Journal of Pedagogical Sociology and Psychology
Volume 5, Issue 2, 2023
https://doi.org/10.33902/jpsp.202320018

Research Article

Psychometric properties career of commitment instrument using classical test theory and graded response model

Muh. Asriadi Am[1,2], Farida Agus Setiawati[1], Samsul Hadi[1] and Edi Istiyono[1]

[1]*Educational Research and Evaluation, Yogyakarta State University, Yogyakarta, Indonesia;* [2]*Indonesian University of Education, Indonesia*

Correspondence should be addressed to Muh Asriadi Am asriadi190197@gmail.com
Received 1 February 2023; Revised 26 July 2023; Accepted 19 August 2023

A good instrument is one that is a measuring tool with good psychometric properties. This exploratory study aims to describe the psychometric properties of the career commitment instrument using the classical test theory approach and the graded response model. The data was obtained from a career commitment questionnaire with 12 statement items and 250 randomly selected respondents. The summated ratings (Likert) method was used for scaling, with five response options. Data were analyzed using R Studio's traditional Graded Response Model Theory Test technique. According to the findings, the career commitment instrument's quality has an estimated reliability of 0.77 (reliable) and a standard error measurement of 3.3. This instrument has a good Index of Endorsements and a Discrimination Index, with a classic and modern approach. Furthermore, the graded response model analysis revealed that 10 items fit and 2 did not. If given to respondents with low ability levels (θ=-2) to high ability levels (θ=2), this instrument provides complete information of 58.93 with a standard error of 1.0. As a result, an instrument is created that is ideal for companies to use in assessing their employees' career commitment.

Keywords: Psychometric properties, classical test theory, graded response model

1. Introduction

Individuals and organizations both benefit from career commitment. Several studies have found that career commitment positively affects performance (M. Najib & Aljanabi, 2020). Career commitment is also an occupational or professional commitment (Fu & Chen, 2015). Given the subtle differences in meaning, these terms are frequently used interchangeably. The term career commitment was chosen for this study because the term career does not only refer to a specific profession or job (Jones et al., 2006). Organizational commitment differs from career commitment. A worker may be dedicated to his job rather than his career (Zhu et al., 2021). Individuals dedicated to their careers are more focused on their careers than their working conditions, coworkers, or even the organization where they work (Cicek et al., 2016). Career commitment is the motivation, attitude and behaviour shown by individuals in a profession in undergoing and surviving in their chosen career(Sultana et al., 2016). Individuals may choose to continue working in an organization if they are dissatisfied with it due to career considerations (Singhal & Rastogi, 2018). Individuals who are highly committed to their careers are typically motivated by their hopes and career objectives (S.R, 2018). Based on the definition of career commitment above, career commitment is an attitude towards a profession or work that includes the development of personal career goals and identification and participation in these goals. Because career commitment motivates people to work hard to advance their careers, a tool is needed to measure it accurately and consistently.

Tests and questionnaires are two instruments commonly used to measure and evaluate variables. The test is a written assessment tool used to record or observe test takers' responses that align with the assessment target (Eleje et al., 2018). Furthermore, the test is a question or a set of tasks designed to gather information about an educational or specific psychological attribute (Debelak & Koller, 2020). The questionnaire evaluation technique assesses the child's overall

personality, including attitudes, behaviour, traits, social attitudes, and others. This instrument was developed through systematic observation, interviews, questionnaires, scale document analysis (both attitude and assessment scales), case studies, and psychometry (van der Lans et al., 2018). Based on this, the best instrument for measuring career commitment is a non-test instrument based on a questionnaire. The questionnaire method was chosen because it will be more efficient and effective in collecting large amounts of data at the right time.

After all courses, testing is required before obtaining a good instrument. There are several methods for analyzing instrument quality. The first method is the most common and has been widely used in education, particularly in research, namely classical test theory (CTT) and item response theory (IRT). The pure score theory is another name for the classical test theory. It is related to the focus of classical test theory research, which seeks to see the pure score of the visible score obtained (Azevedo et al., 2019). Traditional test theory is simpler and easier to grasp. In addition to its understanding, which does not necessitate in-depth knowledge of the statistical distribution function and its mathematical models, this theory has a high practical value in explaining reliability and validity problems. Furthermore, for measurements involving small respondents, such as daily tests in education or measurements in psychology in general, they continue to use the classical test theory approach (Foster, 2020). Nonetheless, this theory has several flaws, including the following: (1) the statistics of the instrument items are highly dependent on the characteristics of the subjects being tested; (2) the estimated ability of the respondent is highly dependent on the test items being tested; (3) the standard error of estimating scores apply to all test takers, so there is no standard error of measurement for each participant and item; and (4) the information presented is limited to answering questions (Yuan et al., 2021).

The reliability, standard error measurement, Index of endorsements, and differentiating power of career commitment instruments can be examined using classical test theory. In classical test theory, the factors that most influence the quality of instrument items are reliability, difficulty level index, and discrimination index. (Bellamkonda & Pattusamy, 2022). The desirability of item characteristics with the purpose and type of the test dramatically influences the instrument item quality. The Index of endorsements (p) in classical item analysis can be calculated in several ways, including: (1) a linear difficulty scale; (2) a bivariate scale; (3) the Davis index; and (4) the proportion of correct answers. The average scale or the proportion of correct answers (p), namely the number of test takers who answered correctly on the item being analyzed compared to the number of respondents, is the simplest and most widely used method (Widyaningsih et al., 2021). While analyzing the instrument as a whole reveals the instrument's reliability and standard measurement errors. The endorsement index is one of the instrument item parameters (Pi), namely the ratio between the complete answer and the number of respondents (Mamun et al., 2022). Indeks daya pembeda suatu butir instrumen berfungsi untuk menentukan dapat tidaknya suatu butir membedakan kelompok dalam aspek yang diukur sesuai dengan perbedaan yang ada pada kelompok itu. The discriminating power index of an instrument item determines whether or not the item can differentiate groups in aspects measured by differences within that group. The discriminating power study aims to examine the ability of specific instrument items to distinguish between respondents with high ability and respondents with low ability. In calculating discriminatory power, three methods are used: (1) discrimination index, (2) correlation index, and (3) alignment index (Sorenson & Hanson, 2021). In this study, differential power is measured using biserial point correlation. The correlation between instrument items and criteria that is not influenced by the Index of endorsements of instrument items is known as biserial correlation. The Index of the discriminating power of instrument items can be used to determine whether an item is good or bad.

