A STUDY OF ASSESSOR TRAINING STRATEGIES AND ITS EFFECT ON ACCURACY, DISCRIMINANT VALIDITY AND INTERRATER RELIABILITY IN UMP

NUR ASHIDA AYU BINTI ABDUL RANI

Thesis submitted in fulfilment of the requirements for the award of the degree of Bachelor of Project Management with (Hons)

Faculty of Industrial Management
UNIVERSITI MALAYSIA PAHANG

DECEMBER 2015

SUPERVISOR'S DECLARATION

I hereby declare that I have checked this project and in my opinion, this project is adequate in terms of scope and quality for the award of the Degree of Bachelor of Project Management with Honours.

Signature :

Name of Supervisor : MOHD HANAFIAH BIN AHMAD

Position : LECTURER

Date :

STUDENT'S DECLARATION

I hereby declare that the work in this project is my own except for quotations and summaries which have been duly acknowledge. The project has not been accepted for any degree and is not concurrently submitted for award of other degree.

Signature :

Name : NUR ASHIDA AYU BINTI ABDUL RANI

ID No : PB12010

Date :

ACKNOWLEDGEMENT

Assalamualaikum W.B.T. I am pleased to point out my sincere gratitude to my supervisor Mohd Hanafiah Bin Ahmad for his guidance, kind of perusal and consideration, and constant support in making this research possible and successfully completed this proposal. I really appreciated his guidance from the initial level until final and give me a chance to develop an understanding of this research thoroughly. Without his advice, comments, and assistance it would be a lot tougher for completion.

Last but not least, I would like to convey special thanks to that person who had contributed to my final year project, whether directly and indirectly. I would like to express the deepest appreciation to my parents for their love and sacrifice throughout my life. I am also grateful to my husband Mohd Hafiz bin Norazman, I cannot find the appropriate words that could properly describe my appreciation for his devotion, faith, and support in my ability to attain my goals. His prayers and advice promote me to stand on my research with flying colors.

Special thanks should be given to my committee members. I would like to acknowledge their comments and suggestions, which was crucial for the successful completion of this study.

ABSTRACT

The purpose of this study was to explore whether the implementation assessor training is suitable to be conducted in University Malaysia Pahang (UMP). Specifically, this study examines the utility of two assessor training strategies; frame of reference training (FOR) and rater error training (RET) in staff selection process. Results are obtained by opinion from 84 assessors in each department in UMP. The two approaches are compared each other and analyzed based on 3 dependent variables. A general framework is presented in term of two training strategies in the effects of accuracy, discriminant validity, and interrater reliability. Several hypothesis was performed and proved that both training are accepted by the assessors in UMP and the accuracy, discriminant validity, and interrater reliability had influenced both training. However, those assessors feels that they prefer to have FOR Training in the future as it is seen as a structured and coordinated training compared to RET.

Tujuan kajian ini adalah untuk menilai sama ada pelaksanaan latihan penilai sesuai yang akan dijalankan di Universiti Malaysia Pahang (UMP). Secara khusus, kajian ini meneliti utiliti dua strategi latihan penilai, rangka latihan rujukan (FOR) dan latihan ralat penilaian (RET) dalam pemilihan pekerja. Keputusan diperolehi melalui pendapat dari 84 penilai dalam setiap jabatan di UMP. Kedua-dua pendekatan dibandingkan antara satu sama lain dan dianalisis berdasarkan 3 pemboleh ubah bersandar. Satu rangka kerja umum dikemukakan dalam tempoh dua strategi latihan dalam kesan ketepatan, kesahihan diskriminan, dan kebolehpercayaan antara penilai. Beberapa hipotesis telah dijalankan dan membuktikan bahawa kedua-dua latihan yang diterima oleh penilai di UMP dan ketepatan, kesahihan diskriminan, dan kebolehpercayaan antara penilai telah mempengaruhi kedua-dua latihan. Walau bagaimanapun, penilai merasakan bahawa mereka lebih suka latihan FOR di masa akan datang kerana ia dilihat sebagai latihan yang berstruktur dan tersusun berbanding RET.

TABLE OF CONTENTS

CHAPTER 2 LITERATURE REVIEW

CHAPTER 3 RESEARCH METHODOLOGY

CHAPTER 4 DATA ANALYSIS

CHAPTER 5 CONCLUSION AND RECOMMENDATION

LIST OF TABLE

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 INTRODUCTION

Understanding appropriate training in an organization is critical to design high-quality outcomes. A topic that are discussed is a study on assessor training strategies and their effects on accuracy, discriminant validity, and interrater reliability. This research explores and compares the assessor training strategies in University Malaysia Pahang (UMP). The purpose of this chapter is to outline the background and rationale of the study, describe its significance, and provide an overview of the objective and research questions.

An introduction of what is the overall content of this research are shown in Chapter 1. It includes an introduction, the background of study, problem statement, research objectives, research questions, the scope of study, expected outcomes, significance of the study, theoretical framework, and lastly operational definition.

The second chapter outlines the most important part that shall address the answer to research questions. That is literature review which combining of an introduction of overall Human Resource Management (HRM) process, understanding where employee selection is needed, what was meaning of the assessor, the assessor training strategies, accuracy, interrater reliability, discriminant validity, research gap, and hypothesized model. This chapter discussed the independent and dependent variables. Chapter 3 in this study comprising of the methodologies which explained more about research method that are used in this study. Fourth chapter comprising all findings and research analysis. And the last chapter illustrates how the researcher would like to conclude and outlined some recommendations to next research related to this topic.

1.2 BACKGROUND OF STUDY

Assessor training is too frequently mentioned as Rater Training Programs in performance management context. In general, this research explores for assessor training strategies practices in the selection process of UMP. This research was to outline which assessor training strategies needed to train selected employees as trained assessors. In human resource, selection process is located in the assessment center. The assessment center methods involved several activities using multiple evaluation techniques (Byham, 1992). The effectiveness of assessor training was measured by its accuracy, discriminant validity and interrater reliability (Lievens, 2001). The major point of this research is, conducting own assessor training is crucial in order to have effective rating accuracy, discriminant validity, and interrater reliability. It may influence largely to the ambiance of selection process (Maimunah, 2011).

As early as the year 1948, the potential measure of assessor training was recognized Bittner (1948). Nevertheless, previous reviews of this training results of various assessor training strategies. According to Spool (1978), it is concluded that assessor training is generally effective. Hence, it shows the need of implementation of assessor training to ensure the selection process are effective for the organization.

The term assessor training indicates the exercise to a person who conducting an assessment. Assessor training is too frequently being mentioned by the professional and managerial level but too little understanding for its importance (Lievens, 2001). These training are not only provided the assessors with needed skills and tools to give a good result but also helps increase satisfaction with the results of employee selection. Assessor training strategies offers people the chance to learn the knowledge and skills required to conduct outcomes-based assessments in their fields of expertise. Completion of this assessor training also signifies that the assessor is able to complete assessments without bias, lending the assessor credibility and legitimacy as a fair practitioner.

A general framework for the study of assessor training strategies is represented in term of two common assessor training strategies (frame-of-reference training and rater error training) and three dependent variables (accuracy, discriminant validity and, interrater reliability). Frame-of-reference training helps improve rating accuracy by familiarizing the raters with various performance dimensions to be assessed. Its goal is to give assessor skills so that they can provide accurate ratings. Meanwhile, the rater error include the most common error that assessors usually did. It can be in two form; intentionally and unintentionally. Thus, rater error training should be conducted to reduce all those errors. This study compares the effects of frame-of-reference training and rater error training towards the three facets.

There a significant difference between recruitment and selection as according to (Maimunah, 2011). Recruitment are occur at the early stage, but selection occur when potential candidates are chosen. Thus, the selection also becomes the most important stage to hire the right person for organization. It is related directly to who will assess the potential candidates. It is the assessor. Despite, this research was to outline the importance of training to the assessors in UMP. To identify effects of implementation of the strategies, a study on effectiveness of assessor training strategies towards the accuracy, inter-reliability, and discriminant validity will be conducted by the researcher.

This comparative research is importance in both practical and conceptual as Lievens (2001) said. Conceptual view stressed on the importance to understand which training serves as the best foot of assessor training while it is pivotal to identify which training strategy contributed to more accurate, valid and reliable ratings during selection process.

1.3 PROBLEM STATEMENT

An effective assessor training strategies should be implement in the field of selection process as a useful approach to get accurate ratings, good interrater reliability and discriminant validity. As Smith (1986) said, the presentation of training strategies and the content of training would influence the effectiveness of assessor training. So, the best suitable strategy should be identified. Smith (1986) also identified a more elaborate framework for the use in assessor training as well as its evaluation division.

In part of that, Lievens (2001) found that only frame-of-reference training has been studied with student populations until no proven studies show the positive results to managerial levels population. He explained that assessor training are influenced by the three dependent variables. However, literature search shows that there has been very limited research carried out in context of staff selection to support this suggestions. Therefore, this study aims to respond to this research gap and contribute to the overall selection process of HRM in UMP.

Instead of giving briefing and exercise to those assessors, effective training should be implemented to ensure the competencies of each assessor. It is very important to know which training suits with situations and style of management in the organization. Hence, the existence of proper training for assessors involved in the assessment center to overcome selection problem is seen as crucial as there still remain uncertain whether frame-of-reference logic was effective or not in UMP. On top of that, according to personal conversations with Human Resource Manager of the university on 8 April 2015, those training are still not yet explored officially in this university.

According to Lievens (2001), assessor in a minimal training record significantly less behavioral description, lower values of rating accuracy, interrater reliability and dimension differentiation. Some organization confused to choose which strategy suits effectively to their assessor; whether to choose only one of the strategies or combining two or more of strategies to be used in an assessment center. It is a big issue that may affect to the future outcomes of the selection process. Thus, the most appropriate strategies are of an intense necessity to develop higher competency of an assessor.

1.4 RESEARCH OBJECTIVES

Below are the objectives that are essential to be achieved for the purpose of completing the ultimate intention of this research which is as follow:

RO1: To evaluate the effects of assessor training strategies towards accuracy, inter-reliability and discriminant validity.

RO2: To compare the effectiveness of two assessor training strategies in University Malaysia Pahang.

1.5 RESEARCH QUESTIONS

The researcher shall answer the following research questions:

RQ1: How do the effects of assessor training strategies towards accuracy, inter-reliability and discriminant validity?

RQ2: Which of the assessor training strategies is successfully implemented?

1.6 SCOPE OF STUDY

The scopes cover for an overview of the suitable training strategies to the assessors and which training are effective in UMP. This research also identified the effects of the training towards accuracy, interrater reliability and discriminant validity of the assessors itself after attended the training. This study focusing on the importance of assessor training to the employee selection of UMP. The significance of each variable are shown in the literature review.

1.7 EXPECTED OUTCOMES

The expected outcomes from this research is the results of the implementation and which assessor training more suitable in UMP. This study was made for obtaining data to support the implementation of assessor training to the assessors in UMP since no past research show prove that assessors in UMP are provided with effective training.

1.8 SIGNIFICANCE OF STUDY

This study will be a significant effort in promoting two specific assessor training strategies to the assessment center of UMP. The expected results also will be beneficial to assessors and the candidates in executing selection of staff when UMP employ the assessor training strategies in their assessment center. By understanding the prominence of assessor training in an organization and the effects of ignoring those training, those assessors be assured of a competitive advantage. Moreover, this study is profitable to UMP's human resource practitioner in developing the assessor training strategies. It will also serve as a small bone in future analysis towards the management of human resource in UMP. In order to design the strategies and business activities in an organization, the internal and external evaluation are both critical. On top of that, this study may help in the future designation of strategies in human resource department of UMP. This research gives a chance to UMP or other organization to study and understand more on the importance of assessor training in an assessment center.

1.9 THEORETICAL FRAMEWORK

A clearly defined independent variable and dependent variable are used in this research. The dependent variable is the variable of primary interest to the researcher. An independent variable is one that influences the dependent variable in either a positive or negative way. Based on hypotheses on Chapter 2, the following framework is developed, further discussion about these hypothesis are discussed in chapter 2. The model are illustrated according to the formation of framework which are Initial Framework, Finalized Framework and Hypothesized Model.

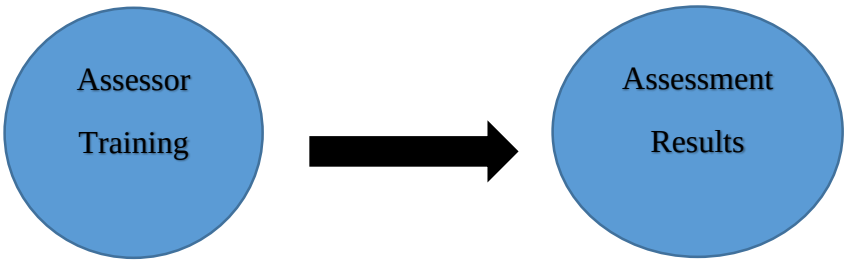

Figure 1.1 Initial Theoretical Framework

The initial framework shows how the independent variable will influence the dependent variable. The three facets (accuracy, discriminant validity, and interrater reliability,) are tested whether they are influenced the two assessor training strategies if performed in UMP. Figure 1.2 show how the elements contained in the Initial Theoretical Framework.

