International Journal of Advances in Psychology Research
ISSN: 2320-7485 (Online)
Journal DOI:10.7439/ijapr
Research Article

Assessing the Unidimensional Structure of a Scale: Exploratory Factor Analysis, Confirmatory Factor Analysis or Rasch Measurement?

Panayiotis Panayides*

Secondary Mathematics Education, Cyprus

***Correspondence Info:**
Panayiotis Panayides
Secondary Mathematics Education, Cyprus
E-mail: p.panayides@cytanet.com.cy

Abstract

This study compares three methods of assessing the dimensionality of a scale: Exploratory Factor Analysis (EFA), Confirmatory Factor Analysis (CFA) and Rasch Measurement by means of an example using real data. The responses of a sample of 475 high school students to a maths test anxiety inventory were used for the purposes of this study. The results demonstrated the superiority of the Rasch approach in both identifying the unidimensional structure of the scale as well as the two components of the construct. More studies in diverse settings are suggested in order to investigate whether the conclusions reached in this study can be generalised.

Keywords: Unidimensionality, structure, EFA, CFA, Rasch.

1. Introduction

1.1 Unidimensionality

In the social sciences, in a similar manner to physical measurement, people's abilities or psychological attributes must be measured through tests or psychological scales one at a time. The estimates of persons' abilities or positions on any psychological construct continuum will be meaningful only if the test or scale used is unidimensional. Thus one of the most important aspects in the validation process of any test or psychological scale is evidence collected to support its unidimensionality.

Bond and Fox [1] draw parallels from physical measurement to explain the meaning of and need for unidimensionality in the social sciences. In attempting to measure physical attributes people use measuring tools that measure one attribute at a time (e.g. a ruler for length, scales for weight, thermometer for temperature etc). If one takes a rectangular block, it has many attributes such as length, breadth, height, weight and density. Any attempt to make estimates about this object's attributes must be made one at a time. This focus of physical measurement on one attribute or dimension at a time is called unidimensionality.

Many psychometricians [2-5] argued that unidimensionality does not implicitly mean only one factor or dimension but rather the presence of a dominant dimension with the possible presence of minor dimensions. The presence of these minor dimensions in the data however, does not necessarily imply substantive multidimensionality. Linacre [6] explains that these extra minor dimensions may reflect different person response styles, different item content areas or could simply be an artefact of test construction. For example, including the identical item several times or many parallel items in a test produces a subset of highly intercorrelated items which may define an extra dimension.

1.2 The Study

This study uses and compares the two types of factor analysis (exploratory and confirmatory) and Rasch measurement in assessing the dimensionality of a scale using real data from the psychological measurement setting. The data included the responses of a sample of high school students to a unidimensional scale in which the dimension of interest is comprised of two correlated components. The aim of this study is therefore to establish which of the three methods best identifies the unidimensional structure of the scale and at the same time separates its two different, distinct but correlated components.

1.3 The two Types of Factor Analysis

Factor analysis is a general method of decomposing the variance of a measure into one or more common factors reflecting what variables share plus additional unique factors which normally describe variance in a measure that cannot be shared with other variables.[7]

There are two major distinct types of factor analysis: exploratory (EFA) and confirmatory (CFA). Despite their many similarities the two types are used in different settings. EFA is used, as the name implies, in an exploratory manner, to explore the data in order to determine the number of factors that account for the covariation (or correlation) between the items when researchers do not have prior evidence to formulate a hypothesis regarding the number of factors underlying the data. More simply, EFA is an investigation of the dimensionality of the data collected using a test or psychometric scale. Whereas CFA is, according to Stapleton [8], a theory-testing method rather than a theory-generating method such as EFA. In CFA the researcher begins with a hypothesis regarding the model, prior to the analysis, based on strong theoretical and/or empirical foundation. With the analysis, the researcher seeks to optimally match the observed and theoretical factor structures in an attempt to determine how well the data set fits the predetermined factor model.