Compensate for the shortcomings of classical test theory. The test must be supplemented with item response theory via the Graded Response Model (GRM). This GRM modelling exists to address flaws in classical test theory (Nur et al., 2020). The GRM model was chosen because it is well suited for items with categorical responses, such as the Likert scale. The GRM model does not require that each item have the same number of response categories. It does not apply to rating

scale models or any other IRT models (Rubio et al., 2007). The GRM model is an extension of the 2-PL Model in which each response category on an item is treated as a dichotomous item, with as many probability curves as response categories (Sethar et al., 2022). In GRM, the b-parameter value for each response category indicates the 50% chance that a randomly selected test taker whose ability level () exactly matches the b-parameter value will score x or higher (Reise et al., 2021).

The Index of endorsements, discrimination index, fit items, and information functions can be used to analyze the quality of career commitment instruments using the Graded Response Model. The endorsement index, discrimination index, and fit items were discovered to help describe the quality of each instrument item (Dai et al., 2021). The ability of the instrument or item to describe the information obtained is referred to as the item information function. Each measurement generates data about the measurement results (Jimam et al., 2019). The desired measurement information is not based on the individual being measured but rather on the measurement focus (Mateucci & Stracqualursi, 2006). This measurement data is based on the instrument's relationship with the individual.

A career commitment instrument that has been tested using traditional and modern techniques can be obtained as a result of this research. The novelty of this study is that it combines classic and modern instrument quality analysis techniques to provide a complete and comprehensive presentation of the psychometric properties of career commitment instruments. So, an agency or company can later use this career commitment instrument to measure its employees' career commitment accurately. As a result, this research aims to describe the psychometric properties of the career commitment instrument using Classical Test Theory (CTT) and Graded Response Model (GRM) approaches.

2. Method

This study takes the form of exploratory research (Creswell, 2012). The Classical Test Theory and Graded Response Model methods will be used in this study to investigate the psychometric properties of the career commitment instrument. The goal is to develop a standard instrument for measuring variables related to career commitment.

2.1. Participants

Research data obtained by research Ingarianti et al. (2019). In this study, 250 people were chosen randomly using a simple random sampling technique. Gender categories and work groups are used to categorize respondents. There were 150 men and 100 women in the gender category. There are 185 employees, 35 self-employed individuals, 18 owners, and 12 investors in the workgroup category.

2.2. Instruments

The Commitment Career Measure (CCM) developed by Carson & Bedeian (1994) was used as a measuring tool in this study. Career commitment has three major dimensions (Carson & Bedeian, 1994); career identity describes an individual's emotional relationship with his chosen career. Career planning includes the process by which individuals determine career development needs and set career goals. Career resilience, which measures a person's persistence in achieving career goals. The summated ratings (Likert) method is used in the career commitment scale, with five response options: strongly disagree (score 1), disagree (score 2), neutral (score 3), agree (score 4), and strongly agree (score 5). (Tabaku & Cerri, 2016). Table 1 describes the measurement model for the Career Commitment variable.

Table 1
Career Commitment Variable Measurement Model

No	*Dimension*	*Item Code*
1	Planning	B1
		B2
		B3
		B4
2	Identity	B5
		B6
		B7
		B8
3	Resilience	B9
		B10
		B11
		B12

2.3. Data Analysis

The application performs classical test analysis, and the Graded Response Model uses R Studio software. The description of each psychometric properties analysis is described as follows.

2.3.1. Reliability estimation

The Alpha Cronbach formula is used to calculate the quality of the instrument items based on the reliability of the statement items in the classical test theory approach (KR-20). Statement items are regarded as reliable if they meet the criteria for the instrument reliability coefficient, as shown in Table 2.

Table 2
Reliability Criteria (adoptef from Rogers & Badham, 2003)

Reliability Value	*Interpretation*
0.0-0.20	Less reliable
>0.20-0.40	Moderately Reliable
>0.40-0.60	Pretty Reliable
>0.60-0.80	Reliable
>0.80-1.00	Very Reliable

2.3.2. Index of endorsements criteria

The Index of Endorsements indicates how many respondents can answer an item with the highest possible score. The Index of Endorsements is interpreted in classical measurement theory using the following criteria (Nima et al., 2020).

Table 3
Category Index of Endorsements classical test theory

Index of Endorsements (IE)	*Interpretation*
IE = 0,00	Very Low
0,00 < IE ≤ 0,30	Low
0,30 < IE ≤ 0,70	Medium
0,70 < IE ≤ 1,00	High
IE = 1	Very High

In the GRM, the item index of endorsements (b) is defined as a point or location on a capability scale where the shaped curve has the steepest slope, the magnitude of which ranges from logit - to logit +, but is usually only -2 logit to 2 logs, making it neither too easy nor too difficult for the

intended test subject. (Hambleton & Swaminathan, 1985; Linden & Hambleton, 1997). As a result, in this study, items are said to have a low level of difficulty (easy items) if b -2.0 logit, a medium level of difficulty (medium item) if b -2.0 logit b 2.0 logit, and a high level of difficulty (high item) if b > 2.0 logit. As a result, the instrument item is said to be "good" if it has an endorsement index of -2 logit ≤ b ≤ +2 logit (Polat, 2022).

