Larcker, 1981; Cheung et al., 2023). However, according to Hair et al., (2010) and Suprapto, Stefany, and Ali (2020), although the minimum recommended AVE is 0.5, a value of 0.4 can still be considered acceptable. If AVE is below 0.5, the convergent validity is acceptable as long as the composite reliability is above 0.6.

Discriminant or divergent validity is about how much a measure does not correlate with other measures that should be different from theoretically. When getting the instrument ready, it is crucial to carefully plan the validation process so that all relevant data can be collected simultaneously. In this process, it is important to state specific hypotheses regarding the relationship between variables clearly. There is the expected nature of the relationship, which can be positive, negative, or there is no relationship at all. There is also the expected relative strength of the association, where stronger and clearer relationships can be shown than others (Silva et al., 2013; Sarmento and Costa, 2019).

Statistical methods have been developed to evaluate discriminant validity by determining if the correlation between two constructs is statistically substantially less than one. The correlation between two constructs is also known as inter-item correlation. A benchmark of 0.85 is often used, but some researchers propose alternative values like 0.9. In other research, a correlation of 0.75 or lower is usually seen as showing no issues with discriminant validity, whereas values close to 0.9 might suggest possible problems (Cheung et al., 2023). To deal with the high correlation between two constructs, which is higher than 0.85, the Heterotrait-Monotrait (HTMT) is introduced. According to Roemer, Schuberth and Henseler (2021), HTMT evaluates the average correlations among various constructs in comparison to those within the same construct. It examines whether two constructs are truly distinct from each other. HTMT values are generally regarded as acceptable when they fall below 0.85.

3.6 Structural Equation Modelling (SEM)

3.6.1 SEM Model Structure

This research utilises Structural Equation Modelling (SEM) as the final statistical method after PCA, EFA, and CFA are completed. SEM can confirm the factor structure identified in the previous section and identify the way the factors influence one another among UTAR freshmen. As mentioned in section 2.7, SEM is typically divided into measurement and structural models. The measurement model, which was tested using CFA in the last section, showed that the model is valid and fits the data well. Therefore, this section will focus on the structural model of SEM, which investigates how the Positive Emotion, Personal Growth, and Social Exploration factors are related and influenced by each other.

According to Civelek (2018), to better understand the structural model mathematically, the equation is expressed as below:

$$Equation\ 1: Factor\ 1 = \beta_1 \cdot Factor\ 2 + Res_1 \tag{9}$$

$$Equation\ 2: Factor\ 1 = \beta_2 \cdot Factor\ 2 + \beta_3 \cdot Factor\ 3 + Res_2 \tag{10}$$

where:

β_i refers to a matrix of coefficients among endogenous variables

Res_j refers to the residuals of the endogenous variables

Except for the distinction between latent and observable variables, variables in SEM are divided into endogenous and exogenous. In structural equation modelling, the distinction between endogenous and exogenous variables is important because a variable can simultaneously play the role of a dependent variable and an independent variable. Endogenous variables are those dependent variables explained by other variables, whereas exogenous variables are those independent variables that are not explained by other variables (Dursun & Kocagöz, 2010; Civelek, 2018). From the equations above, Factor 1 is an endogenous variable for both equations, where Factor 2 and Factor 3 are the exogenous variables.

After setting up the theoretical foundations, the next part will show this research's path diagram and model structure. After that, a series of structural model tests that look at both pairwise and combined relationships among the latent factors will be performed.

3.6.2 SEM Symbols and Measurement Model Example

According to Civelek (2018), the basic symbols used in SEM are shown in Figure 3.17.

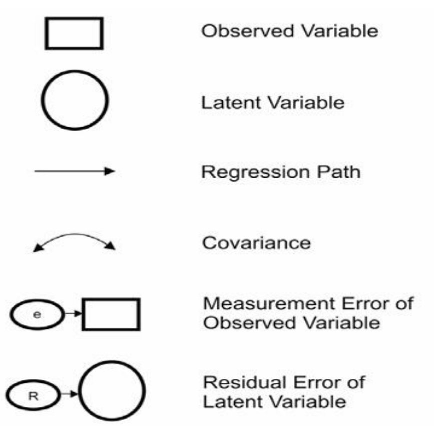


Figure 3.17: Common Diagram Symbols in SEM

The observed variables are represented as a rectangle, while the latent variable is illustrated as an oval or circle. The single-headed arrow represents a regression path to show the prediction direction from one variable to another. In contrast, the double-headed arrow indicates a covariance or correlation between variables. In SEM diagrams, measurement error is also taken into consideration. The small circle marked “e” connected to each observed variable shows its measurement error, which is the part of the variance in the item that is not explained by the latent factor. A residual error term “R” is linked to each latent variable to show the variance still unexplained by other factors in the structural model.

Figure 3.18 shows an example of an SEM path diagram adapted from Vo and Nguyen (2023).

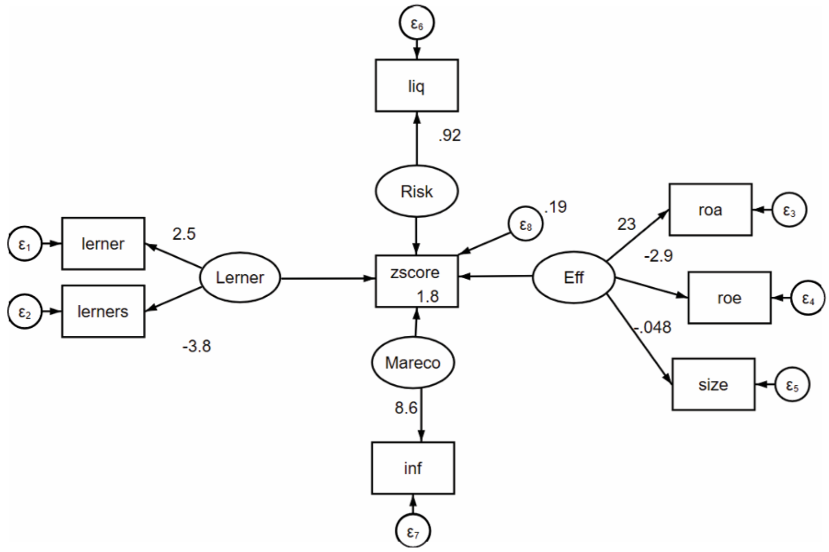


Figure 3.18: Example of a SEM Path Diagram

The diagram eases visualisation and interpretation of the full structural model, as it combines measurement and structural components in a single diagram to explain the complex relationships among the variables.

3.6.3 Structural Models: 6 Pairwise Paths

For this research, six SEM models were constructed to explore the directional relationships between each pair of the three latent variables that came from earlier analyses, which are Positive Emotion, Personal Growth, and Social Exploration. Each model only checks one direction of the relationship at a time so that the changes in one factor that might be able to predict changes in another factor can be observed. These models include:

1. $Positive\ Emotion\ \rightarrow Personal\ Growth$
2. $Personal\ Growth\ \rightarrow Positive\ Emotion$
3. $Positive\ Emotion\ \rightarrow Social\ Exploration$
4. $Social\ Exploration\ \rightarrow Positive\ Emotion$
5. $Personal\ Growth\ \rightarrow Social\ Exploration$
6. $Social\ Exploration\ \rightarrow Personal\ Growth$

The goal of this pairwise approach is to see if changes in one latent construct, like Positive Emotion, can significantly explain or predict the variation in another latent construct, such as Personal Growth. This method provides a closer examination of how each pair of psychological constructs influences one another, which helps to uncover the underlying dynamics affecting the mental well-being of UTAR freshmen.

The Maximum Likelihood Robust (MLR) estimator was used to analyse these models, which is the same estimator applied in the CFA section, since the data did not meet the assumptions of multivariate normality.

The six models mentioned earlier are supported by a hypothetical example from Gupta (2024), which shows how SEM is utilised to explore the relationship between latent variables. In the example, "Customer Satisfaction" (CS) is shown as a latent variable that is measured through indicators such as "Service Quality" and "Delivery Timeliness," while "Loyalty" (L) is another latent construct evaluated by "Repeat Purchase" and "Brand Advocacy."

The author constructs the diagram structure as:

$$CS \rightarrow L$$

This hypothetical model is similar to the current research, where latent factors like Positive Emotion, Personal Growth, and Social Exploration are looked at for how they influence one another. This example serves as a useful guide for understanding similar paths within the SEM framework of this research.

3.6.4 Structural Models: 3 Combined Predictors

Beyond pairwise models, three combined models were assessed. In order to investigate the combined effects of two predictors on a single outcome, each model includes one endogenous latent variable and is predicted by two exogenous latent variables. These models are shown as below:

$$Model\ 1: Personal\ Growth \rightarrow Positive\ Emotion \leftarrow Social\ Exploration$$
$$Model\ 2: Positive\ Emotion \rightarrow Personal\ Growth \leftarrow Social\ Exploration$$
$$Model\ 3: Positive\ Emotion \rightarrow Social\ Exploration \leftarrow Personal\ Growth$$

In the case of a combined model, the structural paths between the latent variables use both left and right arrows. The single-headed arrows represent direct effect paths from exogenous to endogenous latent variables. Using Model 1 as an example, the model means that Personal Growth and Social Exploration predict Positive Emotion. This implies that Positive Emotion is an endogenous latent construct predicted by two separate latent variables, thus acting as a collider in the structural model. The structure of this relationship is underpinned by the concept of a collider, as stated by Kunicki, Smith and Murray (2023).

They illustrated that the structure indicates a joint influence of two X and Y variables on the L variable, thereby classifying L as a collider. Model 1 in the current research follows the same logic, so the Positive Emotion is a collider variable receiving simultaneous input from two predictors. In short, the pairwise models helped to research direct links between two factors, while the combined models looked at how two predictors work together to explain the changes in the outcome variable.

3.6.5 Structural Model Accuracy and Fit Evaluation

After constructing pairwise and combined SEM models, the next step was to estimate the model parameters and check how well they fit. This research used a Maximum Likelihood Robust (MLR) estimator for all SEM models. As discussed earlier, MLR was selected because it works well with large sample sizes and gives more reliable results even when the assumption of multivariate normality is not met. The model fit indices, such as RMSEA, CFI, and TLI, have been calculated and discussed in the CFA section, with all values being within acceptable ranges. The overall model fits the data well, so no adjustments to the model fit right now.

The standardised regression coefficient, β, is one of the most important measures. This number shows how strong the relationship is between two latent constructs and the direction in which they point. If the value is positive, it means that as the factor increases, the other will also increase. If the value is negative, it indicates an inverse relationship. Since the β values in this research were standardised, they are expressed in units of standard deviation, which allows for easier interpretation across different paths in the model (Civelek, 2018).

The standard error shows how much variability or uncertainty surrounds the β estimate. A small standard error indicates that the estimate is quite precise, whereas a large standard error implies more uncertainty. Standard error values play a crucial role in figuring out confidence intervals and assessing whether the estimates hold statistical significance (Civelek, 2018). When the model is applied to different datasets, the standard error may be used to determine how stable the β estimate is.

P-values are used to determine whether the estimated path relationships are statistically meaningful. In SEM, a p-value lower than 0.05 usually shows that the path is statistically significant, suggesting that the relationship between the two latent variables probably does not happen by chance. It is more likely that there is a significant relationship between the constructs if the p-value is large, the β is strong, and the standard error is low (Civelek, 2018).

In SEM, assessing the strength and quality of a model involves not just looking at path coefficients but also considering the extent to which the predictors explain the variance in the endogenous variables. The coefficient of determination, R^2, shows how much variation in a dependent variable can be explained by one or more independent variables in the model (Qin et al., 2021). For instance, in Model 1, where Personal Growth and Social Exploration were used to predict Positive Emotion, the resulting R^2 shows the extent to which the combined impacts of the other two factors may explain the variance in Positive Emotion. According to Ozili (2023), R^2 values greater than 0.5 are acceptable in social science modelling, accounting for over 50% of outcome variance, and R^2 values between 0.51 and 0.90 are empirically strong.

In short, the statistical results, which comprise standard error and p-value, show how reliable the model is and how well it explains the relationships between the factors.

3.7 Software and Coding Environment

i) R Studio

To perform the statistical analysis in this research, R programming language was used due to its flexibility and powerful statistical packages. R programming is used from the data preparation process, data standardisation, normality, reliability test, validation test and multivariate analysis like PCA, EFA, CFA and SEM. Several R packages were used throughout the analysis process:

Table 3.4: R Libraries Used for Data Processing and Analysis

Library	Description
readxl	Use to import the Excel dataset
dplyr, and tidyr	Use to select items for data manipulation
MVN	Mardia's multivariate normality test
ggplot2, and reshape2	Visualizations like correlation heatmaps, scree plots, and histograms
psych	Used to compute skewness and kurtosis, reliability measures, ML estimation and Promax rotation in EFA
GPArotation	Promax rotation in EFA
lavaan	Use for the CFA and SEM process

The complete coding will show in the Appendix H.

ii) SPSS and AMOS

To check whether the output of PCA and EFA is accurate, SPSS was used to reconstruct the whole PCA and EFA process. All variables were loaded into SPSS, and the extraction, rotation method, and number of factors to retain were chosen. After selecting the required criteria for the model, the output was generated easily by SPSS. The SPSS shows that the measures tally with the output from R Studio. Those measures included the correlation between each item, the scree plot, the total variance explained, the reliability tests under PCA and EFA, the communalities value and the factor loadings. The measurements are tally and confirm that the current research outputs are accurate. The output of SPSS will be included in Appendix I.

Next, SPSS and AMOS were used to compute the diagram for the SEM measurement model and structural model. AMOS was used to draw the diagrams using the file with the dataset from SPSS. Firstly, an oval shape was selected to represent the latent variables and linked with the questionnaire items. Then, the arrow was included to show the relationship of the latent variables. The diagram was done by labelling the variables correctly. There is no computational process in SPSS and AMOS for CFA and SEM. This is due to the selection of the estimator. Because of the multivariate normality violation, the Maximum Likelihood of robustness (MLR) was used in the current research. However, SPSS does not offer the MLR packages.

CHAPTER 4

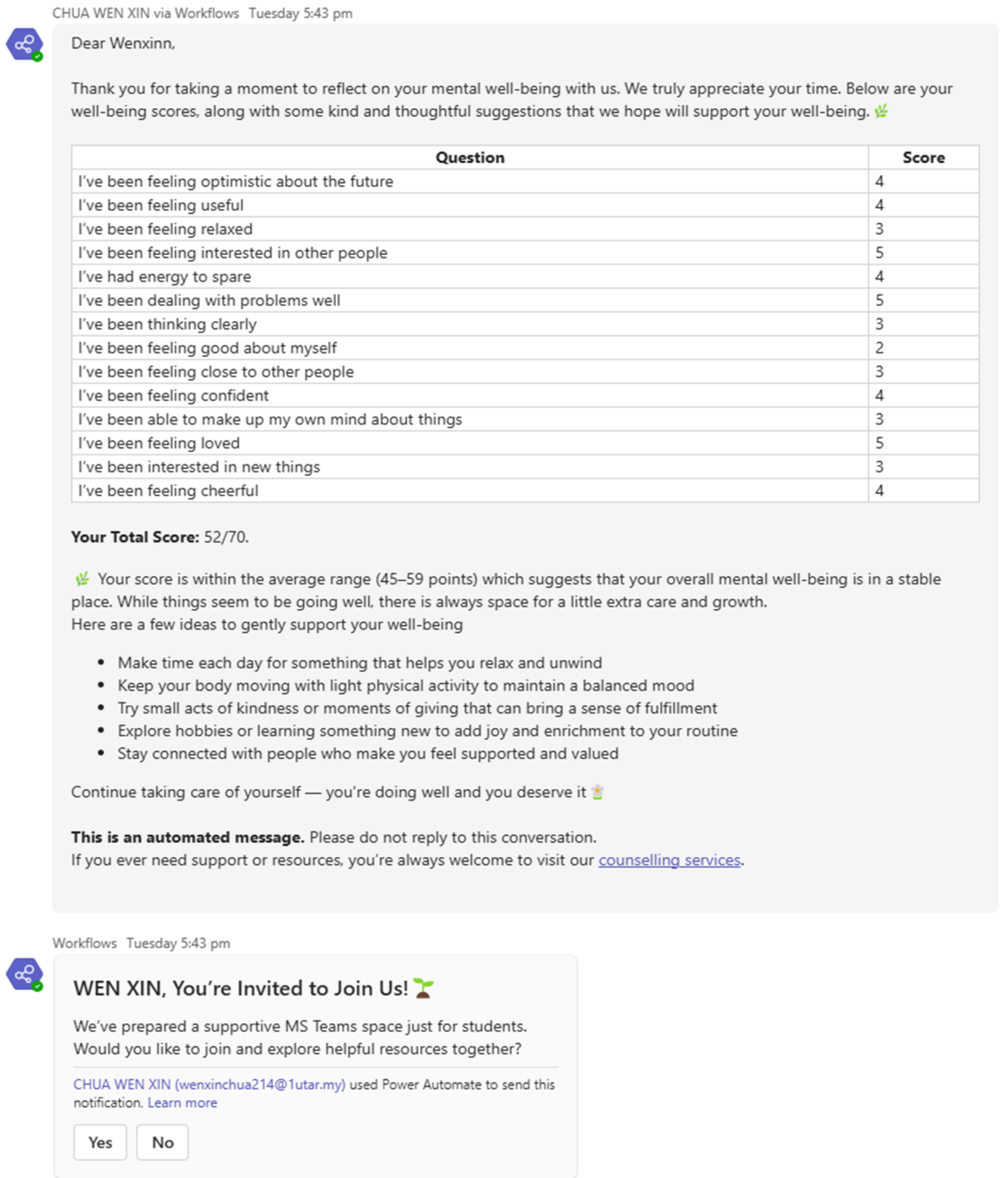
RESULTS AND DISCUSSION

This chapter outlines the research's results and examines them in the context of the research objectives. The report provides an overview of the automated system created with Microsoft Power Automate, detailing the processing of student responses, calculation of scores, and delivery of personalised feedback via Microsoft Teams. Real case outputs from the system are highlighted to demonstrate its functionality and effectiveness. The second part of the chapter examines multivariate analysis, utilising Principal Component Analysis (PCA), Exploratory Factor Analysis (EFA), Confirmatory Factor Analysis (CFA), and Structural Equation Modelling (SEM) to investigate the structure and relationships within the WEMWBS data. The findings are analysed when combined with existing literature to enhance understanding of student mental well-being and the implications for system design.

4.1 Power Automate

4.1.1 Overview of the System Output

This section shows the outputs produced by the Power Automate workflow based on different student scenarios. In each case, the system was supposed to send a personalised message via Microsoft Teams, display the answers in a table, and invite the student to the support group or end the process if there was no response. The screenshots below illustrate what occurred in each scenario. After submitting the Microsoft Forms, the student will immediately receive their scoring result, interpretation of the result according to their well-being group, and access to the UTAR counselling resource, as shown in Figure 4.1.


CHUA WEN XIN via Workflows Tuesday 5:43 pm

Dear Wenxinn,

Thank you for taking a moment to reflect on your mental well-being with us. We truly appreciate your time. Below are your well-being scores, along with some kind and thoughtful suggestions that we hope will support your well-being.

Question	Score
I've been feeling optimistic about the future	4
I've been feeling useful	4
I've been feeling relaxed	3
I've been feeling interested in other people	5
I've had energy to spare	4
I've been dealing with problems well	5
I've been thinking clearly	3
I've been feeling good about myself	2
I've been feeling close to other people	3
I've been feeling confident	4
I've been able to make up my own mind about things	3
I've been feeling loved	5
I've been interested in new things	3
I've been feeling cheerful	4

Your Total Score: 52/70.

Your score is within the average range (45–59 points) which suggests that your overall mental well-being is in a stable place. While things seem to be going well, there is always space for a little extra care and growth.
Here are a few ideas to gently support your well-being

- Make time each day for something that helps you relax and unwind
- Keep your body moving with light physical activity to maintain a balanced mood
- Try small acts of kindness or moments of giving that can bring a sense of fulfillment
- Explore hobbies or learning something new to add joy and enrichment to your routine
- Stay connected with people who make you feel supported and valued

Continue taking care of yourself — you're doing well and you deserve it

This is an automated message. Please do not reply to this conversation.
If you ever need support or resources, you're always welcome to visit our counselling services.

Workflows Tuesday 5:43 pm

WEN XIN, You're Invited to Join Us!

We've prepared a supportive MS Teams space just for students.
Would you like to join and explore helpful resources together?

CHUA WEN XIN (wenxinchua214@1utar.my) used Power Automate to send this notification. Learn more

Yes No

Figure 4.1: Personalised Feedback and Invitation via Microsoft Teams

Next, the respondent was invited to a Supportive Student Space, a Microsoft Teams Group that contains counselling resources and is created by UTAR counsellors, as shown in Figure 4.1. Note that Teams Group is an add-on feature from this research and is not yet a current practice of UTAR. It consists of three types of reactions from the respondents: they are willing to be added, not willing to be added and choose not to reply to the conversation.

Case A: Student Agrees to Join the Resource Group

After the student selects the 'Yes' Option, a notification message prompts the student that it was added to the resource group, as shown in Figure 4.2.

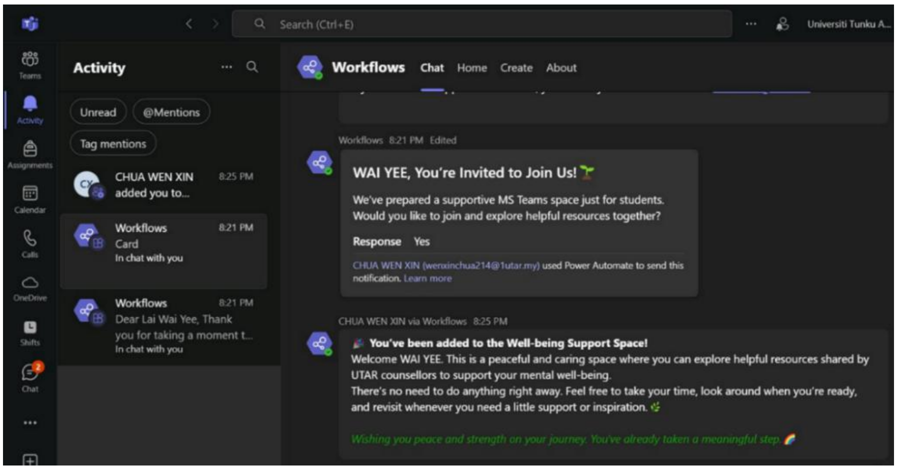


Figure 4.2: Notification of Successful Group Addition in Microsoft Teams

Case B: Student Declines to Join the Resource Group

In Figure 4.1.4, after the student selects the 'No' option, a confirmation message will be sent out to the student, and the message will consist of a hyperlink linked to the UTAR counselling booking appointment Google Form.