1.4 Rasch Measurement

Panayides and Walker [9] describe the reasons why the Rasch models excel in the investigation of the dimensionality of scales. The Rasch models construct a one-dimensional measurement system from ordinal data regardless of the dimensionality of the data. However, more than one latent dimension will always contribute to empirical data. Multidimensionality will become a real concern when the response patterns indicate the presence of two or more dimensions so disparate that it is no longer clear what latent dimension the Rasch dimension operationalizes[6].

Factor analysis is widely used in psychometrics to investigate the dimensionality of empirical data. Like other statistical analyses, factor analysis operates on interval scale scores whereas the scale scores (usually Likert-scale scores) are ordinal by nature. Thus results of studies using these methods to examine whether or not these scales measure a single construct are disputable. Rasch measurement transforms ordinal scores into the logit scale and allows for interval-level measurement [10,11].

Linacre [6] showed that Rasch analysis followed by PCA of standardized residuals was always more effective in identifying multidimensionality than direct factor analysis of the original response-level data. PCA of the standardized residuals is based on the specification of ‘local independence’, which is an assumption of the Rasch model. This asserts that, after the contribution of the measures to the data has been removed, what is left is random, normally distributed noise. Therefore the standardized residuals are modeled to have unit normal distributions which are independent and so uncorrelated. This is testable. If the resulting common factors explain nothing more than random noise across items, then the data conform to the Rasch model. The existence of substantive common factors, however, would indicate departure from unidimensionality.

Apart from the PCA of the standardised residuals, the Rasch models provide two fit statistics (infit and outfit mean squares) aiding the investigation of the dimensionality of the data.

2. Methodology

2.1.1 The Instrument

The psychometric scale used for the comparisons of the three methods of assessing unidimensionality is the Test Anxiety Inventory (TAI). The TAI is a self-reporting psychometric scale, which was developed by Spielberger to measure individual differences in test anxiety as a situation specific personality trait. It consists of 20 items, asking respondents to describe how they generally feel. The items are answered using a 4-point Likert-style scale, scored from 1 (almost never) to 4 (almost always).

Three scores can be derived: worry (8 items), emotionality (8 items) and total (all items combined). Worry is defined as “cognitive concerns about the consequences of failure”, emotionality as “reactions of the autonomic nervous system that are evoked by evaluative stress” and total as a composite of responses to all 20 items.

2.1.2 Validity and Reliability of TAI

The reliability and validity of the TAI scores is supported by several types of evidence provided in the test manual. This evidence includes:

- Test-retest correlations of the Total score of 0.80 or higher over two week time intervals and 0.62 over a six month time interval.
- Alpha reliability estimates of the worry and emotionality factors with median values of 0.88 and 0.90 respectively, indicating satisfactory internal consistency for the 8-item subscales.
- Factor analysis of the 20 TAI items identifying the two strong, distinct factors of worry and emotionality
- Logical patterns of relation between TAI scores and other criterion measures, including positive correlations with six other measures of anxiety and

low-to-moderate negative correlations with measures of study skills, intelligence and ability enhancing the assumption that the scale is indeed unidimensional.

All the evidence presented in the publication manual, together with very similar results presented for the maths TAI by Panayides [12], support the hypothesis that the TAI is a unidimensional scale with two distinct components, those of worry and emotionality.

TAI was modified to apply to the maths test-taking setting. The TAI was then translated into Greek by a psychologist and the Greek version was translated back into English by an independent experienced teacher of English literature, who had not previously seen the original TAI. The two English versions of the inventory were then compared in order to ensure that the translation into Greek did not distort the content or meaning of the items.

2.2 The Sample

The TAI was administered to a sample of 475 high school students after permission from the ministry of education and the head teachers of the five participating schools were obtained. The purpose of the study and the voluntary nature of participation were thoroughly explained. The administration process took about ten minutes to complete.

