# Using The Andrich Rating Scale Model (ARSM) to build a Scale for the Academic Proficiency Among Cairo University Students Psychometric Study[1]

**Prepared by**

**Amr Mohamed Ibrahim Youssef[2]**

**Lecturer of educational psychology – Faculty of Graduates for Education - Cairo University**

## Abstract

The current research aims to prepare a tool to measure the academic proficiency level of Cairo University students, so that it has acceptable psychometric properties, where the tool was applied to students of the Graduate Education - Faculty of Physiotherapy - Faculty of Commerce at Cairo University, according to Andrich's rating scale according to the first dimension of the response theory to the individual. Where it was applied to the number of 577 male and female students, where 100 students were deleted so that the number was 477, where the validity of the tool was confirmed through the traditional and modern method of measurement so that the tool in its final form consisted of 48 items distributed over five dimensions, and the R program and the program were used Winsteps in conducting analysis.

**Keyword's:**

(Academic Proficiency, Andrich Rating Scale Model, Item Response Theory)

## Introduction:

Due to the rapid and successive change of the scientific and technological developments we witness today in our society, which is a feature of our age nowadays, proficiency is one of the most important concepts that should be developed in every individual to be able to contribute to this rapid change and development. The knowledge revolution has become the main feature of the modern age, and the progress of nations is measured by the size and amount of information and knowledge they have, and even how this information is used to achieve academic proficiency. Proficiency is the targeted objective to be achieved at all levels by all educational institutions.

Academic Proficiency concept started by (Chen et al., 1997)to describe the full set of skills, attitudes, and behaviors necessary for academic

---

[1] تم استلام البحث في ٢٠٢٢/٣/١٨ وتقرر صلاحيته للنشر في ٢٠٢٢/٤/٢٤

[2] ت : ٠١١١١١٥٢٥٥٨ Email:dr_amr@cu.edu.eg

success. However, the components that make up the idea of " proficiency" vary among academic instruments and institutions.(Caux et al., 2017), states that academic proficiency is a multidimensional term that contains skills, attitudes, and behavior of the learners that contribute to the academic success, they both agree that academic proficiency and academic ability are both synonymous.

Academic Proficiency is a building that refers to skills, Attitudes and Behaviors that enable the learner to achieve academic success, it also includes the skills of (reading, writing, calculation, problem solving) required for academic success. Some researchers use the term academic Proficiency to mean the academic performance or academic ability (Fortune & Song, 2016; Friedman & Kagan, 2017). It is also defined as a set of skills that help students to employ their abilities and knowledge in an organised manner to reach a distinguished level of achievement. The components of academic Proficiency may vary according to the instrument used by the individual or institution (Gebril & Plakans, 2016). This was agreed upon by (Geide-Stevenson, 2018).

The researchers differed in determining the components of academic proficiency according to the instrument designed for measurement, the point of view of each researcher, as well as the theoretical trends to which he belongs, however, there are those who identified it in two dimensions, three, four, and five components, which differs from a researcher to another. This shows the complexity defining the academic proficiency and the multiplicity of its definitions, thus resulting in complexity in determining its dimensions (Gorzycki et al., 2016).

From the perspectives of the researchers, the components of academic proficiency were identified in five components, as stated in the research of (Gottlieb, 2016; Green, 2020), which consist of: "Effective recall skill, self-management skill for learning, managing and organizing time skill, withstanding academic pressures skill, and the effectiveness of the academic self-esteem." These components include all the dimensions mentioned in both foreign and Arab research, as they illustrate the concept of academic proficiency in a good and comprehensive manner that includes all its aspects which contribute to the appropriate, objective and accurate construction of a measuring instrument for academic proficiency.

A lot of research on academic proficiency have been carried out seeking to find measures, but this research has not provided us with a scale of objective psychometric characteristics. There was therefore a need to find effective scale to measure the academic proficiency that conformed to objective measurement

standards with high psychometric characteristics, which arise the confidence when used to measure academic proficiency among students (Haim, 2018).

The research used the academic proficiency scale or those that sought to build a scale relied on criteria derived from the classical Test Theory (CTT), some are concerned with the item statistics, but this theory was criticized because of its inability to interpret certain issues in psychological metrics, Item Response Theory (IRT) emerged as a result of important developments in psychological and pedagogical measurement theories, models and applications, which had a significant impact on the construction and development of measurement methods and instruments. This new approach has received the attention of researchers because it overcomes many traditional measurement problems, related to the statistics of both the item and the individual. Based on criteria related to Item Response Theory (IRT) has therefore been constructed (Abu Jarad, 2016; Hosseinpour et al., 2019; Hu & Trenkic, 2021).

The theory of Item response is a framework for the current and future trend in the construction and development of scales because of their great effective indicators in terms of test construction, correction and analysis compared to traditional theory. This theory aims to determining the relationship between an individual's performance and the capabilities lie behind such performance. It responds to objective measurement requirements, namely, editing grading the measuring instruments from individual's characteristics, and editing the assessment of individuals' ability from item's characteristics (Aizawa et al., 2020; Ambiel et al., 2015)

This theory assumes features called abilities that lie behind an individual's test performance, where an individual's ability to do the test and his score on the measured features can be predicted from the test, and since these features cannot be directly observed or measured, they are called underlying features (Bradley & Massof, 2018) .It is also based on a set of basic assumptions: (one-dimensional, objective independence, the distinctive curve of the item). It also results in a set of models used to construct the tests, through which statistical indicators of the item can be obtained that do not depend on individuals' characteristics and estimations, neither on the difficulty of the items of the scale(Cadoret et al., 2018; Caux et al., 2017). Models of Item response theory have multiplied, including the one parameter Logistic Model or the so-called Rasch Model. The models developed from Rasch Model have also varied to suit different types of data (J. W. Lang & L. Tay, 2021).

It can be categorized according to the scale items grading into three categories, the first: dichotomous, used when the response to the item takes

binary values (0,1), and the second: polytomous, in which the item's graded response is, for example, the response to some questionnaires and evaluation scales, and the third: related to the continuous response, it is an extension of the multiple response. One of the multiple response models is the "(Andrich Rating Scale Model) (ARSM)(Carrozzino et al., 2021), which is used with data from the rating scale. If the items match this model, the parameters of the items, represented in its difficulty degree, can be assessed independently of the abilities of individuals. Individuals' capabilities can also be assessed independently of the difficulty degree of the items (J. W. Lang & L. Tay, 2021).

The Rating Scale Model is the most appropriate model of Item Response Theory to the academic proficiency scale of university students, where it is used to rate scales with continuous rating. and has been set particularly for the Lykert attitude scale pattern, where the grade value per an item reflects the location of the item on the directional link(Sedoc & Ungar, 2020).

This model was used to analyze data to make sure of its matching because of its suitability for the used grading for constructing the used scale. In particular, the interest in this model is focused on testing the difficulty factors for each response category (Gottlieb, 2016; Sedoc & Ungar, 2020) To achieve an objective scale instrument, it has high psychometric characteristics through which we can assess the proficiency and competency of the university student academically, as well as to reveal most of those factors associated with academic proficiency.

**Research Problem**

The problem of the research was the need for an accurate scale instrument aimed at assessing the academic proficiency of the university student, with good and acceptable psychometric characteristics, and objective scale characteristics. The scale instruments constructed to measure the academic proficiency were procedures whose psychometric characteristics were verified in accordance with classical Rating Scale Theory, which failed to explain some important issues in psychological measurement. Therefore, modern measurement theory (Item Response Theory) has been relied upon in the construction of this instrument, not to mention that the theory of Item Response receives great interest in foreign research and has not received the same attention in Arab research for constructing psychological and educational standards in general, despite its objective advantages in constructing measuring instruments.

**The current research problem can be formulated in the following questions:**

1. What is the degree to which the responses of the academic proficiency scale items match the Andrich Rating Scale Model (ARSM) among Cairo University students?
2. What are the edited values of individual capabilities and the difficulty of the Items resulting from applying the scale of academic proficiency of Cairo University students according to the Andrich Rating Scale Model (ARSM)?
3. What are the psychometric characteristics of the items of the academic proficiency scale edited from individuals and items according to the Andrich Rating Scale Model (ARSM) among Cairo University students?

**Research Objectives:**

**The current research aims to:**

1. Constructing an instrument to measure academic proficiency among undergraduate students according to the " Andrich Rating Scale Model (ARSM) resulting from Rash as one of the models of Item Response theory, with high psychometric characteristics.
2. Introducing some of the programs used to analyze the items of the scale under the Item Response Theory such as (IRTPRO & R).

**Significance of the research:**

**The importance of current research lies in:**

1. To set an objective and progressive scale instrument using one of the models of Item Response Theory, with high psychometric characteristics, which is trustworthy when used to measure academic proficiency, therefore this scale has a special educational value, and in the light of the measurement results are planned for the development of guidance and educational programs, lectures or seminars that can contribute to raising the level of academic proficiency among university students and other educational stages.
2. This research is in line with the modern and contemporary attitude in the field of psychological and educational measurement, where the Andrich Rating Scale Model (ARSM) arising from the Rash Model is used in the graded scale of academic proficiency among undergraduate students. It provides Scale Rating instruments -

Particularly Questionnaires - with good psychometric characteristics and reliable results that encourage researchers to use them.

3. Demonstrate the importance of some of the modern programs used in the analysis of the items of psychological and educational Rating Scales within the framework of the Items Response Theory, including: (IRT PRO-R).

**Research Terms**

**1. Academic Proficiency**

It is defined as a set of skills that a student must have, which help him to employ his or her personal abilities, knowledge and self-suitability to overcome the academic problems, in order to achieve academic success in performing various study tasks, and to reach a distinguished level of education. It is measured with the degree to which the student receives in the scale of academic proficiency and its different dimensions which are (Effective recall skill, self-organization of learning, time management skill, withstanding academic pressures skill, and academic self-effectiveness) (Fortune & Song, 2016).

**2. Item Response Theory**

It is defined as a modern and contemporary attitude in psychological and pedagogical measurement, it assumes that an individual's performance on a test can be interpreted in the light of a feature that distinguishes this performance called the characteristic, which is observed and measured indirectly through an individual's answers to a sample of the test item (Sedoc & Ungar, 2020).

**3. Rasch Model**

It is the simplest model of one-dimensional item response theory, a mono barometer logarithmic model, which aims to objectively measuring of the behavior, addressing the difficulty of the item and assumes an equal distinguishing force between each item. It is used to analyze data from test items based on binary responses (J. W. Lang & L. Tay, 2021).

**4. Andrich Rating Scale Model**

It is one of the models of the item response theory emanating from the Rasch model, it is a comprehensive feature model, used in the analysis of multi-step questionnaire item, with gradients separated by equal distances according to the graded Lykert scale.

**Theoretical framework:**

**First: " Academic Proficiency"**

As a result of the rapid and successive change, which has become a feature of the age in which we live, proficiency is one of the most important concepts that an individual should develop and possess to be able to contribute to this rapid change and development, knowledge has become the key feature of modern times, in which nations' progress is measured by the size and amount of information and knowledge they have, and even how this information is used to access academic proficiency. Proficiency is the main objective and the means to be reached at all levels by all educational institutions. This has made predictive studies one of the most important issues of concern in the educational institutions, and work to uncover the variables that can contribute to this competency (Green, 2020)

(Friedman & Kagan, 2017; Geide-Stevenson, 2018) state that the fields in which an individual can achieve proficiency vary, such as :(behavioral, social, academic, developmental areas), the term is mainly applied to an individual who reflects a significant or clear achievement in one or more fields having the ability to continue to succeed in the future, stated that proficiency has various components or fields, including: (academic, social, home, sports, linguistic, personal... etc). The current research is concerned with academic proficiency because of its association with the academic aspect of undergraduate students (Haim, 2018) .

**The concept of academic proficiency**

The concept of academic proficiency came out by (Ambiel et al., 2015; Bergstrom & Lunz, 1998), to describe the full range of skills and behaviors necessary for academic success, yet the components that make up the idea of "Proficiency" vary among academic instruments and institutions.

Although academic proficiency has been presented in a lot of research as a result scale, the definition of the concept itself remains at least vague and inconsistent. This was explained by the fact that academic proficiency values reflected both the student's performance and the criterion used to assess that performance. Many researchers have not clearly defined the concept, yet they have used academic proficiency in exchange with other models such as academic performance and academic ability (Bradley & Massof, 2018).

The concept of academic proficiency varies according to the views of researchers and the theoretical orientations to which they belong, and education researchers have provided multiple definitions of academic

proficiency. A set of definitions of academic proficiency will therefore be presented according to the opinions and views of researchers, and then come up with a comprehensive definition of it as much as possible. As well as identifying its most important components. Among these definitions, they are a set of study-accepted results that include: (interest in schoolwork, independent learner behaviour, and reflecting a desire to achieve positive designs for the study work(Cadoret et al., 2018). It is also defined as the student's ability to achieve at school, from the point of view of students and teachers(Carrozzino et al., 2021).

It has been defined by many researchers such as (Bergstrom & Lunz, 1998; Bradley & Massof, 2018; Cadoret et al., 2018; Carrozzino et al., 2021; Caux et al., 2017; Chen et al., 1997) as a multidimensional concept consisting of learner skills, trends, attitudes, and behaviors that contribute to teachers' gradings of academic performance or academic competence, and therefore the academic success, They agree that academic proficiency and academic ability are both synonymous.

(Curle et al., 2020; Dev & Qiqieh, 2016)explained that academic proficiency is a set of skills for students that they can perform with the aim of achieving a distinct level of education, improving coping methods, and addressing demands that may be a heavy burden on them, helping to achieve successful performance of academic tasks, It also expresses students' confidence in their ability to succeed in all study tasks, the ability to work hard and enjoy the efficient use of time and organizational skills for academic work, and the good investment of skills and perseverance to achieve the goals.

The previous presentation of definitions of academic proficiency shows the differences in the researchers' views on their definition according to their theoretical orientations. Consequently, their definitions have varied. A comprehensive definition of them can therefore be introduced that combines their different opinions, where academic proficiency is defined as: "A set of skills that a student must have, which helps him to employ his or her abilities, knowledge and personal willingness to overcome academic problems, in order to achieve academic success in performing various study tasks, and to reach a distinguished achievement level" (Dimova, 2017).

**Components of academic proficiency:**

The researchers differed in determining the components of academic competence according to the theoretical orientations to which they belong, some of whom identified four components through their theoretical analysis of the concept of academic proficiency in the light of the psychological

heritage and considered that academic proficiency is an intermediate variable of concepts (motivation for achievement- academic cooperation - academic competition - academic compatibility)(Curle et al., 2020). Also, through the results of the working analysis of students' responses to some psychological measures, it was found that academic competence consisted of four dimensions, which, as he recalls (Dev & Qiqieh, 2016):( The motivation for achievement- academic cooperation, academic competition- academic compatibility) these four dimensions, including academic (study) proficiency for students.

(Erath et al., 2018) argues that the complexity of identifying the components of academic proficiency complicates their definition and includes a wide range of skills that contribute to academic success. It contains three components: (academic skills, academic self-effectiveness, problem-solving skills).

(Friedman & Kagan, 2017; Gebril & Plakans, 2016), explained that academic proficiency is a multidimensional concept consisting of learner skills, attitude and behaviors that contribute to teachers' appreciation of academic performance, combining experimental and theoretical study with a range of components that contribute to academic competence: (academic skills, study skills, academic motivation, personality skills, academic self-concept). Using the factor working analysis of the importance estimates provided by teachers on these five components, the results resulted in five components that contribute to academic mastery and do not directly correspond to the assumed components of the concept's first model: (academic skills, study skills, academic motivation, social skills, academic participation), Although four of them reflect the direct compatibility with four of the five components, the fourth component (academic self-concept) was not supported by the results of this study. One possible explanation is that the concept of self, unlike the other four fields, is essentially internal phenomena, and it is difficult for teachers to assess its stability. To be able to evaluate a concept such as the concept of 0neself, it requires the use of visible behaviours that reflect interesting internal phenomena. It was found that a large number of items of the concept of self-concept included positive participation behaviors such as (asking questions, volunteering to answer- taking leadership in collective situations), because these behaviors can be observed, and have been assumed to reflect the student's confidence in his academic skills. They can be assessed more accurately than the concept of oneself. Thus, although the concept of self can contribute to academic competence, it can be more accurate (and practical) in assessing the academic concept of self through self-esteem rather than teacher reporting.