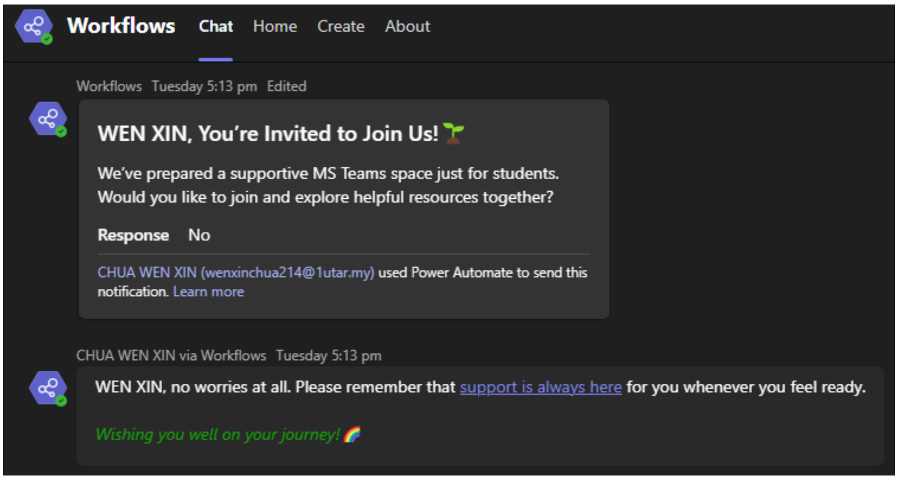


Figure 4.3: Confirmation Message with Counselling Resources for Non-Joining Students

Case C: Student Did Not Respond to the Message

In Figure 4.4, a reminder message is sent to the student after waiting for the student's response for 24 hours. The workflow also informs the student that the chat will be closed, which means the conversation will end after 24 hours.

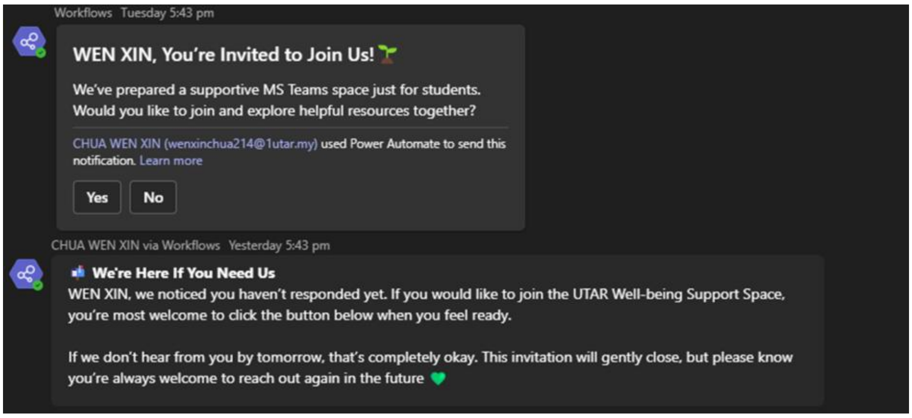
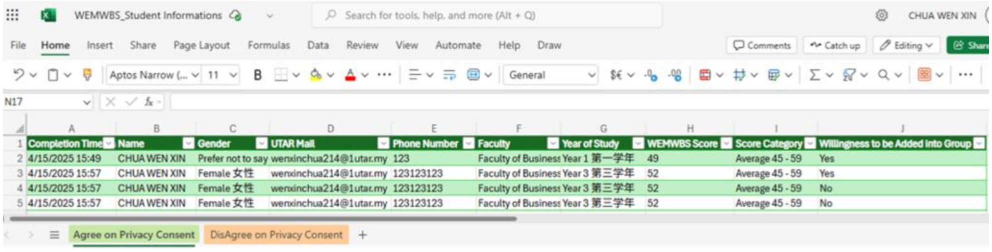


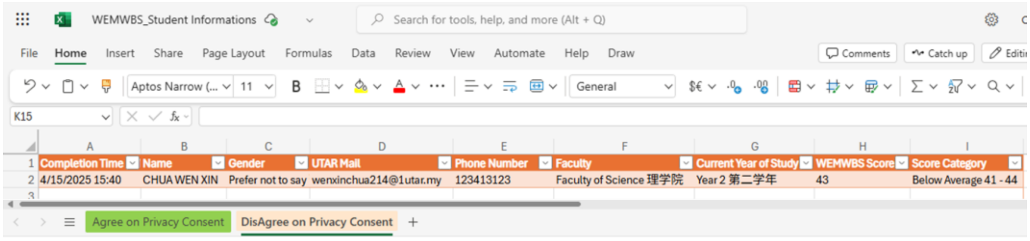
Figure 4.4: Reminder Message for Non-Responding Student

Based on the student's reactions, their response will be recorded in the Excel file automatically. If the students agree with the privacy consent, their information will be stored under the 'Agree on Privacy Consent' sheet, no matter their response. After the student selects 'Yes' or 'No', their response will be immediately recorded in the last column, indicating their willingness to join resource teams. On the other hand, if the student gets no reply, their status will only be updated after the reminder message, which means that 48 hours have passed since they submitted the Microsoft Forms, as shown in Figure 4.5. In contrast, if the student disagrees with the privacy consent, their information will be stored under 'DisAgree on Privacy Consent', as shown in Figure 4.6.



Completion Time	Name	Gender	UTAR Mail	Phone Number	Faculty	Year of Study	WEMWBS Score	Score Category	Willingness to be Added into Group
4/15/2025 15:49	CHUA WEN XIN	Prefer not to say	wenxinchua214@1utar.my	123	Faculty of Business	Year 1 第一学年	49	Average 45 - 59	Yes
4/15/2025 15:57	CHUA WEN XIN	Female 女性	wenxinchua214@1utar.my	123123123	Faculty of Business	Year 3 第三学年	52	Average 45 - 59	Yes
4/15/2025 15:57	CHUA WEN XIN	Female 女性	wenxinchua214@1utar.my	123123123	Faculty of Business	Year 3 第三学年	52	Average 45 - 59	No
4/15/2025 15:57	CHUA WEN XIN	Female 女性	wenxinchua214@1utar.my	123123123	Faculty of Business	Year 3 第三学年	52	Average 45 - 59	No

Figure 4.5: Student Information with Group Participation Response



Completion Time	Name	Gender	UTAR Mail	Phone Number	Faculty	Current Year of Study	WEMWBS Score	Score Category
4/15/2025 15:40	CHUA WEN XIN	Prefer not to say	wenxinchua214@1utar.my	123413123	Faculty of Science 理学院	Year 2 第二学年	43	Below Average 41 - 44

Figure 4.6: Student Records with Disagreement on Privacy Consent

4.1.2 Workflow Insights and User Interaction

The test results from each case scenario indicated that the Power Automate system functioned as expected. Previous studies indicate that providing personalised feedback immediately following depression screening can enhance patient involvement and lessen the severity of depression. In particular, automated and customised responses can encourage users to seek help (Sikorski et al., 2021).

Students could join the support group by clicking 'Yes' or 'No' in the workflow. Research indicates that when digital mental health tools incorporate friendly and empathetic messages while fostering a sense of connection, users may feel more at ease and inclined to share their feelings. Although the chatbot did not directly affect self-disclosure, the social presence made users feel more supported and motivated to engage and share more personal thoughts (Bhatt, 2024).

Additionally, Groot et al., (2022) show that interventions that require little engagement, mainly once or four times the total number of interventions, offer more optimal engagement. This is because people are often most responsive at the beginning of an intervention, and their interest tends to decrease over time. Because of that, this research keeps the conversation short and does not send repeated reminders if a student does not respond.

Overall, Power Automate workflow offered students a caring and timely response while allowing them to engage comfortably. Rather than replacing human support, the workflow was designed to create a gentle first step that bridges the gap between completing a screening and receiving help. It is to emphasise that help is always available for the students. Counsellors can always approach the student if they feel further intervention is needed by referring to the Excel File, which particularly agrees on privacy consent, but refuses to join the group.

4.2 Principal Component Analysis (PCA)

4.2.1 Suitability Test for PCA

To determine the suitability of the WEMWBS dataset adequacy to perform PCA analysis, a few tests were conducted in Chapter 3.

The determinant of the correlation matrix computed was 0.0017. Since the determinant score exceeds 0.00001, multicollinearity is not present, meaning the 14 items do not overlap excessively. Referring to the correlation matrix in the figure below, there are no correlations greater than 0.8, which means no change or removal of variables is needed in this case.

Figure 4.7: Correlation Matrix of the 14 WEMWBS Items

Besides, the Average Inter-Item Correlations (AIIC) were calculated, and they show 0.424, which falls within the acceptable benchmark range from 0.15 to 0.50. This value suggests that the 14 items are moderately correlated, strongly enough to indicate the consistency among items measuring student well-being but not too high as a redundancy. This supports the assumption that the items are related yet distinct, contributing uniquely to the underlying construct.

Bartlett's Test of Sphericity provided a highly significant Chi-square value of 16674.52 with a *p*-value of 0.0000. Since the p-value is less than 0.05, this indicates that the null hypothesis is strongly rejected. The correlation matrix is adequate for factor analysis as the correlation matrix of the variables significantly differs from an identity matrix. Additionally, the Kaiser-Meyer-Olkin (KMO) test resulted in an overall score of 0.9368, a value mostly approaching 1, which suggested a very high adequacy of the sample size for performing factor analysis.

4.2.2 PCA Eigenvalues and Component Retention

The Principal Component Analysis (PCA) generated 14 principal components (PCs), each associated with a distinct eigenvalue and the variance explained by each retained component was analysed and shown in Table 4.1.

Table 4.1: PCA Eigenvalues and Variance Explained

Principal Component	Eigenvalue	Variance Explained (%)	Cumulative Variance Explained (%)
1	6.5830	47.00	47.00
2	1.1342	8.10	55.10
3	0.8649	6.18	61.28
4	0.7475	5.34	66.61
5	0.6523	4.66	71.27
6	0.5963	4.26	75.53
7	0.5731	4.09	79.62
8	0.5503	3.93	83.55
9	0.4855	3.47	87.02
10	0.4530	3.23	90.25
11	0.3726	2.66	92.91
12	0.3651	2.61	95.52
13	0.3387	2.42	97.94
14	0.2888	2.06	100.00

Figure 4.8 shows the scree plot against eigenvalues and the number of principal components.

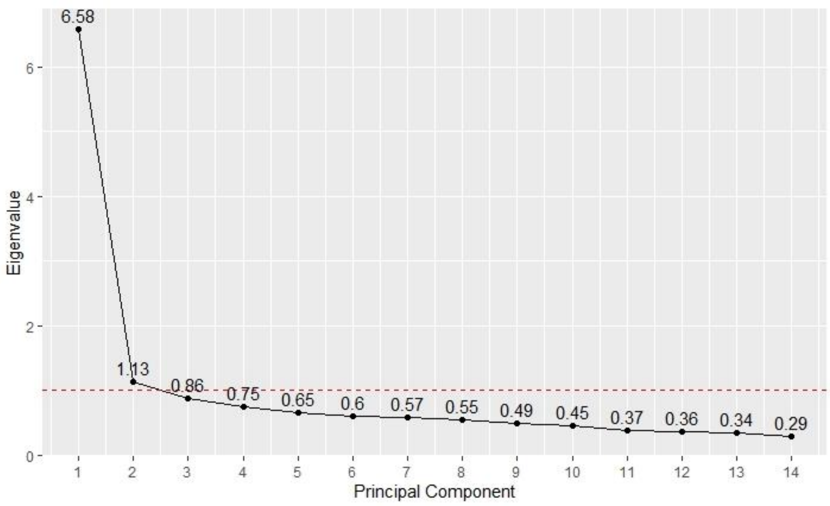


Figure 4.8: Scree Plot of Principal Components

According to Kaiser's criterion, the factors with eigenvalues greater than one should be retained. Hence, in this research, the first two factors would be retained by referring to the eigenvalues of the principal components. However, based on Weismer et al., (2021), it is justifiable to maintain components with eigenvalues slightly below one if they significantly enhance the interpretative clarity and comprehensiveness of the study. Therefore, the third principal component with an eigenvalue of 0.8649 was preserved in this research.

Besides, the variance explained by each retained component was analyzed According to Wasdahl (2024), the first factor of the highest eigenvalue suggests that this factor will be the most influential in explaining the variability for the dataset. The first principal component accounted for the majority of the variance, explaining 47.00% of the total variance. The second component explained an additional 8.10% and the third principal component explained approximately 6.18% of the total variance. These three factors together accounted for 61.28% of the total variation, suggesting that they sufficiently represented important factors of students' well-being.

4.2.3 Cross-Validation of PCA Results

Furthermore, a 10-folds cross-validation was employed to validate the PCA outcome and confirm the robustness of these three components. The results in Table 4.2 showed consistent and stable patterns.

Table 4.2: 10-Folds Cross-Validation Results for PCA

Folds	Explained Variance Ratio	Cumulative Training Accuracy (%)
1	0.47053863, 0.08103842, 0.06282876	61.44
2	0.47478241, 0.08024362, 0.06163263	61.67
3	0.47128907, 0.08016686, 0.06153318	61.3
4	0.47038410, 0.08181380, 0.06221353	61.44
5	0.47098099, 0.08141808, 0.06152963	61.39
6	0.46987890, 0.08068610, 0.06127117	61.18
7	0.46506023, 0.08243363, 0.06229347	60.98
8	0.46576575, 0.08051732, 0.06152541	60.78
9	0.47424979, 0.08000133, 0.06163227	61.59
10	0.46747653, 0.08191578, 0.06166049	61.11

Specifically, Fold 1 exhibited an explained variance ratio of 47.10% for PC1, 8.10% for PC2 and 6.28% for PC3, resulting in a cumulative explained variance of 61.44%. Upon aggregating results across all ten folds, the average explained variance ratio for PC1, PC2, and PC3 were 47.00%, 8.10% and 6.18% respectively. This resulted in an overall cumulative explained variance of 61.29%, reflecting strong consistency and reliability of PCA results through 10-folds cross-validation as it just differs from the origin cumulative explained variance of 0.01%.

4.3 Exploratory Factor Analysis (EFA)

4.3.1 Normality Check

Univariate normality was computed through histograms, skewness, and kurtosis for each of the 14 items of WEMWBS. The histogram, as shown in Figure 4.9, demonstrated approximately symmetrical distribution as it was distributed in a bell shape. Besides, the histogram for all 14 items showed left-skewed, meaning the tail is on the left side of the distribution, it also known as negatively skewed distributions. This is because a majority of UTAR first-year students selected higher response options, which are scored 4 or 5 in the Likert-type WEMBWS questionnaire. The pattern further suggests that the students generally reported high levels of mental well-being across the WEMWBS items.

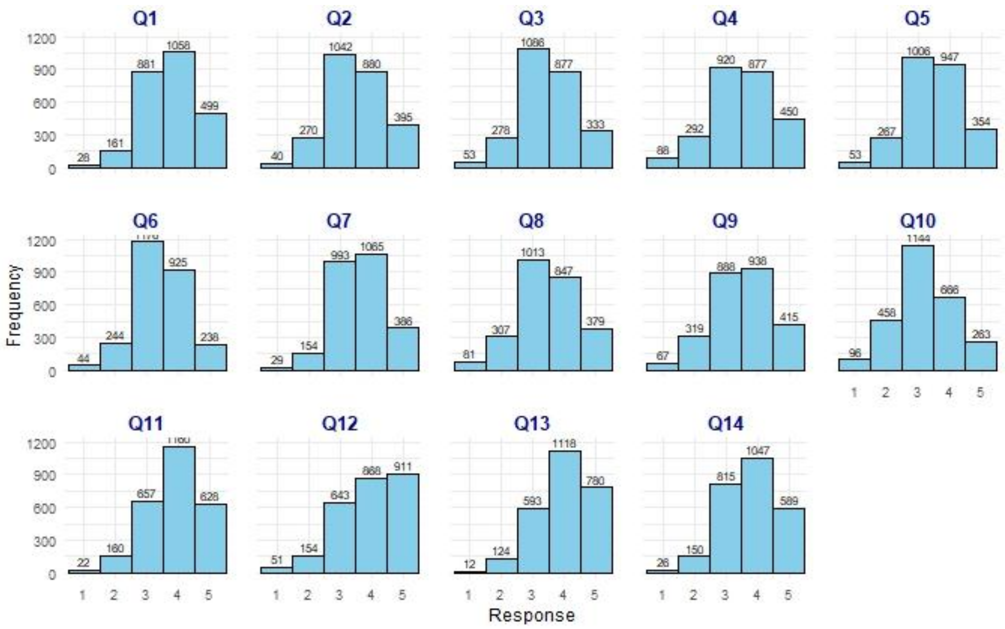


Figure 4.9: Histogram Distribution for All 14 WEMWBS Items

To further confirm the univariate normality, the skewness and kurtosis are calculated and shown in Table 4.3.

Table 4.3: Skewness and Kurtosis of WEMWBS Items

Question	Skewness	Kurtosis
1	-0.289	-0.174
2	-0.117	-0.326
3	-0.141	-0.162
4	-0.31	-0.309
5	-0.227	-0.161
6	-0.138	0.114
7	-0.21	-0.014
8	-0.219	-0.259
9	-0.291	-0.349
10	0.002	-0.27
11	-0.504	-0.062
12	-0.676	-0.103
13	-0.532	-0.184
14	-0.345	-0.235

The skewness values range from -0.676 to 0.002, with the lowest skewness value from question 12 and the highest from question 10. On the other hand, the kurtosis values range from -0.349 to 0.114, with question 9 showing the lowest kurtosis value and question 6 showing the highest. The skewness values indicate a minimal deviation from symmetry, whereas the kurtosis values indicate the responses were neither excessively peaked nor over-flat. All the skewness and kurtosis values fall within the acceptable benchmark range of -1 to +1, proving that the dataset achieves univariate normality.

After achieving univariate normality, multivariate normality was evaluated using Mardia's test. Both of the results of multivariate skewness and kurtosis show a p-value of less than 0.001. This indicates that the null hypothesis is rejected, and the dataset fails to meet the assumption of multivariate normality.

4.3.2 Factor Loadings

Based on Principal Component Analysis (PCA), 14 items were composed and reduced to 3 principal components. Exploratory Factor Analysis (EFA) then further discovers the underlying constructs of these 3 latent factors based on the factor loadings. The factor loadings of 14 questions are shown in Figure 4.3.2.1.

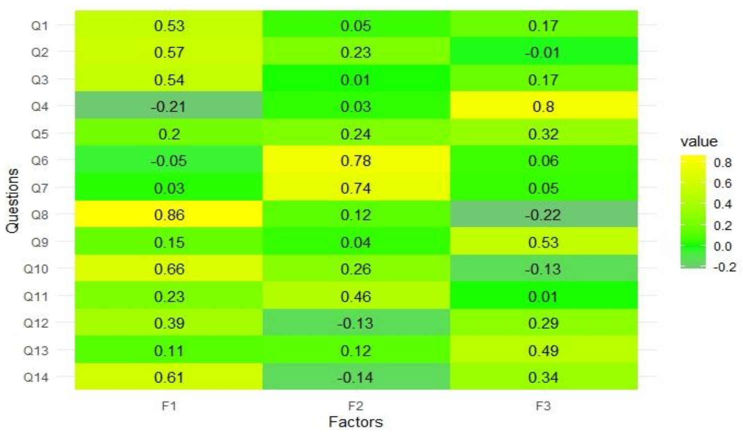


Figure 4.10: Factor Loadings for WEMWBS Items

The factor loadings of each question achieved the minimum benchmark, 0.3, therefore, all the items are retained in the research. The question with the highest factor loading indicates that the question falls in that factor. Factor 1 consists of items Q1, Q2, Q3, Q8, Q10, Q12, and Q14. Factor 2 consists of items Q6, Q7, and Q11. Factor 3 consists of items Q4, Q5, Q9, and Q13.

Table 4.4 displays the original WEMWBS items along with the factor groups that relate to them.

Table 4.4: WEMWBS Items Categorised by Exploratory Factor Group

Question		Factor
1	I've been feeling optimistic about the future	1
2	I've been feeling useful	
3	I've been feeling relaxed	
8	I've been feeling good about myself	
10	I've been feeling confident	
12	I've been feeling loved	
14	I've been feeling cheerful	
6	I've been dealing with problems well	2
7	I've been thinking clearly	
11	I've been able to make up my own mind about things	
4	I've been feeling interested in other people	3
5	I've had energy to spare	
9	I've been feeling close to other people	
13	I've been interested in new things	

4.3.3 Factor Naming and Interpretation

Since WEMWBS is a well-being screening test, the factors will be named based on the concept of well-being. According to Konaszewski et al., (2021), mental well-being involves both hedonic and eudaimonic concepts. According to

Franken et al., (2018), eudaimonic well-being can be explained by the combination of psychological and social well-being.

According to Konaszewski et al., (2021), WEMWBS assesses both hedonic and eudemonic aspects of well-being. Hedonic well-being focuses on positive emotions, joy, happiness, interest, and contentment. In contrast, eudaimonic well-being refers to the psychological functioning associated with self-acceptance, personal growth, autonomy, positive relationships, mastery, and a sense of purpose in life. Social well-being also refers to social integration, acceptance, coherence, contribution, and actualization (Franken et al., 2018).

Factor 1 comprises “optimistic”, “useful”, “relaxed”, “good about myself”, “confident”, “loved”, and “cheerful”. The similarities of all the factors are related to positive emotions, which are part of hedonic well-being.

Factor 2 comprises “dealing with problems well”, “thinking clearly”, and “able to make up my own mind about things”. Those questions focus more on personal growth aspects, part of eudaimonic well-being.

Factor 3 comprises “feeling interested in other people”, “had energy to spare”, “feeling close to other people”, and “interested in new things”. These questions are related to social exploration, which is part of social well-being under eudaimonic well-being.

Therefore, the three factors are named as shown in Table 4.5.

Table 4.5: WEMWBS Items Grouped Under Named Factors

Question		Factor
1	I've been feeling optimistic about the future	Positive Emotion
2	I've been feeling useful	
3	I've been feeling relaxed	
8	I've been feeling good about myself	
10	I've been feeling confident	
12	I've been feeling loved	
14	I've been feeling cheerful	
6	I've been dealing with problems well	Personal Growth
7	I've been thinking clearly	
11	I've been able to make up my own mind about things	
4	I've been feeling interested in other people	Social Exploration
5	I've had energy to spare	
9	I've been feeling close to other people	
13	I've been interested in new things	

4.3.4 Communalities

Communalities were computed and shown in Figure 4.11 to identify the items that shared adequate common variance within their respective factors.