2.3.3. Discrimination index criteria

The discrimination index measures an item's ability to distinguish between respondents with high and low ability to answer questions. In discriminating index research, it can be seen from the Pearson Correlation value. The discriminating Index of instrument items can be divided into four categories based on classical test theory (Himelfarb, 2019), as shown in Table 4.

Table 4
Categories of Discriminating Index of Classical Theory Test Items

Discrimination Index (DI)	*Interpretation*
DI ≥ 0,70	Very Good
0,40 ≤ DI < 0,70	Good
0,20 ≤ DI < 0,40	Enough
DI < 0,20	Bad

In contrast to the typical test theory approach, the instrument items are represented as "a" in the IRT approach with the GRM model. This value of an is theoretically between – ∞ and + ∞. The fundamental value positively correlates with performance on items, with the ability being measured on suitable items and ai between 0 and 2 (Hambleton et al., 1991).

2.3.4. Item fit and information function criteria

Graded Response Model analysis includes item fit analysis and the information function. The chi-square value and the Root Mean Square Error of Approximation (RMSEA) value show the criteria used to determine fit items (Hair et al., 2017). This study used the RMSEA value to determine which items fit. RMSEA is an index value used in large samples to correct the chi-square statistic. The index value that is categorized as acceptable is RMSEA ≤ 0,08 (Kline, 2011). The item and instrument information functions in the Graded Response Model can be seen from the TotalInfo and Proportion values, as well as the Item Information Function (IFF) and Total Information Function (TIF) graphs (Silvia et al., 2021). The greater the peak information that can be obtained, the greater the information value that the item or instrument can provide from the measurements taken.

3. Results and Discussions

3.1. Results of Instrument Quality Analysis with Classical Test Theory

3.1.1. Summary statistic of classical test theory

The recapitulation of the results of the analysis of measuring the quality of career commitment instruments in general with the classical test theory approach can be seen in the table below.

Table 5
Results of Classical Test Theory Analysis

Parameter	*Value*
Number of Items	12
Number of Respondents	250
Reliability (Alpha)	0.77
ScaleMean	41.54
ScalesSD	6.92
Standard Error Measurement (SEM)	3.3

According to Table 5, the processed data consists of 12 statement items from 250 respondents. The analysis using classical test theory yielded a reliability coefficient value of 0.77, which falls into the Reliable category. Furthermore, the Standard Error Measurement value is known to be 3.3.

3.1.2. Index of endorsements with classical test theory

The Index of endorsements indicates the magnitude of the possibility of how well the respondent understands the instrument statement items to answer each statement item correctly. The item difficulty index is interpreted according to the following criteria in classical item measurement theory.

Table 6
Results of Index of Endorsements Distribution with Classical Test Theory

Item Code	*Index of Endorsements*	*Interpretation*
B1	0.814	High
B2	0.818	High
B3	0.582	Medium
B4	0.788	High
B5	0.614	Medium
B6	0.808	High
B7	0.602	Medium
B8	0.586	Medium
B9	0.656	Medium
B10	0.661	Medium
B11	0.717	High
B12	0.662	Medium

According to Table 6, the Index of endorsements analyzed using the classical test theory approach had five items in the high category and seven in the medium category. This table also shows that the overall endorsement index for career commitment instruments is in the medium range. Furthermore, it is known that item B2 has the highest Index of endorsements, while item B3 has the lowest Index of endorsements.

3.1.3. Discrimination Index with Classical Test Theory

In discriminating index research, it can be seen from the Pearson correlation value. The results of the classical test theory-based calculation of the discriminating Index of the items can be broadly classified into three categories, as shown in Table 7.

Table 7
Item Discriminating Index with Classical Theory Tests

Item Code	*Pearson Correlation*	*Interpretation*
B1	0.79	Very Good
B2	0.79	Very Good
B3	0.74	Very Good
B4	0.79	Very Good
B5	0.72	Very Good
B6	0.8	Very Good
B7	0.73	Very Good
B8	0.71	Very Good
B9	0.74	Very Good
B10	0.73	Very Good
B11	0.76	Very Good
B12	0.74	Very Good

According to Table 7, the analysis results of the discrimination index of all items are in the excellent category, with a Pearson Correlation value of 0.7. It indicates that the career commitment instrument has a high level of differentiating power. So that respondents do not answer incorrectly and can distinguish each item from the career commitment instrument.

3.2. Results of Instrument Quality Analysis with the Graded Response Model

3.2.1. Index of endorsements with Graded Response Model

A slice of the response category index of endorsements will be presented in the analysis of the Index of endorsements with the Graded Response Model. Because there are five response categories, the Index of endorsements is divided into four slices in the analysis (b1, b2, b3, b4). Furthermore, the location column contains the average Index of endorsements for all items across all category slices. Table 8 shows how to find each item's Index of endorsements classification.