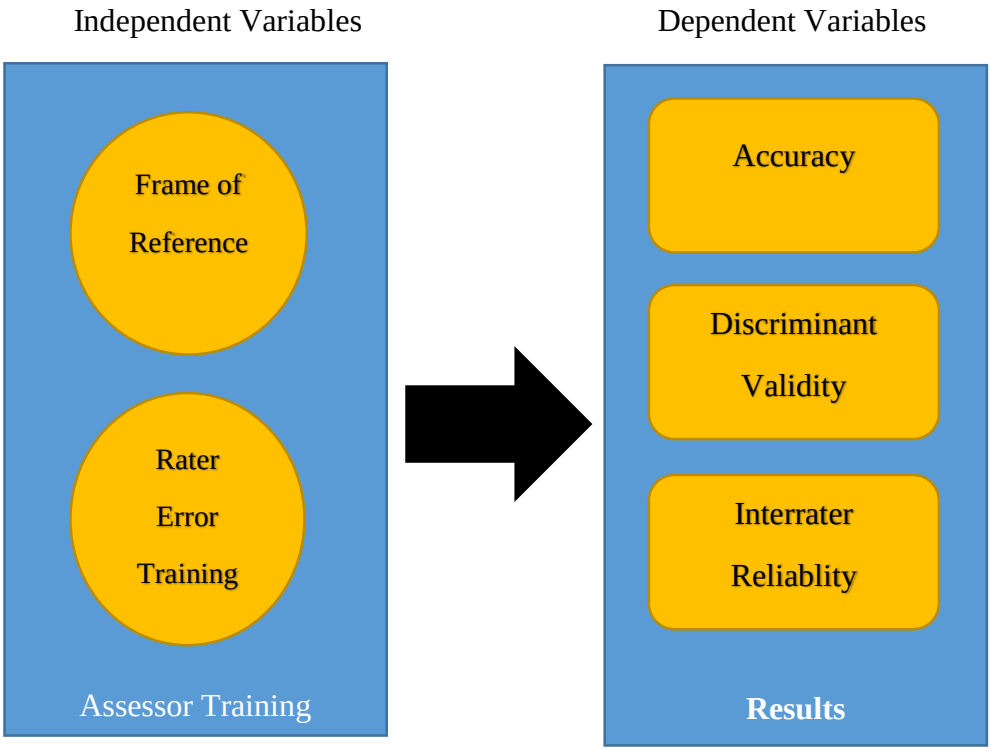

Figure 1.2 Finalized Theoretical Framework

Based on the above Finalized Theoretical Framework, the following hypotheses are developed. Therefore, this study are testing for the hypotheses to obtain the desired results. The hypotheses are illustrated as followed.

H1: Frame of reference training will significantly influence the assessors' ratings accuracy

H2: Frame of reference training will significantly influence the assessors' discriminant validity

H3: Frame of reference training will significantly influence the assessors' interrater reliability

H4: Rater error training will significantly influence the assessors' ratings accuracy

H5: Rater error training will significantly influence the assessors' discriminant validity

H6: Rater error training will significantly influence the assessors' interrater Reliability

H7: Frame of reference training is better than rater error training.

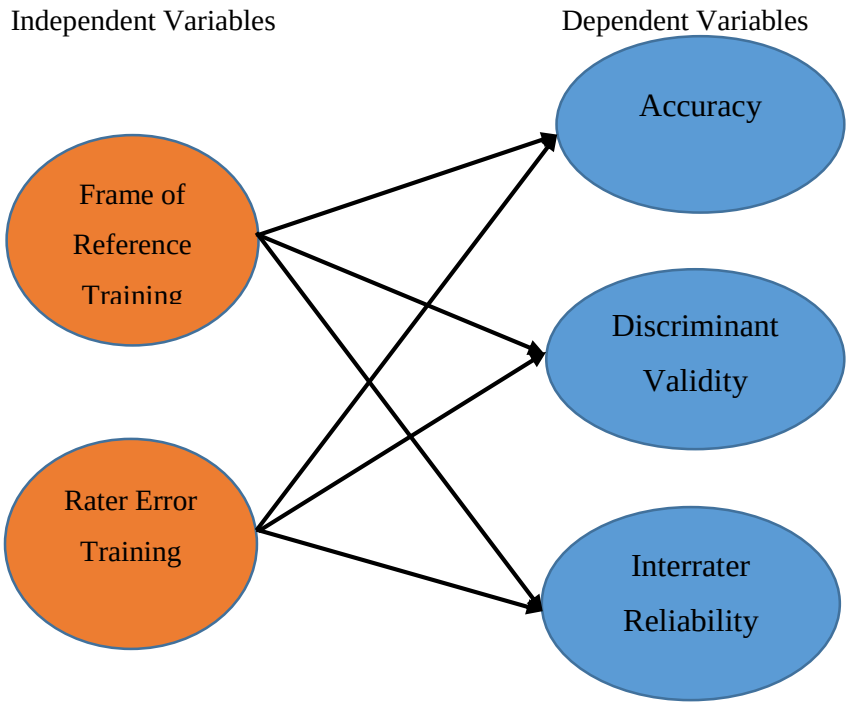

Figure 1.3 Hypothesized Theoretical Model

1.10 OPERATIONAL DEFINITION

i.	Assessor	- Also known as the rater. The person who make an assessment. In this research, an assessor defined as the person who gives ratings to the candidates.
ii.	Accuracy	- The correctness of a scale of ratings in every assessment.
iii.	Reliability	- the ability of a person to come out with assessment result that trustworthy with standard ratings.
iv.	Interrater reliability	- The ability of an assessor to produce a dependable and almost equal result with other assessors.
v.	Validity	- The ability of a person to distinguish result that contain genuineness in its ratings.
vi.	HRM	- Human Resource Management
vii.	UMP	- University Malaysia Pahang, located in Gambang, Pahang, Malaysia.

1.10 SUMMARY

Chapter one introduced to the background of assessor training in an assessment center. The importance of assessor training strategies was highlighted in the problem statement. Then the researcher discussed further of the research objectives, research questions, scope of study, expected outcomes, significance of the study, theoretical framework and elaborated some important operational definition used in this research.

CHAPTER 2

LITERATURE REVIEW

2.1 INTRODUCTION

This chapter tackles about literature review in order to illustrate deep information about the research. This chapter is coordinated as follows. The literature review will give an account about understanding of overall Human Resource Management (HRM) process and several selection methods that can be used by organizations. It is shown the position of where the selection is needed through whole human resource management process. Next topic included some literature about the assessor training in an assessment center and about two training strategies that are focused by the researcher. Then it carries out three facets of assessor training strategies for better understanding of its importance. The expected result that are answered from this research may address the research questions are observed in the following chapters.

The first training is focusing on frame-of-reference training and the second is about rater error training. These two training strategies are expanded more in upcoming pages. This chapter last with the importance of training corresponding to three independent variables: accuracy, discriminant validity, and inter-reliability. Some interesting hypotheses are forecasted through these variables explained on the other pages. According to Lievens (2001), any assessor with limited training will indicates a lower values of rating accuracy, discriminant validity, and interrater reliability. On top of that statement, this research was developed to explore the implementation of assessor training in UMP.

2.1 OVERALL HUMAN RESOURCE MANAGEMENT (HRM) PROCESSES

Organizations have never-ending of human resource management tasks including staff selection. Being the most important part of it are motivated the organizations to identifying of its significance to be delivered in the best ways. Choosing wrong choice of selection may be costly to the organizations. According to a study of Egypt recruitment and selection practices by El-Kot and Leat (2008), different selection methods are used for different types of jobs. It shows the clear importance to have proper selection methods in an organization. An overview of the whole human resource management processes is shown in the following subtopic.

According to Hashim (2010), the British colonization experience of the nineteenth century had most influenced Malaysian HRM practices. Added by Maimunah (2011), the rapid growth of computer and telecommunications technology in Malaysia will impact in major on human resource management. As according to Maimunah (2011), overall HRM process involved from recruitment and selection until industrial relation which can be summarized as the Figure 2.1 below.

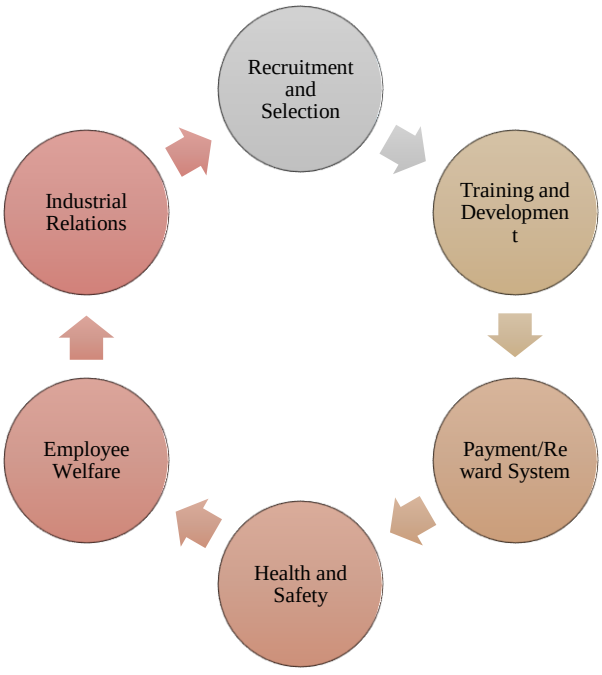

Figure 2.1 Summarized Overall HRM Process

Source: (Maimunah, 2011)

The previous diagram shows how important of selection of employee which appear at the initial stage of HRM in an organization. It is the starting point where the journey of an employee begins. Recruitment and selection is all about predicting future performance. But how can the employee make the right decision to choose the right person for their organization? This issue was answered in this study where researcher focusing on training to be provided to the assessors during selection processes and are measured in terms of their accuracy, discriminant validity, and interrater reliability.

2.1.1 Why selection are important?

As mentioned above, it was very clear picture shown that selection is one of the critical steps in HR process. As the responsibility as HR practitioner, staff selection were one of the important desire to avoid uncertainty and errors (El-Kot and Leat, 2008). More or less, competencies also are one of the reasons to achieve good staff selection. Competencies are used widely and increasingly popular as a fundamental of organization and its resources Knight and Zaklad (2001). Competencies are the physical behavior and represent the ability to perform a certain role at the specific level. It is how they come out a peculiar element of the role.

In this study, competencies are relatively related with quality of employees chosen by the organizations. Lower quality of employee performed as the assessor itself performs lower competencies in the assessment center. While Hagan et al. (2006) mentioned that examining in further research on how different definitions of ‘competency’ affect the design and competency measurement will be more interesting, further understanding for the importance of competencies in an organization should find out.

As past researcher, Fratričová and Rudy (2015) had matched the strategic human resource management with the competitive advantage of an organization, a relative connection formed between the importance of staff selections and the performance of an organization. Competitive advantage is the utmost significance to any companies in order to maintain its position. Moreover, Yahya and Fee-Yean (2015) point out that career commitment are one of the indicator to improve human resource management in an organization. Therefore, the initial selection process stands most importance part to hire the best people for the best position.

However, any organization cannot survive without a proper planning in it. Thus, (Maimunah, 2011) had proposed systematic steps involved in the Recruitment and Selection Processes as shown below.

Figure 2.2 Steps involved in Recruitment and Selection Process

(Maimunah, 2011)

Human resource planning is a continuous process of linking the requirements for manpower with the financial and operational plans of the organization in a particular time period (Maimunah, 2011). As the above steps included in the systematic human resource planning, it is crucial for each organization to focus of the selection process. The highlighted step was parallel to the objective of this research. It shows how selection was important in HRM process. Thus, in order to have a successful selection process, assessor who assessing the applicant are pivotal to be trained effectively by the organization itself. If not, they may do some mistakes during selection. After all, its organization have to take risks appointing the wrong applicants in their company and will impact them in term of cost. Several methods of selection are discussed in the following topic.

2.1.2 Human Resource Staff Selection Methods

Human resource management is one major field that include everything about the humans including staff selection. The potential to choose wrong methods for staff selection would contribute huge downfall of any organization. While this study focusing only the external selection, several selection methods are highlighted to show a clear picture of this literature review. External selection refers to an organization that hiring from the outside. The aim of selection practices is to identify suitable applicants in a short time, legally, and cost-efficiently (Dockalikova and Kashi, 2013). The list below show several selection methods that usually practiced by organizations:

1. Selection Tests
2. Interview
3. Assessment center

2.1.2.1 Selection Tests

Finding the best employees for the organizations are the most challenging to human resource department and specialists. However, there are several methods that can be used in selection. Selection test is one of the important tools to measure how far the candidates performance and background. Two crucial concepts mush be discussed before looking at the various tests available (Maimunah, 2011). The concepts stressed on the validity and reliable of the tests itself to be run. Choosing the most suitable selection tests can be powerful if you have a large amount of candidates to be filtered for the next stage of selection. Nevertheless, most selections tests also can be time-consuming and costly to administer. The most popular types of tests are:

1. Performance tests
2. Aptitude tests
3. Personality tests
4. Intelligence tests
5. Medical tests

2.1.2.2 Interview

An interview is one of the methods used for the selection process of staff in any organization. Even it is summarized by Anderson (1992), many studies have reported the limitations of interviews, Barclay (2001) had outlined several advantages of managing interviews for employment new staff. Other than that, Barclay (2001) also had conducted interviews mainly in terms of reliability and validity. It is associated with this research that focused on the importance of training for the interviewer in terms of accuracy, reliability and validity.