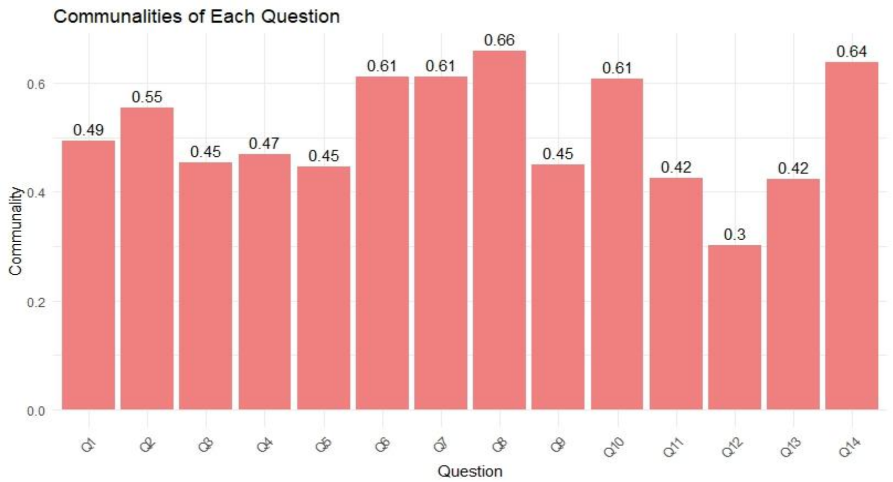


Figure 4.11: Communalities of WEMWBS Items

The communalities ranged from 0.30 to 0.66, with question 12 showing the lowest value of communalities where the question 8 showing the highest value. Questions with high communalities indicate that a large portion of their variance is explained by the underlying latent factors. For example, Q8 ("I've been feeling good about myself") is strongly explained and tied to the core dimensions of Factor 1 (Positive Emotion).

Conversely, Q12 ("I've been feeling loved") with low communality reflects that less of its variance is explained by Factor 1 (Positive Emotion). It makes sense that the Q8 ("I've been feeling good about myself") fully depends on oneself, which is how a person evaluates themselves based on self-satisfaction and self-relationship. On the other hand, Q12 ("I've been feeling loved") depends on how the person interprets the relationships between their loved ones, which can be affected by others' actions.

All the items achieved the benchmark, as all were greater than 0.3, and the majority exceeded 0.4, indicating a good alignment between the items and extracted latent factors. Although question 8 showed a relatively lower communality value, it was retained due to its theoretical significance and the validated status of the WEMWBS questionnaire. This will ensure that the integrity of the questionnaire is maintained for the current research.

4.3.5 Reliability Analysis

To examine the internal consistency within each factor, Cronbach's Alpha, McDonald's Omega, and Composite Reliability (CR) were calculated and shown in Table 4.6.

Table 4.6: Reliability Coefficients for Each Identified Factor

Factor	Cronbach's Alpha	McDonald's Omega	Composite Reliability
1	0.868	0.872	0.799
2	0.778	0.783	0.704
3	0.747	0.748	0.628

These reliability scores exceeded the benchmark, as all of the values for Cronbach's Alpha and McDonald's Omega are greater than 0.7, whereas the Composite Reliabilities are greater than 0.6.

Factor 1 (Positive Emotion) shows the highest scores among Factor 2 (Personal Growth) and Factor 3 (Social Exploration) for all of the calculated reliability tests. This means that the items in factor 1 are very consistent with each other. Interestingly, the communalities section mentioned that the lowest communality value, Q12, falls under Factor 1, which means Factor 1 does not explain Q12 very well. Nevertheless, the presence of the Q12 did not significantly lower the overall reliability of Factor 1. This is likely because the other items in Factor 1 (Positive Emotion) were very consistent and strongly related to each other, which helped to maintain the overall consistency and reliability of Factor 1. In simpler words, even if one item was weakly performed or linked to the group, the other well-performed items will be able to show a high-reliability score.

4.4 Confirmatory Factor Analysis (CFA)

4.4.1 CFA Model Fit Evaluation

Based on the suggested three-factor model that was obtained from the exploratory factor analysis (EFA), confirmatory factor analysis (CFA) was carried out. The three factors are positive emotion, personal growth, and social exploration. Several fit indices were reviewed to see if the model fit the WEMBWS data from UTAR freshmen. Table 4.7 shows the results of fit indices.

Table 4.7: Fit Indices for the Confirmatory Factor Analysis Model

Fit Indices	Statistic	Value		Fit of the Measurement Model
Absolute Fit Indices	GFI	> 0.90	Satisfied	0.927
	RMSEA	> 0.10 $[0.08, 0.10]$ $[0.05, 0.08]$ ≤ 0.05	Bad Suffering Good Very good	0.081
Increment Fit Indices	CFI TLI NFI	< 0.80 $[0.80, 0.90]$ $[0.90, 0.95]$ ≥ 0.95	Bad Suffering Good Very good	0.923 0.905 0.919
	RFI	The better the closer to 1		0.900

The Root Mean Square Error of Approximation (RMSEA) is 0.081, which is under the suffering category, but it is close to the good range ($\leqslant$ 0.08). This suggests that the model is almost acceptable according to this measure. At the same time, the Goodness of Fit Index (GFI), Comparative Fit Index (CFI), Tucker-Lewis Index (TLI), Normed Fit Index (NFI), and Relative Fit Index (RFI) are all in the good range. Thus, the overall model fit is acceptable.

4.4.2 Convergent Validity for CFA

Convergent validity was evaluated using standardised factor loadings, Average Variance Extracted (AVE), and Composite Reliability (CR) to evaluate the effectiveness of the items in measuring their corresponding latent variables.

Standardised factor loadings for each question of the WEMWBS were computed and shown in Table 4.8.

Table 4.8: Standardised Factor Loadings by Factor

Factor	Question	Standardised factor loadings
Positive Emotion	1	0.710
	2	0.748
	3	0.671
	8	0.763
	10	0.752
	12	0.505
	14	0.753
Personal Growth	6	0.757
	7	0.787
	11	0.673
Social Exploration	4	0.573
	5	0.692
	9	0.665
	13	0.654

The standardised factor loadings for the 14 items of WEMWBS are all above 0.5, which shows that each item is acceptable and contributes meaningfully in its respective factor. The standardised factor loadings were between 0.505 and 0.763, with the lowest loadings coming from the Factor Positive Emotion and the highest loadings coming from the Factor Personal Growth.

Item Q7 (“I've been thinking clearly”) showed the highest factor loadings, indicating that this question strongly matches the Personal Growth factor. This shows that having a clear understanding is crucial to how UTAR first-year students reflect on their abilities and make sense of their personal development.

In contrast, item Q12 (“I've been feeling loved”) had the lowest loading, which is part of the Positive Emotion factor. Even though it is above the acceptable benchmark of 0.50, its lower loading indicates that it plays a less significant role in the factor. This aligns with previous EFA findings, where Q12 also showed the lowest communality value of 0.30, meaning that only a small part of its variance was explained by the factor. Even though it had a lower loading, Q12 was kept because it is important for measuring positive emotions and helps maintain the original WEMWBS structure.

Average Variance Extracted (AVE) and Composite Reliability (CR) were computed and shown in Table 4.9.

Table 4.9: AVE and CR by Factor

Factor	AVE	CR
Positive Emotion	0.498	0.872
Personal Growth	0.548	0.784
Social Exploration	0.420	0.742

Although the AVE of Positive Emotion and Social Exploration are below the recommended benchmark of 0.5, according to Suprapto, Stefany, and Ali (2020), a value of 0.4 for AVE can still be considered acceptable if the CR is above 0.6. In this case, three factors showed CR values well above 0.7, meaning all factors have a strong internal consistency. Therefore, all factors are still acceptable for convergent validity across all three constructs.

The lowest AVE value for the Social Exploration factor suggests that the items under this factor do not share as much common variance as those in other factors. This may be because of the Social Exploration factor, composed of a broader nature of the items compared to Personal Growth and Positive Emotion, as it included “feeling interested in other people”, “had energy to spare”, “feeling close to other people”, and “interested in new things”. Although these items are loaded together as a single factor, they may reflect slightly different aspects of social functioning. For example, curiosity and social engagement can be influenced by their friends, cultural factors, and personal social energy. These may contribute to the lower shared variance reflected in the lower AVE value in the Social Exploration factor.

However, the Composite Reliability (CR) values for all three factors are above the recommended benchmark of 0.70. This indicates that despite the slightly lower convergence in AVE, the items still respond consistently as a factor and show strong internal consistency and reliability in each factor. The Positive Emotion factor showed the highest CR value at 0.872, meaning there is strong internal consistency among its items. This may be due to the strong standardised factor loadings supporting the strong CR result. Five out of seven items (Q1, Q2, Q8, Q10, and Q14) had loadings above 0.70, while the other two items (Q3 and Q12) still met the acceptable threshold. High factor loadings indicate that each item captures the underlying constructs of positive emotion. The high item loadings directly affected the calculation of CR, leading to a consistently high CR value. This finding shows that the items in the Positive Emotion factor are connected and consistently measure the same part of well-being like “optimistic”, “useful”, “relaxed”, “good about myself”, “confident”, “loved”, and “cheerful” for UTAR freshmen.

To sum up, the findings from the standardised factor loadings, AVE and CR all indicate acceptable convergent validity for the three factors in the model. Even though there are slightly lower AVE values, the consistently high factor loadings and CR values over 0.70 show that the items in each construct are well-aligned and effectively measure their intended aspects of well-being.

4.4.3 Discriminant Validity for CFA

Discriminant validity checks if each factor in the model is unique, ensuring that the items from different constructs are not too similar to each other. It was conducted by evaluating inter-item correlations and the Heterotrait-Monotrait Ratio (HTMT).

Table 4.10: Inter-item Correlation and HTMT Ratios Between Factors

Factor	Inter-item correlation	Interpretation	HTMT Ratio
Positive Emotion and Personal Growth	0.817	Acceptable (≤ 0.85)	-
Positive Emotion and Social Exploration	0.868	HTMT test needed (≥ 0.85)	0.742
Personal Growth and Social Exploration	0.753	Acceptable (≤ 0.85)	-

Table 4.10 shows that both the Personal Growth and Social Exploration factors and the Positive Emotion and Personal Growth factors have correlation values below 0.85. When the value of the inter-item correlation is less than 0.85, it means that the two factors are sufficiently separate, meaning the items under each factor measure different aspects of psychological well-being. This indicates strong discriminant validity and demonstrates that the model effectively captures various unique dimensions like Positive Emotion, Personal Growth, and Social Exploration in UTAR freshmen.

However, the correlation between Positive Emotion and the Social Exploration factor is slightly above the benchmark of 0.85. This suggests that the two factors may be too closely related, meaning their items might measure similar aspects or consist of overlapping content. It raised concerns about whether this correlation can carry out good discriminant validity. Therefore, the Heterotrait-Monotrait Ratio (HTMT) was used to confirm whether these two factors are distinct enough. The computed HTMT value presented in the table is less than 0.85, indicating that the correlations between Positive Emotion and Social Exploration satisfy the criteria for acceptable discriminant validity. In short, the results from the convergent and discriminant validity indicate that the Positive Emotion, Personal Growth, and Social Exploration factors are statistically valid, well-defined, and conceptually different constructs for evaluating psychological well-being among UTAR freshmen.

4.5 Structural Equation Modelling (SEM)

4.5.1 Measurement Model

As mentioned in the previous section, Structural Equation Modelling (SEM) consists of a measurement and a structural model. The measurement model refers to the Confirmatory Factor Analysis (CFA) and its validity. The CFA results are valid and reliable for the three latent variables. The measurement model diagram, as shown in Figure 4.12, is a graphical method to show the standardised factor loadings and the inter-item correlation between those three latent variables discussed in the previous sections.

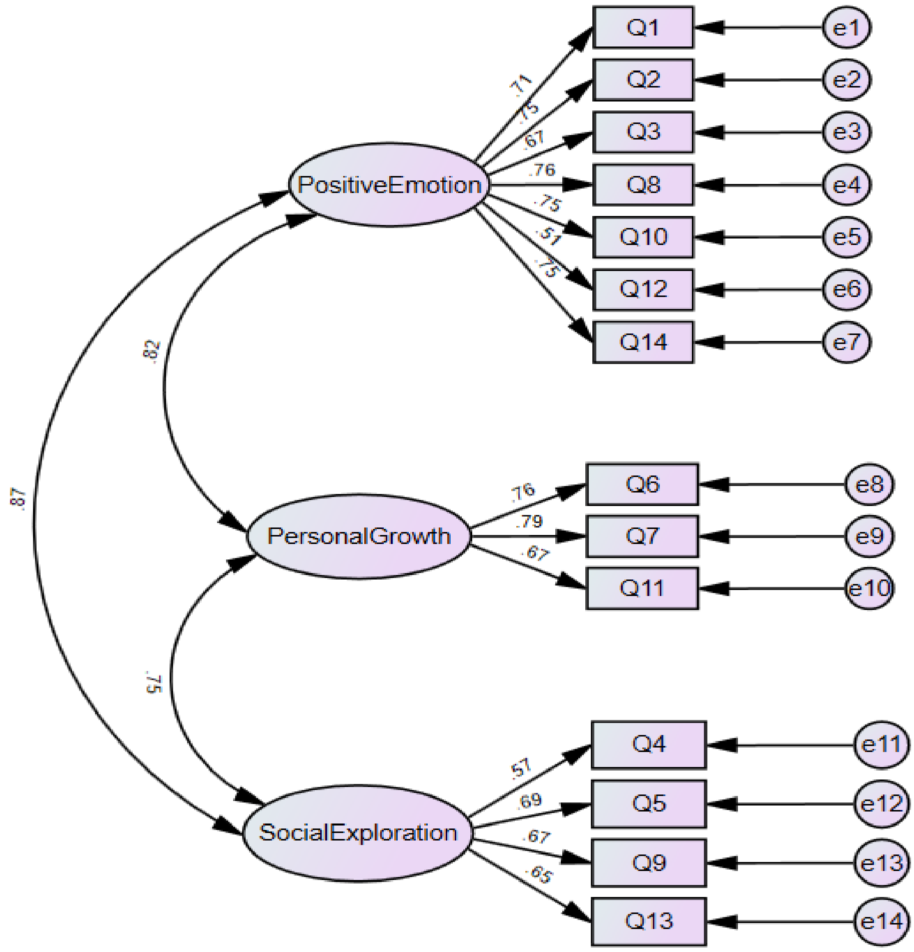


Figure 4.12: Measurement Model for Confirmatory Factor Analysis

Besides, the model fit indices discussed in CFA show that all of the statistical measurements fall in the acceptable range, and the model shows good convergent and discriminant validity. Thus, the models have been proven suitable for further analysis.

4.5.2 Pairwise Structural Path Models

The pairwise structural path analysis investigates the relationship between the three latent factors separately. Six combinations of models with three factors are involved, so six models were tested in Table 4.11.

Table 4.11: Pairwise Structural Path Estimates Between Latent Factors

Model	Structural Path	β	SE	p-value	R^2
1	Positive Emotion → Personal Growth	0.825	0.012	0.000	0.681
2	Personal Growth → Positive Emotion	0.908	0.009	0.000	0.825
3	Positive Emotion → Social Exploration	0.875	0.011	0.000	0.765
4	Social Exploration → Positive Emotion	0.929	0.009	0.000	0.862
5	Personal Growth → Social Exploration	0.872	0.013	0.000	0.761
6	Social Exploration → Personal Growth	0.837	0.013	0.000	0.700

The standardised coefficients (β) were between 0.825 and 0.929, showing strong positive correlations among all pairs. In Model 1, β =0.825 indicates that if UTAR freshmen experience a one standard deviation increase in Positive Emotion, it will lead to an approximate increase of 0.825 standard deviations in Personal Growth.

Besides, the standard errors (SE) for all paths were between 0.009 and 0.013, which are minimal values. This indicates that the estimated path coefficients (β) are stable and accurate. A lower standard error indicates that the β value is much closer to the actual value in the population. In Model 2, the standard error is 0.009. This shows that the connection between Personal Growth and Positive Emotion is reliable, as that outcome has a small variation.

Furthermore, the 0.000 p-value of all paths shows that all the structural paths are statistically significant. For instance, in Model 3, the p-value supports the conclusion that Positive Emotion significantly predicts Social Exploration, and this result is not due to coincidence. Next, the R^2 values in this research for the six pairwise models were between 0.681 and 0.862, above the benchmark of 0.51 in social science research, and this is considered strong and acceptable. For example, in Model 4, Social Exploration predicts Positive Emotion with an R^2

value of 0.862, which means variations in Social Exploration can explain 86.2% of the changes in Positive Emotion.

Overall, the results indicate that Positive Emotion, Personal Growth, and Social Exploration are closely linked, suggesting that well-being is not an isolated concept. If one of these factors had been improved, it might have led to improvements in the other factor as well. For example, a student with a high self-confidence level leads to emotional positivity, influencing the student's participation in university events and making new friends. By exploring new things, the student might discover their interests and continue further learning. The self-learning process is part of personal growth, forming a cycle in which personal growth influences emotional positivity.

4.5.3 Combined Predictor Structural Path Models

After analysing the pairwise relationships between the three latent factors, this section explores how two predictors influence a single outcome in a combined model structure. These models help to better understand whether two aspects of student well-being might interact to influence another area of development. In this research, three combined models were tested.

Table 4.12: Combined SEM Path Estimates with Dual Predictors

Model	Paths	β	SE	p-value	R^2
1	Personal Growth → Positive Emotion	0.379	0.031	0.000	0.815
	Social Exploration → Positive Emotion	0.583	0.031	0.000	0.815
2	Positive Emotion → Personal Growth	0.665	0.058	0.000	0.676
	Social Exploration → Personal Growth	0.176	0.062	0.004	0.676
3	Positive Emotion → Social Exploration	0.761	0.044	0.000	0.759
	Personal Growth → Social Exploration	0.131	0.047	0.006	0.759

$Model\ 1: Personal\ Growth\ \rightarrow Positive\ Emotion\ \leftarrow Social\ Exploration$

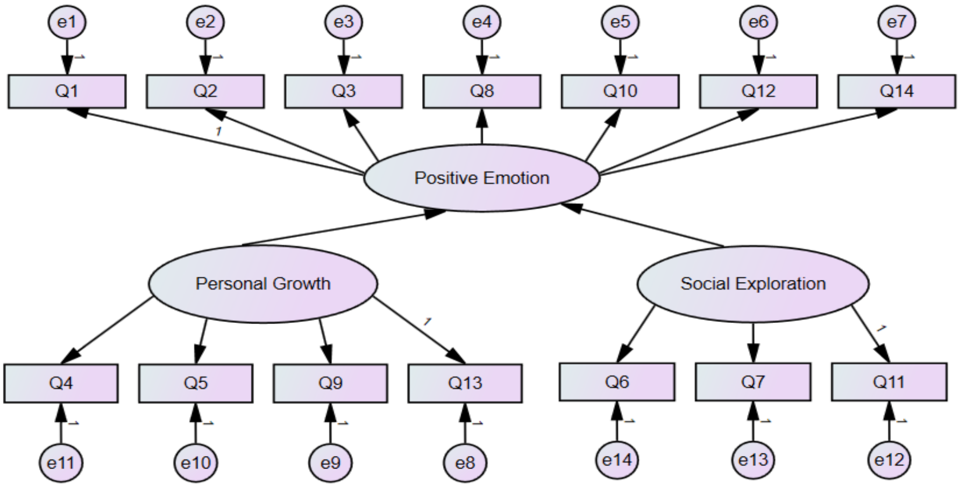


Figure 4.13: Personal Growth and Social Exploration Predict Positive Emotion

In this model, Personal Growth with β = 0.379 and Social Exploration with β = 0.583 were significant predictors of Positive Emotion, with an R^2 = 0.815. This means these two factors explain 81.5% of students' emotional positivity variation, which is an extreme explanatory power. Notably, Social Exploration had a higher β, suggesting that the energy students get from feeling connected to others, being curious, and having social energy plays a slightly stronger role in shaping their emotional positivity than personal growth.

$Model\ 2: Positive\ Emotion\ \rightarrow Personal\ Growth\ \leftarrow Social\ Exploration$

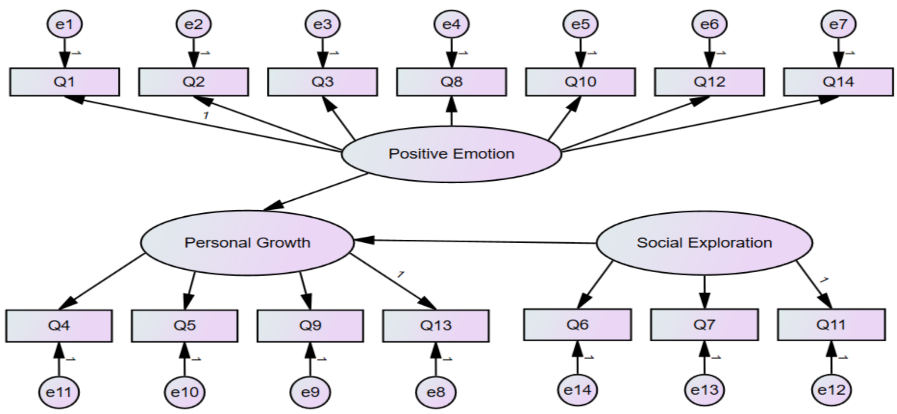


Figure 4.14: Positive Emotion and Social Exploration Predict Personal Growth

In this model, Positive Emotion with β = 0.665 and Social Exploration with β = 0.176 predicted Personal Growth, with an R^2 = 0.676. That means the model explains 67.6% of the changes in Personal Growth. Positive Emotion has a much stronger beta, while Social Exploration's influence is weaker on Personal Growth. This indicates that personal growth, like the ability to deal with problems, is more influenced by students' internal feelings of usefulness and confidence than their social interactions.

$Model\ 3: Positive\ Emotion \rightarrow Social\ Exploration\ \leftarrow Personal\ Growth$

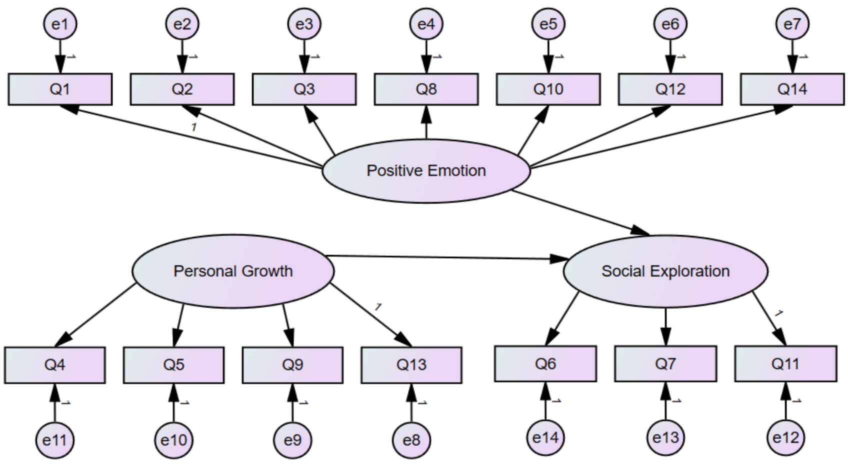


Figure 4.15: Positive Emotion and Personal Growth Predict Social Exploration

In this model, Positive Emotion with β = 0.761 and Personal Growth with β = 0.131 predicted Social Exploration, with an R^2 = 0.759. That means the model explains 75.9% of the variability in Social Exploration. The β values show that Positive Emotion has a more significant impact than Personal Growth on enhancing students' Social Exploration. This is reasonable, as individuals are more inclined to engage, connect, and explore when they feel emotionally positive and confident.