3. Results

3.1 Exploratory Factor Analyses

Table 1 shows the outcomes of EFA. PCA was used (in order to make comparisons easier, since in Rasch measurement PCA is also used) followed by a rotation with the direct oblimin method yielding two correlated factors. The table shows all 20 items of TAI. Before each item number the letters W, E or T appear, indicating whether the item is a worry, emotionality or total (one of the four items that belong to neither the worry nor the emotionality group) item. Few things from this table are noteworthy:

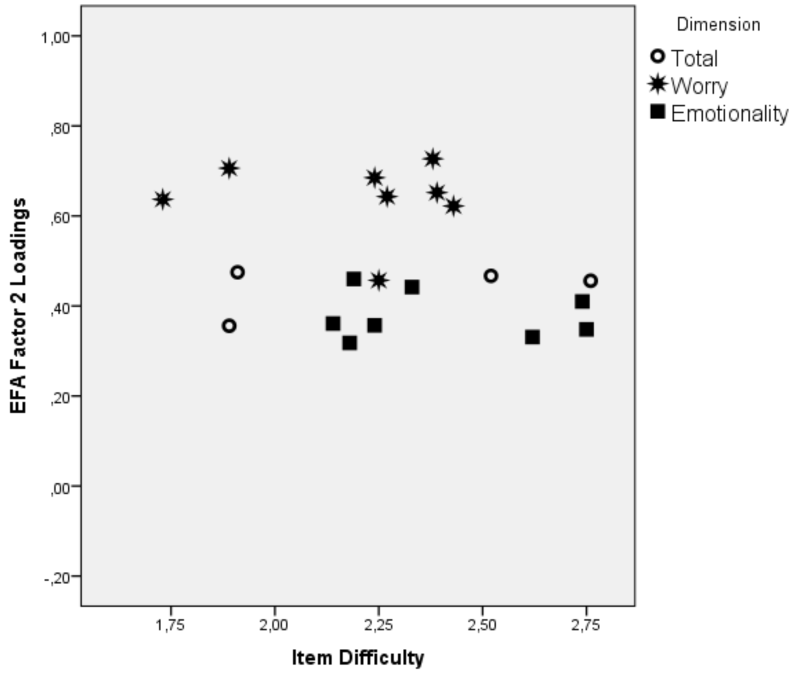
- All items load significantly (> 0.3) on both factors and this strengthens the hypothesis that the factors are correlated.
- All emotionality items have higher loadings on the first factor, which must represent the emotionality factor.
- All worry items, with the exception of item W4 have higher loadings on the second factor, which must represent the worry factor.
- W4 loads significantly on both factors but the loading on the emotionality factor (0.649) is higher than the one on the worry factor (0.486).
- The four total anxiety items seem to have a higher loading with the emotionality factor, but do load significantly with the worry factor too.

Table 1: EFA Factor Loadings.

Statements	Emotion.	Worry
T1. I feel confident and relaxed while taking exams	0.604	0.470
E2. While taking examinations I have an uneasy, upset feeling	0.697	0.410
W3. Thinking about my grade in a course interferes with my work in tests	0.458	0.562
W4. I freeze up on important exams	0.649	0.486
W5. During exams I find myself thinking whether I will ever get through school	0.308	0.500
W6. The harder I work at taking a test, the more confused I get	0.308	0.572
W7. Thoughts of doing poorly interfere with my concentration on tests.	0.572	0.624
E8. I feel very jittery when taking an important test.	0.728	0.446
E9. Even when I'm well prepared for a test, I feel very nervous about it	0.648	0.412
E10. I start feeling very uneasy just before getting a test paper back	0.606	0.463
E11. During tests I feel very tense.	0.783	0.523
T12. I wish examinations did not bother me so much	0.569	0.474
T13. During important tests I am so tense that my stomach gets upset.	0.689	0.423
W14. I seem to defeat myself while working on important tests	0.425	0.651
E15. I feel very panicky when I take an important test	0.781	0.521
E16. I worry a great deal before taking an important examination	0.744	0.394
W17. During tests I find myself thinking about the consequences of failing	0.556	0.668
E18. I feel my heart beating very fast during important tests	0.691	0.400
T19. After an exam is over I try to stop worrying about it but I can't	0.558	0.494
W20. During examinations I get so nervous that I forget facts I really know	0.585	0.628