Training is an important requirement as one of the skills needed for an interviewer (Barclay, 2001). Besides of evaluating candidates' competencies during the interview, their flexibility and self-response would be identified. Although interviews seem to take longer time period for a session, it still be the main choices of organizations in the selection of staffs. Interview commonly defined in two type that is structured and unstructured interview. While structured interview is more familiar with situational and behavioural.

2.1.2.2 Assessment center

Assessment center is very specific in its objectives to evaluate the personnel competencies of candidates on order to make a decision for the best person for the role. An assessment center actually is techniques that bring together behavioural simulation data to report results for selection or promotion, determine training needs, or may catalyse the development of employee, (Jones, 2007). An assessment center not limited only for selection an employee but also helpful in hiring and determining the training needed in the organization.

Assessment center functions as a formal assessment approach that are now used by numerous organization in the world. Recently, researchers have shown an increased interest in the application of assessment center in the most organization. It included a standardized evaluation of behaviour based on multiple evaluations. In common assessment center method, a group of six assesses participate in a series of exercise that related to their job or job level are observed by three or more line managers determines whether they have the skills and abilities necessary to perform that job, (Byham, 1992).

There is a large volume of published studies describing the role of an assessment center. Hence, in order to make effective assessment, proper training should be provided to the assessors in order to create accurate, interrater reliability and discriminant validity when determining which person suits the job (Lievens, 2001).

Usually, throughout the course of the assessment, the assessors will monitor the performance of a person, which can last from a few hours until several days. The wide variety purpose of an assessment center is including selection, screening group of candidates all at one time, or may see as a realistic job preview. Perhaps one important advantage of the assessment center method where it related to future performance, not to current job scope, (Lievens, 2001). Several examples of activities usually practiced in an assessment center are role playing, in-tray assessment, psychometric assessment, and presentations.

2.2 WHO ARE THE ASSESSORS?

In the past 50 years, the present of a team of assessors is needed to assess and evaluation those competencies (Schlebusch and Roodt, 2008). Lievens (2001) said that trained assessors are really important to produce high quality of ratings to the candidates. Therefore, further study is reliable to explore more about the importance of an assessor in an organization. All the observation and ratings by an assessor are critical because they are the physical medium for screening the behavior of candidates into information that are they used to make a decision. Hence, the quality of this output counts on heavily on the assessors' accuracy, reliability and validity of ratings. Gaugler et al. (1987) categorized the moderators of assessment center with several point as shown in Figure 2.3 below.

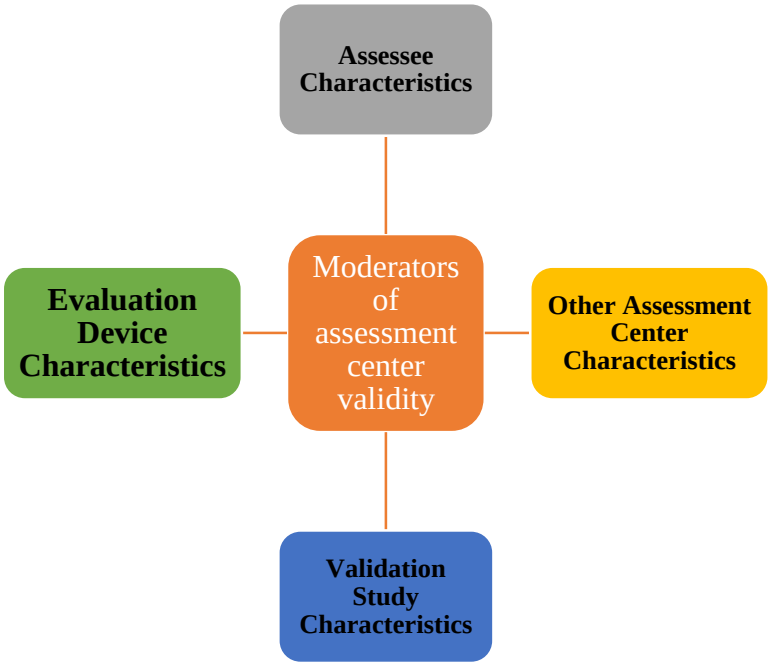

Figure 2.3 The moderators of assessment center validity

Source: (Gaugler et al., 1987)

The categories in Figure 2.3 clearly shown that the need of accurate, reliable and validity on an assessor which lies on Assessor Characteristics is importantly needed. Schlebusch and Roodt (2008) clearly defined that assessors are the groups or individuals who have largest impact directly towards the whole process of human resource management, Wirz et al. (2013) find that there is no significant main effect of the assessor background, somehow it clearly states the strongest effect on rating accuracy is by attending the assessor training.

According to Macan et al. (2011), there the need for implement a true experimental design to assists in choosing an assessor. More or less, it is clear that one crucial element in assessment center is that assessors and their training although many challenges contribute to various debates. The author shall review and synthesize on the accuracy, interrater reliability, and discriminant validity.

2.2.1 Performance Assessment

Performance appraisal generates organizations to administer their overall performance in order to gain their own vision and mission. Despite, they need to bring performance in line with the achievement of corporate strategies and goals. Assessment is always done by the objective of identifying the competency gap of the incumbent and ways to address it via training or job fit. Performance appraisals defined as a system that involves evaluations of employees once in a year without any continuous effort in providing feedback and coaching (Aguinis, 2013). However, performance assessment in this study is the evaluation that focus on assessing the level of achievement with correspondent to the assessor training provided. Performance assessment definition may differ depending on the purpose of it. As Borman (1978) suggest, the performance appraisal should be reliable and accurate if the raters follow effectively with his proposed three steps.

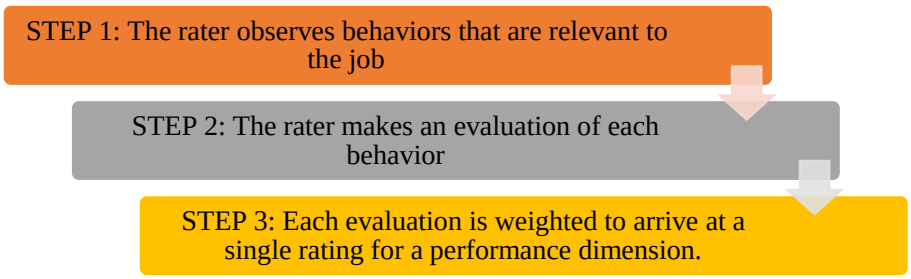

Figure 2.4 Bormann's Model of Performance Appraisal

Source: (Borman, 1978)

2.3 ASSESSOR TRAINING STRATEGIES

The existence of effective assessor training strategies may come out with more powerful results of any organization or sector. The ingredients of this literature review discussed further about two possible selected strategies that are Frame of Reference training and Rater Error Training. These two training are chosen based on the importance towards an organization. Although Lanik and Mitchell Gibbons (2011) said that scaling behaviours and building a common frame of reference are another common part of assessor training, no further research found that it is suitable for all industry.

The objective of training strategies is achieving better performance of an assessor. An assessor must have the relevant skills and experience. Through effective training, the skills for an assessor will be achieved successfully with full cooperation and support of the organization. As Schlebusch and Roodt (2008) said, some of the required skills in observing and evaluating candidates are needed on order to get an accurate rating.

The definition of performance do not include the assessor's behaviour results but only include the assessor's behaviours themselves. On top of that, according to Schlebusch and Roodt (2008) the major goal of training the assessors is to create the personal abilities that allow them to effectively and accurately rate the participants' behaviour. Nowadays, various options of assessor training are provided by authorized organizations in worldwide. However, errors in choosing suitable training strategies may affect the result of an assessor after the training. The main point is, it is crucial to select the most effective strategy and it is depending on the position of vacancies itself.

This study contrasted the effectiveness of two proposed assessor training strategies and its competency with accuracy, interrater reliability and discriminant validity. Each training program needs to be identified it's pro and cons. For example, Smith (1986) point out that the approach used in the presentation of training materials would affect the effectiveness of each training program. As the purposes of assessor training are beneficial, Borman (1978) added that training would increase the discriminant validity on each behaviour of candidates. Another reason why this study will be conducted is following suggestion by Smith (1986) that future studies should identify problems in the assessors' skills.

2.3.1 Frame of Reference Training

Generally, frame-of-reference training present that the assessor be trained to sustain specific standards of performance among the assessors. It demonstrates where the presence of a reference point is compulsory so that there would be an equivalent between the candidates' true scores and the raters' scores. The overall goals is to provide assessors with skills which they can perform accurate ratings of each selected candidates by developing common frame-of-reference (Aguinis, 2013). Focusing on develop mutual understanding among the assessors as defined from Lievens (2002) which is most agreed definition as the mutual understanding

goal is to assists all assessors with same dimension model so they can use it as a landmark when doing an assessment.

By means, frame-of-reference training usually conducted by a professional trainer. This type of training helps assessors use the same standards so that particular assessors will not out of coverage in giving their ratings. There are several steps in conducting frame-of-reference training:

1. The trainers telling the assessors that they will be evaluating four employees in four different dimensions.
2. Appraisal form are given to the assessors and they will read it carefully.
3. Trainer teaches and discuss important dimension to the job and various employees behaviours represent each of the dimension. This step is important to keep baseline theory among all assessors and provide standardize rating for evaluating performance.
4. A videotape of a practice are shown to the assessors and are asked to evaluate the employee's performance by using the provided standard frame-of-reference.
5. Discuss of ratings provided by all assessors involved.
6. Feedback provided by the trainer and explain why the employee should receive the target score he or she created.

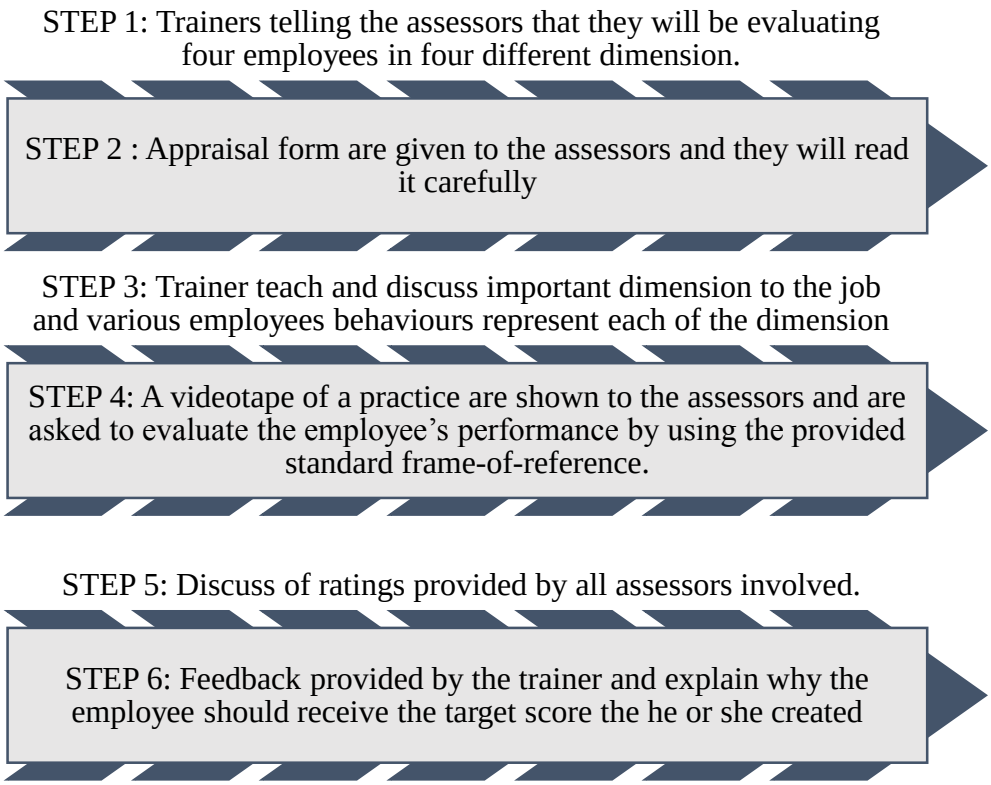

Figure 2.5 Frame of Reference Training Steps

Source: Thayer (2004)

From the previous figure, it could be seen thoroughly that frame-of-reference training would produce the best result, but need longer time and large input and effort to develop. Standard of performance is significant factors to enhance the ability of an assessor to provide effective performance evaluations. The other study pointed out that frame-of-reference training is more correlated to an assessment center process. Smith (1986) stated that combination of performance standard approaches and performance dimensions are the factors of largest improvements in accuracy. As Noonan and Sulsky (2001) researched that highest rating accuracy obtained from combine two training strategies together; frame-of-reference and behavioural training, it is found that frame-of-reference training also affected by interrater reliability factor, thus it is the focus of this study.