These results are consistent with the findings of research by (Luo & Jiao, 2018) through the factor analysis that also resulted in academic competence being a multidimensional concept consisting of learner-oriented skills, attitude and behaviors, which teachers identify as important for academic success, and have resulted in five dimensions: academic skills, academic motivation, study skills, social skills, and academic participation.

Then came the research of (Lawrence et al., 2021; Levander, 2020; McSweeney, 2017; Merwe, 2018; Mezzadri, 2018; Molla & Muche, 2018; Nasseri, 2021), combining these previous dimensions, she explained that the skills, trends and behaviors that contribute to academic proficiency lie in two fields, They include: (reading, mathematics, language skills, critical thinking), and the second dimension: (academic "appointees" possibilities) and include: (study skills, academic motivation, personality skills, academic participation).

Although (Linacre, 1999, 2010; Luo & Jiao, 2018) agreed with previous research in determining academic competence in two key dimensions: (academic skills- and academic aids), but added another dimension , (teaching and learning strategies), to measure both the cognitive and emotional aspects of academic competence, which include: (learning and achievement expectations - management and organization of time - information processing - motivation - self-monitoring - anxiety – Homework and school work - rewards and results - evaluation).

It is clear from what has already been presented in previous foreign research on the components of academic proficiency that its components are determined in two main dimensions: (academic skills- and academic aids).

As a complement to identifying the components of academic proficiency, (Macdonald et al., 2018) identified four components of academic competence, which he called "learning competence", which are: (perseverance, integration, participation, focus).

On the other hand, (Nasseri, 2021) identified five other components of academic proficiency: (recollection skill - self-organization skill for learning - skill in managing and organizing time - withstanding academic pressures skill - the effectiveness of the academic self). Academic competence was identified in three components: (Effective Recall Skill - Time Management Skill - Academic Pressure Management Skill).

Finally, in the light of what has been presented earlier on the identification of the components of academic proficiency, the researcher sees the difference of researchers in determining their components according to the

tool designed for measurement, and the point of view of each researcher, as well as the theoretical orientations to which he belongs, where there are those who identified them in two dimensions, three, four, and five components, they differed from researcher to researcher. This demonstrates the complexity of the definition of academic proficiency and the multiplicity of definitions, and therefore results in complexity in determining its components. It is a multidimensional concept consisting of skills, attitude and behaviors that contribute to teachers' assessments of learners in achieving academic success within the classroom (Neumann et al., 2019).

Thus, the components of academic proficiency can be identified by the researchers' previously presented views as five components, as stated in the research of (Macdonald et al., 2018, 2020; Nyarko et al., 2018), which is: (Effective Recall Skill - Self-Organization Learning Skill - Time Management Skill - Management Skill carrying academic pressures - Academic Self-Effectiveness), where these components are inclusive of all previous components. that have been mentioned both in foreign and Arab research, it demonstrates the concept of academic proficiency well and comprehensively in all its aspects and contributes to the appropriate, objective, and accurate construction of a measuring tool for academic proficiency. These five components can be explained as follows:

**The first dimension:** effective recall skill: defined as a set of qualitative methods, methods and performances used by the student as he learns to organize and accomplish academic tasks and gain knowledge, to achieve academic success such as: (acquiring new behaviors, registration, organization, installation, remembering, summarizing, employing information, solving problems, creating new ideas, criticism, analysis....) These skills can be acquired, learned, and modified at different age stages (Caux et al., 2017).

**The second dimension:** Self-organizing learning skill: defined as a constructive and active process performed by the student using some strategies such as: (setting goals, planning, organizing, monitoring, and adjusting his knowledge, motivation, and behaviors...) Which helps him organize and control his actions, emotions, and ideas in a planned way to achieve the best level of performance, accomplish the tasks of learning, and achieve his academic goals accurately and efficiently (Curle et al., 2020).

**The third dimension:** Time Management and Organization Skill: Defined as the optimal use of time by the student by performing a range of behaviors such as: (recording and analyzing time, setting goals and priorities, planning,

organization, monitoring, evaluation....) To achieve effectiveness and efficiency in the performance of some activities to be achieved within a specific time frame, they can be acquired or learned (Friedman & Kagan, 2017).

**The fourth dimension:** The skill of managing and withstanding academic pressures: defined as a response and compatibility of the student to the stressful attitudes and problems he faces that are related to the academic aspects within and outside the university environment represented in: (time pressures with many study tasks, activities and research ,discussion and sometimes unsatisfactory performance, high family and teacher expectations, examinations and evaluation....) in order to overcome them, and to achieve success in his school life (Gebril & Plakans, 2016).

**The fifth dimension:** Academic Self-Effectiveness: Defined as the student's beliefs and convictions in his own potential, information, confidence in his ability to perform academic tasks successfully and efficiently, and with a certain level of performance and proficiency during his studies to achieve his learning goals (Geide-Stevenson, 2018).

**Measuring academic proficiency:** Due to the increasing need to improve academic proficiency day by day among students, the demand for appropriate academic proficiency assessment has increased, as this assessment can facilitate the preparation of therapeutic strategies for these students with problems in developing their academic proficiency(Gottlieb, 2016; Haim, 2018).

After reading the previous research, the theoretical literature and what has been written about academic proficiency, a set of previous scales used to measure academic proficiency has been compiled, but we need instruments with a higher level of consistency and honesty within the limits of predicting academic success, all of which were built in the light of traditional theory of scales. The stability of these scales was calculated using alpha kronbach coefficient, and half-segmentation. Their authenticity was calculated using link transactions, content reliability, and factor reliability only (Grewe et al., 2021; Hosseinpour et al., 2019).

As a result, psychologists were interested in achieving the objectivity of measurement for psychological tests and scales, in order to get values of individuals independently of the impact of the research sample, values of the items not affected by the test items, characteristics and conditions of the implementation, and to achieve high accuracy in measuring the psychometry characteristics of the tests and psychological scales, so that those

characteristics are not affected by the sample or the length of the test, so that the test can be generalized and applied to another group, the results are not limited to the sample of the study only as in the Traditional Scale Theory.(Levander et al., 2019).

According to the importance of measuring academic proficiency among students, there is a need to deal with them in the light of the theory of Item Response, and there is a need for a listed objective measure that gives accurate and consistent values, and therefore the academic proficiency Rating Scale among the university students will be built and graded using the item response theory according to The Andrich Rating Scale Model (ARSM), which is one of the models of item response theory that is suitable for granulized items of multi- graded (Cordier et al., 2019).

Since there is no Arab research, particularly concerned with constructing the academic proficiency Rating Scale among university students using the theory of item response according to the Andrich Rating Scale Model (ARSM), the scale suffers from the problem of the lack of proper staging, which is the most appropriate example of the item response theory for academic proficiency, as Rasch's model is not appropriate to its analysis, because it is only appropriate for bi-grade items, i.e. the items for which the answer is either (yes or no). The scale of academic proficiency is a five-step measure, where the individual responds on each individual by choosing an alternative of five alternatives, as the response on each individual ranges from (1-5), and the use of the Andrich Rating Scale has been the result of Because the Rasch model is unable to analyze and granulize the items with multiple staging, it also relies on a set of assumptions aimed at achieving the objectivity of measurement, these assumptions are: (one-dimensional - objective independence - the distinctive curve of the item - lack of effect of guesswork - freedom from Speed - linear measurement) (Crowe et al., 2018; Davier et al., 2019).

Based on the significant progress made by modern theory in measurement, in freedom from the impact of individual characteristics on vocabulary features, and from the impact of vocabulary features on the characteristics of individuals when building tests, i.e., tests based on the foundations of this theory, and associated psychometry concepts such as: (features of difficulty, features of discrimination, and parameters of guesswork), do not vary depending on the characteristics of the sample members used to calculate these features. There has been a multiplicity of previous research, particularly foreign research on the use of Item Response Theory in constructing psychological and psychological Rating Scales (Dinić & Raine, 2019; Dougherty et al., 2021).

**Second: "Andrich's Rating Scale Model"**

Models of the item response theory have varied, including the one-parameter logarithmic model or the so-called Rasch model, and the models developed from rush model have multiplied to suit different types of data, including the "Andrich's Rating Scale Model" which is used with data from the appreciation ladder (Davier et al., 2019).

Andrich developed Rasch 's model to Rating Scales that allow us to analyze both tests and questionnaires using a standardized Rating Scale Theory. The Rating Scale variable map helps us visualize the concept and how it is defined by the questionnaire items and can become a very useful instrument in achieving reliability and increase understanding of the concept (Cordier et al., 2019).

Rating Scales require an individual to respond to their choices from a ranked series of categories, and under Item Response Theory, the Rating Scale Model (Andrich, 1978), an extension of the Rasch model commonly used to match the response to the item of the Rating Scale (Cotter et al., 2021; Crowe et al., 2018).

The Rating Scale model is the growth of multi-binary item responses in the same response format, which means that each item has the same and number of responses options. In that model for each item had a barometer for the location of the item (difficulty), as well as parameters with mobile locations (thresholds) that are equal for all items (Clark et al., 2020).

Rasch's multi-binary model was drawn by Andrich after the derivations of Rasch (Rasch, 1961) and Andrich by solving the items appropriate to the overall form of Rasch's model into three thresholds and parameters of discrimination. When deriving the model, Andrich focused on using Lykert Rating Scale in psychometrist for both illustration purposes and helping to interpret the model. The model is sometimes referred to as the grade scale model when the individual (i) has the same number of thresholds and (ii) and in turn, the difference between any threshold location and the average threshold location is equal to or uniform between items. That is, however, a name that can be tricky for the model is far from generalized in an application of so-called rating scales (Cordier et al., 2019; Crowe et al., 2018).

The original conceptual formulation of the Rasch model was dramatically expanded by David Andrich, 1978, when it was suggested that responses to the Lykert method questionnaire items could be arranged and used in a similar way to infer the amount of direction or psychological character of the

respondent. In the Rasch-Andrich Estimate scale model, the conceptual formulation of the individual's difficulty was reformulated to become resistance to approval of the Rating Scale Response category (Gomez et al., 2019).

Andrich (1988) introduced a general formula for expanding Rasch Model, called extended Logarithmic Model (ELM), to suit graded responses, which serves to study reliability within subsets of bi-grade test items, as well as to study attitude questionnaires measured according to the Likert Scale scale, 2005).

The use of the Rating Scale model was the result of Rasch's inability to analyze and gradualism multi-step vocabulary; the grade scale model gradually phases the multi-step items and relies on a set of assumptions aimed at achieving the objectivity of measurement, these assumptions are: (One-dimensional- local independence - singular property curve - freedom from speed)(Hays et al., 2021).

This model is for multi-response tests and assumes that the gradual ness of response categories must be equal in the sense that the values of these categories must increase steadily as they deal with tests construction in a Lykert manner (Cordier et al., 2019)

It is also called the Rasch Multi-Binary Model, a generalization of the Rasch binary model, a measurement model that can be applied in any context, in which the aim is to measure attribute or ability through a process in which Item Response scores are estimated in consecutive correct numbers. For example, the model can be applied in the use of Lykert Scale, assessment scale and educational evaluation of the item that are meant to indicate increased efficiency or attainment levels (Cotter et al., 2021; Crowe et al., 2018).

The primary purpose of developing the Rating Scale Model is to provide researchers with a way to analyze attitude response data using the Item Response Theory. Andrich has expanded the use of the Rasch model in the binarily rating item to include other cases where there are more than two response options available for an individual (Cordier et al., 2019).

(Davier et al., 2019), defines the Model of Rating as a mathematical model used to analyze the item of multi-graded resolutions. As well as (Chiesi et al., 2018; Cordier et al., 2019) as one of the models of modern theory emanating from Rasch Model, it takes the form of multiple responses in gradually separated by equal distances.

(Choi & Asilkalkan, 2019; Clark et al., 2020; Colledani et al., 2018; Cordier et al., 2019) stated that Andrich's Rating Scale Model is a

development of Rasch Model, which is used with data from the Rating Scale. If the item matches this model, the parameters of the item, which are difficult, can be assessed independently of the abilities of individuals. Also, the Individual capabilities can also be assessed independently of the difficulty of the item.

According to (Cordier et al., 2019), the Rating Scale Model is a mono-latent feature, taking the form of polytomous multi-responses in scales separated by equal distances, and defines a set of items that shares the structure of the Rating Scale, where the same response alternatives are chosen for each item, compared to the partial Rating Scale Model that sets each item with its own Rating Scale. (Ágoston et al., 2018)states that the Rating Scale Model is characterized by thresholds that express the boundaries between steps and are fixed through items (Adams et al., 2019).

The idea of the Rating Scale Model is that each item carries an emotional charge that, together with other items, contributes to the composition of a total emotional charge that reflects the individual's attitude in accordance with his or her appreciation of those items, and the model estimates this charge for each item according to the probability statistical function adopted by the model (Ayala, 2018).

An important characteristic of the Rating Scale Model is that it provides means of estimating the stability of an items' measurements in items thresholds and that single parameters are fixed between individuals' samples (i.e. parameter stability). In this current context, a consistent assessment provides a means of determining the extent to which items rating or individual items parameters show changes in rating events. The standard differences of the statistical community or the items set that correspond to the Rating Scale Model have an expected value of (zero) and a standard deviation expected from (1). It is noted that significant data from these expected values indicating more or less consistent values over time than expected (Chen & Ahn, 2020; Cheng et al., 2019) The Andrich Rating Scale Model differs from the partial value model in that the distance between difficulty steps (or levels) from one class to another in each item is the same as the distances in each item (Chen & Ahn, 2020).

In other words, the Rating Scale Model shares the items group in the same Rating scale structure, offering the same response categories to respondents for each item, on contrary, each single item in the partial Value Model has a structure in its own Rating Scale (Bonifay, 2019).

An important characteristic of Andrich's Rating Scale Model is that it allows an individual to assess the stability of the item parameters in all individuals' samples or the stability of an individual's measurements of the items' samples (i.e., determining the stability of barometer values). This feature is useful when comparing two groups of individuals who respond to the same sets of items or equalize two tests (each consisting of different items) taken separately from both groups by one group of individuals. In the current context, stability assessment is useful because it allows an individual to determine the stability of items' parameters and individual measurements between two measurement occasions. The stability of the two parameters estimates (≤≤1φ,2), obtained on two different occasions, is valued by examining the standard differences between values. The standard differences of the statistical community or the items set that correspond to the Rating Scale Model have an expected value (zero) and expected standard deviation (1). A significant deviation from the observed data from these expected values indicates that data are less stable over time is less than expected (Bonifay, 2019).

The Andrich Rating Scale Model is a well-known Rating model in the field of psychological rating and has proven to have many applications on attitude tests or collection tests. This model can be described as a Rasch Scale with more than two answer categories and class scores and is of equal dimensions. However, the Rating Scale Model has proven to be also useful in applications on sociology rating using question batteries that are supposed to be associated with the same direction or social feeling: (not feeling comfortable at home or at work), The underlying variable is called the interviewer, so in many cases the model can be a tool for measuring social evidence. The test score, the total weighted responses to questions, becomes a relative measure of this social guide. The purpose of applying the Rating Scale Model is to consolidate the relationship between separate value test scores and the continuously valued-latent variable (Blanchin et al., 2020).

Using Parscale to restore the barometer to the Rating Scale Model, it was found that, as expected, the sample size appeared to have little impact on the restoration of attribute parameters but affected the restoration of item parameters. The distribution of known attribute levels has a very impact on the restoration of item parameters. It was concluded that the Andrich Grade Ladder model allowed for the use of much smaller calibration samples than we originally recommended for other multi-binary IRT models(Bichi & Talib, 2018).