Figure 1 shows in diagrammatic form the loadings of the items on the worry factor. The one worry item, which is much lower than the other worry items and among the emotionality items, is the fourth item in the TAI (W4). Therefore, the figure shows, with the exception of W4, a satisfactory separation of the worry from the emotionality items. To simplify comparisons, all figures presented in this study represent factor loadings against item difficulty as expressed by the mean item score. Thus, items on the left correspond to more difficult items.

Figure 1: EFA Worry Factor Loadings



The correlation between the two factors is r = 0.535 and the disattenuated correlation is $r_d = 0.627$.

3.2. Confirmatory Factor Analyses

3.2.1 The two-factor model

The first model evaluated was the two-factor model. The researcher fixed all worry items not to correlate with the emotionality factor and the emotionality items not to correlate with the worry factor (zero loadings). The remaining four total items were freed to load on both factors. The fit indices indicated an unsatisfactory fit of the model and then, guided by the standardised residuals and by the modification indices [13] the researcher freed W4, W5, W6 and W14 to load onto the emotionality factor and E10 onto the worry factor. Figure 2 shows the path diagram with the two factors, their loadings with the items and the unique variances of the individual items.

Figure 2: Path Diagram for the 2-Factor Model

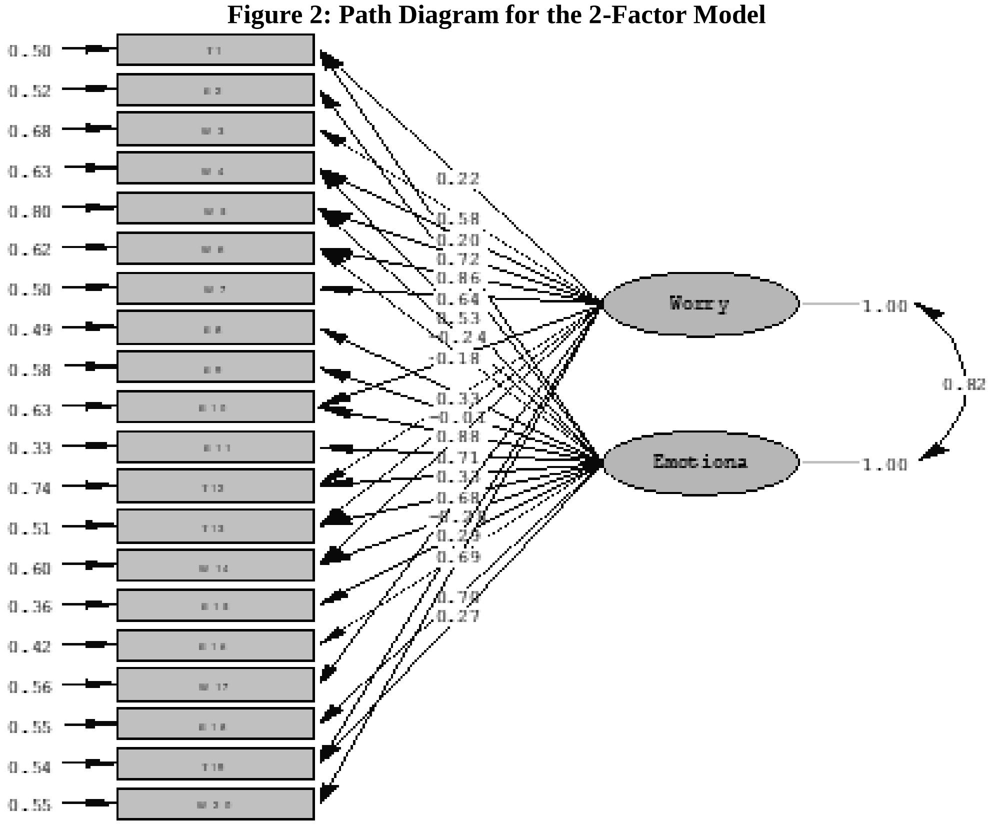


Table 2 displays the values of the three most widely used fit indices, the Standardised Root Mean Square Residual (SRMR), the Root Mean Square Error of Approximation (RMSEA) and the Comparative Fit Index (CFI) as suggested by Brown [13].