2.3.2 Rater Error Training

Rater error training mostly used in performance appraisals because of the most error comes from assessors is unintentional. However, the usage can be narrow down to one of assessor training strategies. The aim of rater error training is to make raters clearly discovered of what rating error that they may produce and also let them come out with a variety of strategies to overcome those errors. As it is clearly depicted in the prior definition given, an effective rater should improve the accuracy of rating by reducing common "rater biases".

There are pretty much of errors can be made by assessor whether it is intentionally or unintentionally errors. It is largely an outcome of the impression that traditional forms of rater training, such as rater error training, have not been successful at increasing rater accuracy even though it minimizes scaling rater errors. Thus, training that will show videotaped vignettes which are designed to reduce rating errors are helpful (Aguinis, 2013). However, one thing needs to remember. A simple awareness of the errors does not conclude that errors will not be made.

This training can be conducted by providing a videotaped vignettes and ask assessors to provide their ratings based on situation shown in the video by filling out the performance appraisal forms. Then, a final comparison will be showed to all assessors between the provided rating by assessors and the correct ratings. Discussion and explanation by the trainer will then provide after that. According to (Aguinis, 2013), rater error training does not guarantee the arising of accuracy in assessors' ratings. Several examples of rater error identified as listed in Table 2.1.

Table 2.1 Examples of Rater Errors likely to be made by assessors

Source: (Aguinis, 2013)

Intentionally Rating Errors	Unintentional Rating Errors
✓ Leniency ✓ Severity ✓ Central Tendency	✓ Similar To Me ✓ Halo ✓ Primacy ✓ Stereotype

In order to minimize all these errors, the need to focus on its' assessors' performance are most crucial. Many assessment centers was a plaque with rater errors as these errors relatively correlated with individual forces. The reduction of rater error results to more effective performance ratings in assessment center and high competency of assessors (Woehr and Huffcutt, 1994). The rater errors are described as following:

1. **Leniency error:** The errors that occur when assessor giving inflated lenient rating scales to the majority of employees rather than actual result merely because of the intention to show high performance of organizations, encouraging people to enhanced quality of work. Sometimes, leniency error occurs to make the upper level directors look good and being as a role model. This rater error may cause from the desire to show highness of something in an evaluation.

2. **Severity error***:* a tendency of an assessor to point out low ratings to almost employees in the organization. However, it may cause of rating declines. Severity error includes the intention of turn out performance evaluation with low rates to the employees. Severity errors occur unintentionally because of inconsistency rating that given lover than the actual performance.
3. **Central tendency error:** This error may occur when an assessor give the average ratings to all employees as a safe position. Despite, this error would have significance effect of the harder to find different factors of each employees.

4. **Similar-to-me error:** An unintentionally errors that have priority for an assessor to access the employees with similar attraction to himself or herself. The attraction may be related to his or her own interest, personality and preferences in life.

5. **Halo error:** A tendency of an assessor to rate employees based on overall consciousness about the employees. It occur when the assessor failed to distinguish between each dimension of performance. For example, the assessor gives high overall rating although the employee's attendance is bad but only perform his job excellently.

6. **Primacy error:** Ratings of the assessor are influenced by the first information obtained about the employees. At the same time, he ignoring the other factors that are more important. For example, in an interview the assessor make a primacy judgment because he seen the low examination result of the candidates and decide that the candidate is not suitable for the position.

7. **Stereotype error:** one of the most focused error that happen because of a priority mindset of grouping factor. This error may cause to low rating performance to a certain group in term of bias, race, gender or physical characteristics. For example, white people' assessors have higher rating scale to the white employee rather that black employee.

As performance ratings are noted to be unreliable by a researcher, these errors above are the evidence that ratings including of both non-systematic and systematic error. Through the present of these errors, several studies identified rater error training to overcome all those errors (Woehr and Huffcutt, 1994). This study will discuss further about the effect of rater error training towards several variables.

2.4 ACCURACY

A study done by Lievens (2001) had shown the prove that trained assessor indicate more accuracy in ratings rather that untrained assessor. However, to evaluate hypothesis in this study, it must be done correctly. Accuracy can be determined by comparing several instruments from various resources. The presence of accuracy in any performance evaluation or performance appraisal is very censorious as it may help in the quality of results itself. Observation is one of assessment that need high and effective accuracy in the appraisals, (Nathan and Alexander, 1985). On top of that statement, Wirz et al. (2013) added the strongest effect of rating accuracy are from the assessor training.

According to Borman's performance appraisal model, Borman (1978) by dealing effectively with his second step of the rating process, the accuracy of performance evaluation of an assessor can be improved simultaneously.

2.5 DISCRIMINANT VALIDITY

Results obtained by Lievens (2001) summarized that discriminant validity of frame of reference result that highest value compared to data-driven training and control training. Discriminant validity has several dimension to be focused; planning and organization, problem analysis and solving and interpersonal sensitivity. Borman (1978) stressed that discriminant validity has highest results compared to other studies, but it differs in each organization. The overall indicator that may affect the assessor training should be analysed. This study will analyse whether discriminant validity may affect overall results of assessor training or not.

2.6 INTERRATER RELIABILITY

Interrater reliability is the ability of several trained or untrained assessor to make average ratings of an employee. Rather than trained psychologists, most assessment center used managerial level as assessors and they will continue with this culture in the assessment center, (Jackson et al., 2005). Interrater reliability is one of the steps taken in assessment center which it help a lot in producing more precise results. The existence of reliability in results is critical. A statement by Jackson et al. (2005) showing that the differences of the presence of interrater reliability are clearly differed under situations of before and after receiving assessor training. Moreover, the result of reliability will change over time, thus continuous training by assessment center are necessary to overcome the problem. Factor such as human resources, technology, physical environment or economic factors may affect the reliability of an assessor. Assessors are able to make better independent evaluations of performance through appropriate assessor training. To determine final evaluation, the training helped a lot in order to increase interrater reliability.

2.7 RESEARCH GAP & HYPOTHESIS

Several interesting hypotheses are derived which expected from the result of this study regarding the effects of the training strategies on managerial assessors of the assessment center. Those hypotheses related closely to rating accuracy, discriminant validity, and interrater reliability. A testable statement would be outperformed in the designation of hypotheses. With using a number of hypotheses, the following are outperform:

Hypothesis 1: Frame of reference training will significantly influence the assessors' ratings accuracy

Hypothesis 2: Frame of reference training will significantly influence the assessors' discriminant validity

Hypothesis 3: Frame of reference training will significantly influence the assessors' interrater reliability

Hypothesis 4: Rater error training will significantly influence the assessors' ratings accuracy

Hypothesis 5: Rater error training will significantly influence the assessors' discriminant validity

Hypothesis 6: Rater error training will significantly influence the assessors' interrater reliability

Hypothesis 7: Frame of reference training is better than rater error training.

Those hypotheses are outperformed based on conceptual models of human judgment and prior past research. Those hypotheses are developed regarding the effects of alternative assessor training strategies towards the assessors in UMP. The research reported here was designed to address this gap. The researcher wanted to obtain information on how assessor training would affect the assessors' accuracy, discriminant validity, and interrater reliability. Which training are the most suitable to be conducted in UMP? Have these training are conducted before in UMP? As well as these questions, the hypothesis and expectations that the researcher outperformed.

Hypothesis 1, 2 and 3 are developed based on previous research by Lievens (2001) where he had proved all those hypothesis are effective to the managerial levels in Ghent University. So then, the researcher conducted this study towards management in UMP only. Hence, it was expected by the researcher to conduct a study on assessor training towards management in UMP. Therefore, Hypotheses 4, 5, and 6 are tested to obtain results of which of the two training strategies are effective based on the three independent variables.

2.9 SUMMARY

UMP is an organization that involves thousand individual as its stakeholders. Therefore, human resources department as the most important in managing resources in UMP needs to revise back their assessment of the assessors. Rather than choosing assessors from the managerial level, other subordinates also could be trained to give the balancing in management. This chapter not only focus to the content and objectives but also as the backbones for this research. The literature review may perform the answer by the results from collected data. In order to explore on how to conduct the research well, refer to next chapter that discussed specific in research methodology.

CHAPTER 3

RESEARCH METHODOLOGY

3.1 INTRODUCTION

The fundamental of a research is to know how to conduct the extended research. Early assumptions are the most important in conducting a research. Then, the analysis from the research answered whether that assumptions are acceptable or not. A set of research method was chosen due to the priority of this research. It is vital to deeply focus on every phase of the study to eliminate overlooked fact and to design great results. Methodologies are defined in many ways. A methodology is a typical process of collecting data and information more precisely for the purpose of logical concern and reasoning.

The study was designed to investigate the relationship of assessor training strategies towards accuracy, inter-reliability and discriminant validity. Primary data are collected by the researcher to quantify this relationship. Generally, research can be divided into two methods which are quantitative and qualitative. Quantitative research method was used in this study. Quantitative methods include questionnaires, field, and statistical data gathered by organization meanwhile the qualitative method include interview, focus group and observations.

This chapter shall represent the portion where population and sampling, data collection method, questionnaires design and research design are introduced. Only one data collection method are used in this research that is a print-out questionnaire. The overall processes of development this paper are discussed in this chapter.

3.2 RESEARCH DESIGN

Research design aims to provide a clear process from initial until the completion of this research conducted. A research design will usually include how the data will be collected, what instruments will be used, how the instruments will be used and the means that seek to analyse the collected data.

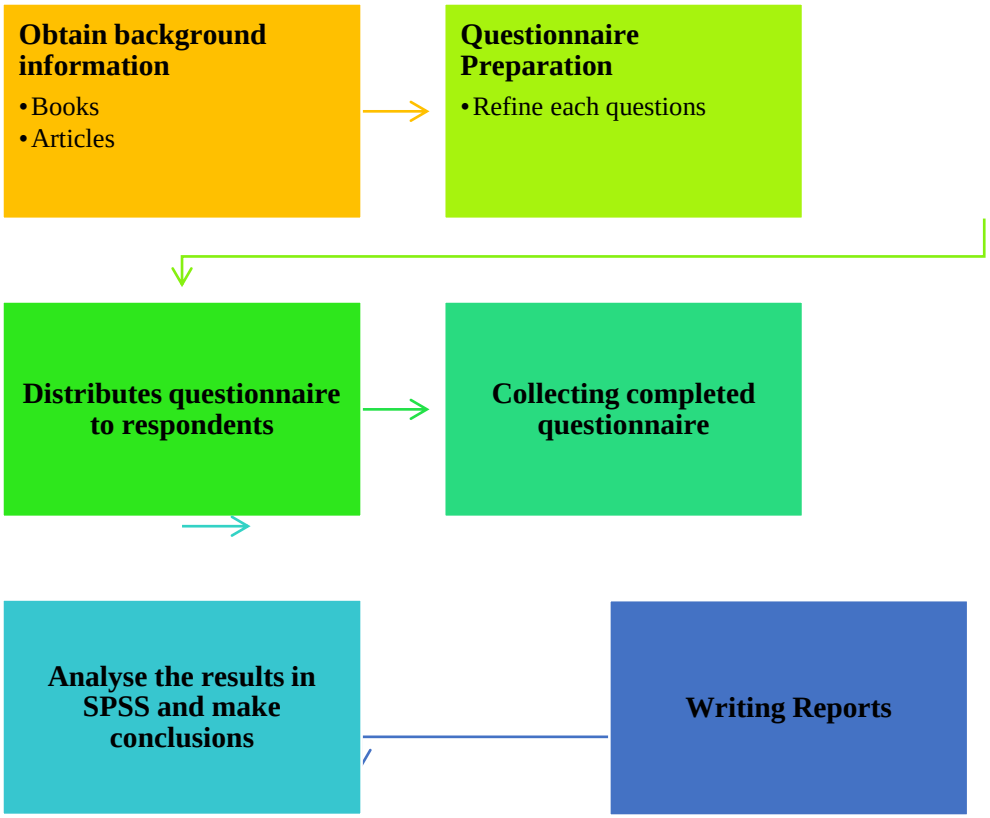

Figure 3.1 Research Process Flow

The research begins with choosing suitable topic and objectives as the first step. Deciding the area of interest was very crucial for the researcher as to maintain in the overall process. During obtain background information process, further confirmation of reasonable of proceeding with the topic. Past research related topic from books and encyclopaedias enhancing the support to proceed more. However, nobody can do well and perfect in anything. Collaborating with experienced researchers are important in any research when refining again about the picked topic in order to narrow down the scope to be more specific and reliable to other research.

Next are questionnaire preparations based on the literature and research objective. Results a gathered in various ways such as questionnaires or interview. Print-out questionnaires are distributed to all targeted populations directly by the researcher. Before the survey set are given, the target respondent was given a short explanation about what the study focusing. It is important in order to obtain specific results. Then, completed questionnaire are analysed by using SPSS and the results are written in the report.