## .Using The Andrich Rating scale model (ARSM) to build a scale

The Andrich model can be expressed in linear logarithmic form as mentioned in (Linacre, 1990, 1999, 2010) as following:

(1) $\text{Log}\ ( P_{nij}/ P_{ni(j-1)})=B_n-D_i-f_j$ …….

(Pnij) is the possibility that an individual (n) faces the item (i) and is observed in category (j) of the order response group (S+M, j= s+1) so that the Rating Scale categories are calculated by (s,s+m) in an upward sequence of orderly numbers. For lucrative forced drafting, S=0 is here. Pni (*j*-1) is the possibility that an individual (n) faces the item (*i*) and the observation in the category (1-*j*). The (B&n) is the ability of the individual (n), (D*i*) is the difficulty of the item (*i*), (F*j*) is the roof of Rasch Andrich which is located at the equal probability point of the categories (1-*j*), (*j*) and group (Fj) called the Rating Scale structure (Alvarenga et al., 2020; Ayala, 2018).

The traditional thing is to identify ($\sum$ (fj)=0) and (*j* =1,m) so that the difficulty of the item is the point on the underlying variable in which the lowest and highest categories are modelled to be equally likely. In the previous equation, the limit ($F_0$) can arise and can be determined ($F_0$ =o) or any appropriate value because it cancels algebraicly. (F*j*) is modeled to be as independent as possible of both items and individuals. It can therefore be conceptually formulated with the structure of the Rating Scale in which each item is involved or as a response method shared by all individuals (Adams et al., 2019; Ágoston et al., 2018).

The Rasch-Andrich's Rating Scale Model is an additional linear model that describes the possibility that a person again responds to a specific Lykert item with a certain category of the Rating scale. The Mathematical Model of this probability contains three parameters: (individual's capacity – item's difficulty - and the difficulty of each step of the scale "i.e. the threshold between adjacent scale levels (Jj.,x-1, x)"). Standardization of the questionnaire data on this model leads to a separate rating of the parameter and standard error per person, item and each step in the context of ratin (Wolfe & Chiu, 1999).

In the Andrich Rating Scale Model, one value in the scale per an item is valued by the same measurement as the attribute barometer. Simultaneously, a set of scale response thresholds is fully assessed. The number of thresholds estimated in the scale is lower than the number of response categories, and Andrich defined the probability of responding in a particular category in

(X=0,1,,.....mi) per item (i) (French & Dodd, 1999).

The Mathematical Model of Andrich's Rating Scale Model is as mentioned by (French & Dodd, 1999; Wolfe & Chiu, 1999; Wolfe et al., 1999), as follows:

$$(2)\quad P_{ix}(\theta) = \frac{\exp \sum_{j=0}^{x}(\theta-(Bi+Jj)}{\sum_{r=0}^{mi} \exp \sum_{j=0}^{r}(\theta-(Bi+Jj)}$$

*when*

$$(3)\quad \sum_{j=0}^{0}\{(\theta - (Bi + Tj) = 0 \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

Where (Px) is likely to respond in a particular response class X to item i, and θ is the attribute level of a particular individual, The Bi is the scale value for the item i, and Tj is the threshold value for the targeted response. As noted, while not in the wording, the distinction of the items is assumed to be consistent in all items. Chart 3 describes the following: activation of the characteristic function of the Rating Scale Items with four response options, the scale value of this item (zero) and the threshold values of the scale are (0.8-), (0.8) (Adams et al., 2019).

In this model, Rumm software or Parscale programming can be used to estimate item's features and evaluating the potential of individuals (Ágoston et al., 2018; Aizawa et al., 2020; Alvarenga et al., 2020) states that all Rasch Scale Model analyses were performed in Winsteps program, Considering the previous research that used Andrich's Scale Model to construct its metrics, it was found that the programs used to analyze metric data were Winsteps, Bigsteps, Rumm 2020-Rumm 2030-Parscale, as well as SPSS to calculate the reliability and validity of scales in traditional ways, as well as to ensure a one-dimensional assumption of the model using analysis factor.

This model assumes that the number of values taken by the items contained in the Rating Scale is equal. When the number of values varies in different items, the estimation process is carried out on each set of items with the same number of values, at which point it is difficult to compare the degree of difficulty of the different items, but the derived ability from is not affected (Adams et al., 2019), The official expressions of the Andrich's Scale Model are presented by (Eckes & Baghaei, 2015) as follows:

$$(4)\quad P_{xni} = \frac{\exp[-\sum_{j=0}^{x}(Ti+x(\theta n-\delta i)])}{\sum_{k=0}^{m} \exp[-\sum_{j=0}^{k} Ti+K\,(\theta n-\delta i)])} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

In this function ($n\theta$) and ($\delta i$) are the individual (n) and item positions (i) respectively, and ($T_i$) is the step site ($J^{,th}$) in each item and (k) is the category. ($x=0, 1, \ldots..m$)

Andrich has developed Rating Scale Model for ranked items' data, which assumes thresholds for an equal items category, and step difficulties in item-order categories governed by a predetermined set of response categories that are repeated in all questions. Since the same alternatives to responses such as: (I don't agree at all, I don't agree, I'm hesitant, I agree, I totally agree) give to all items, it is assumed that the difficulties of the step do not change in the item. In other words, the dimension between (I totally agree) and (I agree) in all the vocabulary in the test is the same. That is, the increase in the level of structure as a result of the approval of (I fully agree), rather than (I agree) is equal in all vocabulary. However, the model does not require that the dimensions between (I do not agree at all), (i do not agree), (hesitant), (I agree), and (I fully agree) equal in one question and the level of increase in structure can be different when the respondent approves (I fully agree) Instead of (I agree) compared to him when he endorses (reluctant) instead of (I do not agree) (Baghaei, 2010).

**Previous Research:**

There are a number of research on the construction of objective psychological and educational Rating instruments according to Andrich's Rating Scale Model, including: Research (Ambiel et al., 2015), which aims to construct a scale for the attitudes of science teachers towards the laboratory. The initial image of the scale was built from (83) single according to the five-year lykert ladder. It was applied to a sample of (224) teachers, and the results indicated a single 58 matching of the assumptions of the Andrich Rating Scale Model emanating from the Rasch's Model. The scale has appropriate psychometric characteristics; the stability factor for the scale (0.98) is valued, the scale has multiple reliability functions, and the scale provides the greatest amount of information to individuals with medium capacity, with average capacity values (0.35), roughly equal to the average item's difficulty values (0.36) thus similar to the value expected by model.

The research by(Allega et al., 2018), which aims to employing the model of the Rating Scale of appreciation in selecting the items of a rating scale to assess the attitude of students of the Faculty of Educational Sciences in Jordan. The research sample consisted of (250) students, and the scale is 38 items, according to lykert's pentagram. The data was analyzed using SPSS and RUMM2020 software. The items' parameters were estimated, and a

statistical calculation matched each item to the Rating Scale Model. The statistical adoption of the item match, and the criteria for linking marks on the item to the overall mark, are based on the scale as criteria in the selection of each Rating Scale's Item.

(Ágoston et al., 2018), conducted a research with the aim of examining the effectiveness of the Rating Scale Model in analyzing the items of the study approach scale among students of the universities of Qatar and Minya, and the Rating Scale was applied to a sample of (600) students, and the results were that al the items of the scale is within the limits of internal and external conformity, the item stability factor (0.98) and the individual stability factor (0.75) and the value of the information function of the scale provides the greatest amount of information at high and low capacity levels.

Also, the research by (Adams et al., 2019), which aims to construct a scale of shyness among university students, and the initial image of the scale was formed of (97) items according to the five-year lykert scale and a sample of (526) students, the results showed that (45) items match the assumptions of Andrich Scale Model, and the scale has appropriate psychometric characteristics; as the stability factor for the scale was 0.97, the scale has multiple reliability aspects, and the scale provided maximum information value with the lowest standard error at average items difficulty, thus it is similar to the predicted value of the model.

(Aizawa et al., 2020), explored the relationship between student mathematics anxiety and achievement by applying the Rasch's Assessment Scale Model for data analysis. The research sample consisted of 79 ninth-grade students from a private school. The results indicate that the Rasch's Assessment Scale analysis provides a more consistent scale of student anxiety, and provides a practical, raw-grade conversion table for anxiety that matches Rasch's scale.

(Almaleki, 2021), aimd to construct an attitude scale for 10th grade students towards professional work, and the initial form of the scale was done on (96) items according to the five-step lykert Scale, it was also applied to a sample of (530) male/female students, the results showed that (46) items match the assumptions of the Andricsh Rating Scale Model, the scale was one-dimensional, and had appropriate psychometric characteristics; the scale also has multiple reliability functions, and the it provides maximum amount of information to individuals with medium capacity, where (individual's capacity = item's difficulty), this corresponds to model expectations.

(Carrozzino et al., 2021), which aims to use the "Andrich Rating Scale Model", in analyzing the data of the attitude scale towards life sciences which

consists of (66) items and the scale was applied to a sample of (680) students from the tenth-grade students in the Directorate of Education of Irbid. The results showed that the number of the items that matched the scale model were 37 items, the final stability factor of the scale was (0.99), and that the correlation factor between scale dimensions and total grade were all statistically significant.

Out of the previous research, the significant progress made by the Theory of Item Response scale while constructing tests and psychological and educational scales, in particular Andrich's Rating Scale Model emanating from the Rasch's model, which is one of the models of Item Response theory. All the results showed the accuracy and objectivity of the Item Response models in the preparation of psychological and educational tests and scales and measuring the level of performance of individuals despite the different quality of the tests and the standards used and their objectives.

Throughout a lot of foreign research that are interested in the field of the effectiveness of the use of the Item Response Theory and associated mathematical models in the construction and preparation of various types of measurement instruments used for the purposes of collecting different types of data, compared to the traditional theory of measurement. Research using the same theory in rating of different test items parameters has also confirmed the excellence of these instruments in terms of the objective characteristics in those instruments used in the field of physical measurement.

As a result of the success made by this theory as well as the associated mathematical models in achieving objectivity in measuring human behavior, and in the construction and preparation of tests, it has achieved tremendous development and progress in various fields of measurement, Many foreign and Arab research have tended to construct, develop or redevelopment some of the world-renowned and commonly used tests and standards in the light of recent measurement theory and statistical analyses of one of the associated mathematical models with the aim of achieving greater objectivity in their measurements.

Statistical analyses of mathematical models associated with this modern theory have included some statistical indicators that are used to verify the accuracy and objectivity of the results and various terms of the test measurement; These include (difficulty, discrimination, guesswork), which can be estimated for each test items as well as the test. The methods and formulas of estimates for these indicators vary depending on the analysis model used, with each of these models relying on variables different from those on which other models depend, which may lead to mixed results in

estimates of the values of those indicators, This may eventually lead to a different accuracy of estimating the capabilities that can be derived from responses to test items according to the different model of analysis used (mono-binary-triple...) parameters, because of the increasing and growing use of applications for the Item Response Theory, in recent times globally, particularly at the Arab level, in various fields of measurement that are interested in constructing and developing new and diverse data collection instruments or the redevelopment of commonly used instruments, as well as the establishment of question banks, An academic efficiency rating scale is to be constructed in the light of one of the models of this theory.

**Research procedures:**

**First: Research approach**

Based on the nature of the research, the objectives it seeks and the data to be obtained and based on the questions the current research sought to answer, the descriptive approach was used to conduct the current research.

**Second: Research sample**

The current research sample was derived in a randomly from undergraduate students from all grades with different specialization from the faculties of: (Graduate Education - Faculty of Physiotherapy - Faculty of Commerce) at Cairo University. Most of the previous research have dealt with undergraduate students. The research sample consisted of (577) students in the first semester of the 2021/2022 academic year, and the following table shows the numerical description of the research sample, their mean was 22.54, standard deviation 2.52.

**Table (2): Numerical sample characterization according to type and specialty n= 577.**

| Academic specialization | Type | | Total |
|---|---|---|---|
| | Male | female | |
| Scientific | 103 | 216 | 319 |
| Literary | 109 | 149 | 258 |
| Total | 212 | 365 | 577 |

**Third: Research Instruments:**

**Academic Proficiency Scale (APS)**

Although a large number of researchers are interested in measuring academic proficiency and having instruments for it, we need instruments with a higher level of stability and reliability within the limits of predicting

academic success, as all of these scales are built in the light of traditional measurement theory.

According to the importance of academic proficiency scale among students, there is a need to deal with it in the light of the theory of Item Response Theory, and there is a dire need for an objective scale that gives accurate and consistent estimates, and in the absence of Arab research that has been interested in constructing the academic proficiency scale among university students using the theory of Item Response, the scale suffers from the problem of lack of proper rating, and therefore the academic proficiency scale will be constructed and rated among university students using the theory of Item Response according to Andrich's Rating Scale Model, , which is the most suitable example of Item Response Theory, which is suitable for the rating of multi-rated items.

Rasch's model is not fit in this analysis, because it is only suitable for two-step item, which means the items with either (yes or no) answer. The scale of academic proficiency is a five-step scale, in which the individual responds to each item by choosing one alternative of five, so the response for each item ranges from (1-5), and the use of the Andrich Rating Scale Model was the result of Rasch's inability to analyze and rate the items. It also relies on a set of assumptions aimed at achieving the objectivity of scaling, which are: (one-dimensional - positional independence - the distinctive curve of individuality - freedom from speed).

The aim of the scale was to use it to identify the skills available to the university student, which helped to employ his or her personal abilities, knowledge and preparations to overcome academic problems, in order to achieve academic success in performing various study tasks and to reach a distinguished achievement level.

**The researcher followed the following procedures by building the scale according to Andrich Rating Scale Model**

**Step 1**: **Determining the dimensions of the academic proficiency scale:**

Identifying the dimensions that are the overall structure of academic proficiency is the key supporting step in constructing the scale, and a key focal point, and this is the most dangerous step in constructing the scale. Therefore, the researcher examined the previous research and scales that dealt with the academic proficiency Rating Scales, and since these Scales are not meet the purpose of the current research, their characteristics differ from those of the research sample, so the academic proficiency Rating Scale is designed to suit the characteristics and purpose of the research sample.

In the light of these sources, the dimensions of academic proficiency have been determined in five key skills: (effective recollection - self-regulation of learning - management and organization of time - management and tolerance of academic pressures - academic self-effectiveness). The relative weight of dimensions from previous research and scales has been determined by calculating the frequency of each dimension in these scales. Most of these scales have been found to have agreed on these five dimensions of academic proficiency.

These dimensions were then presented to a group of juries, faculty professors in the field of specialization, to determine to which extent these dimensions belong to the scale, and the ratio of agreement on the affiliation of these dimensions to the scale was very high.

**Step 2: Crafting the items of the scale:**

Showing the previous literature and research, understanding the theoretical framework and different definitions of dimensions and meaning in each dimension, as well as some scales of academic proficiency, both in the Arab and foreign environment, to determine a procedural definition of each dimension of academic proficiency, new items were formulated to suit the current dimensions, some items were used to suit the nature of the sample, some items were clearly modified and formulated, and the items were formulated appropriately to match specific procedural definition of each dimension of academic proficiency, and to which extent the individual belongs to the dimension. It was taken into consideration that the items were to be clear and specific in meaning, avoid long items. items containing more than one idea, and avoid the development of similar items, as the environment and culture of the research community were considered.

To determine the number of scale items and distribution below each dimension, the relative importance and weight of each dimension were determined, and the percentage calculated by the previous scales as in index (3), and the items were distributed under the five dimensions as follows: (30%: 15%: 20%: 20%: 15%), respectively. Accordingly, the number of items each dimension has been determined. 74 items reflecting the dimensions of academic proficiency have been formulated, and the number of items has been increased to take into account deletion so as not to affect the relative importance of each dimension, thus the number of scale items is (80).