Table 8
Index of Endorsements Results with Graded Response Model

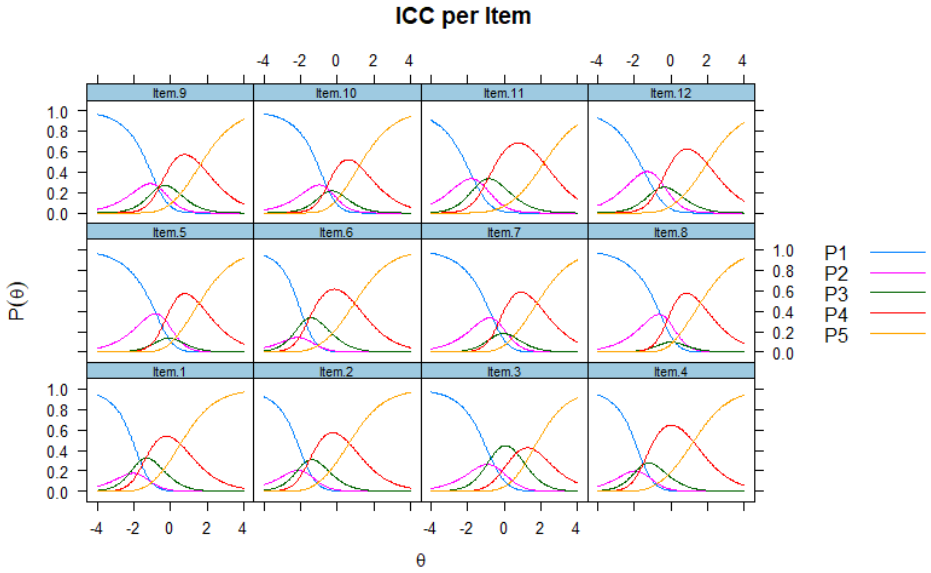
Item Code	*b1*	*b2*	*b3*	*b4*	*Location*	*Description*
B1	6.880003	5.358526	2.501583	-1.28672	3.363349	High
B2	7.380196	5.562963	2.733938	-1.40047	3.569157	High
B3	-1.56093	-0.601	0.89381	2.116624	0.212127	Medium
B4	7.133756	5.36644	2.678455	-2.51852	3.165034	High
B5	-1.18355	-0.24933	0.14747	1.282211	-0.0008	Low
B6	5.794413	4.710392	2.16263	-1.47037	2.799266	High
B7	-1.14094	-0.27874	0.242279	1.423149	0.061437	Medium
B8	-0.97023	-0.06426	0.206603	1.269618	0.110433	Medium
B9	-2.45794	-1.25236	-0.02527	2.41986	-0.32893	Low
B10	-1.52198	-0.73291	-0.05934	1.355785	-0.23961	Low
B11	-6.42449	-3.67731	-1.05404	4.377917	-1.69448	Low
B12	-2.27824	-0.93312	-0.03503	1.978006	-0.3171	Low

According to the results of the Index of endorsements test with the GRM shown in Table 8, five items are in a low category, three in the medium category, and four in the high category. In Item B11, the Index of endorsements or threshold 1 (b1) = -6.42449 means that the respondent must have a minimum ability of -6.42449 to complete category two after category 1. Threshold 2 (b2) = -3.67731 means that the respondent must have a minimum ability of -3.67731 to complete category three after category 2. Threshold 3 (b3) = -1.05404 means that the respondent must have a minimum ability of -1.054. Threshold 4 (b4) = 4.377917 means that respondents must have a minimum ability of 4.377917 to complete category five after category 4. Aside from that, the overall category endorsement index is -1.69448. As a result, it is also known that Item B11 has the lowest endorsement index.

In Item B2, the Index of endorsements or threshold 1 (b1) = 7.380196 means that the respondent must have a minimum ability of 7.380196 to complete category two after category 1. Threshold 2 (b2) = 5.562963 means that the respondent must have a minimum ability of 5.562963 to complete category three after category 2. Threshold 3 (b3) = 2.733938 means that the respondent must have a minimum ability of 2.733938 to complete category four after category 3. Aside from that, the overall endorsement index is 3.569157. So it is also known that Item B2 is the item that has the highest Index of endorsements.

In general, the statement items used to assess career commitment can be done well because they are simple for respondents to complete. As a result, the statement items created meet the ideal criteria for measuring career commitment. The Item Characteristic Curve figure is shown below to reinforce the analysis results.

Figure 1
Item Characteristic Curve



According to Figure 1, Item Characteristic Curves (ICC) show that statement item with a range of -1.69 to -0.0008 have a low index of endorsements. Statement items with a range of 0.06 to 0.21 have an index of endorsements in the medium range. Statement items with a score ranging from 3.16 to 3.57 have a high index of endorsements.

3.2.2. Discrimination index with graded response model

Based on the R studio output that refers to the parameter "a," the discrimination index of the instrument from the Graded Response Model is seen. Table 9 shows the findings of the analysis.

Table 9
Discrimination Index Results with Graded Response Model

Item Code	*a*	*Description*
B1	-0.62079	Not Good
B2	-0.61802	Not Good
B3	1.580409	Good
B4	-0.56122	Not Good
B5	3.561226	Not Good
B6	-0.80347	Not Good
B7	3.260093	Not Good
B8	4.149299	Not Good
B9	1.01252	Good
B10	1.897242	Good
B11	0.573511	Good
B12	1.574507	Good

According to Table 9, there are seven items with not good discrimination index, namely items B1, B2, B4, B5, B6, B7 and B8, and five items with a discrimination index that falls into the excellent category. Generally, the statement items on the career commitment instrument used to measure career commitment have a low discrimination index.

3.2.3. Item fit level

The item fit level of this item is used to determine the item's accuracy with the Model or item fit. The item fit level explains whether or not our item has a normal measuring function. If items are

not appropriate, it indicates that the respondent misunderstood the item statement. The RMSEA column displays item fit. Table 10 shows the results of the item fit analysis.