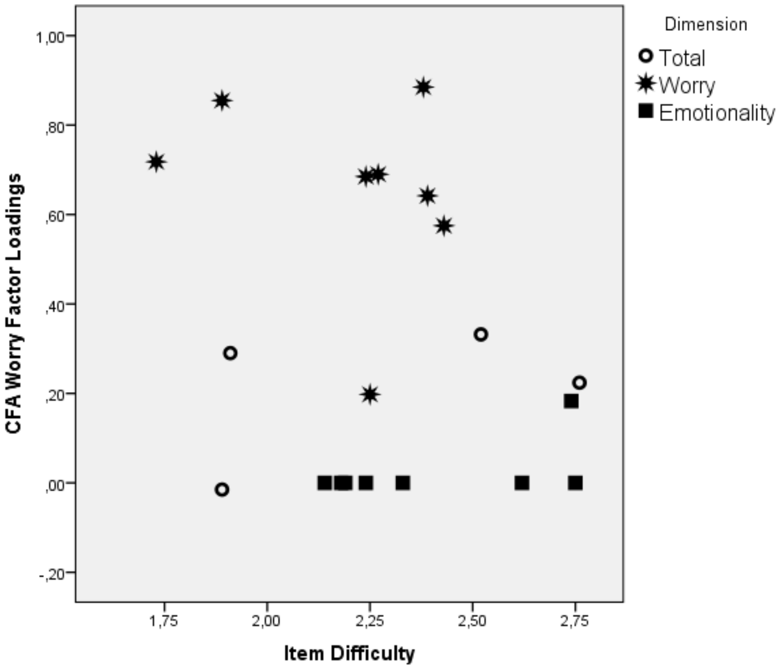
All three indices are within the ranges indicating a good fit of the model to the data.

Table 2: Fit Indices for the 2-Factor Model

Fit index	Index value	Satisfactory ranges
SRMR	0.032	< 0.05, good fit
RMSEA	0.046	< 0.05, good fit
CFI	0.962	> 0.95, good fit

Figure 3 shows the worry factor loadings as estimated by this model.

Figure 3: CFA Worry Factor loadings



In this figure seven of the emotionality items are fixed at zero loadings on the worry factor. The one item that was freed, together with three total items and W4 are approximately positioned between loadings of 0.20 – 0.30. There is a good separation of the worry from the emotionality items.

CFA confirms the 2-component factor structure suggested by theory and results. It also confirms the unidimensionality of the scale with the high disattenuated correlation $r_d = 0.816$ (which corresponds to $r = 0.696$).

3.2.2 Other Models

Considering Kline's [14] criticism of CFA, that if a model is confirmed it does not mean that other models do not fit or even fit better, the researcher tried a 1-factor model and then 3-factor and 4-factor models. In the 3-factor model the four total items (T1, T12, T13 and T19) were fixed at zero loadings on the two factors and were allowed to load on a third factor. Then, since the worry items had generally lower loadings than the emotionality items, the researcher fixed the four worry items with the lowest loadings at zero loadings with the three factors and allowed them to load on a fourth factor. In the 3-factor and the 4-factor models all the factors were highly intercorrelated, as expected, confirming unidimensionality. Some modifications were made in all cases, with the use of the modification indices, in order to improve the fit of the model. In the case of the 1-factor and 4-factor models some pairs of items were allowed to covary. Table 3 shows the fit statistics for all four models evaluated. For comparison purposes the 2-factor model is displayed first.