**Step 3: Presentation to arbitrators (arbitrators' sincerity):**

The scale was presented in its initial form to a group of jury from the faculty professors in the Department of Educational Psychology, and they

were asked to express their opinions on the comprehensiveness of the scale of academic proficiency dimensions, the appropriateness of the specific procedural definition of each dimension, and the accuracy of the linguistic formulation of the items of the scale, and then modify the items as they see it either by deletion or in addition.

In the light of the opinions of the jury, the items no. (5, 20, 21) were deleted in the first subdivision (effective recall skill). Item number (12) in the second subdivision (self-organization learning skill), item number (14,15) in the third subdivision (time management and organization skill), item number (11, 16) in the fourth subdivision (academic pressure management and tolerance skill), and item number (6,11) in the fifth subdivision (academic self-effectiveness).

Following the completion of the arbitration, all the amendments agreed upon by most Jury were made in some of the items of the scale by deleting and modifying some. Most of the Jury also stated that the number of scale items had to be reduced in each dimension so that the respondent could provide reliable responses. Thus, based on the opinions of the Jury, the number of items was reduced by calculating the average number of academic proficiency items in previous scales, and the number of items was distributed to subdivisions in light of their relative weight in previous scales as well, Thus, the scale in its initial form is made up of (56) single.

**Determining How Grades are Valued**

The method of responding to the items of the academic proficiency scale was done by selecting one of the five alternatives: (apply perfectly - apply - apply - apply to some extent - do not apply - do not apply completely), and it takes values (5-4-3-2-1 The highest score a responding student can get on all the items of the scale is (280) degrees, while the lowest score can be (56) degrees, where the upper degree indicates a high level of academic proficiency, while the low score indicates a low score. to the low level of academic proficiency of the university student.

**Step 4: Checking the one-dimensional assumption of the data achieved on the scale:**

The scale was applied in its initial form of (56) items to a sample of (577) students from Cairo University, to verify the one-dimensional assumption of the responses of sample members on the scale, and this assumption represents the first assumptions to be made in any study using one-dimensional Item Response Models, where these models assume the item

ability to interpret an individual's performance in the scale, therefore called one-dimensional models.

A number of indicators by(Colledani et al., 2018), used by (Cordier et al., 2019; Cordier et al., 2018; Cotter et al., 2021)relied upon to indicate one-dimensional indicators:

1. Indices based on Answer Patterns: such as Guttman Retrieval Coefficient, so-called recovery factor, or (Reproducibility Coefficient), Index of Homogeneity (Index of Homogeneity) and (Green Index).
2. Indices based on Reliability: such as Alpha Cronbach stability coefficient (α Cronbach), Coder-Richardson coefficient, as a special case of Alpha Cronbach, the item correlation factor with the total item correlation, as well as the average inter-Item Correlation.
3. Indices based on Factor Analysis: They are indices based on factor analysis, such as: the great value of the Alpha factor used by (Armor) as a one-dimensional indicator, which depends on the great value of the underlying root φ (Eigen Value). Another indicator, McDonald, called Theta and (Curle et al., 2020) called it (omega), which is the minimum stability limit.
4. Indices Based on Latent Trait Models: Wright stated that if there was a set of items, and responses matched Rasch's Model, it was evidence that it was one-dimensional.

To verify a one-dimensional assumption, a number of indices that were selected in the previous context, including:

1- Alpha Kronbach coefficient: The stability factor for the academic proficiency scale was estimated using Cronbach's Alpha method of the scale items (if the item score of the total scale is deleted). Table 3 explains this in detail.

**Table (3): Alpha coefficients for the stability of the Academic Proficiency measure n= 577.**

| Item No | Alpha Coefficient | Item No | Alpha Coefficient | Item No | Alpha Coefficient | Item No | Alpha Coefficient | Item No | Alpha Coefficient |
|---|---|---|---|---|---|---|---|---|---|
| **1** | 0.920 | **13** | 0.921 | **25** | 0.921 | **37** | 0.924 | 49 | 0.918 |
| **2** | 0.921 | **14** | 0.919 | **26** | 0.922 | **38** | 0.930 | 50 | 0.919 |
| **3** | 0.921 | **15** | 0.918 | **27** | 0.924 | **39** | 0.919 | 51 | 0.917 |
| **4** | 0.922 | **16** | 0.923 | **28** | 0.926 | **40** | 0.918 | 52 | 0.922 |
| **5** | 0.924 | **17** | 0.924 | **29** | 0.914 | **41** | 0.914 | 53 | 0.921 |
| **6** | 0.922 | **18** | 0.922 | **30** | 0.915 | **42** | 0.915 | 54 | 0.917 |
| **7** | 0.923 | **19** | 0.921 | **31** | 0.918 | **43** | 0.916 | 55 | 0.928 |
| **8** | 0.924 | **20** | 0.920 | **32** | 0.919 | **44** | 0.922 | 56 | 0.921 |
| **9** | 0.915 | **21** | 0.921 | **33** | 0.922 | **45** | 0.924 | | |
| **10** | 0.913 | **22** | 0.923 | **34** | 0.922 | **46** | 0.926 | | |
| **11** | 0.922 | **23** | 0.924 | **35** | 0.921 | **47** | 0.928 | | |
| **12** | 0.920 | **24** | 0.921 | **36** | 0.923 | **48** | 0.927 | | |
| **Alpha Coefficient** | | 0.929 | | | | | | | |

It is clear from the previous table (3): that the alpha factor of the scale if each single is deleted is less than or equal to the total alpha coefficient of the scale, i.e., all items are constant, since the intervention of the item does not reduce the total stability factor of the scale, and therefore all item is retained in this scale. With the exception of item (38), the intervention of this item was found to reduce the stability factor of the scale, and therefore was deleted. The overall Alpha Cronbach factor of the scale was 0.929, a strong indices of one-dimensional verification.

- Correlation transactions were calculated: between item grades and the overall scale score (if the overall scale grade item score is deleted), table (4) shows correlation coefficients for the Academic Proficiency Scale

**Table (4): Link coefficients for Academic Proficiency Scale n=577**

| Item No | Alpha Coefficient | Item No | Alpha Coefficient | Item No | Alpha Coefficient | Item No | Alpha Coefficient | Item No | Alpha Coefficient |
|---|---|---|---|---|---|---|---|---|---|
| **1** | **0.321 | **13** | **0.430 | **25** | **0.397 | **37** | **0.467 | **49** | **0.483 |
| **2** | **0.325 | **14** | **0.394 | **26** | **0.482 | **38** | **0.197 | **50** | **0.493 |
| **3** | **0.314 | **15** | **0.324 | **27** | **0.480 | **39** | **0.491 | **51** | **0.473 |
| **4** | **0.207 | **16** | **0.298 | **28** | **0.489 | **40** | **0.492 | **52** | **0.481 |
| **5** | **0.201 | **17** | **0.338 | **29** | **0.470 | **41** | **0.498 | **53** | **0.486 |
| **6** | **0.411 | **18** | **0.458 | **30** | **0.489 | **42** | **0.263 | **54** | **0.597 |
| **7** | **0.426 | **19** | **0.576 | **31** | **0.584 | **43** | **0.271 | **55** | **0.488 |
| **8** | **0.349 | **20** | **0.390 | **32** | **0.571 | **44** | **0.462 | **56** | **0.576 |
| **9** | **0.422 | **21** | **0.360 | **33** | **0.477 | **45** | **0.466 | | |
| **10** | **0.424 | **22** | **0.487 | **34** | **0.473 | **46** | **0.479 | | |
| **11** | **0.441 | **23** | **0.484 | **35** | **0.586 | **47** | **0.485 | | |
| **12** | **0.490 | **24** | **0.485 | **36** | **0.567 | **48** | **0.473 | | |

** D at the level (0.01) where the statistical indication of the correlation factor at the degree of freedom (577-2) is (approximately 0.115) where (577) the number of sample stability and reliability.

* D at the level (0.05) where the statistical indication of the correlation factor at the degree of freedom (577-2) is (approximately 0.088) where (577) the number of sample stability and reliability.

From the previous table (4) it is clear that all correlation coefficients between each of the scale items and the overall degree of the scale (if the overall scale grade is deleted) are statistically d, indicating the reliability of all the items of the academic proficiency scale. Correlation transaction values ranged from (٠,٥٨٦-0.179), all of which are significant. This indicates that items share a single latent factor measure expressed by the overall degree that measures academic proficiency. This result can be considered as an indication of the one-dimensional scale.

**3- Using the Factor Analysis**

The factor analysis was used using (Principal Component Analysis) to responses of individuals for items of the Academic Proficiency Scale. The value of (Eigen Value) and the (Explained Variance 0) ratio for each factor were calculated, and table (5) explains this in detail.

**Table (5): Eigen Value interpreted variability ratio and cumulative interpreted variation of each factor n= 577**

| factor | Eigen Value | Explained variation percentage% | Cumulative interpreted variation percentage% |
|---|---|---|---|
| **First** | 11.885 | 21.610 | 21.610 |
| **Second** | 2.816 | 5.101 | 26.711 |
| **Third** | 2.158 | 3.952 | 30.665 |
| **Fourth** | 1.810 | 3.291 | 33.957 |
| **Fifth** | 1.504 | 2.734 | 36.671 |
| **Sixth** | 1.367 | 2.483 | 39.175 |
| **Seventh** | 1.326 | 2.405 | 41.560 |
| **Eighth** | 1.288 | 2.350 | 43.920 |
| **Nineth** | 1.177 | 2.147 | 46.058 |
| **Tenth** | 1.172 | 2.140 | 48.188 |
| **Eleventh** | 1.082 | 1.958 | 50.175 |
| **Twelfth** | 1.048 | 1.915 | 52.076 |
| **thirteenth** | 1.005 | 1.831 | 53.900 |

From the previous table (5) it is clear that: the first factor explains the % 21,61، variation, and that the Eigen Value is equal to (11,885) which is high compared to the rest of the factors, indicating that the scale measures one characteristic. By adopting the value of the Eigen Value test as an indicator of one-dimensional, Lord stated that the items can be one-dimensional, if the Eigen Value of the first factor is significant compared to the Eigen Value of the second factor, and that the ratio of the first Eigen Value to the second factor is large and higher than (1) (Adams et al., 2019; Aizawa et al., 2020) This is achieved in this research. Based on the fact that (Rechase) suggested that the first factor (at least 20%) of the explained discrepancy could be explained, this is an indication of one-dimensionality (Adams et al., 2019; Aizawa et al., 2020). The first factor was also found to be saturated with the most items (51) that was more saturated than (0.3), which is another indication of the one-dimensional of the Academic Proficiency Scale used in this research. Since statistically significant saturation should not be less than (0.3), Saturations are correlations or variations between observed variables and factors (Alvarenga et al., 2020).

**Step 5: Choose the right model:**

After a one-dimensional assumption of the data on the scale was verified, the fifth phase of scale construction, the selection phase of the appropriate model, came (Almaleki, 2021; Ambiel et al., 2015), noted that the step of selecting the appropriate model is one of the most important and accurate steps in constructing the scale, after verifying the one-dimensional assumption, according to the underlying features models(Adams et al., 2019;

Alvarenga et al., 2020; Ayala, 2018). The model chosen to develop the Academic Proficiency Scale to match the data in this research is the Andrich's Model, specifically the so-called Andrich's Rating Scale Model emanating from the Rasch's model. The model was selected as Hambleton (1983) in the research of (Bichi & Talib, 2018; Bock & Gibbons, 2021; Chen & Ahn, 2020) according to a set of considerations as follows:

1. **Statistical assumptions**: The Item Response models are classified into two groups, chosen and differentiated; first: assume the natural distribution of item characteristic curves, called (Normal Ogive Model), and the second: assume Logistic model distribution, as they are difficult to trade off, given the similar distributions of these models. Nevertheless, there are reasons why logistic Models should be preferred over others for their athleticism, and not affected by the responses resulting from the lack of seriousness of the respondents, The arrival of these curves at the Upper Asymptote line is slow.
2. **Response Level:** The Rasch model fits more with Dichotomous, and several models have been developed from this model, as mentioned above, and the most appropriate in this research is Andrich Rating Scale Model because the polytomous scale is the one used. It is suitable for scale data on which the individuals' responses have been based on the five-year lykert Rating Scale from (fully applicable) to (not fully applicable).
3. **The number of statistical features**: If the number of the items varies only in difficulty, the model of the single parameter is chosen, and if the items vary in difficulty and distinguishing, the two-parameter model is chosen, and if the difficulty and discrimination parameters are added to the guesswork parameter, we choose the three-parameter model, but there are also practical foundations and considerations such as: the availability of computer software and the practical experience available. Many researchers prefer the single parameter model, so Andrich Rating Scale was chosen, which is an upgraded image of the Rasch model, for easy use and for the availability of appropriate statistical programs that provide the opportunity to analyze results such as (Bigsteps-Winsteps-Rumm2020-Parscale).

**Step 6: Analysis of scale data:**

After selecting the Andrich Rating Scale Model emanating from Rasch Model, and after correcting students' responses on the scale as follows: [5 (fully applicable), 4 (apply), 3 (fairly applicable), 2 (not applicable), 1 (not fully applicable)]. The data was then entered into the computer memory of

SPSS. The data were analyzed using a program (Winsteps-Bigsteps) to answer current research questions:

**Answer to the first question: What is the degree to which Cairo University students' responses to the items of the Academic Proficiency Scale match Andrich's Rating Scale Model?**

The data was analyzed using IRT PRO-R to verify the degree to which responses match the scale items. The data matching process is done during which individuals who do not match to the form are deleted. To achieve this, each individual's capacity was estimated, and the standard error in measuring capacity The Standardized Information Wieghted Fit Statistics for Persons Intfit (ZSTD), or the average statistics of Mean Square Infit Statistics (MNSQ), a statistical indices of unexpected behaviors affecting responses For items that is close to an individual's capacity level, external matching statistical values have also been estimated, the Standardized Information Weighted Fit Statistics for Persons Outfit (ZSTD), or the average space square statistics Mean Square Outfit Statistics (MNSQ), It is a statistical indicator that is more sensitive to unexpected behaviors that affect responses to vocabulary that are far from an individual's capacity level, for each capacity estimate. Table (6) shows the arithmetic average, the standard deviation of each capacity estimate, the standard error in measuring this capacity, the values of outfit and infit conformity statistics, and the averages of squares for internal and external matching.

**Table (6): Arithmetic average, standard deviation of individual capacity, standard error in capacity estimate, internal and external matching statistics n= 577.**

| | Raw Score | Measure | Model Error | INFIT | | OUTFIT | |
|---|---|---|---|---|---|---|---|
| | | | | MNSQ | ZSTD | MNSQ | ZSTD |
| **Arithmetic average** | 208 | 0.84 | 0.16 | 1.05 | 0.1 - | 1.03 | 0.1 - |
| **Standard deviation** | 25.6 | 0.57 | 0.04 | 0.47 | 2.3 | 0.48 | 2.2 |
| **Highest score** | 268 | 3.22 | 0.50 | 3.50 | 8.6 | 3.76 | 8.6 |
| **Lowest score** | 84 | 1.63- | 0.13 | 0.18 | 7.0 - | 0.20 | 6.9 - |

**REAL RMSE 0.18 ADJ.SD 0.56 SEPARATION 3.37 PERSON RELIABILITY 0.92**

**MODEL RMSE 0.16 ADJ.SD 0.56 SEPARATION 3.64 PERSON RELLIABILITY 0.94**

**S.E. OF PERSON MEAN = 0.03**

From the previous table (6): It is clear that the average calculation of the (MNSQ) is close to one, and that the arithmetic average of infit and outfit matching statistical averages (ZSTD) is close to zero, which is the ideal situation, assumed by the model (zero, 1), while the standard deviation is somewhat far from the one.