Table 10
Item Fit Test Results

Item Code	*RMSEA.S_X2*	*Description*
B1	0.068	Fit
B2	0.06	Fit
B3	0.043	Fit
B4	0.05	Fit
B5	0.082	Not Fit
B6	0.051	Fit
B7	0.08	Fit
B8	0.081	Not Fit
B9	0.049	Fit
B10	0.037	Fit
B11	0.071	Fit
B12	0.053	Fit

Based on the data in Table 10, it is clear that there are ten statement items that fit and only two that do not. B5 and B8 are items that do not fit. Items that do not fit will be corrected or removed from the instrument because they can cause measurement bias or errors. In general, the recommended career commitment instrument comprises only ten statement items that can be used to assess career commitment.

3.2.4. Item information function

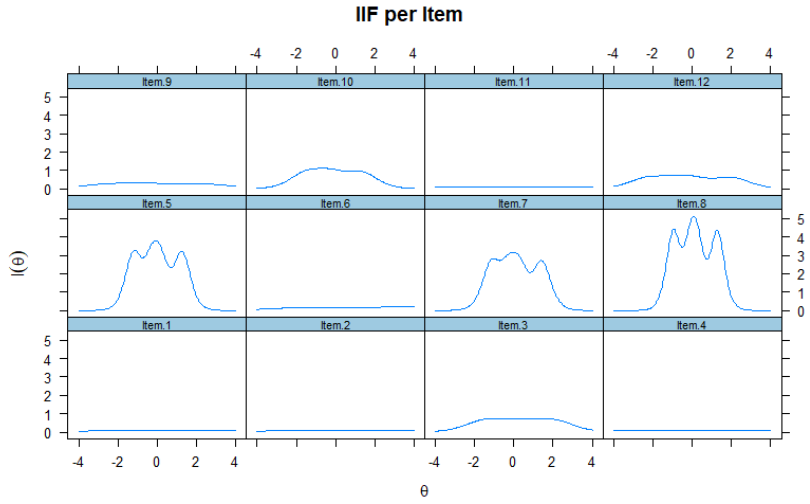
Each measurement yields information about the measurement's outcome. The information function is one factor that influences an instrument's quality in the Graded Response Model. The information function will indicate to whom this instrument is best suited. The desired measurement information does not concern the individual being measured but rather the focus of measurement, particularly the relationship between the instrument and the respondent. The information function indicates the reliability of the measurements. Table 11 shows the measurement information function for each item.

Table 11
Item Information Function

Item Code	*Info (-4,4)*	*TotalInfo*	*Proportion*
B1	0.800302	1.5549	0.514697
B2	0.786859	1.600906	0.491508
B3	4.171754	4.281188	0.974438
B4	0.667855	1.43073	0.466793
B5	11.20973	11.21011	0.999966
B6	1.269849	2.124925	0.597597
B7	10.21751	10.21853	0.9999
B8	13.02205	13.02211	0.999995
B9	2.153476	2.499811	0.861456
B10	4.890037	4.919603	0.99399
B11	0.717915	1.572535	0.456534
B12	4.333799	4.494601	0.964223

Table 11 shows that the items that provide the most information are B8, B7, and B5. B11, B4, B2, and B1 provide minor information. Figure 2 shows the overall information function of the items, which supports the findings.

Figure 2
Item Information Function

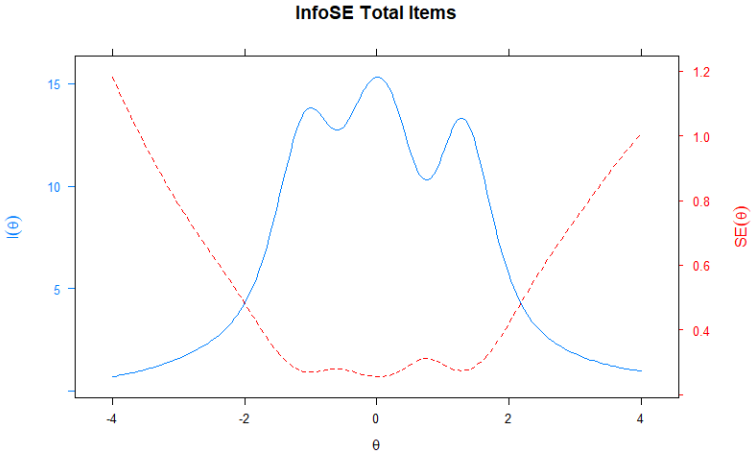


Items B5, B7, and B8 in figure 2 represent an exemplary graph of information items. Items B1, B2, B4, B6, B9, and B11 could provide better information. In the meantime, items B10 and B12 can still provide information, albeit not optimally. According to these findings, the items presented in the instrument still need to be studied in terms of construct or language because the items that can provide the most information are still few.

3.2.5. Total information function

The total information function value can be used to explain the magnitude of the combined contribution of the instrument items in revealing response patterns. The information function will be beneficial in determining which items are appropriate for the Model, allowing items to be selected more easily. The test information function is the sum of all the test information functions. As a result, if the constituent items' information function is high, the information function tested will be classified as high, and vice versa. The total information function also calculates the measurement error. Better the instrument and the lower the measurement error, the higher the total information function. Figure 3 shows the results of the Total Information Function analysis.

Figure 3
Total Information Function



According to Figure 3, the career commitment instrument that provides the most information is 58.93, with a standard error of 1.0 when administered to respondents with low to high ability

levels. The lower limit of the interval is theta -2.0 (low-ability respondents), and the upper limit is theta 2.1 (respondents with high ability). These findings indicate that the instrument performs well in the ability range of -2.0 to 2.1. The instrument is said to be reliable for use, with respondents ranging in ability from low (-2) to high (+2.1). These findings suggest that the item statement is appropriate for determining the level of ability of respondents with low to high abilities. It gives the impression that respondents with low or high abilities will find it simple to use this instrument. This finding is significant for researchers because this instrument can obtain the most information from all ability groups when measuring career commitment.