3.3 POPULATION AND SAMPLING

Population commonly called as a group that containing all the members whereas sample is the selected elements that presents in the population. The population that this research enclosed are identified through discussions with recruitment unit and registration at every office prior to the actual data collection. In this study, the population are 28 departments. Sample respondent chosen is the subset of the population only as it is quite impossible to collect data from the whole population. Sampling size is 84 and it depends on the responses that collected from the researcher. However, by hook or by crook the researcher need to collect the convenient number of sample to obtain sufficient information regarding this research.

The initial target population of this study are the top management of each office in UMP. However, due to the basic of this study focusing only for the assessors, it depends on the office itself. Target population is 100 that are from every faculty, department and center in UMP for every process of staff selection. The focus of this study was in UMP Gambang because the majority of the population are still in UMP Gambang. It is also due to time and logistic constraints for data collection process. The researcher not include only one office, The Registrar Office as they are the owner of the selection system while this study collected for feedback or opinions only. A short discussion between researcher and the representative of each office resulted the list of respondent that should fill the questionnaire. The discussion is very crucial to ensure that data are only from selected assessors. Otherwise, the accuracy of the results will be lower.

Krejcie and Morgan (1970) had come out with a table for determining sample size for a given population. It is important to the researcher to determine whether the sample size is good enough for the accuracy of the results with confidence. The sample size is the number of representative from a population that selected randomly as observations in this study. A table below show the table for determining sample size from populations. Why this table was so important? It may conclude for 95% confidence level of the data collected to be precise or not. For example, this study has sample size with almost 100 so the most appropriate sample needed are a minimum of 80 respondents to prove that the results are valid enough. On top of that, Krejcie and Morgan (1970) also had come out with maximum only 5% margin of error if the sample was 80 out of 100.

Table 3.1 Determination Sample Size

N	S	N	S	N	S	N	S	N	S
10	10	100	80	280	162	800	260	2800	338
15	14	110	86	290	165	850	265	3000	341
20	19	120	92	300	169	900	269	3500	246
25	24	130	97	320	175	950	274	4000	351
30	28	140	103	340	181	1000	278	4500	351
35	32	150	108	360	186	1100	285	5000	357
40	36	160	113	380	181	1200	291	6000	361
45	40	180	118	400	196	1300	297	7000	364
50	44	190	123	420	201	1400	302	8000	367
55	48	200	127	440	205	1500	306	9000	368
60	52	210	132	460	210	1600	310	10000	373
65	56	220	136	480	214	1700	313	15000	375
70	59	230	140	500	217	1800	317	20000	377
75	63	240	144	550	225	1900	320	30000	379
80	66	250	148	600	234	2000	322	40000	380
85	70	260	152	650	242	2200	327	50000	381
90	73	270	155	700	248	2400	331	75000	382
95	76	270	159	750	256	2600	335	100000	384

Note: "N" is population size
"S" is sample size.

Krejcie, Robert V., Morgan, Daryle W., "Determining Sample Size for Research Activities", Educational and Psychological Measurement, 1970.

Source: (Krejcie and Morgan, 1970)

3.3 DATA COLLECTION

One method of data collection of this study was selected to originate the results in a clearer picture. The total length to 2 months was taken to collect the required data. The aim of data collection is to obtain a clear proved results that will be useful for future research.

3.3.1 Print-out questionnaire

A number of printed questionnaire are distributed directly to related targeted respondents of assessors in UMP. Some people may require this method instead of email due to several factors. List of respondent discussed with the representative of each office in UMP Gambang. Only English language are used in the questionnaire. This method was chosen as it is easy to understand terms used in the questionnaire set to avoid confusing to the respondents as the majority of respondent in UMP are the top management and might busy with their own tasks.

3.4 DEVELOPMENT OF INSTRUMENTS

The study employed a referred questionnaire that was personally distributed to the potential assessors. The questions was developed based on research objectives and by referring to the literature review. The questionnaire included only close-ended questions for easier analysing results. Only English language was used in those questionnaire. The questionnaire form was divided into three sections that is Section A, Section B, and Section C.

Section A included the demographic information of the respondents such as age, gender, race, working experience, department or faculty and their position grade. While, Section B focused about the two assessor training strategies that are the frame of reference training and rater error training. It is aimed to answer the first objective of this study. The researcher used Likert scale to measure each question. The Likert scale consists of five levels which is 1 “strong disagree”, 2 “disagree”, 3 “neutral”, 4 “agree” and 5 “strongly agree”.

The following Section C are focused more to the second objective of the study. The questions asking respondents' opinion on how the appearance of assessor training will affect their accuracy, discriminant validity and interrater reliability on ratings for staff selection. The same Likert scale of 1 "strong disagree", 2 "disagree", 3 "neutral", 4 "agree" and 5 "strongly agree" are used in Section C.

3.5 DATA ANALYSIS

The analysis of the survey results was developed using Statistical Package for the Social Sciences (SPSS) software. An analysis on descriptive statistic was conducted on individuals' demographic variables, the age, gender, race, working experience, department or faculty and their position grade. The relationship between dependent variables and dependent variables are analysed by using multiple regressions method. Multiple regressions method are extended from simple linear regressions. It is commonly used when need to forecast the value of a variable based on the value of two or more other variables.

3.6 PILOT STUDY

A pilot questionnaire was tested before proceeding to a grand scale data collection of the study for correct wording and to verify whether the questions are understood by the respondents, to test whether questions are highlighting directly to main points of this research and whether a proper sequence of questions and format was included in the questionnaire.

For this study, a total of 20 lecturers picked randomly in UMP Gambang were given pilot questionnaire for improvement by the researcher. All of these respondents were met face to face for easier commenting and correcting the questionnaire. The testing was carried out for the reliability and consistency of questionnaire. (Dayarathna, 2009).

3.7 SUMMARY

Research methodology function as a process that allow for data collection with corresponding to the objective. Hence, analysis of all data can proceed smoothly with sufficient data. Correct methods in collecting data are the critical factor in this chapter. The researcher needs to think deeply for the methods. Once the questionnaire distributed, nothing could be saved. This chapter deals with research design process that applied a systematic approach to data collection to ensure that once has useful data, the result can be understood, and that the procedures can be carried out by others in the future. As the conclusion, all data that obtained researcher would come out with the output.

CHAPTER 4

DATA ANALYSIS

4.1 INTRODUCTION

Data analysis usage is as old as the calculations used to build the Pyramids. Data analysis can describe customer behaviour over the past month or longer, depending on the firm's needs. Data analysis is known by several different names: analytics, data mining, and statistical analysis (Scarisbrick-Hauser, 2007). It is important to differentiate between these different activities to understand the whole data-analysis process.

In this chapter, an interpretation of the result data using descriptive statistics, reliability analysis, normality, hypothesis testing, and mean analysis had performed. The total of 84 respondents are analysed in this study.

4.2 RELIABILITY ANALYSIS

Reliability usually defined as the consistency of data. Reliability analysis was used to improve existing scales, construct reliable measurement scale and to evaluate for the reliability of each question. Reliability analysis was important to be executed before further analysis could be proceed. The most common used statistics to do reliability analysis is Cronbach's coefficient alpha as the questions in Section B and Section C are using Likert scale.

Tables below shows the level of reliability test for the set of questionnaire. Nunnally (1978) recommends the minimum value of 0.7 for reliable value. A low value of alpha could be due to a low number of questions and poor interrelated with the research. Cronbach's alpha is expressed as a number ranging between 0 and 1.00, with 0 indicating no reliability, and 1.00 indicating perfect reliability. According to Tavakol and Dennick (2011), Cronbach's alpha is expressed as values between 0 and 1, which higher values indicates greater reliability.

Cronbach's alpha values are dependent to the number of items in the scale. Readings underneath 0.50 are thought to be problematic and unsuitable.

Table 4.1 Results of Reliability Tests on the Training that Influence Employee Selection in UMP

Training	Cronbach's Alpha	N of Items
Assessor Training	0.813	4
Frame of Reference Training	0.803	5
Rater Error Training	0.461	5

Table 4.2 Result of Reliability Test on the Variables that Influenced the Assessor Training Strategies

Variables	Cronbach's Alpha	N of Items
Accuracy	0.607	4
Discriminant Validity	0.652	5
Interrater Reliability	0.646	4

Table 4.3 Result of Reliability Test on Section B and Section C

Variables	Cronbach's Alpha	N of Items
Section B	0.732	14
Section C	0.699	13

Table 4.4 Result of Reliability Test of the Questionnaire

Reliability Statistics

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
0.768	0.769	27

Table 4.5 Reliability Analysis of Section B and Section C

Variables	Test reliability	n	Value of Cronbach's Alpha
Assessor Training	Pilot test	10	0.883
	Retest	84	0.813
Frame of Reference Training	Pilot test	10	0.770
	Retest	84	0.803
Rater Error Training	Pilot test		0.761
	Retest	84	0.461
Accuracy	Pilot test	10	0.630
	Retest	84	0.607
Discriminant Validity	Pilot test	10	0.738
	Retest	84	0.652
Interrater Reliability	Pilot test	10	0.640
	Retest	84	0.646

Two data set was run by using SPSS and tabulated in the above table as shown. First data set was run for the pilot testing at the initial stage of this study while second data set used for analyse the results from 84 respondents. The most reliable variable for independent variable is discriminant validity with the value of Cronbach's Alpha of 0.738. The total number of items in this construct is 5. Whereas, the most reliable variable for dependent variable is assessor training with the value of Cronbach's Alpha is 0.813.

4.3 NORMALITY TESTS

Normality data assessment is a prerequisite for many statistical tests because normal data is an underlying assumption in parametric testing. A normality test was used to determine whether a data set is well modelled by normal distribution or not by using SPSS. Moreover, normality test also have several techniques that can be used such as Shapiro-Wilk test, Quantile-Quantile test or as known as Q-Q test, and Skewness-Kurtosis test. However, the technique used to measure normality is Q-Q test only.

The result can be concluded by using Q-Q plot whether the data consisted in this study is normally distributed or not. The dots in Q-Q plot of normally distribution data should fall approximately on the line and not in a curves or S-shape. This is because the general rule of Q-Q plot where normally distribution dots should be approximately and the dots should be a straight line and not forming in any pattern.

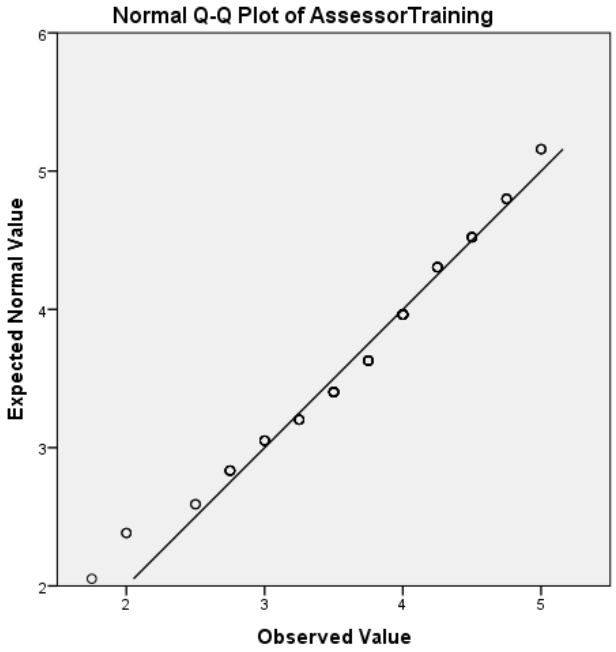

Figure 4.1 Q-Q Plot for Assessor Training

Assessor training is the training that given to the assessors with objective to improve their skills. The figure above shows normality test by using SPSS. While Q-Q plot is used in this test, the above figure is obtained where the Q-Q plot indicates that the data for assessor training is normally distributed. This is because the dots in the Q-Q plot are approximately on the straight line and not forming in any pattern or being in the form of S-shape.

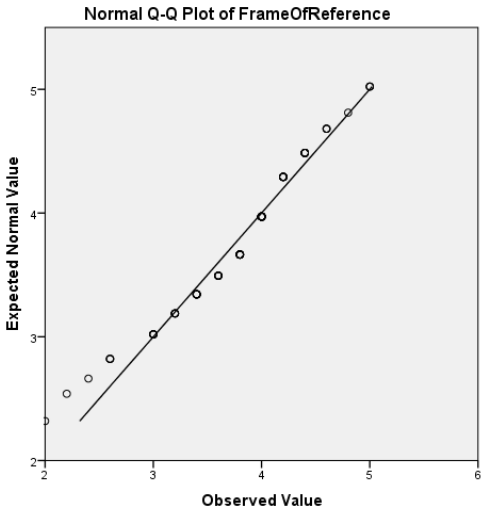

Figure 4.2 Q-Q Plot for Frame of Reference Training

The above figure shows result for normality test on frame of reference training. The data indicates that there is a normality distributed plot for this training because most of the dots in the Q-Q plot are approximately lying on the straight line and not forming in any pattern or being in the form of S-shape.