While examining the infit statistical values of individuals, which indicates that the individual's ability is matched with that of his or her group of individuals in measuring characteristic measured by the scale, if the value of this statistic is more than (+2), (Bradley & Massof, 2018) The ability of the individual is not identical to that of the group of individuals (Chen & Ahn, 2020), it was found that (100) individuals whose observed responses depart from those expected according to their abilities, and these individuals are not identical to the model, because their responses The note departs from the model's expectations. As their outfit statistical values corresponding to their grades exceed (+2) or the values of averages of squares corresponding to their grades above (1), which the model expects as mentioned by (Chiesi et al., 2018; Choi & Asilkalkan, 2019). The results were after individuals who did not conform to the form were deleted as shown in table 7:

**Table (7): Arithmetic average, standard deviation of individual capacity, standard error in capacity estimate, internal and external matching statistics after deletion of non-conforming individuals n=477**

| | Raw Score | Measure | Model Error | INFIT | | OUTFIT | |
|---|---|---|---|---|---|---|---|
| | | | | MNSQ | ZSTD | MNSQ | ZSTD |
| **Arithmetic average** | 208.1 | 1.01 | 0.16 | 1.02 | 0.0 | 1.02 | 0.0 |
| **Standard deviation** | 25.2 | 0.66 | 0.02 | 0.31 | 1.7 | 0.31 | 1.7 |
| **Highest score** | 268 | 3.51 | 0.38 | 2.34 | 5.7 | 2.35 | 6.0 |
| **Lowest score** | 84 | 2.11- | 0.15 | 0.23 | 6.2- | 0.25 | 6.1- |

**REAL RMSE 0.18 ADJ.SD 0.63 SEPARATION 3.57 PERSON RELIABILITY 0.92**

**MODEL RMSE 0.17 ADJ.SD 0.64 SEPARATION 3.84 PERSON RELLIABILITY 0.93**

**S.E. OF PERSON MEAN = 0.04**

After excluding individuals who were not fit to the model, the analysis was re-analyzed to detect non-fit item, where the difficulty parameter for each item was estimated, and the internal statistical values of The standard Information weighted Fit Statistics for Items Infit (ZSTD), statistically expressed by mean Square Infit Statistics (MNSQ), external vocabulary matching statistical values have been estimated, It is statistically expressed by mean Square Outfit Statistics (MNSQ) for each difficulty feature. Table (8)

shows the arithmetic average, standard deviation, standard error in measuring the difficulty parameter, and statistical values of Infit and Outfit of difficulty features.

**Table (8): Arithmetic average, standard deviation, standard error in measuring difficulty parameter, and statistical values of internal and external matching of difficulty parameters.**

| | Raw Score | Measure | Model Error | INFIT | | OUTFIT | |
|---|---|---|---|---|---|---|---|
| | | | | MNSQ | ZSTD | MNSQ | ZSTD |
| **Arithmetic average** | 1771.0 | 0.00 | 0.06 | 1.00 | 0.0 | 1.03 | 0.0 |
| **Standard deviation** | 211.3 | 0.54 | 0.01 | 0.14 | 2.0 | 0.18 | 2.5 |
| **Highest score** | 2172.0 | 1.47 | 0.08 | 1.53 | 8.4 | 1.83 | 8.6 |
| **Lowest score** | 1102.0 | 1.12- | 0.04 | 0.79 | 3.7- | 0.75 | 4.1- |

**REAL RMSE 0.06 ADJ.SD 0.54 SEPARATION 9.50 ITEM RELIABILITY 0.99**

**MODEL RMSE 0.06 ADJ.SD 0.54 SEPARATION 9.68 ITEM RELLIABILITY 0.99**

**S.E. OF ITEM MEAN = 0.07**

From the previous table (8), it is clear that the average calculation of the average internal and external squares (MNSQ) is one, and that the arithmetic average of internal and external matching statistical averages (ZSTD) is zero, which is the ideal situation, which is assumed by the model (zero, 1), while the standard deviation does not come close to the ideal position expected by the model, which is its approach to one, but is moving away from the one.

With regard to items, in the light of the items matching, it was found that (7) items of the model were not identical and somewhat far from the model's expectations, and that they were turbulent values and confused data. The values of the weighted square averages have increased on the right one. For the results of good conformity according to the discrimination indicator, if the values of its binary link coefficient (point-of-charge coefficient) (rpbis) are negative, the special binary correlation factor between the grades observed for the individual, or the individual and the total grades of the items after the individual is deleted, , or the overall grades of the items after deleting the item, and delete the extreme calculated values of the grades, the negative values of the discrimination transactions indicate a bad match, or an estimate in the opposite direction, and the alphabet letters found at the lab are an indication of the good conformity of the items according to the indicator of discrimination. Table (9) shows the values of internal and external conformity statistics, weighted box averages, and the values of discrimination transactions (rpbis) (binary correlation coefficients of individuals' capabilities) for non-conforming items.

**Table (9): non-conforming vocabulary numbers, internal and external match statistics values, weighted box averages and discrimination transaction values.**

| Item | Model Error | Measure | Raw Score | Entry Number | Ptbis Corr | MNSQ | ZSTD | INFIT | OUTFIT |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | MNSQ | ZSTD |
| **A43** | 0.04 | 1.44 | 1102 | 43 | A0.32 | 1.64 | 8.4 | 1.83 | 8.6 |
| **A5** | 0.05 | 0.18 | 1801 | 5 | B0.25 | 1.44 | 6.3 | 1.59 | 6.8 |
| **A42** | 0.06 | 0.46- | 2043 | 42 | C0.28 | 1.13 | 1.8 | 1.28 | 2.8 |
| **A16** | 0.05 | 0.47 | 1638 | 16 | D0.34 | 1.24 | 3.6 | 1.24 | 3.6 |
| **A25** | 0.04 | 0.68 | 1504 | 25 | E0.41 | 1.22 | 3.3 | 1.23 | 3.6 |
| **A4** | 0.05 | 0.26- | 1916 | 4 | F0.33 | 1.13 | 2.2 | 1.19 | 2.7 |
| **A9** | 0.05 | 0.59 | 1537 | 9 | G0.41 | 1.16 | 2.6 | 1.17 | 2.6 |

**Answer to the second question:** What are the liberal values of individuals' abilities and the difficulty of vocabulary resulting from applying the academic Proficiency Rating Scale to Cairo University students according to the Andrich Rating Scale?

After excluding non-conforming individuals, non-conforming items, and for the purpose of ensuring that data from sample members' responses matched the scale, verifying the objectivity of the scale in its final form (48), and verifying the assumptions of the Andrich Rating Scale, the analysis was re-analyzed to obtain estimates free from the difficulty of items and individual capabilities. Table (10) shows the results of the analysis of values free from the capacities of individuals.

**Table (10): Results of analysis of values free from the capabilities of individuals n= 477.**

| | Raw Score | Measure | Model Error | INFIT | | OUTFIT | |
|---|---|---|---|---|---|---|---|
| | | | | MNSQ | ZSTD | MNSQ | ZSTD |
| **Arithmetic average** | 184.0 | 1.03 | 0.18 | 1.00 | 0.1- | 1.01 | 0.1- |
| **Standard deviation** | 22.6 | 0.74 | 0.03 | 0.33 | 1.7 | 0.33 | 1.7 |
| **Highest score** | 235.0 | 3.68 | 0.41 | 2.30 | 5.2 | 2.47 | 5.7 |
| **Lowest score** | 76.0 | 2.17- | 0.16 | 0.27 | 5.5- | 0.27 | 5.4- |

**REAL RMSE 0.20 ADJ.SD 0.70 SEPARATION 3.58 PERSON RELIABILITY 0.93**

**MODEL RMSE 0.18 ADJ.SD 0.70 SEPARATION 3.82 PERSON RELLIABILITY 0.94**

**S.E. OF PERSON MEAN = 0.03**

Table (10) shows that final estimates free of individual capacities ranged from (76-235), average capacity distribution (1.03) logit, the standard deviation (0.74) logit, and the standard error of the average calculation of capacity estimates was (0.03), a value approaching the ideal position assumed by the

model, indicating the accuracy of the location of individuals on the attribute. Table 11 shows the results of the analysis of values free from items difficulty.

**Table (11): Results of analysis of values free from the difficulty of vocabulary n=477.**

| | Raw Score | Measure | Model Error | INFIT | | OUTFIT | |
|---|---|---|---|---|---|---|---|
| | | | | MNSQ | ZSTD | MNSQ | ZSTD |
| **Arithmetic average** | 1791.0 | 0.00 | 0.06 | 1.00 | 0.0 | 1.01 | 0.1 |
| **Standard deviation** | 191.1 | 0.54 | 0.00 | 0.08 | 1.2 | 0.09 | 1.2 |
| **Highest score** | 21720. | 1.39 | 0.07 | 1.13 | 2.2 | 1.29 | 2.8 |
| **Lowest score** | 12480. | 1.08- | 0.05 | 0.78 | 3.5- | 0.78 | 3.9- |

**REAL RMSE 0.06 ADJ.SD 0.53 SEPARATION 9.04 ITEM RELIABILITY 0.99**

**MODEL RMSE 0.06 ADJ.SD 0.53 SEPARATION 9.18 ITEM RELLIABILITY 0.99**

**S.E. OF ITEM MEAN = 0.07**

Table (11) shows that the average capacity distribution was (zero) logit, the standard deviation (0.54) logit, and the values of estimates free from items difficulty ranged from (-1.08) to (1.39), and the standard line of the average calculation of difficulty estimates was (0.07), which is low, indicating the accuracy of the difficulty estimates for the items.

The items parameter values of the scale were estimated in its final form (48) items, using the Unconditional Maximum Likelihood Estimation (UMLE) method, to estimate the most accurate estimates of items' difficulty, to minimize the difficulty of the item scale in its final form, table (12) showing the difficulty values of the items, and the standard error in estimating this difficulty for each item of scale and descending order according to the difficulty of the items (Rating).

**Table (12): The difficulty of vocabulary and its normative errors are arranged downwards according to the difficulty of the items n= 477.**

| Item | Model Error | Measure | Raw Score | Entry Number | Ptbis Corr | INFIT | | OUTFIT | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | | MNSQ | ZSTD | MNSQ | ZSTD |
| **A30** | 0.05 | 1.39 | 1249 | 30 | 0.54 | 1.00 | 0.1 | 1.02 | 0.3 |
| **A41** | 0.06 | 1.09 | 1407 | 41 | 0.49 | 1.06 | 0.9 | 1.05 | 0.9 |
| **A10** | 0.05 | 0.95 | 1453 | 10 | 0.52 | 1.02 | 0.3 | 1.02 | 0.3 |
| **A19** | 0.05 | 0.92 | 1468 | 19 | 0.56 | 0.95 | 0.8- | 0.94 | 0.9- |
| **A49** | 0.05 | 0.92 | 1477 | 49 | 0.50 | 1.08 | 1.4 | 1.07 | 1.2 |
| **A26** | 0.05 | 0.75 | 1521 | 26 | 0.48 | 1.09 | 1.5 | 1.09 | 1.4 |
| **A36** | 0.05 | 0.69 | 1529 | 36 | 0.57 | 0.92 | 1.4- | 0.91 | 1.6- |
| **A46** | 0.06 | 0.65 | 1617 | 46 | 0.49 | 1.01 | 0.2 | 1.01 | 0.2 |
| **A20** | 0.05 | 0.44 | 1631 | 20 | 0.44 | 1.11 | 1.8 | 1.13 | 1.9 |
| **A27** | 0.05 | 0.43 | 1609 | 27 | 0.50 | 1.05 | 0.9 | 1.04 | 0.6 |
| **A28** | 0.05 | 0.40 | 1657 | 28 | 0.51 | 1.00 | 0.0 | 0.98 | 0.2- |
| **A33** | 0.06 | 0.29 | 1691 | 33 | 0.50 | 0.98 | 0.3- | 0.98 | 0.4- |
| **A24** | 0.05 | 0.25 | 1692 | 24 | 0.49 | 1.03 | 0.5 | 1.01 | 0.2 |
| **A55** | 0.06 | 0.19 | 1751 | 55 | 0.51 | 96 | 0.6- | 0.98 | 0.4- |
| **A6** | 0.06 | 0.18 | 1742 | 6 | 0.44 | 1.08 | 1.3 | 1.10 | 1.6 |
| **A23** | 0.06 | 0.13 | 1743 | 23 | 0.43 | 1.08 | 1.3 | 1.10 | 1.5 |
| **A17** | 0.06 | 0.11 | 1952 | 17 | 0.36 | 1.13 | 2.1 | 1.18 | 2.5 |
| **A56** | 0.06 | 0.10 | 1733 | 56 | 0.54 | 0.89 | 1.7- | 0.88 | 1.7- |
| **A12** | 0.06 | 0.09 | 1764 | 12 | 0.50 | 0.96 | 0.6- | 0.95 | 0.7- |

**Continue table (12): The difficulty of items and its normative errors are arranged downwards according to the difficulty of item n=477.**

| **Item** | **Model Error** | **Measure** | **Raw Score** | **Entry Number** | **Ptbis Corr** | **INFIT** | | **OUTFIT** | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | | MNSQ | ZSTD | MNSQ | ZSTD |
| **A35** | 0.06 | 0.08 | 1718 | 35 | 0.56 | 0.88 | 2.0- | 0.88 | 2.0- |
| **A37** | 0.05 | 0.07 | 1751 | 37 | 0.49 | 1.01 | 0.2 | 0.1- | 0.99 |
| **A54** | 0.06 | 0.03 | 1801 | 54 | 0.59 | 0.79 | 3.5- | 3.8- | 0.77 |
| **A45** | 0.06 | 0.00 | 1781 | 45 | 0.47 | 1.00 | 0.0 | 0.1- | 0.99 |
| **A47** | 0.06 | 0.00 | 1748 | 47 | 0.52 | 0.93 | 1.2- | 1.3- | 0.92 |
| **A32** | 0.06 | 0.03- | 1841 | 32 | 0.52 | 0.90 | 1.5- | 1.6- | 0.89 |
| **A29** | 0.06 | 0.05- | 1805 | 29 | 0.50 | 0.94 | 1.0- | 0.7- | 0.96 |
| **A48** | 0.06 | 0.07- | 1832 | 48 | 0.45 | 1.01 | 0.2 | 1.03 | 0.5 |
| **A44** | 0.06 | 0.09- | 1852 | 44 | 0.41 | 1.07 | 1.00 | 1.06 | 0.6 |
| **A53** | 0.06 | 0.10- | 1877 | 53 | 0.46 | 1.00 | 0.0 | 1.01 | 0.2 |
| **A39** | 0.06 | 0.12- | 1814 | 39 | 0.49 | 0.96 | 0.6- | 0.97 | 0.5- |
| **A52** | 0.06 | 0.13- | 1912 | 52 | 0.46 | 0.99 | 0.2- | 1.04 | 0.6 |
| **A7** | 0.06 | 0.25- | 1845 | 7 | 0.46 | 1.00 | 0.1 | 1.01 | 0.2 |
| **A31** | 0.06 | 0.29- | 1921 | 31 | 0.50 | 0.90 | 1.5- | 0.90 | 1.4- |
| **A15** | 0.07 | 0.33- | 2051 | 15 | 0.36 | 1.06 | 0.9 | 1.09 | 1.1 |
| **A51** | 0.06 | 0.35- | 1856 | 51 | 0.52 | 0.88 | 2.0- | 0.90 | 1.7- |
| **A34** | 0.06 | 0.37- | 1974 | 34 | 0.43 | 1.00 | 0.1 | 1.01 | 0.1 |
| **A50** | 0.06 | 0.39- | 1960 | 50 | 0.46 | 0.93 | 1.0- | 0.98 | 0.2- |
| **A13** | 0.06 | 0.45- | 1911 | 13 | 0.46 | 0.98 | 0.4- | 0.95 | 0.7- |

**Continue table (12): The difficulty of items and its standard errors are arranged downwards according to the difficulty of item N= (477).**

| Item | Model Error | Measure | Raw Score | Entry Number | Ptbis Corr | INFIT | | OUTFIT | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | | MNSQ | ZSTD | MNSQ | ZSTD |
| **A21** | 0.06 | 0.46- | 1920 | 21 | 0.43 | 1.02 | 0.3 | 1.00 | 0.1 |
| **A22** | 0.07 | 0.52- | 1964 | 22 | 0.48 | 0.90 | 1.5- | 0.93 | 1.0- |
| **A18** | 0.06 | 0.54- | 1971 | 18 | 0.44 | 0.97 | 0.3- | 1.02 | 0.4 |
| **A11** | 0.06 | 0.62- | 1902 | 11 | 0.46 | 0.99 | 0.2- | 0.95 | 0.7- |
| **A2** | 0.06 | 0.63- | 1914 | 2 | 0.37 | 1.13 | 2.1 | 1.17 | 2.5 |
| **A3** | 0.06 | 0.65- | 2043 | 3 | 0.34 | 1.11 | 1.5 | 1.17 | 2.0 |
| **A14** | 0.06 | 0.75- | 1987 | 14 | 0.33 | 1.14 | 2.1 | 1.21 | 2.7 |
| **A40** | 0.06 | 0.80- | 2012 | 40 | 0.47 | 0.92 | 1.3- | 0.93 | 1.0- |
| **A8** | 0.06 | 0.86- | 2051 | 8 | 0.41 | 1.00 | 0.0 | 0.99 | 0.1- |
| **A1** | 0.08 | 1.09- | 2172 | 1 | 0.27 | 1.06 | 0.8 | 1.28 | 2.1 |

From the previous table (12): It is clear that the matching indicators for each of the scale items were within the Infit and Outfit limits ranging from (0.7-1.4).