4. Discussion

In this study, the psychometric properties of career commitment instruments based on classical test theory can be seen from several aspects: instrument reliability, standard error measurement, Index of endorsement, and discrimination index. Instrument reliability. Standard error measurement and d discrimination index are in a suitable category. However, in the Index of endorsement aspect, items with a low endorsement index are items B1, B2, B4, B6, and B11. Items B1, B2, and B4 are in the Planning dimension. Item B6 is an item that is in the Identity dimension. The description of the psychometric properties of the career commitment instrument based on the graded response model can be seen from several aspects, namely the Index of endorsement, discrimination index, item fit level, and information function. In the Index of endorsement aspect, items with a low endorsement index are items B5, B9, B10, B11, and B12. Item B5 is an item that is in the Planning dimension. Items B9, B10, B11, and B12 are in the Resilience dimension. In the aspect of the discrimination index, items with a poor discrimination index are items B5, B7, and B8.

Items B7 and B8 are items that are in the Identity dimension. In the item-appropriate level aspect, the instrument items that are not fit are items B5 and B8. In the item information function (IFF) aspect, items with low IFF are items B1, B2, B3, B4, B5, B6, B9, B10, B11, and B12. Items B1, B2, B3, and B4 are items that are in the Planning dimension. Items B5, B6, B7, and B8 are items in the Identity dimension. Items B9, B10, B11, and B12 are in the Resilience dimension. These findings indicate that overall items are only B3 items which are relatively stable or have a good value in all aspects.

Meanwhile, the other items need to be improved in each aspect. Therefore it is necessary to check the compiled instruments again. Theoretically, it can be solved by focusing on three aspects of testing, namely material aspects, construct aspects, and language or cultural aspects. In addition, the reliability aspect and the function of information can be emphasized in research. As one type of item characteristic, of course, it is highly desired that the information function value of the test be maximal, and the tests analyzed with classical test theory want a high-reliability value. However, conceptually there is a difference between the reliability in the concept of classical test theory and the information function test by IRT (Himelfarb, 2019). In classical test theory, the item scores that make up the reliability coefficient of the test are not independent of one another (Foster, 2020). Changes in just one item will change all values on the reliability coefficient. It is not the case for the test information function. In IRT, items are independent of one another, so changing an item only changes the information function of the test and does not change the value of the information function of the other items. According to Hambleton & Swaminathan (1985), the measurement of the test information function is more accurate when compared to the use of reliability because: (1) the form depends only on the items in the test, (2) it has an estimated measurement error at each level of ability. The information function of IRT is inversely related to uncertainty. It means that the higher the uncertainty, the lower the value of the information function test (Dorfman & Kalugin, 2020). Conversely, the lower the uncertainty, the higher the value of the information function test (Hu et al., 2022).

This study's findings also show a significant difference between the item response theory and the graded response model in terms of assessing good or fit items. Point of view of item response theory on the Index of endorsement aspect, the trend of items is in the medium and high categories. Meanwhile, in the aspect of the discrimination index, all items are in an excellent

category. It contrasts the Index of endorsement aspect in the graded response model point of view because the items are distributed into three categories, namely low, medium and high. Items that fall into the medium category in the grain response theory are instead items in the weak category in the graded response model. Whereas in the discrimination index, from the view of the graded response model, the item tends to be divided into two, namely, good and evil. If we look at it more in-depth, it turns out that the classical statistical test item theory depends on the characteristics of the respondents who fill out the instrument. The estimated ability of the respondents is very dependent on the items worked on, and the information presented is limited to the form of the answers given without regard to the pattern of respondents' answers (Scotti di Uccio et al., 2019).

Whereas in the multilevel response model, the parameters of the item items and the test takers do not influence each other, making it possible to see the contribution of the item items when the item items are added or subtracted by the test kit. In addition, the standard Error of Measurement (SEM) has different values between scores (or response patterns) but is common among populations (Debelak & Koller, 2020). Referring to this comparison, the researcher concludes that these two instrument test formats complement one another. Even if we review it again based on the item response theory and the graded response model, the career commitment instrument still needs to improve, especially in terms of the quality of the statement items. Therefore, using these two types of instrument test analysis will improve the quality of the career commitment instrument.

The findings in this study are supported by the research of Bellamkonda & Pattusamy (2022), that utilizing classical test theory and item response theory can produce excellent instruments. It is also in line with research conducted by Yuan et al. (2021) that the classical test theory and item response theory complement each other in producing high-quality instrument items.

5. Conclusion

The results showed that the quality of the career commitment instrument through the classical test theory approach obtained a reliability coefficient of 0.77, which was included in the reliable category and a standard error measurement of 3.3. The Index of endorsement with classical test theory has five items in the low category and seven in the medium category. At the same time, the Index of endorsement with the graded response model obtained five items in the low category, three in the medium category, and four in the high category. Discrimination index analysis using classical test theory shows that all items are in the excellent category. In contrast, the discrimination index with a graded response model shows five good items and seven not-good items. With the graded response model analysis, 10 fit items and two not fit items were also obtained. Items that provide complete information in making a career commitment are items B8, B7, and B5. The career commitment instrument provides 58.93 pieces of information with a standard error of 1.0 when given to respondents with low ability levels (theta -2.0) to respondents with high abilities (theta 2.1). The overall analysis of both the classical test theory approach and the graded response model supports that the career commitment instrument has good psychometric properties. So it is feasible to use to measure career commitment. Even so, this study still has limitations. Namely, several items still have to be reviewed in terms of construct and language because they still have an index of endorsement, discrimination index, item fit and information functions that still need to be improved or not optimal. Recommendations that can be given from this study are that researchers can add dimensions or factors based on career commitment theory from different experts or literature. Researchers also suggest that users or future researchers consider or review items that still need improvement if they feel these items can cause bias in measurement.