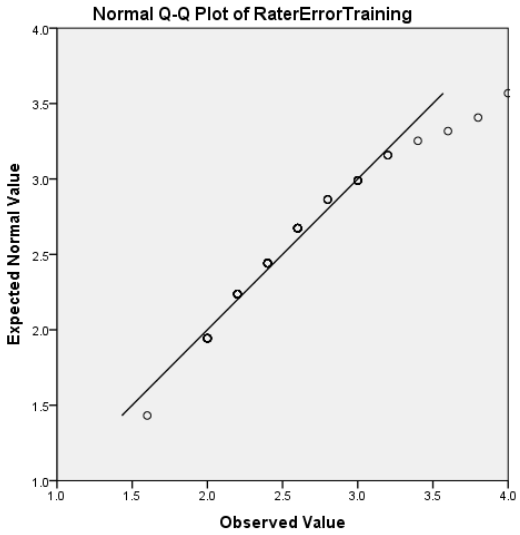

Figure 4.3 Q-Q Plot for Rater Error Training

Rater error training is one of the training strategies for assessors. The above figure shows the dots in the Q-Q plot above indicates that the data is normally distribution. Most of the dots in the Q-Q plot are approximately lying on the straight line and not forming in any pattern or being in the form of S-shape.

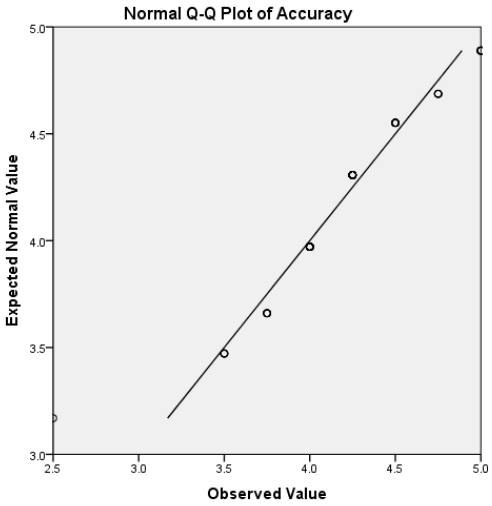

Figure 4.4 Q-Q Plot for Accuracy

Accuracy in every assessment is meaningful as for the results. Therefore, the questions must be normally distributed to obtain better outcomes. Figure 4.4 demonstrate the normally distributed Q-Q plots for accuracy. Most of the dots in the Q-Q plot are approximately lying on the straight line and not forming in any pattern or being in the form of S-shape. Consequently, it shows that the data is strong enough to be analysed.

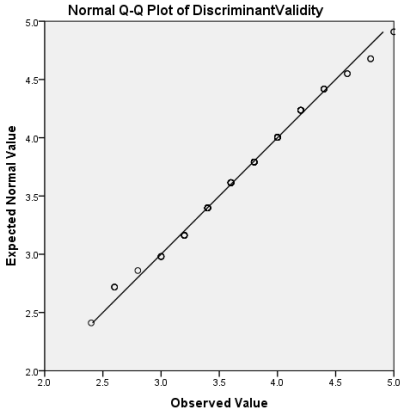

Figure 4.5 Q-Q Plot for Discriminant Validity

The second independent variable in this study is the discriminant validity. Figure 4.5 shows normal distribution as for discriminant validity. Some statisticians prefer to use their experience to make a subjective judgement about the data from plots/graphs.

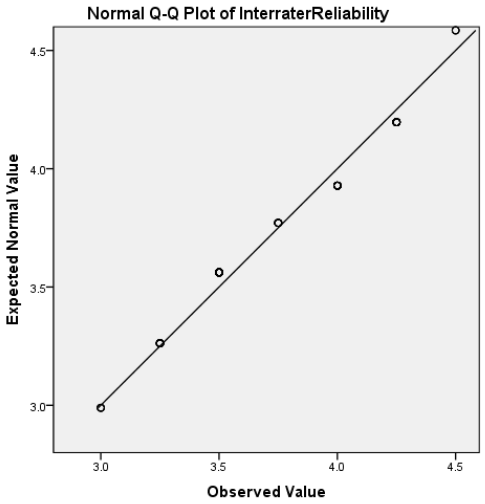

Figure 4.6 Q-Q Plot for Interrater Reliability

Interrater reliability is one of the independent variables of this study. Normal distribution showed in the above figure. Therefore, can be assumed that data for interrater reliability is normally distributed. This is because most of the dots in the Q-Q plot are approximately lying on the straight line and not forming in any pattern or being in the form of S-shape.

4.4 DESCRIPTIVE ANALYSIS

Descriptive statistics can be used to summarize the data in SPSS. Descriptive analysis are used by the researcher to determine the percentage and frequency of each category in demographic part of the questionnaire. Moreover, descriptive analysis also used to obtain analysis of mean, mode, median and standard deviation of the demographic section. Findings from this study several conclusions as shown in the tables below for each of the demographic variable with percentage, frequency and cumulative percentage for each question.

Table 4.6 Demographic Frequency of Respondents' Age

Age

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	21-30	15	17.9	17.9	17.9
	31-40	52	61.9	61.9	79.8
	41-50	12	14.3	14.3	94.0
	51-60	5	6.0	6.0	100.0
	Total	84	100.0	100.0	

According to Table 4.6, majority of respondents are in the age of 31-40 with frequency of 52 out of 84. It shows that most of respondents are in the middle manager position where the respondents are starts to involve in maturity stage.

Table 4.7 Demographic Frequency of Respondents' Gender

Gender

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Male	49	58.3	58.3	58.3
	Female	35	41.7	41.7	100.0
	Total	84	100.0	100.0	

Based on Table 4.7, majority of the respondents are male with value of 58.3%. More than half was male with number of 49 respondents. Meanwhile, 41.7% more are female respondents with 35 person.

Table 4.8 Demographic Frequencies of Respondents' Race

Race

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Malay	79	94.0	94.0	94.0
	Chinese	3	3.6	3.6	97.6
	Indian	1	1.2	1.2	98.8
	Others	1	1.2	1.2	100.0
	Total	84	100.0	100.0	

According to Table 4.8 above, it is shown that 94% of respondents is Malay. It may be influenced by location of the university itself where already known as Islamic country that is Malaysia. The frequency of Malay respondents are 79 out of 84 respondents in total. The rest respondents are from Chinese, Indian and other races.

Table 4.9 Demographic Frequencies of Respondents' No of years working in UMP

Years

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 year and below	4	4.8	4.8	4.8
	2-5 years	22	26.2	26.2	31.0
	6-10 years	30	35.7	35.7	66.7
	More than 10 years	28	33.3	33.3	100.0
	Total	84	100.0	100.0	

According to Table 4.9 above, it can be seen that majority of respondents are having working experience in UMP for 6-10 years as it is the highest frequency. At this number of working experience, it can be conclude that those respondents are person who experienced well with system in UMP so that the data are more accurate. Through analysis by using SPSS, it is shown that most of respondents are experienced well with UMP's environment. Therefore, it is assumed that data collected are more precise and confident to be reliable and valid. Respondents' perceptions are very important in order to obtain correct data analysis.

Table 4.10 Demographic Frequencies of Respondents' Working Grade

Grade

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Below 29	3	3.6	3.6	3.6
	41	21	25.0	25.0	28.6
	44	27	32.1	32.1	60.7
	45	6	7.1	7.1	67.9
	48	3	3.6	3.6	71.4
	51	11	13.1	13.1	84.5
	52	7	8.3	8.3	92.9
	Above 54	6	7.1	7.1	100.0
	Total	84	100.0	100.0	

Working grade for government employees are very important. It shows their achievement and differentiate of status of each employee. Based on this research, analysis of grade also needed in order to convince that the data are good enough to trusts. Based on analysis, it shows that most of respondents are entitled with average of grade 44 as shown in Table 4.10. Thus, it means that they have longer working experience and better skills in their own areas.

Table 4.11 Statistics of the Demographic Data

Statistics

		Age	Gender	Race	Years	Department	Grade
N	Valid	84	84	84	84	84	84
	Missing	0	0	0	0	0	0
Mean		3.08	3.08	1.42	1.10	2.98	13.07
Median		3.00	3.00	1.00	1.00	3.00	12.00
Mode		3	3	1	1	3	7
Std. Deviation		0.752	.748	.496	.428	.891	8.301

Table 4.11 show the statistics of the demographic data obtained from SPSS. Zero missing value means that no questions that have not answered by respondents.

4.5 MEAN ANALYSIS

Table 4.12 are analysed based on the first objective of this study that is to explore the effects of accuracy, discriminant validity and interrater reliability. Mean analysis is another measurement that uses to see which the factors give most affect to the employee performance in a company. Mean of each of the factor will be determined and rank accordingly.

Table 4.12 Mean of the accuracy, discriminant validity and interrater reliability

EVALUATION OF ACCURACY				
1. The frame of reference training will help very much in rating accuracy.	4.14	4	4.19	1
2. More training needed to improve my accuracy in assessment.	4.18	3		
3. The accuracy of ratings depends on experiences with the standard performance criteria.	4.19	2		
4. Assessors with less training may give less accurate ratings.	4.26	1		
EVALUATION OF DISCRIMINANT VALIDITY				
1. The frame of reference training should be provided to assessors to improve our discriminant validity skills.	4.04	1	3.73	3
2. Discriminant validity will affect overall rating scales.	3.95	2		
3. I was unable to differentiate between positive and negative behaviours of candidates without better training.	3.29	5		

4. There are no such proven training that support if my ratings are good enough.	3.65	4		
5. I frequently feel my ratings are will be improved if I am familiarized with standard performance criteria.	3.71	3		
EVALUATION OF INTERRATER RELIABILITY				
1. Consistent rating scales can be obtained with frame of reference training.	4.12	4	4.18	2
2. A standard reference helped assessors to give higher between assessors.	4.16	3		
3. During the selection process, my ratings sometimes differ than the other assessors.	4.19	2		
4. I need to have training with standardize criteria to improve reliability.	4.26	1		

The above Table 4.12 shows the mean results and ranking for three independent variables of this study; accuracy, discriminant validity, and interrater reliability. Mean results are obtained from SPSS outputs of collected data from respondents. This mean analysis also indicates which items are more important to support the hypothesis.

Starts with accuracy variables, the results shows that the fourth item is the most important. According to this item, assessors with less training will give less accuracy in their ratings. It is corresponding to the initial analysis by researcher related to assessor training. From this item, it can be considered that most respondents believe that they need more training to have better accuracy in giving ratings during selection process. Moreover, the second ranking supported the Hypothesis 1 where the frame of reference training is influenced by ratings accuracy.

Next, looking forward to the second independent variable; discriminant validity. The first item have the highest means among all items in the construct. It shows that respondents support that frame of reference training would influence their discriminant validity skills. This result clearly supported Hypothesis 2 where it is assumed that frame of reference training is significantly influenced by the discriminant validity.

According to interrater reliability of the assessors, most respondents believe that they need frame of reference training to have better interrater reliability. This mean result is corresponding to Hypothesis 3 where the frame of reference training is influenced by interrater reliability.

By comparing the three construct, rating accuracy had achieved the first ranking among them. This result indicates that the respondents are concerning more to accuracy rather than interrater reliability and discriminant validity.

Figure 4.13 Mean Results for Assessor Training Strategies

Assessor Training	Mean	Rank	Total Mean	Total Rank
ASSESSOR TRAINING				
1. Better performance may be achieved through assessor training.	3.82	1		
2. Through effective training, skills of an assessor will be achieved.	3.73	3	3.73	2
3. Before publishing an advertisement of, assessors must be familiarized with standardized criteria by training.	3.75	2		
4. Assessor training should be provided in UMP	3.63	4		
FRAME OF REFERENCE TRAINING				
1. We should be given training that train assessors with respect to common evaluative standards.	3.25	5		
2. A trainer should train us on how to assign ratings for an evaluation by using standard criteria.	3.82	4		
			3.78	1

3. A trainer should have training that discuss various employee behaviours to improve ratings.	3.83	3		
4. All the assessors should attend training on making ratings using new standard dimensions.	3.94	2		
5. We believe that assessors tend to be more confident after completion of training programs with the performance standard.	4.06	1		
RATER ERROR TRAINING				
1. I feel like only familiarizing with common error is enough to have correct ratings.	2.46	3		
2. When assessors realized that they did some mistakes in the selection process, top management encourages them to share what they have learned with other assessors.	2.43	5		
3. I don't need this training to reduce my unintentionally rating error.	2.63	1		
4. By this training, I will be alert to different errors only and I will be a good assessor.	2.57	2	2.50	3
5. Assessors can improve the assessment ratings by having reducing error training only.	2.40	4		

Based on Table 4.13, the first construct is about the assessor training. The first rank shows that respondents give an opinion where better performance may be achieved through assessor training. It can be seen that they assessor training is important for them to achieve better results in assessing candidates.

Results from the frame of reference training construct shows that the first ranking indicates that respondents believe that assessors tend to be more confident after completion of training programs with the performance standard. It is corresponding to the initial assumption by the researcher where the assessor training is important in UMP.