**Other indicators indicating the realization of other assumptions include:**

(a) Equal discrimination of the items or (equal items in discriminatory capacity), this result confirms the clear convergence in the values of bilateral association coefficients (rpbis), which reflect a convergence in items discrimination transactions. Thus, this is an indication of the realization of a model assumption, namely that the item has an almost equal discriminatory capacity.(Alvarenga et al., 2020; Ayala, 2018; Ayis et al., 2018), that in order to achieve the assumption of equal discrimination indicators and conformity to the model, their values must be within range limits (average discrimination transactions = +0.15), and because the average discrimination transactions were (0.43), the matching limits ranged from (0.27- 0.59), so the values of the special bilateral engagement transactions were all within the range, the standard deviation value of these transactions was also small, an indication of the realization of this assumption. This is confirmed by the parallel of the characteristic curves of the items.

(b) With regard to the assumption of a lower guessing indicator, there is no direct way to determine whether an individual has answered the item by guessing (random selection) or not. But when the non-linear slope lines of the grades are drawn on the scale (the characteristic curve of the individual), it was observed that the lower asymptote is close to zero in all items, this occurs when the individual does not respond randomly. It is an indication of the good

item matching of the model, and it cannot be said that the speed factor played a role to answer the item, as no respondent left any item unanswered.

**Answer to the third question:** What are the psychometric characteristics of the items of the Academic Proficiency Scale free from individuals and items according to Andrich Rating Scale Model among Cairo University students?

After providing these indicators that indicate the realization of the model assumptions and the good matching of vocabulary to Andrich Rating Scale. Indicators of the psychometric characteristics of the scale items must be provided in its final form (48) items.

**Stability Semantics:**

The stability coefficients of the scale were estimated in two ways: the first: using traditional methods of measurement, and the second: using the Item Response Theory.

**First: using the Traditional Theory of Measurement:**

(a) The stability factor for the Academic Proficiency Scale was calculated in its final form (48) items for each of its dimensions, i.e., after deleting items and individuals that do not conform to the Andrich Rating Scale. The sample (477) students at Cairo University, each dimension independently by calculating Cronbach's Alpha factor for the items of each subdivision (If the individual score is deleted from the total degree of the dimension to which the item belongs). Table (13) shows Alpha coefficients for the stability of the Academic Proficiency Scale:

**Table (13): Alpha coefficients for the stability of the Academic Proficiency Scale n= 477.**

| Effective recall skill | | Self-organizing skill for learning | | Managing and organizing time skill | | managing and withstanding academic pressures Skill | | Academic self-effectiveness | |
|---|---|---|---|---|---|---|---|---|---|
| Item | Alpha Coeff | item | Alpha Coeff | item | Alpha Coeff | Item | Alpha Coeff | Item | Alpha Coeff |
| **1** | 0.782 | 19 | 0.627 | 26 | 0.834 | 39 | 0.747 | 50 | 0.779 |
| **2** | 0.777 | 20 | 0.645 | 27 | 0.828 | 40 | 0.757 | 51 | 0.768 |
| **3** | 0.780 | 21 | 0.649 | 28 | 0.828 | 41 | 0.755 | 52 | 0.784 |
| **6** | 0.775 | 22 | 0.669 | 29 | 0.819 | 44 | 0.758 | 53 | 0.769 |
| **7** | 0.771 | 23 | 0.677 | 30 | 0.831 | 45 | 0.746 | 54 | 0.752 |
| **8** | 0.771 | 24 | 0.681 | 31 | 0.828 | 46 | 0.739 | 55 | 0.783 |
| **10** | 0.778 | | | 32 | 0.830 | 47 | 0.737 | 56 | 0.784 |
| **11** | 0.776 | | | 33 | 0.839 | 48 | 0.753 | | |
| **12** | 0.769 | | | 34 | 0.842 | 49 | 0.748 | | |
| **13** | 0.771 | | | 35 | 0.823 | | | | |
| **14** | 0.780 | | | 36 | 0.827 | | | | |
| **15** | 0.778 | | | 37 | 0.836 | | | | |
| **17** | 0.776 | | | | | | | | |
| **18** | 0.767 | | | | | | | | |
| **G. Alfa Coeff** | 0.788 | | 0.691 | | 0.845 | | 0.768 | | 0.799 |

From the previous table (13): It is clear that the alpha factor of the scale if each item is deleted is less than or equal to the alpha coefficient of the sub transmission to which the individual belongs, i.e., all items are constant, as the intervention of the items does not reduce the total stability factor of the sub transmission to which the item belongs, and therefore all items are retained in this scale.

The stability of sub dimensional degrees and the overall stability of the Academic Proficiency Scale were calculated for the retained items, in two ways: The alpha factor calculation for Cronbach, and the second: the calculation of the subdivision stability factor in the half-retail manner of "Spearman/Brown", the results were as in table (14), which shows the sub dimensional stability factors and the overall stability of the Academic Proficiency Scale.

**Table (14): Subdivision stability and total stability of the Academic Proficiency Scale n=477.**

| No | | Stability factor | | |
|---|---|---|---|---|
| | | Alpha L Kronbach | Retail Half and Spearman/Brown Correction | |
| | | | Two-half test way. | individual and couple items way |
| 1 | **Effective recall skill** | 0.789 | 0.768 | 0.813 |
| 2 | **Self-organizing skill for learning** | 0.698 | 0.598 | 0.748 |
| 3 | **Managing and organizing time skill** | 0.844 | 0.809 | 0.872 |
| 4 | **managing and withstanding academic pressures Skill** | 0.768 | 0.775 | 0.764 |
| 5 | **Academic self-effectiveness** | 0.779 | 0.746 | 0.806 |
| | | 0.938 | 0.898 | 0.949 |

Table (14) shows that the subdivision stability coefficients of the Academic Proficiency Scale in both ways (Alpha L Cronbach, Spearman/Brown Half Hash) are high, indicating the stability of all subdivisions of the Academic Proficiency Scale.

(**b**) Internal consistency was calculated by correlation coefficients between the degree of each item and the total degree of the sub transaction to which the items belonged. Table (15) shows the internal consistency of the Academic Proficiency Scale

**Table (15): Internal consistency of Academic Proficiency Scale n= (477).**

| Effective recall skill | | Self-organizing skill for learning | | Managing and organizing time skill | | managing and withstanding academic pressures Skill | | Academic self-effectiveness | |
|---|---|---|---|---|---|---|---|---|---|
| **Item** | **Alpha Coeff** | **item** | **Alpha Coeff** | **item** | **Alpha Coeff** | **Item** | **Alpha Coeff** | **Item** | **Alpha Coeff** |
| **1** | $^{**}0.424$ | 19 | $^{**}0.693$ | 26 | $^{**}0.602$ | 39 | $^{**}0.596$ | 50 | $^{**}0.648$ |
| **2** | $^{**}0.495$ | 20 | $^{**}0.671$ | 27 | $^{**}0.664$ | 40 | $^{**}0.502$ | 51 | $^{**}0.687$ |
| **3** | $^{**}0.454$ | 21 | $^{**}0.617$ | 28 | $^{**}0.655$ | 41 | $^{**}0.556$ | 52 | $^{**}0.628$ |
| **6** | $^{**}0.526$ | 22 | $^{**}0.568$ | 29 | $^{**}0.647$ | 44 | $^{**}0.543$ | 53 | $^{**}0.704$ |
| **7** | $^{**}0.572$ | 23 | $^{**}0.588$ | 30 | $^{**}0.631$ | 45 | $^{**}0.622$ | 54 | $^{**}0.766$ |
| **8** | $^{**}0.566$ | 24 | $^{**}0.617$ | 31 | $^{**}0.534$ | 46 | $^{**}0.663$ | 55 | $^{**}0.648$ |
| **10** | $^{**}0.525$ | | | 32 | $^{**}0.638$ | 47 | $^{**}0.676$ | 56 | $^{**}0.633$ |
| **11** | $^{**}0.518$ | | | 33 | $^{**}0.536$ | 48 | $^{**}0.573$ | | |
| **12** | $^{**}0.577$ | | | 34 | $^{**}0.488$ | 49 | $^{**}0.622$ | | |
| **13** | $^{**}0.554$ | | | 35 | $^{**}0.597$ | | | | |
| **14** | $^{**}0.453$ | | | 36 | $^{**}0.717$ | | | | |
| **15** | $^{**}0.475$ | | | 37 | $^{**}0.577$ | | | | |
| **17** | $^{**}0.492$ | | | | | | | | |
| **18** | **0.596 | | | | | | | | |

** Eff at level (0.01)

Table (15) of the previous table shows that the correlation coefficients between each item and the overall degree of dimension to which the items belong are statistically significant, indicating the internal consistency of the Academic Proficiency Scale. The results showed that the scale has a high degree of internal consistency, which indicated that the scale has high internal consistency (as an indicator of stability).

**Second: Using the Item Response Theory:**

After verifying that the scale vocabulary matched its final form of the Andrich Rating scale model, the liberal values of both items difficulty and individual capabilities were obtained, and through these values two types of transactions were obtained: person Reliability and Item Reliability. Stability in modern theory means accuracy in estimating the location of both individuals and items on the attribute connection, The accuracy of the items in defining this caller can be determined by the calculation of item separation index, which is defined as the ratio between the standard deviation of progressive, liberal item values, and the average standard error of these values. The separation factor for the scale items has reached its final form (9.04), which is more than (2), and therefore is sufficient to define the attribute connection scale, as well as the value of the Person Separation Index (3.58), which is more than (2), Therefore, the sample of individuals is also sufficient to separate items, and through these transactions (Gp, Gi), the stability factor for both items and individuals has been calculated, according to the following mathematical formula mentioned by (Bichi & Talib, 2018; Bonifay, 2019; Choi & Asilkalkan, 2019)

(5) (R= G2/1+ G2).....

Where (G): Symbolizes the separation coefficient, (R) stability factor, stability coefficients for both items and individuals (0.99-0.93) respectively They are two high values, the first of which indicates the adequacy of the items sample in the separation of individuals, and therefore in the distinction between the performance levels of these individuals, and the second indicates the adequacy of the sample of individuals in the separation of items, and thus in the definition of the attribute that this items scale. It should be noted that stability transaction values in this way reward the stability coefficient values in the Kuder-Richardson (KR-20) manner in the Traditional Theory, particularly since the value of the stability factor in this way represents the minimum stability factor as mentioned by (Chen & Ahn, 2020; Chiesi et al., 2018).

The number of distinct statistical layers for both items and individuals were also determined using the following mathematical formula:

(6) H= (4G+ 1)/3.....

Where H: Number of statistical layers, G: separation coefficient, the number of statistical layers for both items and individuals (12.37 To 5.11) respectively. The first value indicates the ability of item to show individual differences between individuals in the degree to which they have the characteristic significantly, and the second indicates that there are (5) levels in these individuals.

Other indicators refer to the stability of the scale using the Item Response Theory include information function, which can be used to ensure the accuracy of the assessment of the parameters of the items and the ability parameter, and its importance comes from the fact that the contribution of each items to the test information function is determined independently of the rest of the test items, but in the Traditional Theory, the items' contribution to the stability and reliability of the test is not determined independently of the rest of the test items. The information function of the scale is an amount that is inversely proportional to the standard error of the scale, Thus, the information function is an indicator of the stability of the scale, because by increasing the standard error of estimate, stability is lower and vice versa (Clark et al., 2020; Cordier et al., 2019).

(Crowe et al., 2018) states that the information function of the item can be defined using the Andrich Rating Scale Model emanating from Rasch's Model as follows:

(7) $I(\theta) = Pi(\theta)Qi(\theta)$.....

In general, the information function of the test will be greater than the information function of the items, The test therefore measures the capacity with a high degree of accuracy more than an item. The information function curve of the scale results from the gyration of items curves on top of each other according to the relationship:

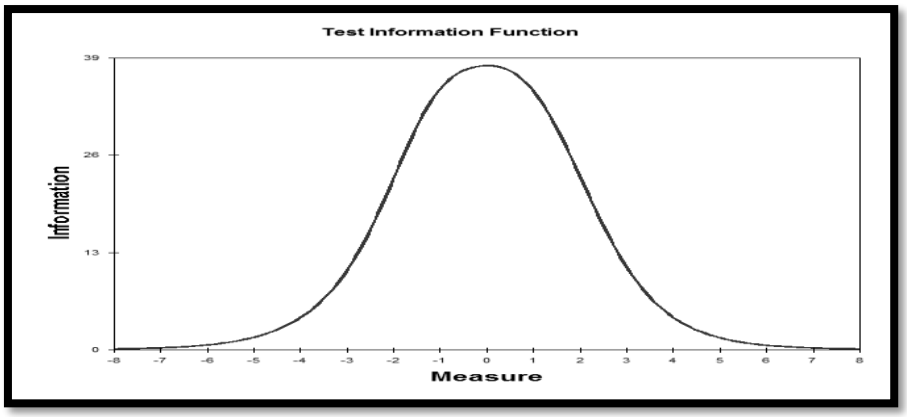
(8) $I(\theta) = \sum_g Ig\ (\theta)$ ... ... ... ... ... ... ... ... ... ... ...

$I(\theta)$: The amount of information for testing at the capacity level is $\theta$, $\sum_g Ig\ (\theta)$the sum of the information functions of the items of that scale at the same capacity level ($\theta$). Therefore, an increase in the number of items gives a small standard line S.E. $\theta$, and the decrease in the value of the standard error in estimating capacity at the capacity level ($\theta$) increases the amount of information to test according to the following relationship:

$$(9)\ I(\theta) = \frac{1}{\sqrt{S.E(\theta)}} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

Thus, the decrease in the value of the standard error in the estimate of capacity increases the value of the stability factor, as does traditional theory. What distinguishes this method from traditional methods is that the assessment of stability in traditional theory is linked to the sample, which is a poor characteristic of stability assessment, and leads to a collective estimation of individual's errors on scale markers, the so-called standard measurement error. While modern theory provides us with an estimate of the standard error of measurement at each level from capacity levels, we can determine the extent to which each item contributes to determining the accuracy of the scale as Test, 1971 (Dinić & Raine, 2019; Dougherty et al., 2021).

Furthermore, the amount of information for testing was estimated at each specific capacity level, using (Winsteps), after selecting (48) items spread over the attribute connection using the items map, covering the difficulty of the items in the extent to which progressive values are distributed, from (-1.09) logit to (1.38) logit, and figure (١) shows the graph to assess the amount of information for the scale at each level of capacity for individuals.