Author contributions: All authors are agreed with the results and conclusions.

Declaration of interest: No conflict of interest is declared by authors.

Funding: No funding source is reported for this study.

References

Azevedo, J. M., Oliveira, E. P., & Beites, P. D. (2019). Using learning analytics to evaluate the quality of multiple-choice questions: a perspective with classical test theory and item response theory. *International Journal of Information and Learning Technology*, *36*(4), 322–341. https://doi.org/10.1108/IJILT-02-2019-0023

Bellamkonda, N., & Pattusamy, M. (2022). Validation of fear of COVID-19 scale in India: Classical test theory and item response theory approach. *International Journal of Mental Health and Addiction*, *20*(4), 2400–2407. https://doi.org/10.1007/s11469-021-00521-2

Carson, K. D., & Bedeian, A. G. (1994). Career commitment: Construction of a measure and examination of its psychometric properties. *Journal of Vocational Behavior*, *44*(3), 237–262. https://doi.org/10.1006/jvbe.1994.1017

Cicek, I., Karaboga, T., & Sehitoglu, Y. (2016). A New Antecedent of Career Commitment: Work to Family Positive Enhancement. *Procedia - Social and Behavioral Sciences*, *229*, 417–426. https://doi.org/10.1016/j.sbspro.2016.07.152

Creswell, J. W. (2012). *Educational Research: planning, conducting, and evaluating quantitative and qualitative research.* Pearson.

Dai, S., Vo, T. T., Kehinde, O. J., He, H., Xue, Y., Demir, C., & Wang, X. (2021). Performance of polytomous IRT models with rating scale data: an investigation over sample size, instrument length, and missing data. *Frontiers in Education*, *6*(September), 1–18. https://doi.org/10.3389/feduc.2021.721963

Debelak, R., & Koller, I. (2020). Testing the local independence assumption of the Rasch model with Q3-Based nonparametric model tests. *Applied Psychological Measurement*, *44*(2), 103–117. https://doi.org/10.1177/0146621619835501

Dorfman, L. Y., & Kalugin, A. Y. (2020). Resources, potentials and academic achievements of students: Part 1. Differentiation of resources and potentials [Соотношение ресурсов, потенциалов и академических достижений студентов сообщение 1. дифференциация ресурсов и потенциалов]. *Obrazovanie i Nauka*, *22*(4), 64–88. https://doi.org/10.17853/1994-5639-2020-4-64-88

Eleje, L. I., Onah, F. E., & Abanobi, C. C. (2018). Comparative study of Classical Test Theory and Item Response Theory using diagnostic quantitative economics skill test item analysis results. *European Journal of Educational & Social Sciences*, *3*(1), 57–75.

Foster, R. C. (2020). A generalized framework for classical test theory. *Journal of Mathematical Psychology*, *96*, 102330. https://doi.org/10.1016/j.jmp.2020.102330

Fu, J. R., & Chen, J. H. F. (2015). Career commitment of information technology professionals: The investment model perspective. *Information and Management*, *52*(5), 537–549. https://doi.org/10.1016/j.im.2015.03.005

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2017). Multivariate Data Analysis (MVDA). In *Pearson*. https://doi.org/10.1002/9781118895238.ch8

Hambleton, R. K., & Swaminathan, H. (1985). Item Response theory: principles and applications. *Springer.*

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). Fundamentals of item response theory. In *SAGE Publications, Inc.* (Vol. 29, Issue 07). https://doi.org/10.5860/choice.29-4185

Himelfarb, I. (2019). A primer on standardized testing: History, measurement, classical test theory, item response theory, and equating. *Journal of Chiropractic Education*, *33*(2), 151–163. https://doi.org/10.7899/JCE-18-22

Hu, S., Cai, W., Gao, T., & Wang, M. (2022). An automatic residual-constrained and clustering-boosting architecture for differentiated heartbeat classification. *Biomedical Signal Processing and Control*, *77*. https://doi.org/10.1016/j.bspc.2022.103690

Ingarianti, T. M., Fajrianthi, & Purwono, U. (2019). Adaptasi Instrumen Komitmen Karier. *Jurnal Psikologi*, *18*(2), 199–217.

Jimam, N. S., Ahmad, S., & Ismail, N. E. (2019). Psychometric Classical theory test and item response theory validation of patients' knowledge, attitudes and practices of uncomplicated Malaria instrument. *Journal of Young Pharmacists*, *11*(2), 186–191. https://doi.org/10.5530/jyp.2019.11.39

Jones, M. L., Zanko, M., & Kriflik, G. (2006). On the antecedents of career commitment. *Proceedings of the Australian and New Zealand Academy of Management Conference*, 1–22.

Kline, R. B. (2011). Principles and practice of structural equation modeling. In *The Guilford Press* (Vol. 245). The Guilford Press. https://doi.org/10.1097/00003086-198908000-00042

Linden, W. J. van der, & Hambleton, R. K. (1997). Handbook of Modern Item Response Theory. In *Springer Science+Business Media, LLC*. Springer Science+Business Media, LLC. https://doi.org/10.1007/978-1-4757-2691-6

M.Najib, H. H., & Aljanabi, A. R. A. (2020). The Mediation Role of Career Adaptability between Career Commitment and Career Motivation: An Empirical Investigation. *Information Management and Business Review*, *12*(1), 27–40.