The next construct shows mean results for the frame of reference training where first ranking is the statement of 'We believe that assessors tend to be more confident after completion of training programs with the performance standard'. It was great evidence to show that most respondents trust that they need frame of reference training in UMP.

Compared to frame of reference training, most respondents responds to rater error training with lower results. However, the item of 'I don't need this training to reduce my unintentionally rating error' got the highest mean. It can be analysed that most respondents doesn't feel that rater error training is enough for them to reduce their unintentionally errors only.

Lastly, mean results obtained by comparing construct frame of reference and rater error training, researcher concluded that frame of reference is better than rater error training to support for Hypothesis 7.

4.6 CORRELATION ANALYSIS

In order to prove the hypothesis developed, several multiple regression analysis was conducted to test the relationship between independent variables and dependent variables. This study contained 7 hypothesis as below:

Hypothesis 1: Frame of reference training will significantly influence the assessors' ratings accuracy

Hypothesis 2: Frame of reference training will significantly influence the assessors' discriminant validity

Hypothesis 3: Frame of reference training will significantly influence the assessors' interrater reliability

Hypothesis 4: Rater error training will significantly influence the assessors' ratings accuracy

Hypothesis 5: Rater error training will significantly influence the assessors' discriminant validity

Hypothesis 6: Rater error training will significantly influence the assessors' interrater reliability

Hypothesis 7: Frame of reference training is better than rater error training.

The Hypothesis 7 had answered in previous report of Mean Analysis. Therefore, the rest of hypotheses are tested by using correlation technique. Correlation is a technique for investigating the relationship between two quantitative and continuous variables. In order to determine whether two variables are linearly connected to each other, bivariate correlation can be used. Correlation is a bivariate analysis that measures the strengths of association between two variables. When the value of the correlation coefficient lies around ± 1, then it is said to be a perfect degree of association between the two variables. As the correlation coefficient value goes towards 0, the relationship between the two variables will be weaker.

Pearson Correlation Analysis was conducted to test the quality of the relationship between two variables. The Pearson's coefficient are always ubiquitous while considering other indices of relationship between two variables (Bobko, 2001). Based on the results of Correlation Analysis, 6 hypotheses are tested and the results had discussed in the following section.

Table 4.14 Strength of Correlation Coefficient Value

Size Correlation Coefficient (r)	Correlation Strength
1.00	Perfect
0.80 to 0.90	Very Strong
0.50 to 0.80	Strong
0.30 to 0.50	Moderate
0.10 to 0.30 > 0.10	Modest Weak
0.0	No correlation

Source: Dancey and Reidy's (2004)

Table 4.14 had illustrates for the strength of correlation coefficient value that is proposed by Dancey and Reidy's (2004). The positive values represents for positive relationship between two variables and negative values represents for negative relationship. This study used Pearson Correlation Coefficient method to determine the relationship between independent variables and dependent variables.

Table 4.15 Pearson Correlation for Dependent Variables

Correlations

		Frame Of Reference	Accuracy	Discriminant Validity	Interrater Reliability
Frame Of Reference	Pearson Correlation	1	.218*	.332**	.097
	Sig. (2-tailed)		.047	.002	.378
	N	84	84	84	84
Accuracy	Pearson Correlation	.218*	1	.401**	.023
	Sig. (2-tailed)	.047		.000	.834
	N	84	84	84	84
Discriminant Validity	Pearson Correlation	.332**	.401**	1	.170
	Sig. (2-tailed)	.002	.000		.122
	N	84	84	84	84
Interrater Reliability	Pearson Correlation	.097	.023	.170	1
	Sig. (2-tailed)	.378	.834	.122	
	N	84	84	84	84

*. Correlation is significant at the 0.05 level (2-tailed).

**. Correlation is significant at the 0.01 level (2-tailed).

The above table shows the Pearson Correlation results between the first independent variable and 3 dependent variables. The first value of Pearson correlation is 0.218 that is between Frame of Reference Training and accuracy. Therefore, there is a modest weak positive relationship between these two variables and correlation is significant at the 0.01 level. This means that changes in one variable are correlated with changes in the second variable. Based on the statistically significant correlation, the value is 0.047 where it is less than 0.01 level. It can be concluded that there is a statistically correlation between Frame of Reference Training and Accuracy. Hypothesis 1 was proved supported through this significant.

The Pearson correlation value between frame of reference training and discriminant validity is 0.332. This results interpret that there is moderate positive correlation between them and was supported for Hypothesis 2. The value of Significant is 0.02 where it is less than the provided significant value. On top of this, there is a statistically correlation between Frame of Reference Training and Discriminant Validity. Thus, Hypothesis 2 is supported.

Interrater reliability and Freame of Reference Training have Pearson correlation coefficient of 0.097 value. It is near to 0 value, so it can be concluded that there is a weak positive correlation relationship between these two variables. The significant value is 0.378 which is greater than 0.01 significant level as given. It shows that no statistically correlation between them. This means that changes in Frame of Reference Training are correlated with changes in the interrater reliability. Hypothesis 3 is supported.

Table 4.16 Pearson Correlation for Independent Variables

Correlations

		Rater Error Training	Accuracy	Discriminant Validity	Interrater Reliability
Rater Error Training	Pearson Correlation	1	.149	.059	.172
	Sig. (2-tailed)		.177	.592	.119
	N	84	84	84	84
Accuracy	Pearson Correlation	.149	1	.401**	.023
	Sig. (2-tailed)	.177		.000	.834
	N	84	84	84	84
Discriminant Validity	Pearson Correlation	.059	.401**	1	.170
	Sig. (2-tailed)	.592	.000		.122
	N	84	84	84	84
Interrater Reliability	Pearson Correlation	.172	.023	.170	1
	Sig. (2-tailed)	.119	.834	.122	
	N	84	84	84	84

**. Correlation is significant at the 0.01 level (2-tailed).

The above table shows the Pearson Correlation results between the second independent variable and 3 dependent variables. The first value of Pearson correlation is 0.149 that is between Rater Error Training and accuracy. Therefore, there is weak positive relationship between these two variables and correlation is significant at the 0.01 level. This means that changes in one variable are correlated with changes in the second variable. Based on the statistically significant correlation, the value is 0.177 where it is more than 0.01 level. It can be concluded that there is no statistically correlation between Rater Error Training and Accuracy. Hypothesis 4 was proved supported through this significant.

The Pearson correlation value between rater error training and discriminant validity is 0.59. This results interpret that there is low positive correlation between them and was supported for Hypothesis 5. The value of Significant is 0.592 where it is more than the provided significant value. On top of this, there is no statistically correlation between Rater Error Training and Discriminant Validity. Thus, Hypothesis 5 is supported.

Interrater reliability and Rater Error Training have Pearson correlation coefficient of 0.172 value. It is near to 0 value, so it can be concluded that there is a weak positive correlation relationship between these two variables. The significant value is 0.119 which is greater than 0.01 significant level as given. It shows that no statistically correlation between them. This means that changes in Rater Error Training are not correlated with changes in the interrater reliability. The Hypothesis 6 is supported based on the results.

4.7 SUMMARY OF FINDINGS

Table below shows the summary of analysis of this research. It is a summary of the effects of assessor training strategies towards accuracy, inter-reliability and discriminant validity and summary of the effectiveness of two assessor training strategies in University Malaysia Pahang. The first summary answered the first objective of developing this research and answering the first research question, while the second summary indicates the answer of second research question and achieved second objective.

Table below shows the summary of the effects of assessor training strategies towards accuracy, inter-reliability and discriminant validity. From the analysis of data above, the findings are illustrated in Table 4.17.

Table 4.17 Findings of Study

Hypothesis		Correlation/ Mean Results	Analysis Result	Analysis Method
H1	Frame of reference training will significantly influence the assessors' ratings accuracy	0.218	Supported	Correlation Analysis
H2	Frame of reference training will significantly influence the assessors' discriminant validity	0.332	Supported	Correlation Analysis
H3	Frame of reference training will significantly influence the assessors' interrater reliability	0.097	Supported	Correlation Analysis
H4	Rater error training will significantly influence the assessors' ratings accuracy	0.149	Supported	Correlation Analysis
H5	Rater error training will significantly influence the assessors' ratings discriminant validity	0.059	Supported	Correlation Analysis
H6	Rater error training will significantly influence the assessors' interrater reliability	0.172	Supported	Correlation Analysis
H7	Frame of reference training is better than rater error training.	3.78	Supported	Mean Analysis

Table below shows the summary of effectiveness of Assessor Training Strategies in UMP.

Table 4.18 Level and score used to determine the effects of assessor training strategies towards accuracy, discriminant validity and inter-reliability

Level	**Score**
High	0.01 – 1.00
Moderate	1.01 – 2.00

Table 4.19 Result of the Effectiveness of Assessor Training in UMP

Assessor Training	**Total mean**	**Average Mean Score**	**Level**
Frame of Reference Training	3.78	1	High
Rater Error Training	2.50	2	Moderate

The Table 4.19 clearly shows that the second research objective are answered very well where frame of reference training are more effective for the assessors in UMP. On top of that, this results also answered for fourth hypothesis developed by the researcher. Designation of the hypothesis can be seen in Chapter 2. It can be concluded that frame of reference training strategy is more effective than the rater error training based on the literature review. This is based on the total mean of frame of reference is higher.

4.8 DISCUSSION OF FINDINGS

Since overall results had shown in previous section, they need to be discussed and analysed comprehensively. Various methods and techniques are used to tests them. Hence, the purpose of this discussion is to conclude all those results according to the research objective and whether the research questions are answered or not.

The first objective of this research is to explore the effects of accuracy, discriminant validity and interrater reliability. Throughout this study, the variables had influenced by assessor training are accuracy, discriminant validity and interrater reliability. All hypotheses are supported by the previous research by (Lievens, 2001). Those results are shown in Table 4.17 and are analysed by using Pearson correlation coefficient. It shows that the 6 hypothesis related to assessor training and dependent variables was supported. As conclusion, both frame of reference training and rater error training are accepted to be executed in assessment center of UMP. First objective was achieved for this study.

By comparing the three construct of dependent variables, rating accuracy had achieved the first ranking among them. This result indicates that the respondents are concerning more to accuracy rather than interrater reliability and discriminant validity.

Second objective is concerning about the effective of assessor training strategies in UMP. These two training strategies was analysed by comparing their means using ranking technique. The total mean of frame of reference training is higher than rater error training. Based on Table 5.3, there was evidently shows that the second research objective is answered very well where frame of reference training are more effective for the assessors in UMP.

4.9 CONCLUSION

This chapter had compute the project data collection of the effects of assessor training strategies towards accuracy, inter-reliability and discriminant validity and the assessor training strategies. Based on overall results, some conclusions are made by the researcher. It is concluded that Frame of Reference Training shows more effective rather than rater error training. Moreover, it is also concluded that accuracy, inter-reliability and discriminant validity had given large effects towards Frame of Reference Training. It shows that Frame of Reference Training is important to the assessors in UMP.

CHAPTER 5

CONCLUSION AND RECOMMENDATION

5.1 INTRODUCTION

Throughout this chapter, overall analysis of the results had compiled together in order to obtain the research objective. There will be several discussions on recapitulation of the study, limitation, recommendation and the end of this chapter will be the conclusion. The limitation and recommendation are highlighted for future research in performance related to the assessor training strategies.

5.2 CONTRIBUTION

The contribution of the current investigation are at least beneficial for twofold. First, the focused group that is assessors in UMP and second is the organization itself that is UMP. As for the assessors, more exposure and awareness to the importance of the assessor training for them is one of the contribution by this study. They might seeks more information of what type of training needed.

On top of that, this study gives guidelines to other educational organizations in the world that they also need to focus on training to the assessors in order to choose high quality employees and next would improve their management attribute. Besides, this study will benefits to UMP itself to recognize which type of training that is more suitable for their assessors.

Finally, this study gives huge contribution for the researchers to know more about the practice of selection process in UMP. The evidence from this study presented about two assessor training that is suitable to be executed in UMP.

5.2 LIMITATION

Despite the contribution of our study, conclusions should be interpreted with certain limitations in mind. Limitation perpetually becomes the barriers to ensure that this study can be progress smoothly. Throughout this study, some limitations are faced by the researcher. However, it could be improved in further studies. Firstly, lack of cooperation from the respondents had drag time period to run this study. This is because most of the target respondents are from the top management. Hence, they do not have much time to fill up the questionnaires due to their work constraints.

Moreover, due to logistic limitation the researcher could not focus to the entirely whole university and just collect data from UMP Gambang only. However, all those data in this study still considered as valid because only 3 faculty located in UMP Pekan as for now.

Luckily the researcher choose to distribute the questionnaires personally to each assessors. Even though it takes quite longer to distribute and wait for the respondents to return back, more than 80 responds are collected successfully. According to Krejcie and Morgan (1970), if the population is 100, then at least 80 responds need to have to be valid. However, limitation of the quality of data cannot be avoided. Some respondents give excuses not to answer the questions and just give random assessment.