**Form (1): Information function for testing**

**Figure 1** shows that the values of the amount of information provided by the scale are as great as possible at capacity (0.35) logit, i.e., the scale gives more information about individuals with intermediate academic proficiency, while the amount of information provided by the scale is as low as possible at the lowest capacity values, which means that the scale provides little data about the individuals with intermediate academic proficiency which reflects

the model's expectations. The amount of information provided by the item is as great as possible when (b = θ), This is for both the single-parameter and the two-parameter models, which means the accuracy of the ability estimate at the middle of the characteristic level (capacity scale), while at the parties, the amount of information for the scale is reduced. The average capacity value for individuals was (1.02). This is confirmed by the standard error value of the scale, which was the lowest possible at the average progressive value of the scale items. The standard error is small as individuals' ability values approach the item difficulty values, so the amount of information increases with the decrease in the standard error. This in turn increases the stability factor of the scale.

**(Gomez et al., 2019),** stated that there is a relation between the scale's information function and stability. If the scale's information function increases, the number of standard errors decreases, which leads to increasing the stability(Hays et al., 2021; Immekus et al., 2019).

**Indices of Validity:**

Cronbach, (1971) states that the most important thing for developers of psychological scales is the (Construct Validity), and with the multiplicity of methods and procedures used in the studies of the validity of these scales, they can be classified into three main groups. Logical Analysis, Correlational Techniques and Experimental Techniques (Kean et al., 2018; J. Lang & L. Tay, 2021).

(Shavelson et al., 1991), states that the most commonly used methods include Correlational Techniques which is used in the study of the validity of psychological scales (Factor Analysis), Multitrati Matrix, Multimethod and Regression Analysis(Alvarenga et al., 2020; Ayala, 2018). The procedures for demonstrating the validity of this research instrument were carried out through logical validity procedures and the use of correlational techniques.

(**a**) With regard to logical validity, it has been verified based on the theoretical analysis by defining the concept and dimensions of Academic Proficiency. By defining the items and the method of formulating and evaluating them by a group of Jury. Based on their opinions, the items of the scale were modified. Logical analysis is one of the methods and procedures used to verify the validity of the construction of the scale instruments.

(**b**) The Correlational Techniques used in the verification procedures for the validity of the scale instruments in its final form are (48) items, in this research:

## .Using The Andrich Rating scale model (ARSM) to build a scale

**1. The first method:** The factor analysis method was used to determine the factor structure of this scale. The basic component method has been used, the Eigen Value has been calculated, the ratio of interpreted variation per factor. Table (16) explains this in detail.

**Table (16): Eigen Values, interpreted variability ratio and cumulative interpreted variation of each N= (477) factor.**

| factor | Eigen root | Explained variation percentage% | Cumulative interpreted variation percentage% |
|---|---|---|---|
| 1st | 3.618 | 72.380 | 72.380 |
| 2nd | 0.442 | 8.821 | 81.202 |
| 3rd | 0.346 | 6.896 | 88.098 |
| 4th | 0.334 | 6.677 | 94.776 |
| 5th | 0.261 | 5.225 | 100.000 |

Table (16) shows that the first factor explains the (72,380%) variation, and that its **Eigen** value is (3,618), which is high compared to the rest of the factors, indicating that the scale measures a single feature.

**- The second method:** Calculating the correlation factors between the item's grades and the overall degree of the sub dimension to which the item belongs (if the item score is deleted from the total degree of the dimension to which the item belongs), and after deleting items that do not conform to the Rating Scale Model in order to know the extent to which each item of the scale contributes to what the sub dimension to which it belongs. Table (17) shows correlation coefficients for academic proficiency.

**Table (17): Correlation coefficients of the academic proficiency scale between the single degree and the overall degree of the subdivision to which the item belongs n=477.**

| Effective recall skill | | Self-organizing skill for learning | | Managing and organizing time skill | | managing and withstanding academic pressures Skill | | Academic self-effectiveness | |
|---|---|---|---|---|---|---|---|---|---|
| **Item** | **Alpha Coeff** | **item** | **Alpha Coeff** | **item** | **Alpha Coeff** | **Item** | **Alpha Coeff** | **Item** | **Alpha Coeff** |
| **1** | $^{**}0.349$ | 19 | $^{**}0.481$ | 26 | $^{**}0.493$ | 39 | $^{**}0.448$ | 50 | $^{**}0.519$ |
| **2** | $^{**}0.375$ | 20 | $^{**}0.435$ | 27 | $^{**}0.566$ | 40 | $^{**}0.367$ | 51 | $^{**}0.558$ |
| **3** | $^{**}0.345$ | 21 | $^{**}0.436$ | 28 | $^{**}0.565$ | 41 | $^{**}0.408$ | 52 | $^{**}0.467$ |
| **6** | $^{**}0.389$ | 22 | $^{**}0.388$ | 29 | $^{**}0.559$ | 44 | $^{**}0.397$ | 53 | $^{**}0.561$ |
| **7** | $^{**}0.464$ | 23 | $^{**}0.374$ | 30 | $^{**}0.530$ | 45 | $^{**}0.488$ | 54 | $^{**}0.647$ |
| **8** | $^{**}0.467$ | 24 | $^{**}0.395$ | 31 | $^{**}0.435$ | 46 | $^{**}0.537$ | 55 | $^{**}0.488$ |
| **10** | $^{**}0.385$ | | | 32 | $^{**}0.556$ | 47 | $^{**}0.567$ | 56 | $^{**}0.489$ |
| **11** | $^{**}0.465$ | | | 33 | $^{**}0.425$ | 48 | $^{**}0.428$ | | |
| **12** | $^{**}0.478$ | | | 34 | $^{**}0.383$ | 49 | $^{**}0.449$ | | |
| **13** | $^{**}0.478$ | | | 35 | $^{**}0.504$ | | | | |
| **14** | $^{**}0.378$ | | | 36 | $^{**}0.637$ | | | | |
| **15** | $^{**}0.371$ | | | 37 | $^{**}0.477$ | | | | |
| **17** | $^{**}0.372$ | | | | | | | | |
| **18** | $^{**}0.466$ | | | | | | | | |

** D at the level (0.01) where the statistical indication of the correlation factor at the degree of freedom (477-2) is (approximately 0.128) where (477) the number of sample stability and honesty.

* D at the level (0.05) where the statistical indication of the correlation factor at the degree of freedom (477-2) is (approximately 0.098) where (477) the number of sample stability and validity.

From the previous table (17): it is clear that all correlation coefficients between each item of the scale, and the overall degree of the sub dimension to which the item belongs (if the item score of the total degree of the dimension to which the item belongs is deleted) is statistically significant, that indicates the validity of all the items of the academic proficiency scale.

**- Method 3:** Calculating correlations between the item's score and the overall scale score (if the overall scale score is deleted), and after deleting items that do not match the Rating Scale Model to see how each item of the scale items contributes to what the scale measures. Table (18) shows correlation coefficients for academic proficiency.

**Table (18): Correlation coefficients for the academic mastery scale between the individual score and the overall score of the scale.**

| Item | Alpha factor | Item | Alpha factor | item | Alpha factor | Item | Alpha factor |
|---|---|---|---|---|---|---|---|
| **1** | $^{**}0.390$ | 18 | $^{**}0.565$ | 32 | $^{**}0.557$ | 48 | $^{**}0.438$ |
| **2** | $^{**}0.354$ | 19 | $^{**}0.566$ | 33 | $^{**}0.582$ | 49 | $^{**}0.422$ |
| **3** | $^{**}0.386$ | 20 | $^{**}0.459$ | 34 | $^{**}0.498$ | 50 | $^{**}0.495$ |
| **6** | $^{**}0.493$ | 21 | $^{**}0.445$ | 35 | $^{**}0.572$ | 51 | $^{**}0.553$ |
| **7** | $^{**}0.466$ | 22 | $^{**}0.563$ | 36 | $^{**}0.568$ | 52 | $^{**}0.489$ |
| **8** | $^{**}0.437$ | 23 | $^{**}0.451$ | 37 | $^{**}0.493$ | 53 | $^{**}0.475$ |
| **10** | $^{**}0.508$ | 24 | $^{**}0.490$ | 39 | $^{**}0.511$ | 54 | $^{**}0.639$ |
| **11** | $^{**}0.488$ | 26 | $^{**}0.457$ | 40 | $^{**}0.499$ | 55 | $^{**}0.518$ |
| **12** | $^{**}0.572$ | 27 | $^{**}0.498$ | 41 | $^{**}0.455$ | 56 | $^{**}0.557$ |
| **13** | $^{**}0.492$ | 28 | $^{**}0.584$ | 44 | $^{**}0.428$ | | |
| **14** | $^{**}0.398$ | 29 | $^{**}0.585$ | 45 | $^{**}0.493$ | | |
| **15** | $^{**}0.466$ | 30 | $^{**}0.581$ | 46 | $^{**}0.489$ | | |
| **17** | $^{**}0.551$ | 31 | $^{**}0.578$ | 47 | $^{**}0.588$ | | |

**** D at the level (0.01) where the statistical indication of the correlation factor at the degree of freedom (477-2) is (approximately 0.128) where (477) the number of sample stability and validity.**

*** D at the level (0.05) where the statistical indication of the correlation factor at the degree of freedom (477-2) is (approximately 0.098) where (477) the number of sample stability and validity.**

From the previous table (18): it is clear that all correlation coefficients between each of the scale items and the overall degree of the scale (if the overall scale score is deleted) are statistically significant, which indicates the validity of all the items of the academic proficiency scale.

The results of previous correlation factors showed that all correlation factors values were statistically significant (at level 0.01), and that the correlation between the item and its dimension was consistently greater than the factor correlation with the instrument. This provides evidence of the effectiveness of the scale items, as it measures what the dimension measures, and measures what the scale measures. The items of the scale are consistent in measuring a general feature in line with the assumptions of the modern theory.

**4. Method 4**: the correlations between the total score of the sub dimension and the overall scale, as each of these dimensions measures a facet of academic proficiency, and table (19) shows the coefficients of the correlation of the

academic proficiency scale between the total score of the sub dimension and the overall score of the scale.

**Table (19): Academic mastery scale correlation coefficients between the total score of the sub dimension and the overall score of the scale.**

| Academic proficiency | Effective recall skill | Self-organizing skill for learning | Managing and organizing time skill | managing and withstanding academic pressures Skill | Academic self-effectiveness | Total degree |
|---|---|---|---|---|---|---|
| **Effective recall skill** | - | **0.642 | **0.687 | **0.848 | **0.658 | **0.878 |
| **Self-organizing skill for learning** | | – | **0.697 | **0.688 | **0.597 | **0.799 |
| **Managing and organizing time skill** | | | – | **0.774 | **0.648 | **0.898 |
| **managing and withstanding academic pressures Skill** | | | | – | **0.728 | **0.879 |
| **Academic self-effectiveness** | | | | | – | **0.846 |
| **Total degree** | | | | | | – |

(**) Function at level (0.01)

From the previous table (19): all correlations between the total score of sub dimension and the overall score of the scale are statistically significant, indicating the validity of all the items of the academic proficiency scale.

**The final form of the scale:** From the previous procedures, the validity and stability of the scale of academic proficiency and internal consistency in the traditional and modern methods of measurement, and its validity to measure academic proficiency among students of Cairo University, where the scale in its final form consists of (48) items distributed over the five dimensions.

**Research recommendations:**

Based on the results of the current research, a set of recommendations has been made, which we hope will be considered to be carried out and utilized, and these recommendations are:

1) Pay attention to academic proficiency and its components for students and work to develop them for their role in raising the level of confidence among students, which is reflected positively in the level of their resilience and way of thinking, and thus achieving their goals and aspirations.

2) The establishment of centers for psychological guidance and talented people within Egyptian universities that improve the skills of students in many academic fields such as academic proficiency skills, which will have a great impact on them after graduation.
3) Holding training and educational courses within the framework of the development of faculty members and how they are guided by the problems they can face with students
4) Introducing the faculty member to academic pressures and how to deal with them skillfully through training and education processes, which reflects on him positively and on his attitude within the institution with his students, and ultimately leads to control and control of pressures within his work.
5) Providing incentives for faculty members to publish scientific research on academic proficiency and ways to develop and deal with it.

**Proposed research:**

This research paves the way for researchers in psychology and education, in light of the theoretical framework on the current research variable, and its results, and in terms of what we see as an update of the research that there is a set of variables that call for attention to their future research and study such as:

1) Aware educational support and its relationship to academic proficiency and academic resilience among pupils with learning difficulties.
2) Study the differences in academic proficiency with other variables such as: (age, experience, type of family education, general culture, types of support and school promotion) in different samples.
3) Due to the novelty of models of individual response theory, and their concepts and foundations that differ from the concepts and foundations of traditional theory, it is proposed to conduct other research to verify the characteristics of individual response theory models, especially with the fourth model, which is the negligence model in theory, in order to help more accurately design, build and develop psychological and educational measures in various fields.
4) Studying the impact of the use of Item Response Models on the validity and stability of the academic proficiency scale among high school students.

**Reference**

Abu Jarad, H. Y. (2016). Using Andrich Model in Constructing the Scale of the Attitudes of Faculty Staff Members Towards Students' Evaluation of Educational Practices at University Level. *Journal of Al-Quds Open University for Educational & Psychological Research & Studies*, *4*(15), 1.

Adams, D., Bolt, D., Deng, S., Smith, S., & ... (2019). Using multidimensional item response theory to evaluate how response styles impact measurement. *British Journal of …*. https://doi.org/10.1111/bmsp.12169

Ágoston, C., Urbán, R., Richman, M., & Demetrovics, Z. (2018). Caffeine use disorder: An item-response theory analysis of proposed DSM-5 criteria. *Addictive Behaviors*. https://www.sciencedirect.com/science/article/pii/S0306460318300753

Aizawa, I., Rose, H., Thompson, G., & ... (2020). Beyond the threshold: Exploring English language proficiency, linguistic challenges, and academic language skills of Japanese students in an English medium …. *Language Teaching …*. https://doi.org/10.1177/1362168820965510

Allega, O., Leng, X., Vaccarino, A., Skelly, M., & ... (2018). Performance of the biological rhythms interview for assessment in neuropsychiatry: an item response theory and actigraphy analysis. *Journal of Affective …*. https://www.sciencedirect.com/science/article/pii/S0165032717311540

Almaleki, D. (2021). Examinee characteristics and their impact on the psychometric properties of a multiple choice test according to the item response theory (IRT). *Engineering, Technology & Applied Science Research*. http://etasr.com/index.php/ETASR/article/view/4056

Alvarenga, M., Santos, T., & Andrade, D. (2020). Item Response Theory-based validation of a short form of the Disordered Eating Attitude Scale (DEAS-s) to a Brazilian sample. *Cadernos de Saúde Pública*.

https://www.scielo.br/j/csp/a/Z9QTdrr9nWjk78FB9fc4ctn/abstract/?lang=en

Ambiel, R. A., Noronha, A. P. P., & de Francisco Carvalho, L. (2015). Analysis of the professional choice self-efficacy scale using the Rasch-Andrich rating scale model. *International Journal for Educational and Vocational Guidance*, *15*(3), 205-219.

Ayala, R. D. (2018). Item response theory and Rasch modeling. *The reviewer's guide to quantitative methods in the …*. https://books.google.com/books?hl=en&lr=&id=8zz3DwAAQBAJ&oi=fnd&pg=PA145&dq=item+response+theory&ots=pNhCQ41K3I&sig=jVr_Z2zDW4DMSKPOTW63WyrfzCg

Ayis, S., Ayerbe, L., Ashworth, M., & Wolfe, C. (2018). Evaluation of the hospital anxiety and depression scale (HADS) in screening stroke patients for symptoms: item response theory (IRT) analysis. *Journal of affective disorders*. https://www.sciencedirect.com/science/article/pii/S0165032717311321

Baghaei, P. (2010). A comparison of three polychotomous Rasch models for super-item analysis. *Psychological Test and Assessment Modeling*, *52*(3), 313.