The combined models demonstrate that even when both predictors are analysed together, each carries its weight, as some are stronger than others, even when they work together. Based on the combined SEM models, Positive Emotion was identified as the most significant factor among the three dimensions of well-being. The results indicated significant predictive effects on Personal Growth (β = 0.665) and Social Exploration (β = 0.761), suggesting that students with higher levels of emotional positivity are more inclined to experience personal development and participate in social activities.

To sum up, the SEM pairwise models showed the relationship between each latent factors, demonstrating how improvement in one area can positively affect others. The combined models offered a more detailed perspective by demonstrating how two factors can work together to influence a third, which helps to identify which influences are more significant. Together, these results provide a complete picture of how UTAR freshmen's well-being interacts.

CHAPTER 5

CONCLUSION

5.1 Conclusion

Overall, the research has enhanced the mental health support for UTAR students by integrating digital automation and psychological data analysis. The Warwick–Edinburgh Mental Well-being Scale (WEMWBS) was used to create an automated workflow using Microsoft Power Automate. The system calculates scores, sorts well-being categories, and provides personalised feedback via Microsoft Teams. Students' privacy consent is checked before proceeding, and their data is stored on the main page or a separate sheet. The system respects privacy, supports data analysis, and assists counsellors by organising cases according to consent and willingness. While the system helps streamline the process, it is not meant to replace counsellors' roles. Instead, it shortens the time between finishing the screening and getting feedback, making help more available and faster.

Principal Component Analysis (PCA) revealed three key components that explained 61.28% of the total variance. Exploratory Factor Analysis (EFA) using the Maximum Likelihood method confirmed three underlying factors, which were identified as Positive Emotion, Personal Growth, and Social Exploration. Reliability tests, including Cronbach's Alpha, McDonald's Omega, and Composite Reliability, showed values above the standard threshold, confirming that the items within each factor were consistent. Confirmatory Factor Analysis (CFA) was conducted using the Maximum Likelihood Robust estimation method. The model achieved acceptable fit indices. Convergent and discriminant validity were confirmed through standardised factor loadings, Average Variance Extracted, Composite Reliability, inter-item correlations, and HTMT values. Finally, Structural Equation Modelling (SEM) was used to examine the relationships between the three factors. The results showed that all three factors were significantly related to one another. Positive Emotion was found to be the most influential factor. It strongly predicted both Personal Growth and Social Exploration, with standardised coefficients of 0.665 and 0.761, respectively. This suggests that promoting emotional positivity among students may have the most widespread effect on their overall well-being.

Essentially, all the research objectives were met successfully. An automated workflow was created to screen UTAR freshmen for mental health issues and connect them with helpful tools. Although after the student submits the Microsoft Forms, the Microsoft Forms will automatically collect the responses. However, using Power Automate Workflow can capture the student's responses towards the resources group. Besides, the workflow enables the data to be stored on different Excel sheets, as it divides the Excel sheet based on the student privacy agreement. Therefore, counsellors can easily view the student responses and quickly identify the students who are not willing or do not respond to the resource group but who agreed to be contacted by the university. The counsellor can take further action to support students with this behaviour, particularly those with lower well-being scores. This ensures that the student who needs support can be supported in a shorter time period.

5.2 Recommendations for Further Research

Firstly, it is recommended to include individuals with higher education levels such as postgraduate students or those in their senior academic years as targeted respondents in future studies, given their potential exposure to increased stress and challenges. Furthermore, it would be valuable to draw comparisons between students across various academic years, such as those in their first year versus those in their final year, as this may uncover unique patterns in the WEMWBS questionnaire or discover the variations in their reactions to automated support systems.

Secondly, future research might include more psychological assessments such as the Depression, Anxiety and Stress Scale (DASS-21). This addition would help researchers gain a better understanding of students' mental health. For instance, looking into how stress and anxiety levels affect well-being that could help create more focused and effective interventions.

Thirdly, even though Microsoft Power Automate provided instant feedback and greatly reduced counsellor workload, it is important to acknowledge that it is a helpful tool and not a replacement for face-to-face communication. Future research might improve the automated system by adding more advanced digital tools, like interactive chatbots or AI-driven counselling assistants, which can offer immediate personalised interactions based on what students share. For example, using conversational AI that replies instantly might boost how engaged students are and their readiness to use support resources.

Fourthly, this research analysed the data based on certain methodological choices. For example, Principal Component Analysis (PCA) and Exploratory Factor Analysis (EFA) use specific extraction and rotation methods. Even though these gave some good insights, trying different methods might lead to different outcomes. Future studies might investigate how strong the methods are by comparing different extraction techniques. Additionally, the results from Confirmatory Factor Analysis (CFA) and Structural Equation Modelling (SEM) are significantly influenced by the chosen model-fit criteria. Researchers can test different fit indices and share sensitivity analyses to show how strong the identified models are. To find out if there are significant differences in the correlations between well-being characteristics between other student groupings, such as males and females, multi-group SEM analysis might also be carried out. For example, variations in the emotional and social responses of male and female students could highlight the need for tailored support resources based on gender.

Lastly, conducting longitudinal studies could significantly contribute to overcoming this limitation. For instance, tracking the same group of students during their undergraduate years could help identify specific periods when mental health support might be most beneficial. Further research could look into whether the well-being factor structure stays the same over longer periods or if new factors come up over time because of shifting academic pressures or personal situations.

References

Afrin, S., Roksana, S. and Akram, R., 2024. AI-Enhanced Robotic Process Automation: A Review of Intelligent Automation Innovations. *IEEE Access*, p. 1.

Alavi, M. et al., 2020. Chi-square for model fit in confirmatory factor analysis. *Journal of Advanced Nursing*, 76(9), pp. 2209–2211.

Arifin, S. et al., 2022. The Prevalence of Attitudes toward Seeking Counseling Help among Malaysian University Students. *International journal of academic research in business & social sciences*, 12(11).

Austin, J. D. et al., 2020. A structural equation Modelling approach to understanding pathways linking survivorship care plans to survivor-level outcomes. *Journal of Cancer Survivorship*, 14(6), pp. 834–846.

Baggyalakshmi, N., Dhanya, R. and Revathi, R., 2024. HR Onboarding Kit. *International academic journal of innovative research*, 11(1), pp. 27–38.

Bhatt, S., 2024. Digital Mental Health: Role of Artificial Intelligence in Psychotherapy. *Annals of Neurosciences*.

Bollen, K. A., 1989. Introduction to Structural Equation Models with Latent Variables. *Wiley*.

Braman, C. R. and Azzam, T., 2023. Consequences of survey modification in a program evaluation: An exploratory research on evaluation study. *Evaluation and Program Planning*, 98, p. 102274.

Brereton, R. G., 2025. Principal Component Analysis: Standardisation. *Journal of Chemometrics*, 39(1).

Briggs, S. R., Cheek, J. M., 1986. The role of factor analysis in the development and evaluation of personality scales. *J. Pers*. 54, pp. 106–148.

Brooks, J. L. et al., 2022. Factor analysis of the Center for Epidemiological Studies Depression Scale in American Indian women. *Research in Nursing & Health*, 45(6), pp. 733–741.

Brown, T. A., 2015. Confirmatory factor analysis for applied research. *Guilford Press*.

Caputo, A. et al., 2023. Employer Attractiveness: Two Instruments to Measure Employer Branding and Reputation. *SAGE Open*, 13(3).

Cattel, R. B., 1973. Factor analysis. *Greenwood Press*.

Cattell, R. B., 1966. The Scree Test For The Number Of Factors. *Multivariate Behavioral Research*, 1(2), pp. 245–276.

Cheung, G. W. et al., 2023. Reporting reliability, convergent and discriminant validity with structural equation Modelling: A review and best-practice recommendations. *Asia Pacific Journal of Management*, pp. 1–39.

Child, D., 2006. The essentials of factor analysis. *Continuum*.

Chua, B. S. et al., 2021. Psychological Distress, Relationship Quality, and Well-Being Among Malaysian Couples During the COVID-19 Pandemic. *Asia-Pacific Journal of Public Health*, 33(5), p. 10105395211014322.

Cilar, L., Pajnkihar, M. and Stiglic, G., 2020. Validation of the Warwick-Edinburgh Mental Well-being Scale among nursing students in Slovenia. *Journal of Nursing Management*, 28(6), pp. 1335–1346.

Civelek, M. E., 2018. Essentials of Structural Equation Modelling. *Essentials of Structural Equation Modelling*. Available at: https://ssrn.com/abstract=3338325.

Cong, L. M. et al., 2024. Psychological difficulties and the needs for psychological services for high school students. *Discover Mental Health*, 4(1).

Ding, D. Y., 2023. Transitioning to Microsoft Power Platform: An Excel User Guide to Building Integrated Cloud Applications in Power BI, Power Apps, and Power Automate. in, pp. 379–438.

Dursun, Y., Kocagoz, E., 2012. Yapısal Eşitlik Modellemesi ve Regresyon: Karşılaştırmalı Bir Analiz. *Erciyes Üniversitesi İ.İ.B.F. Dergisi* (35), 1-17.

Fabrigar, L. R. and Wegener, D. T., 2012. Exploratory factor analysis. *Oxford University Press*.

Fernández-Cardero, Á. et al., 2024. An Approach to Understand the Main Factors Contributing to the Variability in People with Overweight or Obesity Using Principal Components Analysis.

Field, A., 2013. Discovering statistics using IBM SPSS statistics. *SAGE*.

Fornell, C. and Larcker, D. F., 1981. Evaluating Structural Equation Models with Unobservable Variables and Measurement Error. *Journal of Marketing Research*, 18(1), pp. 39–50.

Franken, K. et al., 2018. Validation of the Mental Health Continuum-Short Form and the dual continua model of well-being and psychopathology in an adult mental health setting. *J Clin Psychol*, 74(12), pp. 2187–2202.

Fung, S.-F., 2019. Psychometric evaluation of the Warwick-Edinburgh Mental Well-being Scale (WEMWBS) with Chinese University Students. *Health and Quality of Life Outcomes*, 17(1), p. 46.

Grewal, R., Cote, J. A. and Baumgartner, H., 2004. Multicollinearity and measurement error in structural equation models: Implications for theory testing. *Marketing Science*, 23, pp. 519–529.

Groot, J. et al., 2022. The effectiveness of fully automated digital interventions to promote mental well-being in the general population: A systematic review and meta-analysis (Preprint). *JMIR mental health*.

Gupta, R., 2024. A Comprehensive Analysis of Path Analysis and Structural Equation Modelling: Concepts & Applications.

Guttman, L., 1954. Some necessary conditions for common-factor analysis. *Psychometrika*, 19, pp. 149–161.

Haider, S. et al., 2022. Translation, Validation, and Psychometric Evaluation of the Diabetes Quality-of-Life Brief Clinical Inventory: The Urdu Version. *Journal of multidisciplinary healthcare*, 15, pp. 955–966.

Hair, J. F. et al.,1998. Multivariate Data Analysis. *Prentice Hall*, 5th.

Hair, J. F. et al., 2010. Multivariate Data Analysis. *Pearson*.

Hair, J. et al., 2009. Multivariate data analysis. *Prentice-Hall*.

Haron@shafiee, H. I., Abd Halim, M. S. and Ismail, M., 2023. Assessment Model for Determinant Factor Constructs in Edu-Tourism Using Confirmatory Factor Analysis (CFA). *International Journal of Sustainable Development and Planning*, 18(7), pp. 2037–2043.

Haşılloğlu, S. B. and HASILOGLU-CIFTCILER, M., 2023. What should be the measure of conformity to normal distribution (normality) test in Likert type digital and face-to-face survey data? *internet uygulamalari ve yönetimi*.

Hatem, G. et al., 2022. Normality Testing Methods and the Importance of Skewness and Kurtosis in Statistical Analysis. *BAU Journal - Science and Technology*, 3(2), p. Article 7.

Hock, R. S. et al, 2012. A new resolution for global mental health. *The Lancet*, 379(9824), pp. 1367–1368.

Hooper, D., Coughlan, J. and Mullen, M. R., 2008. Structural equation modelling: Guidelines for determining model fit. *Electronic Journal of Business Research Methods*, 6(1), pp. 53–60.

Huppert, F., 2009. Psychological well-being: evidence regarding its causes and consequences†. *Appl Psychol Health Well Being*, 1(2), pp. 137–64.

Ismail, A. and Kahwa, K. M., 2020. Prevalence, associated factors, and help seeking behavior related to psychological distress among international students at Universiti Kebangsaan Malaysia. *Malaysian Journal of Public Health Medicine*, 20(2), pp. 215–223.

Jöreskog, K. G. and Sörbom, D., 1989. LISREL 7: A Guide to the Program and Applications.

Jurewicz, I., 2015. Mental health in young adults and adolescents – supporting general physicians to provide holistic care. *Clinical Medicine*, 15(2), pp. 151–154.

Kaiser, H. F.,1970. A second generation little jiffy. *Psychometrika*, 35, pp. 401–415.

Kannan, R. and Chin, J.-J., 2022. Backup Automation Using Power Automate for Malaysian Vaccination Centres. 1(1), pp. 23–34.

Karimian, Z. and Chahartangi, F., 2024. Development and validation of a questionnaire to measure educational agility: a psychometric assessment using exploratory factor analysis. *BMC Medical Education*, 24(1).

Kline, R. B., 2005. Principles and practice of structural equation modelling. *Guilford Press*.

Konaszewski, K. et al., 2021. Factor structure and psychometric properties of a Polish adaptation of the Warwick-Edinburgh Mental Wellbeing Scale. *Health and Quality of Life Outcomes*, 19(1), pp. 1–11.

Koushede, V. et al., 2019. Measuring mental well-being in Denmark: Validation of the original and short version of the Warwick-Edinburgh mental well-being scale (WEMWBS and SWEMWBS) and cross-cultural comparison across four European settings. *Psychiatry Research-neuroimaging*, 271, pp. 502–509.

Kunicki, Z. J., Smith, M. L. and Murray, E. J., 2023. A Primer on Structural Equation Model Diagrams and Directed Acyclic Graphs: When and How to Use Each in Psychological and Epidemiological Research. *Advances in methods and practices in psychological science*, 6(2), pp. 1–14.

Kyriazos, T. A., 2023. Applied Psychometrics: Estimator Considerations in Commonly Encountered Conditions in CFA, SEM, and EFA Practice. *Psychology*, 14(05), pp. 799–828.

Li, G. and Qin, Y., 2024. An Exploration of the Application of Principal Component Analysis in Big Data Processing. *Applied mathematics and nonlinear sciences*.

Loong, Y. Q. et al., 2024. The impact of social stigma and loss of face on mental health help-seeking behavior among university students in malaysia: examining mediating relationships. *International Journal of Education, Psychology and Counseling*, 9(56), pp. 01–23.

Lumumba, V. et al., 2024. Comparative Analysis of Cross-Validation Techniques: LOOCV, K-folds Cross-Validation, and Repeated K-folds Cross-Validation in Machine Learning Models. *American Journal of Theoretical and Applied Statistics*, 13(5), pp. 127–137.

Ma, S., Zhu, Y. and Bresnahan, M., 2022. Chinese International Students' Face Concerns, Self-Stigma, Linguistic Factors, and Help-Seeking Intentions for Mental Health. *Health Communication*, pp. 1–9.

Maccallum, R. C., Browne, M. W. and Sugawara, H. M., 1996. Power analysis and determination of sample size for covariance structure modeling of fit involving a particular measure of model. *Psychological Methods*, 13(2), pp. 130–149.

Merkulova, E. A. et al., 2023. Using PCA Machine Learning Approach Based on Psychological Questionnaires and Spectral Characteristics of the EEG to Separate the Healthy Participants and Participants with Major Depressive Disorder. pp. 1740–1745.

Meydan, C. H., Sen, H., 2011. Yapısal Eşitlik Modellemesi AMOS Uygulamaları. Ankara: Detay Yayıncılık.

Microsoft, 2025, *Released versions for Power Automate - Release Notes*. [online] Available at: https://learn.microsoft.com/en-us/power-platform/released-versions/power-automate [Accessed 13 April 2025].

Microsoft, n.d., *Microsoft Power Automate – Process Automation Platform* [online] Available at: https://www.microsoft.com/en-us/power-platform/products/power-automate [Accessed 13 April 2025].

Mir, I. A. et al., 2023. Determinants and predictors of mental health during and after COVID-19 lockdown among university students in Malaysia. *PLOS ONE*, 18(1), p. e0280562.

Mohtar, L. E. et al., 2024. Students' Self-Regulated Learning in Physics: Exploratory Factor Analysis (EFA) and Confirmatory Factor Analysis (CFA). 3(1).

Morrison, T. G., Morrison, M. A. and McCutcheon, J. M., 2017. Best Practice Recommendations for Using Structural Equation Modelling in Psychological Research. *Psychology*, 08(9), pp. 1326–1341.

Mwange, A., Chiseyeng'i, J. and Matoka, W., 2023. Business Research Methods: Theoretical Demystification of The Use of Multivariate Analysis Techniques in Research. *Journal of Education and Practice*, 14(21), pp. 56–63.

Napitupulu, D., Kadar, J. A. and Jati, R. K., 2017. Validity Testing of Technology Acceptance Model Based on Factor Analysis Approach. *Indonesian Journal of Electrical Engineering and Computer Science*, 5(3), pp. 697–704.

National Health and Morbidity Survey, 2022, *Ministry of Health Malaysia Institute for Public Health.* [online] Available at: https://iku.gov.my/nhms-mch [Accessed 13 April 2025].

National Health and Morbidity Survey, 2023, *Ministry of Health Malaysia Institute for Public Health.* [online] Available at: https://iku.gov.my/nhms-2023 [Accessed 13 April 2025].

Nizar, J. et al., 2019. Assessment model for construct occupational accident using confirmatory factor analysis. *E3S Web of Conferences*, 90, p. 03004.

Omura, S. et al., 2022. Exploratory factor analysis determines latent factors in Guillain–Barré syndrome. *Dental science reports*, 12(1).

Ozili, P. K., 2023. The Acceptable R-Square in Empirical Modelling for Social Science Research. *Advances in knowledge acquisition, transfer and management book series*, pp. 134–143.

Pallant, J., 2010. SPSS survival manual: A step by step guide to data analysis using SPSS. *Open University Press/McGrawHill*, 4th Edition.

Peng, G. et al., 2023. Association between depressive symptoms and lung function in the United States adults without pulmonary diseases: A cross-sectional study from NHANES. *Journal of Affective Disorders*, 325, pp. 787–793.

Perera, B. P. R. et al., 2022. Measuring mental well-being in Sri Lanka: validation of the Warwick Edinburgh Mental Well-being Scale (WEMWBS) in a Sinhala speaking community. *BMC Psychiatry*, 22(1).

Phakiti, A., 2018. Exploratory Factor Analysis, pp. 423–457.

Portela, D. M. P., 2012. Contributo das Técnicas de Análise Fatorial para o Estudo do Programa "Ocupação Científica de Jovens nas Férias.". *Universidade aberta*, p. 2010.

Qin, L. et al., 2021. R-squared of a Latent Interaction in Structural Equation Model: A Tutorial of Using R. *International Journal of Statistics and Probability*, 10(3), p. 69.

Rattanakijsuntorn, W., 2023. Application of EFA and AHP in the Last-Mile Delivery Service in Thailand. pp. 0513–0517.

Raykov, T. and Widaman, K. F., 1995. Issues in Structural Equation Modeling Research. *Structural Equation Modeling*, 2, pp. 289–318.

Roemer, E., Schuberth, F. and Henseler, J., 2021. HTMT2– an improved criterion for assessing discriminant validity in structural equation Modelling. *Industrial Management and Data Systems*, 121(12), pp. 2637–2650.

Roesel, I. et al., 2024. The Patient Activation Measure-13 (PAM-13) in an oncology patient population: psychometric properties and dimensionality evaluation. *Health and Quality of Life Outcomes*, 22(1), p. 39.

Ronaldson, A. et al., 2020. Using structural equation modelling in routine clinical data: Depression, diabetes, and use of Accident & Emergency (Preprint). *JMIR medical informatics*.

Ruggeri, K. et al., 2020. Well-being is more than happiness and life satisfaction: a multidimensional analysis of 21 countries. *Health and Quality of Life Outcomes*, 18(1), pp. 1–16.

Saidi, N. et al., 2024. Reviewing the influence of mental health and coping strategies on academic performance. *International Journal of Education, Psychology and Counseling*, 9(54), pp. 431–446.

Salleh, M. R., 2018. The Burden of Mental Illness: An Emerging Global Disaster. 3(1), p. 5.

Samuels, P., 2017. Advice on Exploratory Factor Analysis.

Santos, J. J. A. dos et al., 2015. Adaptation and cross-cultural validation of the Brazilian version of the Warwick-Edinburgh mental well-being scale. *Revista Da Associacao Medica Brasileira*, 61(3), pp. 209–214.

Sarmento, R. P. and Costa, V., 2019. Confirmatory Factor Analysis - A Case study. *arXiv: Applications*.

Shah, A. et al., 2021. Principal component analysis based construction and evaluation of cryptocurrency index. *Expert Systems With Applications*, 163, p. 113796.

Shaharudin, S. M. and Ahmad, N., 2017. Choice of Cumulative Percentage in Principal Component Analysis for Regionalization of Peninsular Malaysia Based on the Rainfall Amount. *Communications in Computer and Information Science*. Edited by M. Mohamed Ali et al., 752.

Shrestha, N., 2021. Factor Analysis as a Tool for Survey Analysis. *American Journal of Applied Mathematics and Statistics*, 9, pp. 4–11.

Sikorski, F. et al., 2021. The efficacy of automated feedback after internet-based depression screening: Study protocol of the German, three-armed, randomised controlled trial DISCOVER. *Internet Interventions*, 25, p. 100435.

Silva, R. P. A., Macêdo, L. C. B. and Silva, I. L. R., 2013. Avaliação das Características Psicométricas dos Questionários Utilizados nos Periódicos da

Area Contábil: Um Estudo Longitudinal Compreendido no Período 2003-2012. *XX Congresso Brasileiro de Custos*.