5.3 RECOMMENDATION

Some recommendations are outlined to further improvements related to the topic of this study. The assessment center effectiveness is recommended to be maximized. This is due to only a small proportion of the management in human resource of UMP provide assessors with any information regarding their performance. Providing the assessors with at least a minimal quantity of information by training would give better results for selection process. Moreover, few assessors report considering issues of not sure that their ratings are good enough or not. This issue is due to lack of information delivery from human resource department.

Next is, the researcher find difficulties in outlining the initial target population. This is because some websites of the faculty and departments in UMP are outdated and need to be improved. Number of assessors of each departments and faculty is not consistent to each other. Further clarification need to gain by the research during the distribution process.

In terms of technique used, various techniques are recommended to obtain high quality of results for the next future research. Better results can be obtain if interview technique is used to acquire the opinions from the assessors. However, that researcher must really expert in analysing the results then.

5.4 CONCLUSION

As a conclusion, the objectives of this study had achieved successfully. There was an effects of assessor training strategies towards accuracy, inter-reliability and discriminant validity. The frame of reference training are proved to be better rather than the rater error training. It should be notice that training to the assessors in UMP is significance in order to achieve better accuracy, discriminant validity and interrater reliability of the assessors. The 7 hypotheses are tested and answered greatly in the data analysis chapter. Therefore, the researcher suggests that these two training need to be executed in UMP and focus for the assessors.

REFERENCES

AGUINIS, H. 2013. *Performance Management*, Pearson Education, Limited.

ANDERSON, N. R. 1992. Eight Decades of Employment Interview Research: A Retrospective Meta-review and Prospective Commentary. *European Work and Organizational Psychologist,* 2**,** 1-32.

BARCLAY, J. M. 2001. Improving selection interviews with structure: organisations' use of "behavioural" interviews. *Personnel Review,* 30**,** 81-101.

BITTNER, R. H. 1948. Developing an Industrial Merit Rating Procedure. *Personnel Psychology,* 1**,** 403-432.

BOBKO, P. 2001. *Correlation and regression: Applications for industrial organizational psychology and management*, Sage Publications.

BORMAN, W. C. 1978. Exploring Upper Limits of Reliability and Validity in Job Performance Ratings. *Journal of Applied Psychology,* 63**,** 135-144.

BYHAM, W. C. 1992. *The Assessment Center Method and Methodology: New Applications and Technologies,* Development Dimensions International.

DOCKALIKOVA, I. & KASHI, K. 2013. Selection of Employees: Multiple Attribute Decision Making Methods in Personnel Management. *Proceedings of the European Conference on Management, Leadership & Governance***,** 367-375.

EL-KOT, G. & LEAT, M. 2008. A survey of recruitment and selection practices in Egypt. *Education, Business and Society: Contemporary Middle Eastern Issues,* 1**,** 200-212.

FRATRIČOVÁ, J. & RUDY, J. 2015. Get Strategic Human Resource Management Really Strategic: Strategic HRM in Practice. *International Journal of Management Cases,* 17**,** 149-155.

GAUGLER, B. B., ROSENTHAL, D. B., THORNTON III, G. C. & BENTSON, C. 1987. Journal of Applied Psychology Monograph - Meta-Analysis of Assessment Center Validity. *Journal of Applied Psychology,* 72**,** 493-511.

HAGAN, C. M., KONOPASKE, R., BERNARDIN, H. J. & TYLER, C. L. 2006. Predicting assessment center performance with 360-degree, top-down, and customer-based competency assessments. *Human Resource Management,* 45**,** 357-390.

HASHIM, J. 2010. Human resource management practices on organisational commitment: The Islamic perspective. *Personnel Review,* 39**,** 785-799.

JACKSON, D. J. R., ATKINS, S. G., FLETCHER, R. B. & STILLMAN, J. A. 2005. Frame of Reference Training for Assessment Centers: Effects on Interrater Reliability When Rating Behaviors and Ability Traits. *Public Personnel Management,* 34**,** 17-30.

JONES, R. G. 2007. Assessment Centers in Human Resource Management: Strategies for Prediction, Diagnosis, and Development. *Personnel Psychology,* 60**,** 249-252.

KNIGHT, R. M. & ZAKLAD, A. 2001. The HR Scorecard: Linking People, Strategy, and Performance. *Human Resource Planning,* 24**,** 50-51.

KREJCIE, R. V. & MORGAN, D. W. 1970. Determining sample size for research activities. *Educ Psychol Meas*.

LANIK, M. & MITCHELL GIBBONS, A. 2011. Guidelines for Cross-Cultural Assessor Training in Multicultural Assessment Centers. *Psychologist-Manager Journal (Taylor & Francis Ltd),* 14**,** 221-246.

LIEVENS, F. 2001. Assessor Training Strategies and Their Effects on Accuracy, Interrater Reliability, and Discriminant Validity. *Journal of Applied Psychology,* 86**,** 255-264.

LIEVENS, F. 2002. An examination of the accuracy of slogans related to assessment centres. *Personnel Review,* 31**,** 86-102.

MACAN, T., MEHNER, K., HAVILL, L., MERIAC, J. P., ROBERTS, L. & HEFT, L. 2011. Two for the Price of One: Assessment Center Training to Focus on Behaviors Can Transfer to Performance Appraisals. *Human Performance,* 24**,** 443-457.

MAIMUNAH, A. 2011. *Human Resource Management: Principles and Practices,* Oxford University Press.

NATHAN, B. R. & ALEXANDER, R. A. 1985. The Role of Inferential Accuracy in Performance Rating. *Academy of Management Review,* 10**,** 109-115.

NOONAN, L. E. & SULSKY, L. M. 2001. Impact of Frame-of-Reference and Behavioral Observation Training on Alternative Training Effectiveness Criteria in a Canadian Military Sample. *Human Performance,* 14**,** 3-26.

NUNNALLY, J. 1978. Psychometric methods. New York: McGraw-Hill.

SCARISBRICK-HAUSER, A. 2007. Data analysis and profiling. *Direct Marketing: An International Journal,* 1**,** 114-116.

SCHLEBUSCH, S. & ROODT, G. 2008. *Assessment Centres: Unlocking Potential for Growth,* Knowres Pub.

SMITH, D. E. 1986. Training Programs for Performance Appraisal: A Review. *Academy of Management Review,* 11**,** 22-40.

SPOOL, M. D. 1978. TRAINING PROGRAMS FOR OBSERVERS OF BEHAVIOR: A REVIEW. *Personnel Psychology,* 31**,** 853-888.

TAVAKOL, M. & DENNICK, R. 2011. Making sense of Cronbach's alpha. *International journal of medical education,* 2**,** 53.

THAYER, P. W. 2004. Developing Organizational Simulations: A Guide for Practitioners and Students (Book). *Personnel Psychology,* 57**,** 536-538.

WIRZ, A., MELCHERS, K. G., LIEVENS, F., DE CORTE, W. & KLEINMANN, M. 2013. Trade-Offs Between Assessor Team Size and Assessor Expertise in Affecting Rating Accuracy in Assessment Centers. *Trade-offs entre tamaño del equipo evaluador y pericia del evaluador y su efecto sobre la precisión de la valoración en los assessment centers.,* 29**,** 13-20.

WOEHR, D. J. & HUFFCUTT, A. I. 1994. Rater training for performance appraisal: A quantitative review. *Journal of Occupational & Organizational Psychology,* 67**,** 189-205.

YAHYA, K. K. & FEE-YEAN, T. 2015. ENHANCING CAREER COMMITMENT: THE INFLUENCE OF HUMAN RESOURCE MANAGEMENT PRACTICES. *International Journal of Business & Society,* 16**,** 237-246.

APPENDIX A : QUESTIONNAIRE

Bachelor Degree Final Year Project

A study of assessor training strategies and its effect on accuracy, discriminant validity and interrater validity in UMP

SECTION A: <u>Demographic Information</u>

Please tick (／) one answer for the following question.

1. **Age**
   - ☐ 20 and below
   - ☐ 21 – 30
   - ☐ 31 – 40
   - ☐ 41-50
   - ☐ 51-60
2. **Gender**
   - ☐ Male
   - ☐ Female
3. **Race**
   - ☐ Malay
   - ☐ Chinese
   - ☐ Indian
   - ☐ Others: _____
4. **No. of years working in this company**
   - ☐ 1 year and below
   - ☐ 2 – 5 years
   - ☐ 6 – 10 years
   - ☐ More than 10 years
5. **Department and Faculty**

   _____

6. **Position Grade**

   - ☐ Below 29
   - ☐ 41
   - ☐ 44
   - ☐ 45
   - ☐ 48
   - ☐ 51
   - ☐ 52
   - ☐ Above 54

*The following section B, C and D are asking your **OPINION** regarding training to assessors, accuracy, discriminant validity, and interrater reliability.*

<u>SECTION B:</u>

Please indicate your level of agreement or disagreement with the following statements.

1=Strongly Disagree 2=Disagree 3=Neutral 4=Agree 5=Strongly Agree

B1: ASSESSOR TRAINING STRATEGIES

The term assessor training indicates the exercise to the person who involved during selection process.

B1.1	**ASSESSOR TRAINING**					
1	Better performance may be achieved through assessor training.	1	2	3	4	5
2	Through effective training, skills of an assessor will be achieved.	1	2	3	4	5
3	Before publishing an advertisement of, assessors must be familiarized with standardized criteria by training.	1	2	3	4	5
4	Assessor training should be provided in UMP	1	2	3	4	5
B1.2	**FRAME OF REFERENCE TRAINING**					
1	We should be given training that train assessors with respect to common evaluative standards.	1	2	3	4	5
2	A trainer should train us on how to assign ratings for an evaluation by using standard criteria.	1	2	3	4	5
3	A trainer should have training that discuss various employee behaviors to improve ratings.	1	2	3	4	5
4	All the assessors should attend training on making ratings using new standard dimensions.	1	2	3	4	5
5	We believe that assessors tend to be more confident after completion of training programs with the performance standard.	1	2	3	4	5

B1.3	**REDUCING ERROR TRAINING** Training provided by viewing examples of common errors potentially made by assessors.					
1	I feel like only familiarizing with common error is enough to have correct ratings.	1	2	3	4	5
2	When assessors realized that they did some mistakes in the selection process, top management encourages them to share what they have learned with other assessors.	1	2	3	4	5
3	I don't need this training to reduce my unintentionally rating error.	1	2	3	4	5
4	By this training, I will be alert to different errors only and I will be a good assessor.	1	2	3	4	5
5	Assessors can improve the assessment ratings by having reducing error training only.	1	2	3	4	5

SECTION C:

Please indicate your level of agreement or disagreement with the following statements.

1=Strongly Disagree 2=Disagree 3=Neutral 4=Agree 5=Strongly Agree

C1	**EVALUATION OF ACCURACY** Ability to give actual ratings to candidates.					
1	The frame of reference training will help very much in rating accuracy.	1	2	3	4	5
2	More training needed to improve my accuracy in assessment.	1	2	3	4	5
3	The accuracy of ratings depends on experiences with the standard performance criteria.	1	2	3	4	5
4	Assessors with less training may give less accurate ratings.	1	2	3	4	5

C2	**EVALUATION OF DISCRIMINANT VALIDITY** Ability to differentiate positive and negative ratings of candidates according to valid criteria.					
1	The frame of reference training should be provided to assessors to improve our discriminant validity skills.	1	2	3	4	5
2	Discriminant validity will affect overall rating scales.	1	2	3	4	5
3	I was unable to differentiate between positive and negative behaviors of candidates without better training.	1	2	3	4	5
4	There are no such proven training that support if my ratings are good enough.	1	2	3	4	5
5	I frequently feel my ratings are will be improved if I am familiarized with standard performance criteria.	1	2	3	4	5

C3	**EVALUATION OF INTERRATER RELIABILITY** Ability to give consistent ratings between assessors.					
1	Consistent rating scales can be obtained with frame of reference training.	1	2	3	4	5
2	A standard reference helped assessors to give higher between assessors.	1	2	3	4	5
3	During the selection process, my ratings sometimes differ than the other assessors.	1	2	3	4	5
4	I need to have the training to improve reliability.	1	2	3	4	5

Thank you for your cooperation.

APPENDIX B - GANTT CHART FOR FINAL YEAR PROJECT 1 & 2

No.	**Research activity**	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	Briefing by FYP coordinator														
2	Meeting with supervisor, discussing project title and objective														
3	Deciding the topic and objectives														
4	Getting supervisor's approval of the topic and objectives														
5	Preparing project research proposal														
6	Process of chapter 1,2,3														
7	Submitting draft chapter 1,2,3														
8	Correcting and editing 1,2,3														
9	Preparing cover page, content, preference list, questionnaire														
10	Submitting full report of FYP 1														

11	Preparing slide for oral presentation														
12	Presenting the FYP 1														

No.	**Research activity**	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	Data Collection														
2	CHAPTER 4 Introduction														
3	Reliability Analysis														
4	Normality Analysis														
5	Descriptive Analysis														
6	Mean Analysis														
7	Correlation Analysis														
8	Summary and Discussion of Findings														
9	CHAPTER 5 Introduction														
10	Contribution and Limitation														
11	Recommendation and Conclusion														
12	Submit draft report														
13	Submit final report														