Mamun, M. A., Alimoradi, Z., Gozal, D., Manzar, M. D., Broström, A., Lin, C. Y., Huang, R. Y., & Pakpour, A. H. (2022). Validating insomnia severity index (Isi) in a bangladeshi population: Using classical test theory and rasch analysis. *International Journal of Environmental Research and Public Health*, *19*(1). https://doi.org/10.3390/ijerph19010225

Mateucci, M., & Stracqualursi, L. (2006). Student assessment via graded response model. *STATISTICA*, *66*(4), 435–447.

Nima, A. Al, Cloninger, K. M., Lucchese, F., Sikström, S., & Garcia, D. (2020). Validation of a general subjective well-being factor using Classical Test Theory. *PeerJ*, *2020*(6), 1–29. https://doi.org/10.7717/peerj.9193

Nur, L., Nurani, L. A., Suryana, D., & Ahmad, A. (2020). Rasch model application on character development instrument for elementary school students. *International Journal of Learning, Teaching and Educational Research*, *19*(3), 437–459. https://doi.org/10.26803/ijlter.19.3.24

Polat, M. (2022). Comparison of Performance measures obtained from foreign language tests according to item response theory vs classical test theory. *International Online Journal of Education and Teaching*, *9*(1), 471–485.

Reise, S. P., Du, H., Wong, E. F., Hubbard, A. S., & Haviland, M. G. (2021). Matching IRT Models to patient-reported outcomes constructs: the graded response and log-logistic models for scaling depression. *Psychometrika*, *86*(3), 800–824. https://doi.org/10.1007/s11336-021-09802-0

Rogers, G., & Badham, L. (2003). Evaluation in schools: Getting started with training and implementation. In *Routledge*. Routledge. https://doi.org/10.4324/9780203393314

Rubio, V. J., Aguado, D., Hontangas, P. M., & Hernández, J. M. (2007). Psychometric properties of an emotional adjustment measure: An application of the graded response model. *European Journal of Psychological Assessment*, *23*(1), 39–46. https://doi.org/10.1027/1015-5759.23.1.39

S.R, D. (2018). A Study on career commitment of teaching profession in Chennai City. *Journal of Management (JOM)*, *5*(3), 45–51.

Scotti di Uccio, U., Colantonio, A., Galano, S., Marzoli, I., Trani, F., & Testa, I. (2019). Design and validation of a two-tier questionnaire on basic aspects of quantum mechanics. *Physical Review Physics Education Research*, *15*(1), 1–25. https://doi.org/10.1103/physrevphyseducres.15.010137

Sethar, W. A., Pitafi, A., Bhutto, A., Nassani, A. A., Haffar, M., & Kamran, S. M. (2022). Application of Item Response Theory (IRT)-Graded Response Model (GRM) to Entrepreneurial Ecosystem Scale. *Sustainability (Switzerland)*, *14*(9), 1–27. https://doi.org/10.3390/su14095532

Silvia, P. J., Rodriguez, R. M., Beaty, R. E., Frith, E., Kaufman, J. C., Loprinzi, P., & Reiter-Palmon, R. (2021). Measuring everyday creativity: A Rasch model analysis of the Biographical Inventory of Creative Behaviors (BICB) scale. *Thinking Skills and Creativity*, *39*, 100797. https://doi.org/10.1016/j.tsc.2021.100797

Singhal, H., & Rastogi, R. (2018). Psychological capital and career commitment: the mediating effect of subjective well-being. *Management Decision*, *56*(2), 458–473. ,https://doi.org/10.1108/MD-06-2017-0579

Sorenson, B., & Hanson, K. (2021). Using Classical Test Theory and Rasch Modeling to Improve General Chemistry Exams on a per Instructor Basis. *Journal of Chemical Education*, *98*(5), 1529–1538. https://doi.org/10.1021/acs.jchemed.1c00164

Sultana, R., Yousaf, A., Khan, I., & Saeed, A. (2016). Probing the interactive effects of career commitment and emotional intelligence on perceived objective/subjective career success. *Personnel Review*, *45*(4), 724–742. https://doi.org/10.1108/PR-11-2014-0265

Tabaku, E., & Cerri, S. (2016). An Assessment of Service Quality and Customer Satisfaction in the Hotel Sector. *Tourism & Hospitality Industry 2016, 12*(1), 480–489.

van der Lans, R. M., van de Grift, W. J. C. M., & van Veen, K. (2018). Developing an Instrument for teacher feedback: using the Rasch model to explore teachers' development of effective teaching strategies and behaviors. *Journal of Experimental Education*, *86*(2), 247–264. https://doi.org/10.1080/00220973.2016.1268086

Widyaningsih, S. W., Yusuf, I., Prasetyo, Z. K., & Istiyono, E. (2021). The development of the hots test of physics based on modern test theory: Question modeling through e-learning of moodle lms. *International Journal of Instruction*, *14*(4), 51–68. https://doi.org/10.29333/iji.2021.1444a

Yuan, T., Honglei, Z., Xiao, X., Ge, W., & Xianting, C. (2021). Measuring perceived risk in sharing economy: A classical test theory and item response theory approach. *International Journal of Hospitality Management*, *96*(April), 102980. https://doi.org/10.1016/j.ijhm.2021.102980

Zhu, D., Kim, P. B., Milne, S., & Park, I. J. (2021). A Meta-Analysis of the Antecedents of Career Commitment. *Journal of Career Assessment*, *29*(3), 502–524. https://doi.org/10.1177/1069072720956983