Skinner, C. J., Holt, D. and Smith, T. M. F., 1989. Domain Means, Regression and Multivariate Analysis. *Analysis of Complex Surveys*, pp. 59–87.

Smith, O. R. F. et al., 2017. Measuring mental well-being in Norway: validation of the Warwick-Edinburgh Mental Well-being Scale (WEMWBS). *BMC Psychiatry*, 17(1), p. 182.

Stefana, A. et al., 2025. Psychological, psychiatric, and behavioral sciences measurement scales: best practice guidelines for their development and validation. *Frontiers in Psychology*, 15.

Steiger, J. H., 2007. Understanding the limitations of global fit assessment in structural equation modelling. *Personality and Individual Differences*, 42(5), pp. 893–898.

Stevens, J. P., 2002. Applied multivariate statistics for the social sciences.

Stewart-Brown, S., 2016. Population level: wellbeing in the general population. *Wellbeing recovery mental health*. Edited by M. Salde, L. Oades, and A. Jarden, pp. 215–30.

Stewart-Brown, S. et al., 2011. ‘The Warwick–Edinburgh Mental Well-being Scale (WEMWBS): a valid and reliable tool for measuring mental well-being in diverse populations and projects. *J Epidemiol Community Health*, 65(Suppl 2), pp. A38-9.

Suprapto, W., Stefany, S. and Ali, S., 2020. Service Quality, Store Image, Price Consciousness, and Repurchase Intention on Mobile Home Service. 76, p. 01056.

Sürücü, L., YIKILMAZ, İ. and Maşlakçı, A., 2024. Exploratory Factor Analysis (EFA) in Quantitative Researches and Practical Considerations. *Gümüşhane Üniversitesi Sağlık Bilimleri Dergisi*, 13(2), pp. 947–965.

Tamil, A. M. et al., 2023. Depressive symptoms among adults: Baseline findings of PURE Malaysia cohort study. *Heliyon*.

Tennant, R. et al., 2006. Monitoring positive mental health in Scotland: validating the Affectometer 2 scale and developing the Warwick–Edinburgh Mental Well-being Scale for the UK. *Edinb NHS Health Scotl*.

Thartori, V. and Nordin, M. S., 2019. Structural Equation Modeling and Relationships Between Mental Wellbeing, Resilience and Self-stigma. *Research in World Economy*, 10(2), p. 129.

Thuryrajah, A. E. and Jeyakumar, R., 2017. Factors Determining the University Counselling Services Effectiveness. *Business and Economics Journal*, 8(4), pp. 1–6.

van der Aalst, W. M. P., Bichler, M. and Heinzl, A., 2018. Robotic Process Automation. *Bus Inf Syst Eng*, 60, pp. 269–27

Verma, J., 2013. Data analysis in management with SPSS software. *Springer*.

Victor-Edema, U. A., 2023. Comparative analysis of some approaches to multivariate normality test. *Faculty of Natural and Applied Sciences Journal of Scientific Innovations*, 4(2), pp. 154–164.

Vo, T. L. U. and Nguyen, T. H. N., 2023. Application of SEM in analysing the effect of competitiveness, operating effectiveness and risk management on the stability of Vietnamese construction corporations. *Tạp chí Khoa học và Công nghệ Việt Nam*, 65(1), pp. 11–23.

Wahyuni, E. et al., 2024. Enhancing Mental Health Literacy in University; Interactions between Student Initiatives and Counselor Strategies. *Health Educ Health Promot*, 12(2), pp. 215–223.

Warwick Medical School, 2020. *WEMWBS: 14-item vs 7-item scale*. [Online]. Available at:

https://warwick.ac.uk/fac/sci/med/research/platform/wemwbs/about/wemwbsvsswemwbs/ [Accessed: 10 April 2025].

Wasdahl, A., 2024. Machine Credibility: How News Readers Evaluate Ai-generated Content. *Interactions: UCLA Journal of Education and Information Studies*, 19(1).

Watkins, M. W., 2021. A Step-by-Step Guide to Exploratory Factor Analysis with R and Rstudio. *Routledge*.

Weismer, S. E. et al., 2021. A preliminary epidemiologic study of social (pragmatic) communication disorder in the context of developmental language disorder. International Journal of Language & Communication Disorders, 56(6), pp. 1235–1248.

Wilimitis, D. and Walsh, C. G., 2023. Practical Considerations and Applied Examples of Cross-Validation for Model Development and Evaluation in Health Care: Tutorial. *JMIR AI*, 2(1), p. e49023.

Wong, T.-T. and Yeh, P.-Y., 2020. Reliable Accuracy Estimates from k -folds Cross Validation. *IEEE Transactions on Knowledge and Data Engineering*, 32(8), pp. 1586–1594.

Yong, A. G. and Pearce, S., 2013. A beginner’s guide to factor analysis: Focusing on exploratory factor analysis. *Tutorials in Quantitative Methods for Psychology*, 9(2), pp. 79–94.

Youngstrom, E. A. et al., 2017. Evidence-based assessment as an integrative model for applying psychological science to guide the voyage of treatment. *Clin. Psychol. Sci. Pract.*, 24, pp. 331–363.

Yu, M. et al., 2021. Is too much as bad as too little? The S-curve relationship between corporate philanthropy and employee performance. *Asia Pacific Journal of Management*, 39, pp. 1511–1534.

Yulia, A. et al., 2021. Response Actions of Malaysian Universities and Colleges in Managing Student Mental Health: A Systematic Review. pp. 1–16.

Appendices

Appendix A: The Warwick–Edinburgh Mental Well-being Scale (WEMWBS) Questionnaire

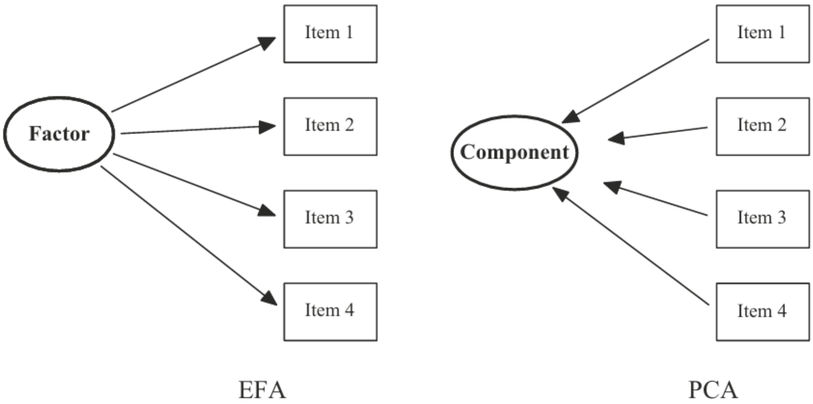
The Warwick–Edinburgh Mental Well-being Scale (WEMWBS)

Below are some statements about feelings and thoughts.
Please tick the box that best describes your experience of each over the last 2 weeks

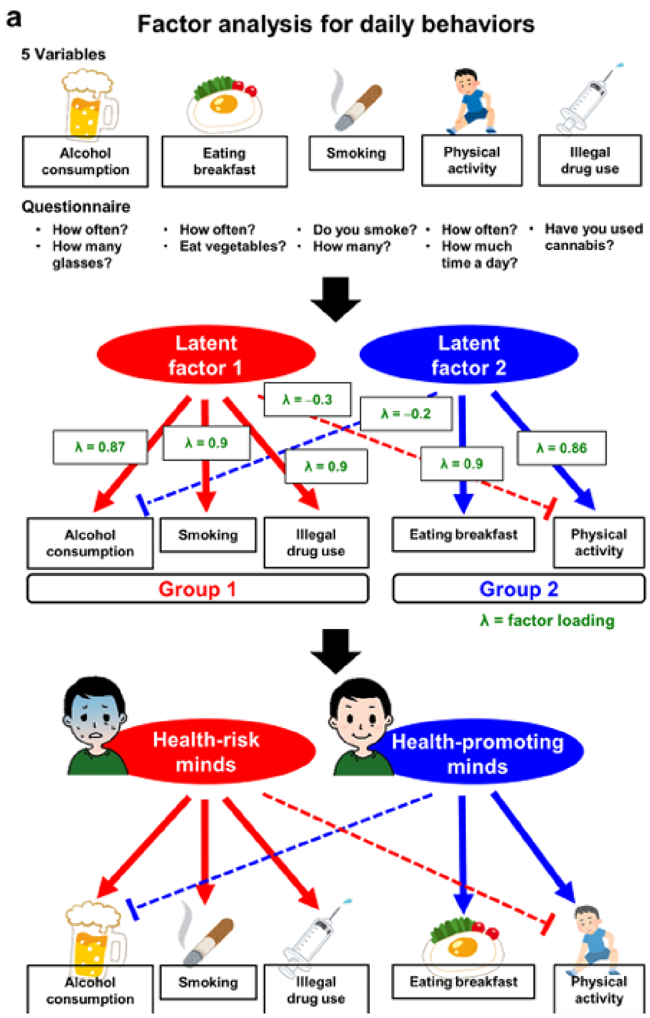
STATEMENTS	None of the time	Rarely	Some of the time	Often	All of the time
I've been feeling optimistic about the future	1	2	3	4	5
I've been feeling useful	1	2	3	4	5
I've been feeling relaxed	1	2	3	4	5
I've been feeling interested in other people	1	2	3	4	5
I've had energy to spare	1	2	3	4	5
I've been dealing with problems well	1	2	3	4	5
I've been thinking clearly	1	2	3	4	5
I've been feeling good about myself	1	2	3	4	5
I've been feeling close to other people	1	2	3	4	5
I've been feeling confident	1	2	3	4	5
I've been able to make up my own mind about things	1	2	3	4	5
I've been feeling loved	1	2	3	4	5
I've been interested in new things	1	2	3	4	5
I've been feeling cheerful	1	2	3	4	5

Warwick–Edinburgh Mental Well-being Scale (WEMWBS)

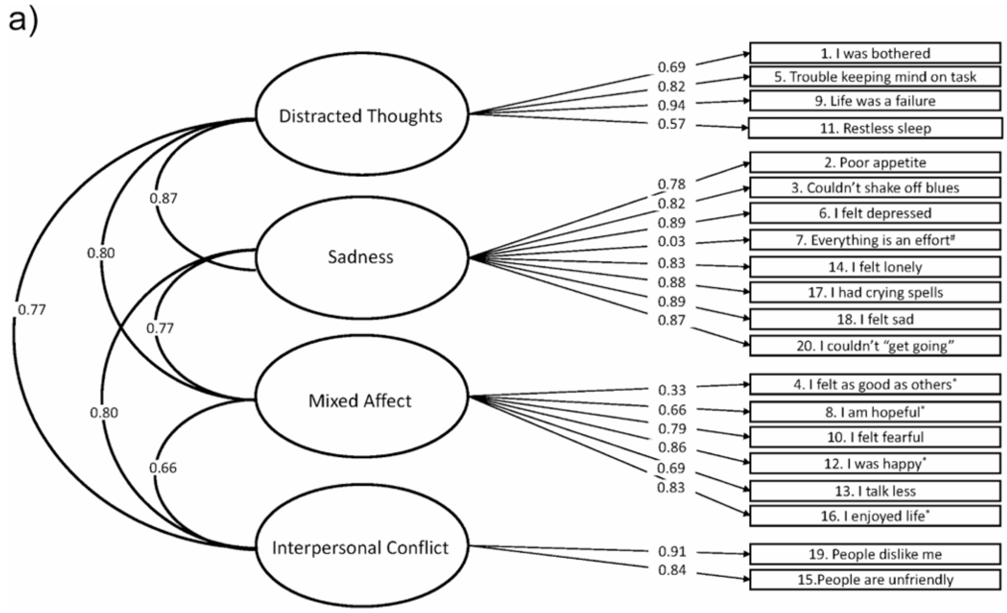
Appendix B: Conceptual Difference Between Principal Component Analysis (PCA) and Exploratory Factor Analysis (EFA)



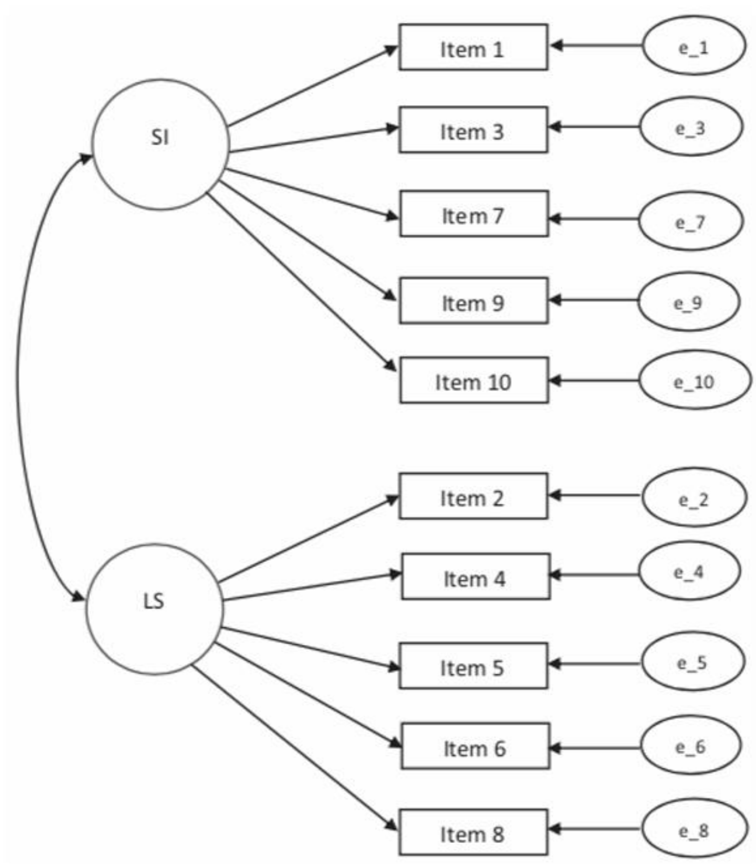
Appendix C: Illustrative Example of Factor Analysis in Health Behaviour Research



Appendix D: Example of a Confirmatory Factor Analysis (CFA) Path Diagram



Appendix E: Example of a Measurement Model Representing Two Latent Constructs



Appendix F: Scoring Content for Each Well-being Category in the Power Automate Workflow

{x} Score 14-40

Parameters Settings Code view About

Scoring Content

Value *

🌱 Your score falls within the very low range (14–40 points). This may mean you're going through a tough time right now, and that's completely okay. Please remember, you are not alone in this. Life can sometimes feel heavy, but taking the first step to check in with yourself is already something brave and meaningful.

Here are a few gentle ideas that might bring a little light into your day
- Talk to someone you trust, like a close friend, a kind family member, or a counsellor who can listen with care
- Try adding a short walk or some light movement to your day. Even a few minutes can gently lift your mood
- Give yourself a quiet moment with things that comfort you like music, reading, or simply resting
- Doing something kind, no matter how small, can bring warmth and connection
- Explore a new hobby or interest that sparks even a little curiosity or joy. It might offer a refreshing shift

You truly deserve support, peace, and care just as much as anyone else 🌷

{x} Score 41-44

Parameters Settings Code view About

Name *

Scoring Content

Value *

🧡 Your score falls in the below-average range (41–44 points), which suggests there may be areas where your well-being could be gently improved. The encouraging part is that even small and consistent changes can make a meaningful difference in how you feel. You've already taken a thoughtful step by checking in with yourself.

Here are a few soft suggestions to support your journey:
- Reach out to someone you trust and let them know how you're feeling
- Spend a few minutes each day doing things that calm or uplift you
- Try light physical activity like a short walk to gently boost your mood
- Do something kind for others as even small acts of care can bring connection and warmth
- Explore a new interest or hobby that sparks curiosity or joy

You deserve kindness, support, and the time to grow at your own pace 🌷

Score 45-59

Parameters | Settings | Code view | About

Name *

Scoring Content

Value *

🌿 Your score is within the average range (45–59 points) which suggests that your overall mental well-being is in a stable place. While things seem to be going well, there is always space for a little extra care and growth.

Here are a few ideas to gently support your well-being
- Make time each day for something that helps you relax and unwind
- Keep your body moving with light physical activity to maintain a balanced mood
- Try small acts of kindness or moments of giving that can bring a sense of fulfillment
- Explore hobbies or learning something new to add joy and enrichment to your routine
- Stay connected with people who make you feel supported and valued

Continue taking care of yourself — you're doing well and you deserve it 🌼

Score 60-70

Parameters | Settings | Code view | About

Name *

Scoring Content

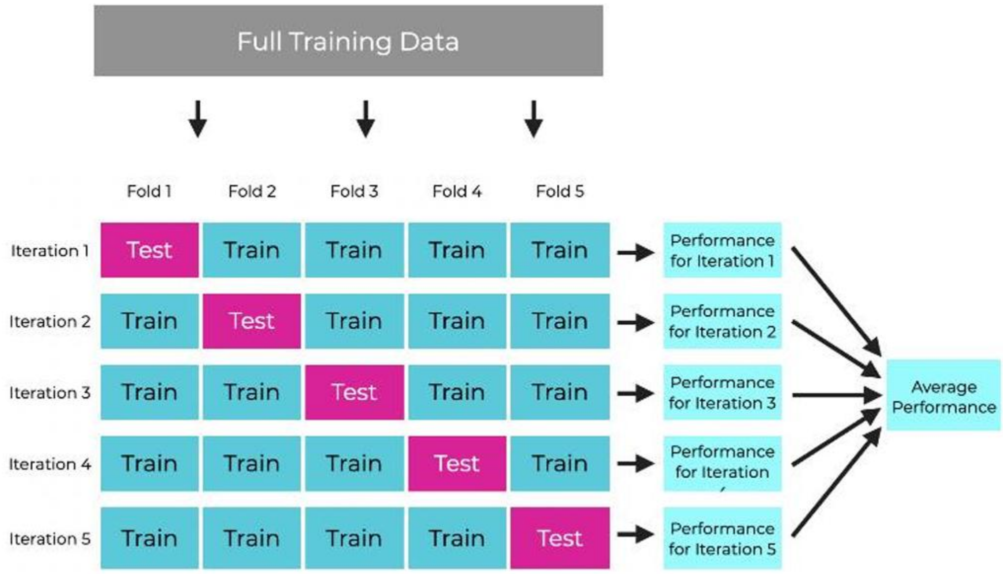
Value *

🌞 Your score falls within the above-average range (60–70 points) which indicates a high level of mental well-being compared to your peers. It's truly wonderful to see you thriving and taking care of yourself.

To continue nurturing this positive state
- Keep enjoying the activities that bring you joy and help you relax each day
- Explore new interests or learning opportunities to add meaning and excitement to your life
- Maintain regular physical activity to support both your body and your mind
- Share your positivity with others as offering help or encouragement can be deeply fulfilling
- Stay close to the people who uplift you as supportive relationships are a beautiful source of strength

Keep shining — your journey of well-being is something to be proud of ☀️

Appendix G: Illustration of 5-Folds Cross-Validation Process for Model Evaluation



Appendix H: R Code

This section consists of two parts. Part A is the raw R code, and Part B is the R code with outputs and plots.