Bergstrom, B. A., & Lunz, M. E. (1998). Rating Scale Analysis: Gauging the Impact of Positively and Negatively Worded Items.

Bichi, A., & Talib, R. (2018). Item Response Theory: An Introduction to Latent Trait Models to Test and Item Development. *International Journal of Evaluation and Research in …*. https://eric.ed.gov/?id=EJ1183755

Blanchin, M., Guilleux, A., Hardouin, J., & ... (2020). … of structural equation modelling, item response theory and Rasch measurement theory-based methods for response shift detection at item level: A simulation study. *Statistical Methods in …*. https://doi.org/10.1177/0962280219884574

Bock, R., & Gibbons, R. (2021). *Item response theory*. John Wiley & Sons.

Bonifay, W. (2019). *Multidimensional item response theory*. books.google.com. https://books.google.com/books?hl=en&lr=&id=ijDEDwAAQBAJ&oi=fnd&pg=PT12&dq=item+response+theory&ots=Ad6eA3eXIk&sig=VthSaeCWHei8hV69-VzDlWNhT3s

Bradley, C., & Massof, R. W. (2018). An improved method for estimating measures from visual function questionnaires. *Investigative Ophthalmology & Visual Science*, *59*(9), 4144-4144.

Cadoret, G., Bigras, N., Duval, S., Lemay, L., & ... (2018). The mediating role of cognitive ability on the relationship between motor proficiency and early academic achievement in children. *Human movement …*. https://www.sciencedirect.com/science/article/pii/S0167945717304670

Carrozzino, D., Christensen, K. S., Mansueto, G., & Cosci, F. (2021). Construct validity of the Smoker Complaint Scale: A clinimetric analysis using Item Response Theory (IRT) models. *Addictive Behaviors*, *117*, 106849.

Caux, B. C. d., Lam, C., Lau, R., & ... (2017). Reflection for learning in doctoral training: Writing groups, academic writing proficiency and reflective practice. *Reflective …*. https://doi.org/10.1080/14623943.2017.1307725

Chen, S.-K., Hou, L., Fitzpatrick, S. J., & Dodd, B. G. (1997). The effect of population distribution and method of theta estimation on computerized adaptive testing (CAT) using the rating scale model. *Educational and psychological measurement*, *57*(3), 422-439.

Chen, Z., & Ahn, H. (2020). Item response theory based ensemble in machine learning. *International Journal of Automation and Computing*. https://doi.org/10.1007/s11633-020-1239-y

Cheng, S., Liu, Q., Chen, E., Huang, Z., Huang, Z., & ... (2019). DIRT: Deep learning enhanced item response theory for cognitive diagnosis. *Proceedings of the 28th …*. https://doi.org/10.1145/3357384.3358070

Chiesi, F., Morsanyi, K., Donati, M., & ... (2018). *Applying item response theory to develop a shortened version of the need for cognition scale*.

ncbi.nlm.nih.gov. https://www.ncbi.nlm.nih.gov/pmc/articles/pmc7171511/

Choi, Y., & Asilkalkan, A. (2019). R packages for item response theory analysis: Descriptions and features. *Measurement: Interdisciplinary Research …*. https://doi.org/10.1080/15366367.2019.1586404

Clark, D., Donnellan, M., Durbin, C., & ... (2020). Using item response theory to evaluate the Children's Behavior Questionnaire: Considerations of general functioning and assessment length. *Psychological …*. https://doi.org/10.1037/pas0000883

Colledani, D., Anselmi, P., & Robusto, E. (2018). *Using item response theory for the development of a new short form of the Eysenck Personality Questionnaire-Revised*. frontiersin.org. https://doi.org/10.3389/fpsyg.2018.01834

Cordier, R., Munro, N., Wilkes-Gillan, S., Speyer, R., & ... (2019). *Applying Item Response Theory (IRT) modeling to an observational measure of childhood pragmatics: The Pragmatics Observational Measure-2*. frontiersin.org. https://doi.org/10.3389/fpsyg.2019.00408

Cordier, R., Speyer, R., Schindler, A., Michou, E., & ... (2018). *Using rasch analysis to evaluate the reliability and validity of the swallowing quality of life questionnaire: an item response theory approach*. Springer. https://doi.org/10.1007/s00455-017-9873-4

Cotter, K., Chen, D., Christensen, A., & ... (2021). Measuring art knowledge: Item response theory and differential item functioning analysis of the Aesthetic Fluency Scale. *… , Creativity, and the …*. https://psycnet.apa.org/record/2021-45481-001

Crowe, M., Sleep, C., Carter, N., Campbell, W., & ... (2018). Self-esteem and narcissism: An item response theory analysis of curvilinearity. *Personality and …*. https://www.sciencedirect.com/science/article/pii/S0191886918300825

Curle, S., Yuksel, D., Soruç, A., & Altay, M. (2020). Predictors of English Medium Instruction academic success: English proficiency versus first language medium. *System*.

https://www.sciencedirect.com/science/article/pii/S0346251X20307387

Davier, M. v., Yamamoto, K., Shin, H., & ... (2019). Evaluating item response theory linking and model fit for data from PISA 2000–2012. *... : Principles, Policy & ...*. https://doi.org/10.1080/0969594x.2019.1586642

Dev, S., & Qiqieh, S. (2016). The relationship between English language proficiency, academic achievement and self-esteem of non-native-English-speaking students. *International Education Studies*. https://eric.ed.gov/?id=EJ1099424

Dimova, S. (2017). Life after oral English certification: The consequences of the Test of Oral English Proficiency for Academic Staff for EMI lecturers. *English for Specific Purposes*. https://www.sciencedirect.com/science/article/pii/S0889490616302137

Dinić, B., & Raine, A. (2019). An item response theory analysis and further validation of the Reactive–Proactive Aggression Questionnaire (RPQ): The Serbian adaptation of the RPQ. *Journal of Personality Assessment*. https://doi.org/10.1080/00223891.2019.1573430

Dougherty, L., Galano, M., Chad-Friedman, E., & ... (2021). Using item response theory to compare irritability measures in early adolescent and childhood samples. *...*. https://doi.org/10.1177/1073191120936363

Eckes, T., & Baghaei, P. (2015). Using testlet response theory to examine local dependence in C-tests. *Applied Measurement in Education*, *28*(2), 85-98.

Erath, K., Prediger, S., Quasthoff, U., & Heller, V. (2018). *Discourse competence as important part of academic language proficiency in mathematics classrooms: The case of explaining to learn and learning to ...*. Springer. https://doi.org/10.1007/s10649-018-9830-7

Fortune, T., & Song, W. (2016). Academic achievement and language proficiency in early total Mandarin immersion education. *Journal of Immersion and Content-Based ...*. https://doi.org/10.1075/jicb.4.2.02for

French, G. A., & Dodd, B. G. (1999). Parameter recovery for the rating scale model using PARSCALE. *Journal of outcome measurement*, *3*(2), 176-199.

Friedman, D., & Kagan, O. (2017). Academic writing proficiency of Russian heritage speakers: A comparative study. *Heritage Language Education*. https://doi.org/10.4324/9781315092997-13

Gebril, A., & Plakans, L. (2016). Source-based tasks in academic writing assessment: Lexical diversity, textual borrowing and proficiency. *Journal of English for Academic Purposes*. https://www.sciencedirect.com/science/article/pii/S1475158516300716

Geide-Stevenson, D. (2018). Does English proficiency affect academic performance? *International Review of Economics Education*. https://www.sciencedirect.com/science/article/pii/S1477388017300464

Gomez, R., Stavropoulos, V., Beard, C., & ... (2019). *Item response theory analysis of the recoded internet gaming disorder scale-short-form (IGDS9-SF)*. Springer. https://doi.org/10.1007/s11469-018-9890-z

Gorzycki, M., Howard, P., Allen, D., Desa, G., & ... (2016). An exploration of academic reading proficiency at the university level: A cross-sectional study of 848 undergraduates. *Literacy Research and …*. https://doi.org/10.1080/19388071.2015.1133738

Gottlieb, M. (2016). *Assessing English language learners: Bridges to educational equity: Connecting academic language proficiency to student achievement*. books.google.com. https://books.google.com/books?hl=en&lr=&id=RNCtCwAAQBAJ&oi=fnd&pg=PP1&dq=academic+proficiency&ots=b-VPw4SmjY&sig=lqxSNAykE48-yGUmH7mwkxdNpjo

Green, S. (2020). *Scaffolding academic literacy with low-proficiency users of English*. Springer. https://doi.org/10.1007/978-3-030-39095-2

Grewe, J., Kohler, D., Colver, M., & ... (2021). Improving university student course performance using reflective exam analysis: An approach to

developing academic proficiency. *... of Teaching and ...*. https://psycnet.apa.org/record/2021-42255-001

Haim, O. (2018). Academic proficiency performance in second and third language: the role of school context. *Language, Culture and Curriculum*. https://doi.org/10.1080/07908318.2018.1427759

Hays, R., Spritzer, K., & Reise, S. (2021). *Using item response theory to identify responders to treatment: Examples with the patient-reported outcomes measurement information system (PROMIS®) ...*. Springer. https://doi.org/10.1007/s11336-021-09774-1

Hosseinpour, N., Biria, R., & Rezvani, E. (2019). Promoting academic writing proficiency of Iranian EFL learners through blended learning. *Turkish Online Journal of ...*. https://dergipark.org.tr/en/pub/tojde/issue/49972/640525

Hu, R., & Trenkic, D. (2021). The effects of coaching and repeated test-taking on Chinese candidates' IELTS scores, their English proficiency, and subsequent academic achievement. *International Journal of Bilingual Education and ...*. https://doi.org/10.1080/13670050.2019.1691498

Immekus, J., Snyder, K., & Ralston, P. (2019). *Multidimensional item response theory for factor structure assessment in educational psychology research*. frontiersin.org. https://doi.org/10.3389/feduc.2019.00045

Kean, J., Bisson, E., Brodke, D., Biber, J., & Gross, P. (2018). An introduction to item response theory and rasch analysis: application using the Eating Assessment Tool (EAT-10). *Brain Impairment*. https://www.cambridge.org/core/journals/brain-impairment/article/an-introduction-to-item-response-theory-and-rasch-analysis-application-using-the-eating-assessment-tool-eat10/05EF29268426C47255D9B0C2A66403D5

Lang, J., & Tay, L. (2021). The science and practice of item response theory in organizations. *Annual Review of Organizational Psychology ...*. https://doi.org/10.1146/annurev-orgpsych-012420-061705

Lang, J. W., & Tay, L. (2021). The science and practice of item response theory in organizations. *Annual Review of Organizational Psychology and Organizational Behavior*, *8*, 311-338.

Lawrence, J., Knoph, R., McIlraith, A., & ... (2021). Reading Comprehension and Academic Vocabulary: Exploring Relations of Item Features and Reading Proficiency. *Reading Research …*. https://doi.org/10.1002/rrq.434

Levander, S. (2020). Construction of educational proficiency in academia: Peer review of educational merits in academic recruitment in sweden. *Education Inquiry*. https://doi.org/10.1080/20004508.2020.1843234

Levander, S., Forsberg, E., & Elmgren, M. (2019). *The meaning-making of educational proficiency in academic hiring: a blind spot in the black box*. Taylor & Francis. https://doi.org/10.1080/13562517.2019.1576605

Linacre, J. M. (1990). Modelling Rating Scales.

Linacre, J. M. (1999). Investigating rating scale category utility. *Journal of outcome measurement*, *3*(2), 103-122.

Linacre, J. M. (2010). Predicting responses from Rasch measures. *Journal of Applied Measurement*, *11*(1), 1.

Luo, Y., & Jiao, H. (2018). Using the Stan program for Bayesian item response theory. *Educational and psychological measurement*, *78*(3), 384-408.

Macdonald, K., Milne, N., Orr, R., & Pope, R. (2018). Relationships between motor proficiency and academic performance in mathematics and reading in school-aged children and adolescents: a systematic review. *International Journal of …*. https://www.mdpi.com/321148

Macdonald, K., Milne, N., Orr, R., & Pope, R. (2020). *Associations between motor proficiency and academic performance in mathematics and reading in year 1 school children: a cross-sectional study*. Springer. https://doi.org/10.1186/s12887-020-1967-8

McSweeney, M. (2017). I text English to everyone: Links between second-language texting and academic proficiency. *Languages*. https://www.mdpi.com/205410

Merwe, D. V. d. (2018). Aspects of academic language proficiency of intermediate phase teacher education students. *South African Journal of childhood education*. https://doi.org/10.4102/sajce.v8i1.555

Mezzadri, M. (2018). *Testing academic language proficiency*. books.google.com. https://books.google.com/books?hl=en&lr=&id=Wn9IDwAAQBAJ&oi=fnd&pg=PR5&dq=academic+proficiency&ots=4QgWthE6nk&sig=lv6brB-KBoW5ma-eXpUu0byYv7Y

Molla, E., & Muche, M. (2018). *Impact of cooperative learning approaches on students' academic achievement and laboratory proficiency in biology subject in selected rural schools …*. hindawi.com. https://www.hindawi.com/journals/edri/2018/6202484/

Nasseri, M. (2021). Is postgraduate English academic writing more clausal or phrasal? Syntactic complexification at the crossroads of genre, proficiency, and statistical modelling. *Journal of English for Academic Purposes*. https://www.sciencedirect.com/science/article/pii/S1475158520304744

Neumann, H., Padden, N., & ... (2019). Beyond English language proficiency scores: Understanding the academic performance of international undergraduate students during the first year of study. *... Education Research & ...*. https://doi.org/10.1080/07294360.2018.1522621

Nyarko, K., Kugbey, N., Kofi, C., Cole, Y., & ... (2018). English reading proficiency and academic performance among lower primary school children in Ghana. *SAGE ...*. https://doi.org/10.1177/2158244018797019

Sedoc, J., & Ungar, L. (2020). Item response theory for efficient human evaluation of chatbots. Proceedings of the First Workshop on Evaluation and Comparison of NLP Systems,

Shavelson, R. J., Baxter, G. P., & Pine, J. (1991). Performance assessment in science. *Applied Measurement in Education*, *4*(4), 347-362.

Wolfe, E. W., & Chiu, C. (1999). Measuring pretest-posttest change with a Rasch Rating Scale Model. *Journal of outcome measurement*, *3*(2), 134-161.

Wolfe, E. W., Chiu, C. W., & Reckase, M. D. (1999). Changes in secondary teachers' perceptions of barriers to portfolio assessment. *Assessing Writing*, *6*(1), 85-105.

# استخدام نموذج التقدير لأندريش في بناء مقياس للاتقان الأكاديمي لدى طلاب جامعه القاهرة دراسة سيكومترية

**عمرو محمد إبراهيم يوسف**

**مدرس علم النفس التربوى – كلية الدراسات العليا للتربية – جامعه القاهرة**

**المخلص**:

يهدف البحث الحالي إلى إعداد أداة لقياس مستوى الإتقان الأكاديمي لطلاب جامعة القاهرة، بحيث يكون لها خصائص القياس النفسي المقبولة، حيث تم تطبيق الأداة على طلاب الدراسات العليا – كلية العلاج الطبيعي – كلية التجارة بجامعة القاهرة، وفقًا لمقياس تصنيف أندريش وفقًا للبعد الأول لنظرية الاستجابة للفرد. حيث تم تطبيقه على عدد ٥٧٧ طالبا وطالبة حيث تم حذف ١٠٠ طالب ليصبح العدد ٤٧٧ حيث تم التأكد من صلاحية الأداة من خلال طريقة القياس التقليدية والحديثة بحيث تكون الأداة في شكلها النهائي يتكون من ٤٨ عنصرًا موزعة على خمسة أبعاد، وتم استخدام برنامج R والبرنامج Winsteps في إجراء التحليل.