Part A: Raw R Code

```
############################################################
# 0. Install and Load Required Libraries
############################################################

# install.packages("reshape2")
# install.packages("igraph")
# install.packages("recipes")
# install.packages("caret")
# install.packages("lavaan")
# install.packages("MVN")


library(readxl)
library(dplyr)
library(ggplot2)
library(reshape2)
library(psych)
library(igraph)
library(lavaan)
library(tidyr)
library(dplyr)
library(MVN)
library(GPArotation)
############################################################
# 1. Data Loading and Preparation
############################################################
n_factors <- 3

data <- read_excel("C:/Users/lenov/Downloads/FYP First
Sem/FYP_Dataset_LatestUpToMar25_OnlyFirstYearStudent.xlsx")
question_cols <- paste0("Q", 1:14)
data_questions <- data %>% select(all_of(question_cols))

data_questions <- data_questions %>%
  mutate(across(everything(), ~ ifelse(is.na(.), mean(., na.rm = TRUE), .)))

data_scaled <- scale(data_questions)

############################################################
# 2. Univariate and Multivariate Normality Check
############################################################
#Skewness, Kurtosis and Histogram for Univariate Normality Check

normality_summary <- data.frame(
  Skewness = round(sapply(data_questions, skew), 3),
  Kurtosis = round(sapply(data_questions, kurtosi), 3)
)
rownames(normality_summary) <- question_cols  # Set row names as Q1-Q14

cat("\nSkewness and Kurtosis for Each Variable:\n")
print(normality_summary)

skew_range <- range(normality_summary$Skewness)
kurtosis_range <- range(normality_summary$Kurtosis)

cat("\nRange of Skewness: [", skew_range[1], ",", skew_range[2], "]\n")
cat("Range of Kurtosis: [", kurtosis_range[1], ",", kurtosis_range[2], "]\n")


#All 14 Ques Histogram in one page
data_long <- data_questions %>%
  mutate(ID = row_number()) %>%
  pivot_longer(cols = starts_with("Q"), names_to = "Question", values_to =
"Response")

data_long$Question <- factor(data_long$Question, levels = paste0("Q", 1:14))
```

```r
ggplot(data_long, aes(x = factor(Response))) +
  geom_bar(fill = "skyblue", color = "black", width = 1) +
  geom_text(stat = "count", aes(label = ..count..), vjust = -0.4, size = 2.2)
+
  facet_wrap(~ Question, ncol = 5) +
  scale_x_discrete(limits = as.character(1:5)) +
  labs(title = "Histograms of All Questions", x = "Response", y = "Frequency")
+
  theme_minimal(base_size = 9) +
  theme(
    plot.title = element_text(size = 14, face = "bold", hjust = 0.5),
    strip.text = element_text(size = 10, face = "bold", color = "darkblue"),
    axis.text.x = element_text(size = 7),
    axis.text.y = element_text(size = 7),
    panel.spacing = unit(1, "lines")
  )


# Mardia's Test for Multivariate Normality Check
cat("\n[Mardia's Test]\n")
mardia_result <- mvn(data = data_questions, mvnTest = "mardia")
print(mardia_result$multivariateNormality)


############################################################
# 3. Correlation Matrix and Heatmap
############################################################
corr_matrix <- cor(data_scaled)
melted_corr <- melt(corr_matrix)
p_corr <- ggplot(data = melted_corr, aes(x = Var1, y = Var2, fill = value)) +
  geom_tile() +
  geom_text(aes(label = round(value, 2))) +
  scale_fill_gradient2(low = "blue", high = "red", mid = "white", midpoint =
0) +
  theme_minimal() +
  labs(x = NULL, y = NULL) +
  ggtitle("Correlation Matrix of 14 WEMWBS Items")
print(p_corr)

############################################################
# 4. Determinant and Average Inter-Item Correlation (AIIC)
############################################################
det_corr <- det(corr_matrix)
cat("\nDeterminant of the correlation matrix (Determinant Score):", det_corr,
"\n")
upper_tri <- corr_matrix[upper.tri(corr_matrix)]
aiic <- mean(upper_tri)
cat("Average Inter-Item Correlation (AIIC):", aiic, "\n")

############################################################
# 5. Data Adequacy Testing: Bartlett's Test and KMO
############################################################
bartlett_result <- cortest.bartlett(corr_matrix, n = nrow(data_scaled))
cat("\nBartlett's Test:\n")
print(bartlett_result)

kmo_result <- KMO(corr_matrix)
cat("\nKaiser-Meyer-Olkin (KMO) Test:\n")
print(kmo_result)

############################################################
# 6. Principal Component Analysis (PCA)
############################################################
pca_res <- prcomp(data_scaled, center = TRUE, scale. = TRUE)
eigenvalues <- pca_res$sdev^2
explained_variance_ratio <- eigenvalues / sum(eigenvalues)
cumulative_explained <- cumsum(explained_variance_ratio)

cat("\nEigenvalues from PCA:\n")
for (i in 1:length(eigenvalues)) {
  cat(paste("Principal Component", i, ":", round(eigenvalues[i], 4), "\n"))
}

############################################################
# 7. Scree Plot
############################################################

df_scree <- data.frame(PC = 1:length(eigenvalues), Eigenvalue = eigenvalues)
p_scree <- ggplot(df_scree, aes(x = PC, y = Eigenvalue)) +
  geom_line() +
  geom_point() +
  geom_hline(yintercept = 1, linetype = "dashed", color = "red") +
  geom_text(aes(label = round(Eigenvalue, 2)), vjust = -0.5) +
  ggtitle("Scree Plot") +
  xlab("Principal Component") +
  ylab("Eigenvalue") +
  scale_x_continuous(breaks = 1:length(eigenvalues))
```

```r
print(p_scree)

cat("\nExplained Variance Ratio (in %):\n")
for (i in 1:length(explained_variance_ratio)) {
  cat(paste("PC", i, ":", round(explained_variance_ratio[i] * 100, 2), "%\n"))
}
cat("\nCumulative Explained Variance (in %):\n")
for (i in 1:length(cumulative_explained)) {
  cat(paste("PC1 to PC", i, ":", round(cumulative_explained[i] * 100, 2),
"%\n"))
}

############################################################
# 8. Cumulative Explained Variance Plot
############################################################
df_cum <- data.frame(
  Components = 1:length(cumulative_explained),
  Cumulative = cumulative_explained * 100
)

p_cum <- ggplot(df_cum, aes(x = Components, y = Cumulative)) +
  geom_line(color = "darkgreen", linewidth = 1) +
  geom_point(color = "darkgreen", size = 2) +
  geom_text(
    aes(label = paste0(round(Cumulative, 2), "%")),
    vjust = -0.5, size = 3
  ) +
  labs(
    title = "Cumulative Variance Explained by PCA Components",
    x = "Number of Components",
    y = "Cumulative Explained Variance (%)"
  ) +
  theme_minimal(base_size = 11) +
  scale_y_continuous(limits = c(0, 105), breaks = seq(0, 100, 10)) +
  scale_x_continuous(breaks = 1:14)

print(p_cum)

############################################################
# 9. K-Folds Cross Validation for PCA
############################################################
set.seed(123)
k_folds <- 10
n_factors <- 3
n <- nrow(data_scaled)
fold_indices <- sample(rep(1:k_folds, length.out = n))

fold_results <- matrix(0, nrow = k_folds, ncol = n_factors)
training_accuracies <- numeric(k_folds)

for (k in 1:k_folds) {
  test_idx <- which(fold_indices == k)
  train_data <- data_scaled[-test_idx, ]

  pca_model <- prcomp(train_data, center = TRUE, scale. = TRUE)
  explained_var <- (pca_model$sdev)^2 / sum((pca_model$sdev)^2)
  explained_ratio_top <- explained_var[1:n_factors]
  cumulative_accuracy <- sum(explained_ratio_top) * 100

  fold_results[k, ] <- explained_ratio_top * 100
  training_accuracies[k] <- cumulative_accuracy

  cat(sprintf("Fold %2d: Explained Variance Ratio: [%0.8f, %0.8f, %0.8f],
Training Accuracy (cumulative): %0.2f%%\n",
              k, explained_ratio_top[1], explained_ratio_top[2],
explained_ratio_top[3], cumulative_accuracy))
}

avg_explained <- colMeans(fold_results)
cat("\n[Cross-Validation] Average Explained Variance Ratio across folds:\n")
print(round(avg_explained, 2))

cumulative_explained <- cumsum(avg_explained)
cat("\nCumulative Explained Variance:", round(cumulative_explained, 2), "\n")
cat("Average Training Accuracy (cumulative explained variance):",
round(mean(training_accuracies), 2), "%\n")


############################################################
# 10. Exploratory Factor Analysis (EFA)
############################################################
# Perform EFA using maximum likelihood and promax rotation
efa_res <- fa(data_scaled, nfactors = n_factors, fm = "ml", rotate = "promax")
cat("\nFactor Loadings:\n")
print(efa_res$loadings)

loadings_mat <- unclass(efa_res$loadings)
```

```r
loadings_df <- as.data.frame(loadings_mat)
loadings_df$Question <- rownames(loadings_df)
loadings_df$Question <- factor(loadings_df$Question, levels = paste0("Q",
14:1))
colnames(loadings_df)[1:n_factors] <- paste0("F", 1:n_factors)
melted_loadings <- melt(loadings_df, id.vars = "Question")

# Heatmap Plot
p_loadings <- ggplot(melted_loadings, aes(x = variable, y = Question, fill =
value)) +
  geom_tile() +
  geom_text(aes(label = round(value, 2))) +
  scale_fill_gradient2(low = "blue", high = "yellow", mid = "green", midpoint
= 0) +
  labs(title = "Factor Loadings Heatmap", x = "Factors", y = "Questions") +
  theme_minimal()
print(p_loadings)

############################################################
# 11. Communalities
############################################################
communalities <- efa_res$communality
communalities_df <- data.frame(
  Question = factor(question_cols, levels = paste0("Q", 1:14)),
  Communality = communalities
)

cat("\nCommunalities for each question:\n")
print(communalities_df, row.names = FALSE)

p_communalities <- ggplot(communalities_df, aes(x = Question, y =
Communality)) +
  geom_bar(stat = "identity", fill = "lightcoral") +
  geom_text(aes(label = round(Communality, 2)), vjust = -0.5) +
  ggtitle("Communalities of Each Question") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
print(p_communalities)



############################################################
# 12. Dynamic Grouping of Items by Factor
############################################################
loadings_mat <- as.matrix(efa_res$loadings)
factor_groups <- list()
for (i in 1:n_factors) {
  factor_groups[[paste0("Factor", i)]] <- c()
}
for (i in 1:nrow(loadings_mat)) {
  max_index <- which.max(abs(loadings_mat[i, ]))
  factor_groups[[paste0("Factor", max_index)]] <-
c(factor_groups[[paste0("Factor", max_index)]],
                                                      rownames(loadings_mat)[i])
}
cat("\nDynamic grouping of items by factor:\n")
print(factor_groups)

############################################################
# 13. McDonald's Omega for Each Factor Group
############################################################
cat("\nMcDonald's Omega for each factor group:\n")
for (factor in c("Factor1", "Factor2", "Factor3")) {
  items <- factor_groups[[factor]]
  if (length(items) < 2) {
    cat(paste(factor, ": Skipped (only 1 item in this factor group).\n"))
    next
  }
  omega_res <- omega(data_questions[, items], nfactors = 1, plot = FALSE)
  cat(paste(factor, ": Omega =", round(omega_res$omega.tot, 3), "\n"))
}


############################################################
# 14. Cronbach's Alpha for Dynamically Grouped Items
############################################################
cat("\nCronbach's Alpha for dynamically grouped items:\n")
for (factor in c("Factor1", "Factor2", "Factor3")) {
  items <- factor_groups[[factor]]
  if (length(items) >= 2) {
    alpha_res <- alpha(data_questions[, items])
    cat(paste(factor, ": α =", round(alpha_res$total$raw_alpha, 3), "\n"))
  } else {
    cat(paste(factor, ": Skipped (only 1 item)\n"))
  }
}
```

```r
############################################################
# 15. Composite Reliability for Each Factor Group
############################################################
cat("\nComposite Reliability for each factor group:\n")
for (factor in c("Factor1", "Factor2", "Factor3")) {
  items <- factor_groups[[factor]]
  if (length(items) < 2) {
    cat(paste(factor, ": Skipped (only 1 item in this factor group).\n"))
    next
  }
  factor_index <- as.numeric(gsub("Factor", "", factor))
  indices <- which(rownames(loadings_mat) %in% items)
  group_loadings <- loadings_mat[indices, factor_index]
  cr <- (sum(group_loadings))^2 / ((sum(group_loadings))^2 + sum(1 -
group_loadings^2))
  cat(paste(factor, ": Composite Reliability =", round(cr, 3), "\n"))
}


############################################################
# 16. Confirmatory Factor Analysis (CFA)
############################################################
cfa_model <- '
  Factor1 =~ Q1 + Q2 + Q3 + Q8 + Q10 + Q12 + Q14
  Factor2 =~ Q6 + Q7 + Q11
  Factor3 =~ Q4 + Q5 + Q9 + Q13
'

fit_cfa <- cfa(cfa_model, data = data_questions, estimator = "MLM")

# =========================================
# CFA Fit Summary
# =========================================
cfa_fit <- fitMeasures(fit_cfa, c("gfi", "rmsea", "cfi", "tli", "nfi", "rfi",
"chisq", "df"))

# Compute Chi-square/df (Chi-square Test of Independence)
chi_cfa_ratio <- round(cfa_fit["chisq"] / cfa_fit["df"], 3)

# Print Fit Summary
cat(sprintf(

"GFI: %0.3f\nRMSEA: %0.3f\nCFI: %0.3f\nTLI: %0.3f\nNFI: %0.3f\nRFI: %0.3f\nChi
-square/df: %0.3f\n",
  cfa_fit["gfi"], cfa_fit["rmsea"],
  cfa_fit["cfi"], cfa_fit["tli"], cfa_fit["nfi"], cfa_fit["rfi"],
chi_cfa_ratio
))


# Extract standardized factor loadings
load_table <- parameterEstimates(fit_cfa, standardized = TRUE) %>%
  filter(op == "=~") %>%
  select(Factor = lhs, Item = rhs, Std_Loading = std.all)


# --- Standardized Loadings Table ---
cat("\n========================================\nStandardized Factor
Loadings\n========================================\n")
load_table <- parameterEstimates(fit_cfa, standardized = TRUE) %>%
  filter(op == "=~") %>%
  select(Factor = lhs, Item = rhs, Std_Loading = std.all)
print(load_table)

# --- Inter-Item Correlations (CFA) ---
cat("\n========================================\nInter-Item
Correlations\n========================================\n")
latent_corrs <- standardizedSolution(fit_cfa) %>%
  filter(op == "~~", lhs != rhs, lhs %in% c("Factor1", "Factor2",
"Factor3")) %>%
  select(Factor1 = lhs, Factor2 = rhs, Correlation = est.std)
print(latent_corrs)

std_sol <- standardizedSolution(fit_cfa)
# Extract the loadings per factor
get_items <- function(factor_name) {
  std_sol %>%
    filter(op == "=~", lhs == factor_name) %>%
    pull(rhs)
}

# Compute correlations between all items from F2 and F3 (Heterotrait-
Heteromethod)
items_f2 <- get_items("Factor2")
items_f3 <- get_items("Factor3")
```

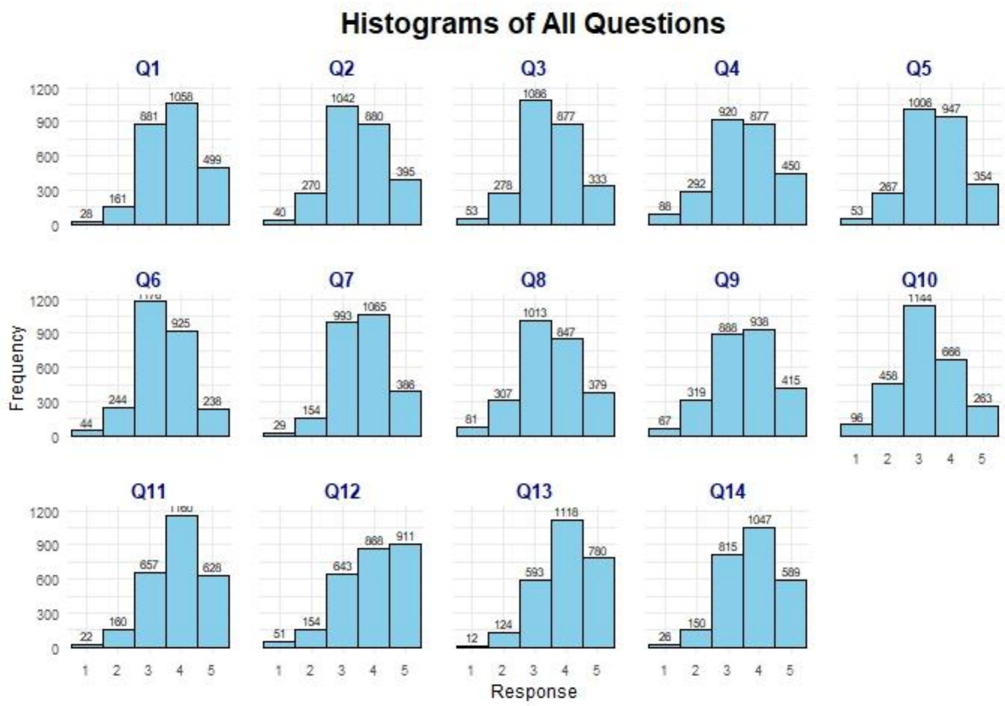
```r
data_f2 <- data_questions[, items_f2]
data_f3 <- data_questions[, items_f3]

ht_corrs <- as.matrix(cor(data_f2, data_f3))
ht_avg <- mean(abs(ht_corrs))

mt_f2_corrs <- cor(data_f2)
mt_f2_avg <- mean(abs(mt_f2_corrs[upper.tri(mt_f2_corrs)]))

mt_f3_corrs <- cor(data_f3)
mt_f3_avg <- mean(abs(mt_f3_corrs[upper.tri(mt_f3_corrs)]))

# HTMT ratio
htmt_f2_f3 <- ht_avg / sqrt(mt_f2_avg * mt_f3_avg)

cat("HTMT Ratio (Factor2 vs Factor3):", round(htmt_f2_f3, 3), "\n")

# --- AVE and Composite Reliability ---
cat("\n========================================\nAVE and Composite
Reliability\n========================================\n")
ave_cr <- load_table %>%
  group_by(Factor) %>%
  summarise(
    AVE = mean(Std_Loading^2),
    CR = (sum(Std_Loading))^2 / ((sum(Std_Loading))^2 + sum(1 -
Std_Loading^2))
  )
print(ave_cr)

############################################################
# 17. 6 Pairwise SEM Structural Models
############################################################

measurement_model <- '
  Factor1 =~ Q1 + Q2 + Q3 + Q8 + Q10 + Q12 + Q14
  Factor2 =~ Q6 + Q7 + Q11
  Factor3 =~ Q4 + Q5 + Q9 + Q13
'

path_pairs <- list(
  c("Factor1", "Factor2"),
  c("Factor2", "Factor1"),
  c("Factor1", "Factor3"),
  c("Factor3", "Factor1"),
  c("Factor2", "Factor3"),
  c("Factor3", "Factor2")
)

sem_results <- data.frame(
  Model = character(),
  Path = character(),
  Std_Estimate = numeric(),
  SE = numeric(),
  p_value = numeric(),
  R2 = numeric(),
  RMSEA = numeric(),
  CFI = numeric(),
  stringsAsFactors = FALSE
)

cat("\n==========================\nPairwise SEM Structural Models
\n==========================\n")

# Loop through all 6 cases
for (i in seq_along(path_pairs)) {
  from <- path_pairs[[i]][1]
  to   <- path_pairs[[i]][2]

  sem_model <- paste0(
    measurement_model,
    "\n", to, " ~ ", from
  )

  fit <- sem(sem_model, data = data_questions, estimator = "MLM")
  fit_std <- standardizedSolution(fit)
  fit_index <- fitMeasures(fit, c("rmsea", "cfi"))

  path_row <- fit_std %>% filter(op == "~" & lhs == to & rhs == from)

  r2_val <- inspect(fit, "r2")[to]

  sem_results[i, ] <- list(
    Model = paste("Model", i),
    Path = paste(from, "→", to),
    Std_Estimate = round(path_row$est.std, 3),
    SE = round(path_row$se, 3),
    p_value = round(path_row$pvalue, 3),
```

```r
    R2 = round(r2_val, 3),
    RMSEA = round(fit_index["rmsea"], 3),
    CFI = round(fit_index["cfi"], 3)
  )
}

print(sem_results)

############################################################
# 18. 3 Combined Predictors for SEM
############################################################
combined_models <- list(
  "F1_predicted_by_F2F3" = '
    Factor1 =~ Q1 + Q2 + Q3 + Q8 + Q10 + Q12 + Q14
    Factor2 =~ Q6 + Q7 + Q11
    Factor3 =~ Q4 + Q5 + Q9 + Q13

    Factor1 ~ Factor2 + Factor3
  ',
  "F2_predicted_by_F1F3" = '
    Factor1 =~ Q1 + Q2 + Q3 + Q8 + Q10 + Q12 + Q14
    Factor2 =~ Q6 + Q7 + Q11
    Factor3 =~ Q4 + Q5 + Q9 + Q13

    Factor2 ~ Factor1 + Factor3
  ',
  "F3_predicted_by_F1F2" = '
    Factor1 =~ Q1 + Q2 + Q3 + Q8 + Q10 + Q12 + Q14
    Factor2 =~ Q6 + Q7 + Q11
    Factor3 =~ Q4 + Q5 + Q9 + Q13

    Factor3 ~ Factor1 + Factor2
  '
)

for (model_name in names(combined_models)) {
  cat("\n========================================\nSEM Path Model:",
model_name, "\n========================================\n")

  fit <- sem(combined_models[[model_name]], data = data_questions, estimator =
"MLM")
  sem_paths <- standardizedSolution(fit)
  path_effects <- sem_paths[sem_paths$op == "~", c("lhs", "rhs", "est.std",
"se", "pvalue", "ci.lower", "ci.upper")]

  r2_val <- inspect(fit, "r2")
  outcome <- unique(path_effects$lhs)  # get the dependent variable name
(Factor1, Factor2, or Factor3)

  path_effects$R2 <- round(r2_val[outcome], 3)
  print(path_effects)
}
```

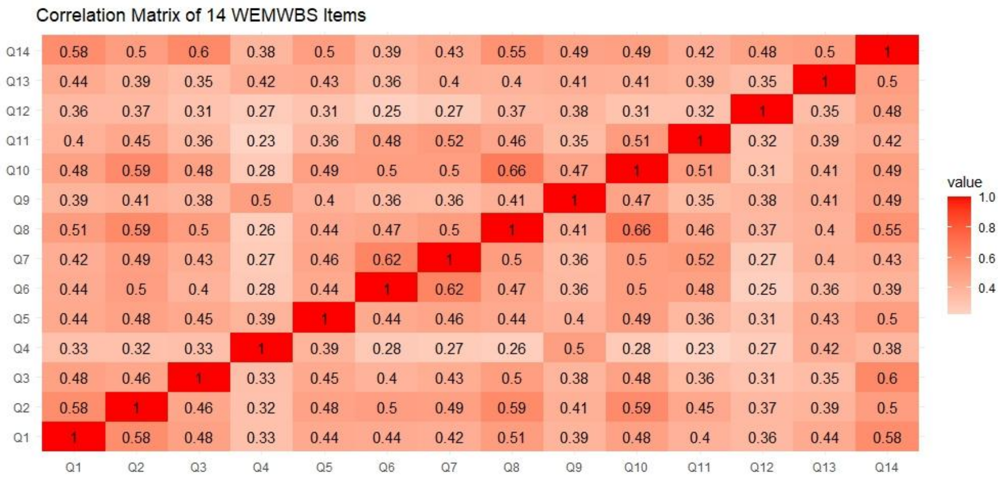
Part B: R Code Outputs and Plots

```
> ############################################################
> # 0. Install and Load Required Libraries
> ############################################################
>
> # install.packages("reshape2")
> # install.packages("igraph")
> # install.packages("recipes")
> # install.packages("caret")
> # install.packages("lavaan")
> # install.packages("MVN")
>
>
> library(readxl)
> library(dplyr)
> library(ggplot2)
> library(reshape2)
> library(psych)
> library(igraph)
> library(lavaan)
> library(tidyr)
> library(dplyr)
> library(MVN)
> library(GPArotation)


> ############################################################
> # 1. Data Loading and Preparation
> ############################################################
> n_factors <- 3
>
> data <- read_excel("C:/Users/lenov/Downloads/FYP First Sem/FYP_Dataset_Lates
tUpToMar25_OnlyFirstYearStudent.xlsx")
> question_cols <- paste0("Q", 1:14)
> data_questions <- data %>% select(all_of(question_cols))
>
> data_questions <- data_questions %>%
+   mutate(across(everything(), ~ ifelse(is.na(.), mean(., na.rm = TRUE), .)))
>
> data_scaled <- scale(data_questions)
>
> ############################################################
> # 2. Univariate and Multivariate Normality Check
> ############################################################
> #Skewness, Kurtosis and Histogram for Univariate Normality Check
>
> normality_summary <- data.frame(
+   Skewness = round(sapply(data_questions, skew), 3),
+   Kurtosis = round(sapply(data_questions, kurtosi), 3)
+ )
> rownames(normality_summary) <- question_cols  # Set row names as Q1-Q14
>
> cat("\nSkewness and Kurtosis for Each Variable:\n")

Skewness and Kurtosis for Each Variable:
```

```
> print(normality_summary)
    Skewness Kurtosis
Q1    -0.289   -0.174
Q2    -0.117   -0.326
Q3    -0.141   -0.162
Q4    -0.310   -0.309
Q5    -0.227   -0.161
Q6    -0.138    0.114
Q7    -0.210   -0.014
Q8    -0.219   -0.259
Q9    -0.291   -0.349
Q10    0.002   -0.270
Q11   -0.504   -0.062
Q12   -0.676   -0.103
Q13   -0.532   -0.184
Q14   -0.345   -0.235
>
> skew_range <- range(normality_summary$Skewness)
> kurtosis_range <- range(normality_summary$Kurtosis)
>
> cat("\nRange of Skewness: [", skew_range[1], ",", skew_range[2], "]\n")

Range of Skewness: [ -0.676 , 0.002 ]
> cat("Range of Kurtosis: [", kurtosis_range[1], ",", kurtosis_range[2], "]\n
")
Range of Kurtosis: [ -0.349 , 0.114 ]
>
>
> #All 14 Ques Histogram in one page
> data_long <- data_questions %>%
+   mutate(ID = row_number()) %>%
+   pivot_longer(cols = starts_with("Q"), names_to = "Question", values_to = "
Response")
>
> data_long$Question <- factor(data_long$Question, levels = paste0("Q", 1:14))
> ggplot(data_long, aes(x = factor(Response))) +
+   geom_bar(fill = "skyblue", color = "black", width = 1) +
+   geom_text(stat = "count", aes(label = ..count..), vjust = -0.4, size = 2.
2) +
+   facet_wrap(~ Question, ncol = 5) +
+   scale_x_discrete(limits = as.character(1:5)) +
+   labs(title = "Histograms of All Questions", x = "Response", y = "Frequency
") +
+   theme_minimal(base_size = 9) +
+   theme(
+     plot.title = element_text(size = 14, face = "bold", hjust = 0.5),
+     strip.text = element_text(size = 10, face = "bold", color = "darkblue"),
+     axis.text.x = element_text(size = 7),
+     axis.text.y = element_text(size = 7),
+     panel.spacing = unit(1, "lines")
+   )
>
>
```

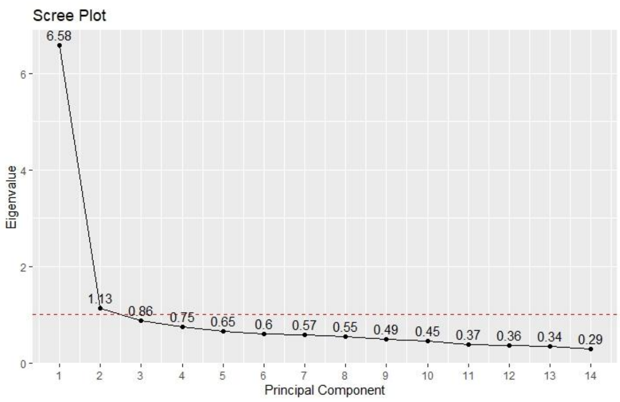



```
> # Mardia's Test for Multivariate Normality Check
> cat("\n[Mardia's Test]\n")

[Mardia's Test]
> mardia_result <- mvn(data = data_questions, mvnTest = "mardia")
> print(mardia_result$multivariateNormality)
             Test        Statistic               p value Result
Mardia Skewness 2911.12587612046 4.54556969004242e-313     NO
Mardia Kurtosis  82.188668943424                     0     NO
              MVN             <NA>                  <NA>     NO
>
> ############################################################
> # 3. Correlation Matrix and Heatmap
> ############################################################
> corr_matrix <- cor(data_scaled)
> melted_corr <- melt(corr_matrix)
> p_corr <- ggplot(data = melted_corr, aes(x = Var1, y = Var2, fill = value))
+
+   geom_tile() +
+   geom_text(aes(label = round(value, 2))) +
+   scale_fill_gradient2(low = "blue", high = "red", mid = "white", midpoint =
 0) +
+   theme_minimal() +
+   labs(x = NULL, y = NULL) +
+   ggtitle("Correlation Matrix of 14 WEMWBS Items")
> print(p_corr)
```



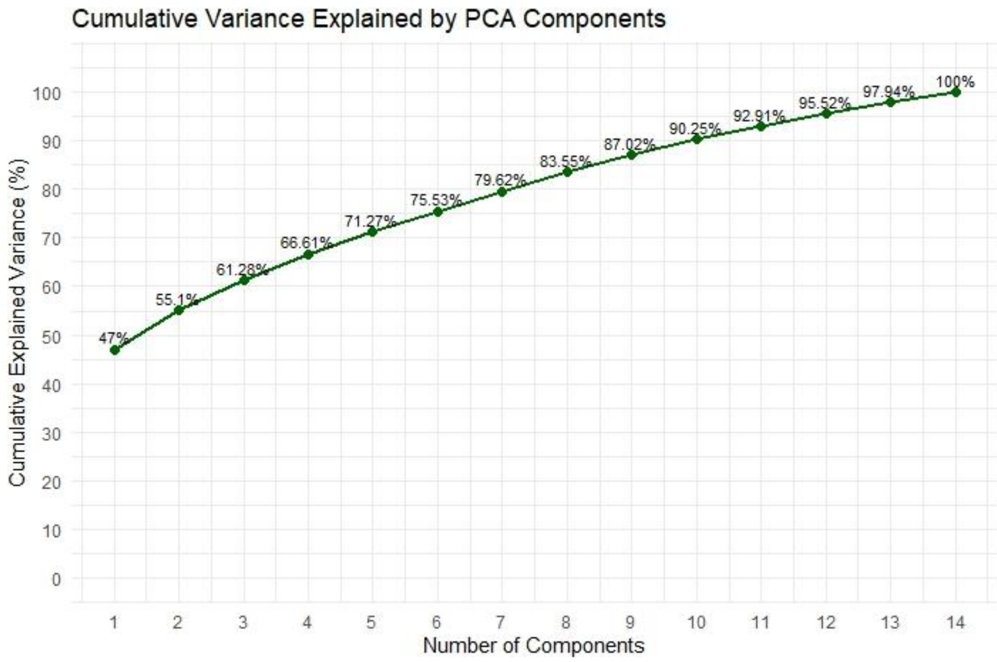

```
> ############################################################
> # 4. Determinant and Average Inter-Item Correlation (AIIC)
> ############################################################
> det_corr <- det(corr_matrix)
> cat("\nDeterminant of the correlation matrix (Determinant Score):", det_cor
r, "\n")

Determinant of the correlation matrix (Determinant Score): 0.001724003
> upper_tri <- corr_matrix[upper.tri(corr_matrix)]
> aiic <- mean(upper_tri)
> cat("Average Inter-Item Correlation (AIIC):", aiic, "\n")
Average Inter-Item Correlation (AIIC): 0.4243146
>
> ############################################################
> # 5. Data Adequacy Testing: Bartlett's Test and KMO
> ############################################################
> bartlett_result <- cortest.bartlett(corr_matrix, n = nrow(data_scaled))
> cat("\nBartlett's Test:\n")

Bartlett's Test:
> print(bartlett_result)
$chisq
[1] 16674.52

$p.value
[1] 0

$df
[1] 91

>
> kmo_result <- KMO(corr_matrix)
> cat("\nKaiser-Meyer-Olkin (KMO) Test:\n")

Kaiser-Meyer-Olkin (KMO) Test:
> print(kmo_result)
Kaiser-Meyer-Olkin factor adequacy
Call: KMO(r = corr_matrix)
Overall MSA =  0.94
MSA for each item = 
  Q1   Q2   Q3   Q4   Q5   Q6   Q7   Q8   Q9  Q10  Q11  Q12  Q13  Q14 
0.95 0.95 0.94 0.89 0.96 0.93 0.93 0.94 0.93 0.93 0.95 0.95 0.95 0.92
```

```
> ############################################################
> # 6. Principal Component Analysis (PCA)
> ############################################################
> pca_res <- prcomp(data_scaled, center = TRUE, scale. = TRUE)
> eigenvalues <- pca_res$sdev^2
> explained_variance_ratio <- eigenvalues / sum(eigenvalues)
> cumulative_explained <- cumsum(explained_variance_ratio)
>
> cat("\nEigenvalues from PCA:\n")

Eigenvalues from PCA:
> for (i in 1:length(eigenvalues)) {
+   cat(paste("Principal Component", i, ":", round(eigenvalues[i], 4), "\n"))
+ }
Principal Component 1 : 6.5805
Principal Component 2 : 1.1337
Principal Component 3 : 0.8645
Principal Component 4 : 0.7472
Principal Component 5 : 0.6521
Principal Component 6 : 0.596
Principal Component 7 : 0.5729
Principal Component 8 : 0.55
Principal Component 9 : 0.4854
Principal Component 10 : 0.4529
Principal Component 11 : 0.3725
Principal Component 12 : 0.365
Principal Component 13 : 0.3385
Principal Component 14 : 0.2887
>
> ############################################################
> # 7. Scree Plot
> ############################################################
>
> df_scree <- data.frame(PC = 1:length(eigenvalues), Eigenvalue = eigenvalues)
> p_scree <- ggplot(df_scree, aes(x = PC, y = Eigenvalue)) +
+   geom_line() +
+   geom_point() +
+   geom_hline(yintercept = 1, linetype = "dashed", color = "red") +
+   geom_text(aes(label = round(Eigenvalue, 2)), vjust = -0.5) +
+   ggtitle("Scree Plot") +
+   xlab("Principal Component") +
+   ylab("Eigenvalue") +
+   scale_x_continuous(breaks = 1:length(eigenvalues))
> print(p_scree)
```



```
> cat("\nExplained Variance Ratio (in %):\n")

Explained Variance Ratio (in %):
> for (i in 1:length(explained_variance_ratio)) {
+   cat(paste("PC", i, ":", round(explained_variance_ratio[i] * 100, 2), "%\n
"))
+ }
PC 1 : 47 %
PC 2 : 8.1 %
PC 3 : 6.18 %
PC 4 : 5.34 %
PC 5 : 4.66 %
PC 6 : 4.26 %
PC 7 : 4.09 %
PC 8 : 3.93 %
PC 9 : 3.47 %
PC 10 : 3.23 %
PC 11 : 2.66 %
PC 12 : 2.61 %
PC 13 : 2.42 %
PC 14 : 2.06 %
> cat("\nCumulative Explained Variance (in %):\n")

Cumulative Explained Variance (in %):
> for (i in 1:length(cumulative_explained)) {
+   cat(paste("PC1 to PC", i, ":", round(cumulative_explained[i] * 100, 2), "%
\n"))
+ }
PC1 to PC 1 : 47 %
PC1 to PC 2 : 55.1 %
PC1 to PC 3 : 61.28 %
PC1 to PC 4 : 66.61 %
PC1 to PC 5 : 71.27 %
PC1 to PC 6 : 75.53 %
PC1 to PC 7 : 79.62 %
PC1 to PC 8 : 83.55 %
PC1 to PC 9 : 87.02 %
PC1 to PC 10 : 90.25 %
PC1 to PC 11 : 92.91 %
PC1 to PC 12 : 95.52 %
PC1 to PC 13 : 97.94 %
PC1 to PC 14 : 100 %


> ############################################################
> # 8. Cumulative Explained Variance Plot
> ############################################################
> df_cum <- data.frame(
+   Components = 1:length(cumulative_explained),
+   Cumulative = cumulative_explained * 100
+ )
>
> p_cum <- ggplot(df_cum, aes(x = Components, y = Cumulative)) +
+   geom_line(color = "darkgreen", linewidth = 1) +
+   geom_point(color = "darkgreen", size = 2) +
+   geom_text(
+     aes(label = paste0(round(Cumulative, 2), "%")),
+     vjust = -0.5, size = 3
```

```
+   ) +
+   labs(
+     title = "Cumulative Variance Explained by PCA Components",
+     x = "Number of Components",
+     y = "Cumulative Explained Variance (%)"
+   ) +
+   theme_minimal(base_size = 11) +
+   scale_y_continuous(limits = c(0, 105), breaks = seq(0, 100, 10)) +
+   scale_x_continuous(breaks = 1:14)
>
> print(p_cum)
```



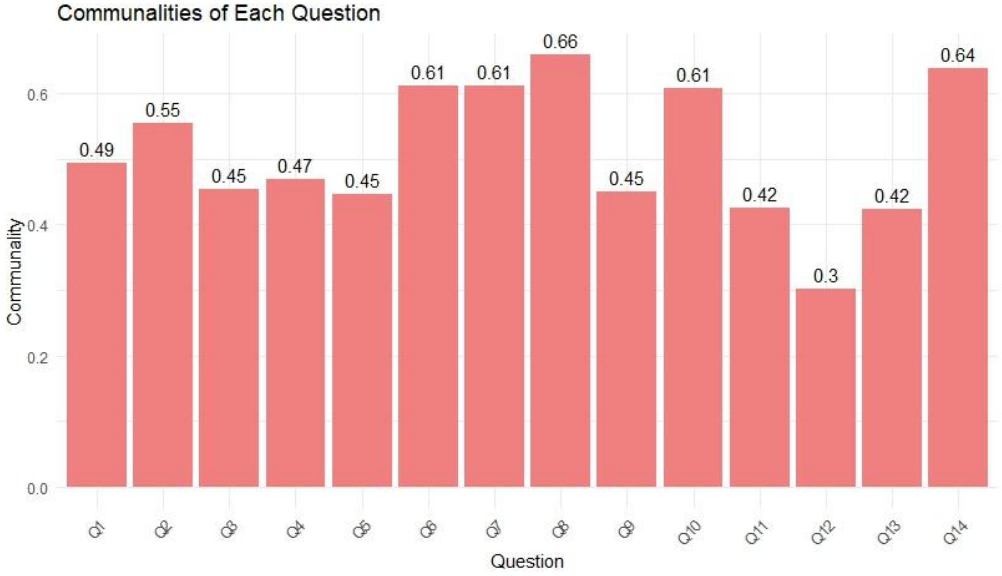

```
> ############################################################
> # 9. K-Folds Cross Validation for PCA
> ############################################################
> set.seed(123)
> k_folds <- 10
> n_factors <- 3
> n <- nrow(data_scaled)
> fold_indices <- sample(rep(1:k_folds, length.out = n))
>
> fold_results <- matrix(0, nrow = k_folds, ncol = n_factors)
> training_accuracies <- numeric(k_folds)
>
> for (k in 1:k_folds) {
+   test_idx <- which(fold_indices == k)
+   train_data <- data_scaled[-test_idx, ]
+
+   pca_model <- prcomp(train_data, center = TRUE, scale. = TRUE)
+   explained_var <- (pca_model$sdev)^2 / sum((pca_model$sdev)^2)
+   explained_ratio_top <- explained_var[1:n_factors]
+   cumulative_accuracy <- sum(explained_ratio_top) * 100
+
+   fold_results[k, ] <- explained_ratio_top * 100
+   training_accuracies[k] <- cumulative_accuracy
+
```

```
+   cat(sprintf("Fold %2d: Explained Variance Ratio: [%0.8f, %0.8f, %0.8f], Tr
aining Accuracy (cumulative): %0.2f%%\n",
+               k, explained_ratio_top[1], explained_ratio_top[2], explained_r
atio_top[3], cumulative_accuracy))
+ }
Fold  1: Explained Variance Ratio: [0.47053863, 0.08103842, 0.06282876], Train
ing Accuracy (cumulative): 61.44%
Fold  2: Explained Variance Ratio: [0.47478241, 0.08024362, 0.06163263], Train
ing Accuracy (cumulative): 61.67%
Fold  3: Explained Variance Ratio: [0.47128907, 0.08016686, 0.06153318], Train
ing Accuracy (cumulative): 61.30%
Fold  4: Explained Variance Ratio: [0.47038410, 0.08181380, 0.06221353], Train
ing Accuracy (cumulative): 61.44%
Fold  5: Explained Variance Ratio: [0.47098099, 0.08141808, 0.06152963], Train
ing Accuracy (cumulative): 61.39%
Fold  6: Explained Variance Ratio: [0.46987890, 0.08068610, 0.06127117], Train
ing Accuracy (cumulative): 61.18%
Fold  7: Explained Variance Ratio: [0.46506023, 0.08243363, 0.06229347], Train
ing Accuracy (cumulative): 60.98%
Fold  8: Explained Variance Ratio: [0.46576575, 0.08051732, 0.06152541], Train
ing Accuracy (cumulative): 60.78%
Fold  9: Explained Variance Ratio: [0.47424979, 0.08000133, 0.06163227], Train
ing Accuracy (cumulative): 61.59%
Fold 10: Explained Variance Ratio: [0.46747653, 0.08191578, 0.06166049], Train
ing Accuracy (cumulative): 61.11%
>
> avg_explained <- colMeans(fold_results)
> cat("\n[Cross-Validation] Average Explained Variance Ratio across folds:\n")

[Cross-Validation] Average Explained Variance Ratio across folds:
> print(round(avg_explained, 2))
[1] 47.00  8.10  6.18
>
> cumulative_explained <- cumsum(avg_explained)
> cat("\nCumulative Explained Variance:", round(cumulative_explained, 2), "\n
")

Cumulative Explained Variance: 47 55.11 61.29
> cat("Average Training Accuracy (cumulative explained variance):", round(mean
(training_accuracies), 2), "%\n")
Average Training Accuracy (cumulative explained variance): 61.29 %
```

```
> ###########################################################
> # 10. Exploratory Factor Analysis (EFA)
> ###########################################################
> # Perform EFA using maximum likelihood and promax rotation
> efa_res <- fa(data_scaled, nfactors = n_factors, fm = "ml", rotate = "promax
")
> cat("\nFactor Loadings:\n")

Factor Loadings:
> print(efa_res$loadings)

Loadings:
    ML1    ML3    ML2
Q1   0.532         0.168
Q2   0.574  0.227
Q3   0.536         0.165
Q4  -0.209         0.805
Q5   0.196  0.238  0.324
Q6          0.781
Q7          0.737
Q8   0.864  0.124 -0.218
Q9   0.148         0.527
Q10  0.660  0.264 -0.132
Q11  0.233  0.456
Q12  0.395 -0.132  0.286
Q13  0.105  0.120  0.490
Q14  0.614 -0.145  0.344

                 ML1   ML3   ML2
SS loadings    2.788 1.613 1.598
Proportion Var 0.199 0.115 0.114
Cumulative Var 0.199 0.314 0.428
>
> loadings_mat <- unclass(efa_res$loadings)
> loadings_df <- as.data.frame(loadings_mat)
> loadings_df$Question <- rownames(loadings_df)
> loadings_df$Question <- factor(loadings_df$Question, levels = paste0("Q", 1
4:1))
> colnames(loadings_df)[1:n_factors] <- paste0("F", 1:n_factors)
> melted_loadings <- melt(loadings_df, id.vars = "Question")
>
> # Heatmap Plot
> p_loadings <- ggplot(melted_loadings, aes(x = variable, y = Question, fill =
 value)) +
+   geom_tile() +
+   geom_text(aes(label = round(value, 2))) +
+   scale_fill_gradient2(low = "blue", high = "yellow", mid = "green", midpoin
t = 0) +
+   labs(title = "Factor Loadings Heatmap", x = "Factors", y = "Questions") +
+   theme_minimal()
> print(p_loadings)
```


```
> ############################################################
> # 11. Communalities
> ############################################################
> communalities <- efa_res$communality
> communalities_df <- data.frame(
+   Question = factor(question_cols, levels = paste0("Q", 1:14)),
+   Communality = communalities
+ )
>
> cat("\nCommunalities for each question:\n")

Communalities for each question:
> print(communalities_df, row.names = FALSE)
 Question Communality
       Q1   0.4943155
       Q2   0.5546125
       Q3   0.4534470
       Q4   0.4699522
       Q5   0.4456393
       Q6   0.6112023
       Q7   0.6118007
       Q8   0.6585652
       Q9   0.4500943
      Q10   0.6071861
      Q11   0.4246734
      Q12   0.3011049
      Q13   0.4244732
      Q14   0.6392482
>
> p_communalities <- ggplot(communalities_df, aes(x = Question, y = Communalit
y)) +
+   geom_bar(stat = "identity", fill = "lightcoral") +
+   geom_text(aes(label = round(Communality, 2)), vjust = -0.5) +
+   ggtitle("Communalities of Each Question") +
```

```
+   theme_minimal() +
+   theme(axis.text.x = element_text(angle = 45, hjust = 1))
> print(p_communalities)
>
```




```
> ############################################################
> # 12. Dynamic Grouping of Items by Factor
> ############################################################
> loadings_mat <- as.matrix(efa_res$loadings)
> factor_groups <- list()
> for (i in 1:n_factors) {
+   factor_groups[[paste0("Factor", i)]] <- c()
+ }
> for (i in 1:nrow(loadings_mat)) {
+   max_index <- which.max(abs(loadings_mat[i, ]))
+   factor_groups[[paste0("Factor", max_index)]] <- c(factor_groups[[paste0("F
actor", max_index)]],
+                                                       rownames(loadings_mat)
[i])
+ }
> cat("\nDynamic grouping of items by factor:\n")

Dynamic grouping of items by factor:
> print(factor_groups)
$Factor1
[1] "Q1"   "Q2"   "Q3"   "Q8"   "Q10"  "Q12"  "Q14"

$Factor3
[1] "Q4"   "Q5"   "Q9"   "Q13"

$Factor2
[1] "Q6"   "Q7"   "Q11"
```

```
> ###########################################################
> # 13. McDonald's Omega for Each Factor Group
> ###########################################################
> cat("\nMcDonald's Omega for each factor group:\n")

McDonald's Omega for each factor group:
> for (factor in c("Factor1", "Factor2", "Factor3")) {
+   items <- factor_groups[[factor]]
+   if (length(items) < 2) {
+     cat(paste(factor, ": Skipped (only 1 item in this factor group).\n"))
+     next
+   }
+   omega_res <- omega(data_questions[, items], nfactors = 1, plot = FALSE)
+   cat(paste(factor, ": Omega =", round(omega_res$omega.tot, 3), "\n"))
+ }
Omega_h for 1 factor is not meaningful, just omega_t
Factor1 : Omega = 0.872
Omega_h for 1 factor is not meaningful, just omega_t
Factor2 : Omega = 0.783
Omega_h for 1 factor is not meaningful, just omega_t
Factor3 : Omega = 0.748
Warning messages:
1: In schmid(m, nfactors, fm, digits, rotate = rotate, n.obs = n.obs,  :
  Omega_h and Omega_asymptotic are not meaningful with one factor
2: In schmid(m, nfactors, fm, digits, rotate = rotate, n.obs = n.obs,  :
  Omega_h and Omega_asymptotic are not meaningful with one factor
3: In schmid(m, nfactors, fm, digits, rotate = rotate, n.obs = n.obs,  :
  Omega_h and Omega_asymptotic are not meaningful with one factor
>
>
> ###########################################################
> # 14. Cronbach's Alpha for Dynamically Grouped Items
> ###########################################################
> cat("\nCronbach's Alpha for dynamically grouped items:\n")

Cronbach's Alpha for dynamically grouped items:
> for (factor in c("Factor1", "Factor2", "Factor3")) {
+   items <- factor_groups[[factor]]
+   if (length(items) >= 2) {
+     alpha_res <- alpha(data_questions[, items])
+     cat(paste(factor, ": α =", round(alpha_res$total$raw_alpha, 3), "\n"))
+   } else {
+     cat(paste(factor, ": Skipped (only 1 item)\n"))
+   }
+ }
Factor1 : α = 0.868
Factor2 : α = 0.778
Factor3 : α = 0.747
```

```
> ###########################################################
> # 15. Composite Reliability for Each Factor Group
> ###########################################################
> cat("\nComposite Reliability for each factor group:\n")

Composite Reliability for each factor group:
> for (factor in c("Factor1", "Factor2", "Factor3")) {
+   items <- factor_groups[[factor]]
+   if (length(items) < 2) {
+     cat(paste(factor, ": Skipped (only 1 item in this factor group).\n"))
+     next
+   }
+   factor_index <- as.numeric(gsub("Factor", "", factor))
+   indices <- which(rownames(loadings_mat) %in% items)
+   group_loadings <- loadings_mat[indices, factor_index]
+   cr <- (sum(group_loadings))^2 / ((sum(group_loadings))^2 + sum(1 - group_l
oadings^2))
+   cat(paste(factor, ": Composite Reliability =", round(cr, 3), "\n"))
+ }
Factor1 : Composite Reliability = 0.799
Factor2 : Composite Reliability = 0.704
Factor3 : Composite Reliability = 0.628
>
> ###########################################################
> # 16. Confirmatory Factor Analysis (CFA)
> ###########################################################
> cfa_model <- '
+   Factor1 =~ Q1 + Q2 + Q3 + Q8 + Q10 + Q12 + Q14
+   Factor2 =~ Q6 + Q7 + Q11
+   Factor3 =~ Q4 + Q5 + Q9 + Q13
+ '
>
> fit_cfa <- cfa(cfa_model, data = data_questions, estimator = "MLM")
>
> # =========================================
> # CFA Fit Summary
> # =========================================
> cfa_fit <- fitMeasures(fit_cfa, c("gfi", "rmsea", "cfi", "tli", "nfi", "rfi
", "chisq", "df"))
>
> # Compute Chi-square/df (Chi-square Test of Independence)
> chi_cfa_ratio <- round(cfa_fit["chisq"] / cfa_fit["df"], 3)
>
> # Print Fit Summary
> cat(sprintf(
+   "GFI: %0.3f\nRMSEA: %0.3f\nCFI: %0.3f\nTLI: %0.3f\nNFI: %0.3f\nRFI: %0.3f\
nChi-square/df: %0.3f\n",
+   cfa_fit["gfi"], cfa_fit["rmsea"],
+   cfa_fit["cfi"], cfa_fit["tli"], cfa_fit["nfi"], cfa_fit["rfi"], chi_cfa_ra
tio
+ ))
GFI: 0.927
RMSEA: 0.081
CFI: 0.923
TLI: 0.905
NFI: 0.919
RFI: 0.900
Chi-square/df: 18.345
```

```
>
>
> # Extract standardized factor loadings
> load_table <- parameterEstimates(fit_cfa, standardized = TRUE) %>%
+   filter(op == "=~") %>%
+   select(Factor = lhs, Item = rhs, Std_Loading = std.all)
>
>
> # --- Standardized Loadings Table ---
> cat("\n========================================\nStandardized Factor Loadin
gs\n========================================\n")

========================================
Standardized Factor Loadings
========================================
> load_table <- parameterEstimates(fit_cfa, standardized = TRUE) %>%
+   filter(op == "=~") %>%
+   select(Factor = lhs, Item = rhs, Std_Loading = std.all)
> print(load_table)
    Factor Item Std_Loading
1  Factor1   Q1       0.710
2  Factor1   Q2       0.748
3  Factor1   Q3       0.671
4  Factor1   Q8       0.763
5  Factor1  Q10       0.752
6  Factor1  Q12       0.505
7  Factor1  Q14       0.753
8  Factor2   Q6       0.757
9  Factor2   Q7       0.787
10 Factor2  Q11       0.673
11 Factor3   Q4       0.573
12 Factor3   Q5       0.692
13 Factor3   Q9       0.665
14 Factor3  Q13       0.654
>
> # --- Inter-Item Correlations (CFA) ---
> cat("\n========================================\nInter-Item Correlations\n=
========================================\n")

========================================
Inter-Item Correlations
========================================
> latent_corrs <- standardizedSolution(fit_cfa) %>%
+   filter(op == "~~", lhs != rhs, lhs %in% c("Factor1", "Factor2", "Factor3
")) %>%
+   select(Factor1 = lhs, Factor2 = rhs, Correlation = est.std)
> print(latent_corrs)
  Factor1 Factor2 Correlation
1 Factor1 Factor2       0.817
2 Factor1 Factor3       0.868
3 Factor2 Factor3       0.753
>
> std_sol <- standardizedSolution(fit_cfa)
> # Extract the loadings per factor
> get_items <- function(factor_name) {
+   std_sol %>%
+     filter(op == "=~", lhs == factor_name) %>%
+     pull(rhs)
```

```
+ }
>
> # Compute correlations between all items from F2 and F3 (Heterotrait-Heterom
ethod)
> items_f2 <- get_items("Factor2")
> items_f3 <- get_items("Factor3")
>
>
> data_f2 <- data_questions[, items_f2]
> data_f3 <- data_questions[, items_f3]
>
> ht_corrs <- as.matrix(cor(data_f2, data_f3))
> ht_avg <- mean(abs(ht_corrs))
>
> mt_f2_corrs <- cor(data_f2)
> mt_f2_avg <- mean(abs(mt_f2_corrs[upper.tri(mt_f2_corrs)]))
>
> mt_f3_corrs <- cor(data_f3)
> mt_f3_avg <- mean(abs(mt_f3_corrs[upper.tri(mt_f3_corrs)]))
>
> # HTMT ratio
> htmt_f2_f3 <- ht_avg / sqrt(mt_f2_avg * mt_f3_avg)
>
> cat("HTMT Ratio (Factor2 vs Factor3):", round(htmt_f2_f3, 3), "\n")
HTMT Ratio (Factor2 vs Factor3): 0.742
>
> # --- AVE and Composite Reliability ---
> cat("\n=========================================\nAVE and Composite Reliabil
ity\n=========================================\n")

=========================================
AVE and Composite Reliability
=========================================
> ave_cr <- load_table %>%
+   group_by(Factor) %>%
+   summarise(
+     AVE = mean(Std_Loading^2),
+     CR = (sum(Std_Loading))^2 / ((sum(Std_Loading))^2 + sum(1 - Std_Loading^
2))
+   )
> print(ave_cr)
# A tibble: 3 × 3
  Factor    AVE    CR
  <chr>   <dbl> <dbl>
1 Factor1 0.498 0.872
2 Factor2 0.548 0.784
3 Factor3 0.420 0.742
>
```

```
> ###########################################################
> # 17. 6 Pairwise SEM Structural Models
> ###########################################################
>
> measurement_model <- '
+   Factor1 =~ Q1 + Q2 + Q3 + Q8 + Q10 + Q12 + Q14
+   Factor2 =~ Q6 + Q7 + Q11
+   Factor3 =~ Q4 + Q5 + Q9 + Q13
+ '
>
> path_pairs <- list(
+   c("Factor1", "Factor2"),
+   c("Factor2", "Factor1"),
+   c("Factor1", "Factor3"),
+   c("Factor3", "Factor1"),
+   c("Factor2", "Factor3"),
+   c("Factor3", "Factor2")
+ )
>
> sem_results <- data.frame(
+   Model = character(),
+   Path = character(),
+   Std_Estimate = numeric(),
+   SE = numeric(),
+   p_value = numeric(),
+   R2 = numeric(),
+   RMSEA = numeric(),
+   CFI = numeric(),
+   stringsAsFactors = FALSE
+ )
>
> cat("\n===========================\nPairwise SEM Structural Models \n=======
====================\n")

===========================
Pairwise SEM Structural Models
===========================
>
> # Loop through all 6 cases
> for (i in seq_along(path_pairs)) {
+   from <- path_pairs[[i]][1]
+   to   <- path_pairs[[i]][2]
+
+   sem_model <- paste0(
+     measurement_model,
+     "\n", to, " ~ ", from
+   )
+
+   fit <- sem(sem_model, data = data_questions, estimator = "MLM")
+   fit_std <- standardizedSolution(fit)
+   fit_index <- fitMeasures(fit, c("rmsea", "cfi"))
+
+   path_row <- fit_std %>% filter(op == "~" & lhs == to & rhs == from)
+
+   r2_val <- inspect(fit, "r2")[to]
+
+   sem_results[i, ] <- list(
+     Model = paste("Model", i),
```

```
+     Path = paste(from, "→", to),
+     Std_Estimate = round(path_row$est.std, 3),
+     SE = round(path_row$se, 3),
+     p_value = round(path_row$pvalue, 3),
+     R2 = round(r2_val, 3),
+     RMSEA = round(fit_index["rmsea"], 3),
+     CFI = round(fit_index["cfi"], 3)
+   )
+ }
>
> print(sem_results)
    Model              Path Std_Estimate    SE p_value    R2 RMSEA   CFI
1 Model 1 Factor1 → Factor2        0.825 0.012       0 0.681 0.081 0.922
2 Model 2 Factor2 → Factor1        0.908 0.009       0 0.825 0.088 0.907
3 Model 3 Factor1 → Factor3        0.875 0.011       0 0.765 0.081 0.922
4 Model 4 Factor3 → Factor1        0.929 0.009       0 0.862 0.084 0.917
5 Model 5 Factor2 → Factor3        0.872 0.013       0 0.761 0.088 0.907
6 Model 6 Factor3 → Factor2        0.837 0.013       0 0.700 0.084 0.917
>
> ###########################################################
> # 18. 3 Combined Predictors for SEM
> ###########################################################
> combined_models <- list(
+   "F1_predicted_by_F2F3" = '
+     Factor1 =~ Q1 + Q2 + Q3 + Q8 + Q10 + Q12 + Q14
+     Factor2 =~ Q6 + Q7 + Q11
+     Factor3 =~ Q4 + Q5 + Q9 + Q13
+
+     Factor1 ~ Factor2 + Factor3
+   ',
+   "F2_predicted_by_F1F3" = '
+     Factor1 =~ Q1 + Q2 + Q3 + Q8 + Q10 + Q12 + Q14
+     Factor2 =~ Q6 + Q7 + Q11
+     Factor3 =~ Q4 + Q5 + Q9 + Q13
+
+     Factor2 ~ Factor1 + Factor3
+   ',
+   "F3_predicted_by_F1F2" = '
+     Factor1 =~ Q1 + Q2 + Q3 + Q8 + Q10 + Q12 + Q14
+     Factor2 =~ Q6 + Q7 + Q11
+     Factor3 =~ Q4 + Q5 + Q9 + Q13
+
+     Factor3 ~ Factor1 + Factor2
+   '
+ )
>
> for (model_name in names(combined_models)) {
+   cat("\n========================================\nSEM Path Model:", model_
name, "\n========================================\n")
+
+   fit <- sem(combined_models[[model_name]], data = data_questions, estimator
 = "MLM")
+   sem_paths <- standardizedSolution(fit)
+   path_effects <- sem_paths[sem_paths$op == "~", c("lhs", "rhs", "est.std",
"se", "pvalue", "ci.lower", "ci.upper")]
+
+   r2_val <- inspect(fit, "r2")
```

```
+   outcome <- unique(path_effects$lhs)  # get the dependent variable name (Fa
ctor1, Factor2, or Factor3)
+ 
+   path_effects$R2 <- round(r2_val[outcome], 3)
+   print(path_effects)
+ }

=========================================
SEM Path Model: F1_predicted_by_F2F3
=========================================
       lhs     rhs est.std    se pvalue ci.lower ci.upper    R2
15 Factor1 Factor2   0.379 0.031      0    0.317    0.440 0.815
16 Factor1 Factor3   0.583 0.031      0    0.522    0.644 0.815

=========================================
SEM Path Model: F2_predicted_by_F1F3
=========================================
       lhs     rhs est.std    se pvalue ci.lower ci.upper    R2
15 Factor2 Factor1   0.665 0.058  0.000    0.552    0.778 0.676
16 Factor2 Factor3   0.176 0.062  0.004    0.055    0.297 0.676

=========================================
SEM Path Model: F3_predicted_by_F1F2
=========================================
       lhs     rhs est.std    se pvalue ci.lower ci.upper    R2
15 Factor3 Factor1   0.761 0.044  0.000    0.676    0.847 0.759
16 Factor3 Factor2   0.131 0.047  0.006    0.038    0.223 0.759
```

Appendix I: SPSS Output for PCA and EFA

Notes

Output Created		29-MAR-2025 21:39:10
Comments		
Input	Data	C:\Users\lenov\Downloads\FYP First Sem\ML_Promax_3var_WITH Kaiser Nor.sav
	Active Dataset	DataSet1
	Filter	<none>
	Weight	<none>
	Split File	<none>
	N of Rows in Working Data File	2627
Missing Value Handling	Definition of Missing	MISSING=EXCLUDE: User-defined missing values are treated as missing.
	Cases Used	LISTWISE: Statistics are based on cases with no missing values for any variable used.
Syntax		FACTOR /VARIABLES Q1 Q2 Q3 Q4 Q5 Q6 Q7 Q8 Q9 Q10 Q11 Q12 Q13 Q14 /MISSING LISTWISE /ANALYSIS Q1 Q2 Q3 Q4 Q5 Q6 Q7 Q8 Q9 Q10 Q11 Q12 Q13 Q14 /PRINT INITIAL CORRELATION DET KMO EXTRACTION ROTATION /FORMAT BLANK(0.3) /PLOT ROTATION /CRITERIA FACTORS(3) ITERATE(25) /EXTRACTION ML /CRITERIA KAISER

		ITERATE(25) /ROTATION PROMAX(4).
Resources	Processor Time	00:00:00.25
	Elapsed Time	00:00:00.17
	Maximum Memory Required	25128 (24.539K) bytes

[DataSet1] C:\Users\lenov\Downloads\FYP First Sem\ML_Promax_3var_WITH Kaiser Nor.sav

Correlation Matrix[a]

		Q1	Q2	Q3	Q4	Q5	Q6	Q7
Correlation	Q1	1.000	.578	.485	.326	.442	.437	.423
	Q2	.578	1.000	.459	.323	.481	.499	.489
	Q3	.485	.459	1.000	.329	.448	.395	.427
	Q4	.326	.323	.329	1.000	.392	.282	.269
	Q5	.442	.481	.448	.392	1.000	.443	.464
	Q6	.437	.499	.395	.282	.443	1.000	.620
	Q7	.423	.489	.427	.269	.464	.620	1.000
	Q8	.505	.592	.501	.255	.441	.465	.503
	Q9	.388	.407	.381	.496	.403	.359	.356
	Q10	.484	.590	.480	.277	.490	.499	.500
	Q11	.402	.451	.356	.227	.364	.484	.516
	Q12	.362	.365	.312	.270	.308	.246	.273
	Q13	.436	.388	.347	.423	.428	.358	.404
	Q14	.580	.503	.598	.381	.501	.395	.431

Correlation Matrix[a]

		Q8	Q9	Q10	Q11	Q12	Q13	Q14
Correlation	Q1	.505	.388	.484	.402	.362	.436	.580
	Q2	.592	.407	.590	.451	.365	.388	.503
	Q3	.501	.381	.480	.356	.312	.347	.598
	Q4	.255	.496	.277	.227	.270	.423	.381
	Q5	.441	.403	.490	.364	.308	.428	.501
	Q6	.465	.359	.499	.484	.246	.358	.395

	Q7	.503	.356	.500	.516	.273	.404	.431
	Q8	1.000	.409	.664	.457	.368	.398	.554
	Q9	.409	1.000	.470	.351	.376	.414	.486
	Q10	.664	.470	1.000	.509	.315	.405	.491
	Q11	.457	.351	.509	1.000	.316	.390	.417
	Q12	.368	.376	.315	.316	1.000	.346	.477
	Q13	.398	.414	.405	.390	.346	1.000	.501
	Q14	.554	.486	.491	.417	.477	.501	1.000

a. Determinant = .002

Descriptive Statistics

	N	Skewness		Kurtosis	
	Statistic	Statistic	Std. Error	Statistic	Std. Error
Q1	2627	-.289	.048	-.170	.095
Q2	2627	-.117	.048	-.322	.095
Q3	2627	-.141	.048	-.158	.095
Q4	2627	-.310	.048	-.306	.095
Q5	2627	-.227	.048	-.157	.095
Q6	2627	-.139	.048	.119	.095
Q7	2627	-.210	.048	-.009	.095
Q8	2627	-.219	.048	-.255	.095
Q9	2627	-.291	.048	-.345	.095
Q10	2627	.002	.048	-.266	.095
Q11	2627	-.504	.048	-.058	.095
Q12	2627	-.676	.048	-.099	.095
Q13	2627	-.532	.048	-.179	.095
Q14	2627	-.346	.048	-.231	.095
Valid N (listwise)	2627				

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.937
Bartlett's Test of Sphericity	Approx. Chi-Square	16674.520
	df	91
	Sig.	<.001

Communalities

	Initial	Extraction
Q1	.479	.494
Q2	.528	.555
Q3	.444	.453
Q4	.341	.470
Q5	.417	.446
Q6	.476	.611
Q7	.504	.612
Q8	.561	.659
Q9	.419	.450
Q10	.574	.607
Q11	.398	.425
Q12	.283	.301
Q13	.384	.424
Q14	.583	.639

Extraction Method: Maximum Likelihood.

Total Variance Explained

Factor	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	6.581	47.004	47.004	6.102	43.583	43.583
2	1.134	8.098	55.102	.644	4.604	48.186
3	.865	6.175	61.277	.400	2.859	51.045
4	.747	5.337	66.614			
5	.652	4.658	71.272			
6	.596	4.257	75.529			
7	.573	4.092	79.622			
8	.550	3.929	83.550			
9	.485	3.467	87.017			
10	.453	3.235	90.252			
11	.373	2.661	92.913			
12	.365	2.607	95.520			
13	.339	2.418	97.938			
14	.289	2.062	100.000			

Factor Matrix[a]

	Factor		
	1	2	3
Q1	.692		
Q2	.734		
Q3	.657		
Q4	.483	.367	.319
Q5	.655		
Q6	.673	-.318	
Q7	.693		
Q8	.755		
Q9	.608		
Q10	.751		
Q11	.624		
Q12	.497		
Q13	.602		
Q14	.744		

Extraction Method: Maximum Likelihood.[a]

a. 3 factors extracted. 5 iterations required.

Goodness-of-fit Test

Chi-Square	df	Sig.
706.747	52	.000

Pattern Matrix[a]

	Factor		
	1	2	3
Q1	.532		
Q2	.566		
Q3	.536		
Q4			.771
Q5			.318
Q6		.786	
Q7		.741	
Q8	.844		
Q9			.513
Q10	.646		
Q11		.456	
Q12	.402		
Q13			.476
Q14	.621		.351

Extraction Method: Maximum Likelihood.
Rotation Method: Promax with Kaiser Normalization.
a

a. Rotation converged in 7 iterations.

Structure Matrix

	Factor		
	1	2	3
Q1	.690	.532	.569
Q2	.729	.634	.512
Q3	.661	.491	.546
Q4	.387	.342	.677
Q5	.597	.563	.590
Q6	.558	.781	.449
Q7	.589	.781	.463
Q8	.798	.622	.453
Q9	.554	.451	.654
Q10	.757	.664	.473
Q11	.570	.631	.422
Q12	.502	.315	.490
Q13	.539	.475	.628
Q14	.754	.494	.695

Extraction Method: Maximum Likelihood.
Rotation Method: Promax with Kaiser Normalization.

Factor Correlation Matrix

Factor	1	2	3
1	1.000	.724	.685
2	.724	1.000	.548
3	.685	.548	1.000

Extraction Method: Maximum Likelihood.
Rotation Method: Promax with Kaiser Normalization.

Reliability

McDonald's Omega for Factor 1

McDonald's Omega	N of Items
.869	7

McDonald's Omega for Factor 2

McDonald's Omega	N of Items
.780	3

McDonald's Omega for Factor 3

McDonald's Omega	N of Items
.748	4

Cronbach's Alpha for Factor 1

Cronbach's Alpha	N of Items
.868	7

Cronbach's Alpha for Factor 2

Cronbach's Alpha	N of Items
.778	3

Cronbach's Alpha for Factor 3

Cronbach's Alpha	N of Items
.747	4

Universiti Tunku Abdul Rahman			
Form Title : Supervisor's Comments on Originality Report Generated by Turnitin for Submission of Final Year Project Report (for Undergraduate Programmes)			
Form Number: FM-IAD-005	Rev No.: 0	Effective Date: 01/10/2013	Page No.: 1 of 1

FACULTY OF SCIENCE

Full Name(s) of Candidate(s)	Chua Wen Xin
ID Number(s)	21ADB02307
Programme / Course	Bachelor of Science (Hons) Statistical Computing and Operation Research
Title of Final Year Project	Enhancing UTAR Freshman Mental Health Screening: Integrating Power Automate to Improve Counselling Support

Similarity	**Supervisor's Comments** **(Compulsory if parameters of originality exceeds the limits approved by UTAR)**
Overall similarity index: 10 % **Similarity by source** Internet Sources: 9% Publications: 5% Student Papers: 5%	
Number of individual sources listed of more than 3% similarity: 0	

Parameters of originality required and limits approved by UTAR are as follows:

(i) **Overall similarity index is 20% and below , and**
(ii) **Matching of individual sources listed must be less than 3% each , and**
(iii) **Matching texts in continuous block must not exceed 8 words**

Note: Parameters (i) – (ii) shall exclude quotes, bibliography and text matches which are less than 8 words.

Note Supervisor/Candidate(s) is/are required to provide softcopy of full set of the originality report to Faculty/Institute

Based on the above results, I hereby declare that I am satisfied with the originality of the Final Year Project Report submitted by my student(s) as named above.

Signature of Supervisor

Name: Dr Lim Huai Tein

Date: 30 April 2025

FYP_Turnitin Check.docx

ORIGINALITY REPORT

10%	9%	5%	5%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	eprints.soton.ac.uk Internet Source	1%
2	essay.utwente.nl Internet Source	1%
3	Eunice Magalhães, Shalhevet Attar-Schwartz. "Social Support of Young People in and after Residential Care: Is Someone There for You? - An International Perspective", Routledge, 2025 Publication	1%
4	eventscribe.com Internet Source	1%
5	www.pubfacts.com Internet Source	1%
6	www.frontiersin.org Internet Source	1%
7	www.gov.im Internet Source	1%
8	unclekentang.com	

	Internet Source	1 %
9	docta.ucm.es Internet Source	1 %
10	repository.nwu.ac.za Internet Source	1 %
11	umpir.ump.edu.my Internet Source	1 %
12	www.mdpi.com Internet Source	1 %

Exclude quotes	On	Exclude matches	< 8 words
Exclude bibliography	On		