Table 3: Fit Indices for the Various Models

	Models			
Fit index	2-factor	1-factor	3-factor	4-factor
SRMR	0.0322	0.0347	0.0333	0.0332
RMSEA	0.0462	0.0402	0.0459	0.0354
CFI	0.962	0.974	0.961	0.977

The best SRMR value is 0.0322 (the 2-factor model), RMSEA is 0.0354 and CFI is 0.977 (the 4-factor model). However all fit indices take values well into the acceptable ranges indicating good fit, therefore all four models fit the data very well.

3.3 Rasch Measurement

For direct comparisons of the three methods the researcher does not report the item point measure correlations or the item infit and outfit values which are also tools for assessing unidimensionality by identifying possible departures from it. Therefore only PCA of the standardised residuals is presented. Table 4 shows the results of this analysis.

Table 4: Table of Standardized Residual Variance (in Eigenvalue units)

		Empirical	Modeled
Total raw variance in observations	36.4	100.0%	100.0%
Raw variance explained by measures	16.4	45.1%	45.2%
Raw variance explained by persons	12.3	33.9%	34.0%
Raw Variance explained by items	4.1	11.2%	11.2%
Raw unexplained variance (total)	20.0	54.9% (100.0%)	54.8%
Unexplained variance in 1st contrast	2.4	6.7%	12.2%

The variance explained by the measures is 45.1% of the total variance in the standardised residuals. The unexplained variance in the first contrast (factor) has an eigenvalue of 2.4 and this is only 12.2% of the unexplained variance and just 6.7% of the total variance. The ratio of the variance explained by the measures to the variance explained by the first factor is approximately 7:1.

Linacre [15] suggests also looking at the content of the items at the top (with the highest positive loadings on the contrast) and bottom (with the highest negative loadings on the contrast). The number of items to look at depends on the eigenvalue. In this case the eigenvalue is 2.4 so one should look at the two top items and the two bottom ones. If these items are different enough, in content, to be considered different dimensions then Linacre suggests splitting the data into separate analyses and comparing the two sets of person measures. If the items are part of the same dimension then no action is necessary. Table 5 shows the item loadings on the first factor extracted through the PCA of standardised residuals.

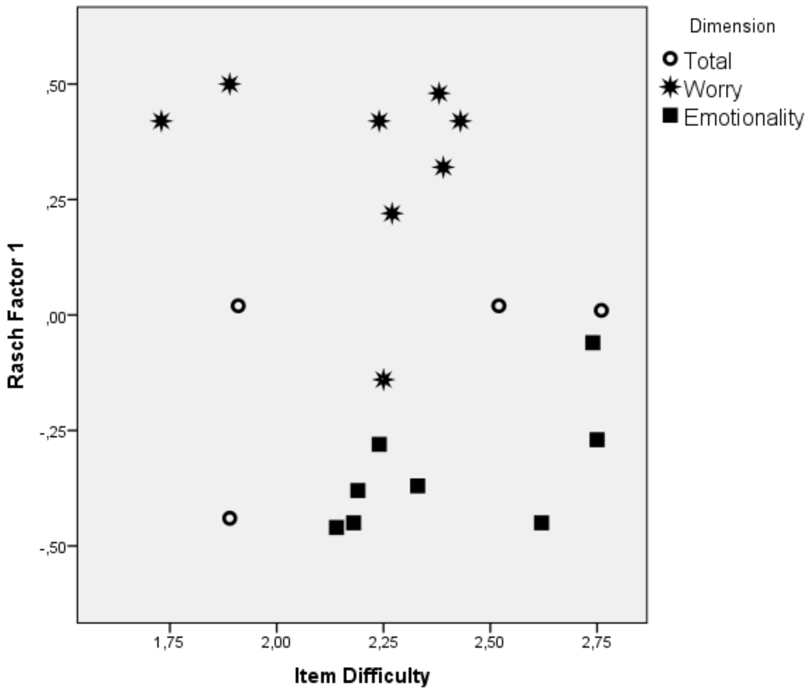
Table 5: Item Loadings on Factor 1 Extracted by PCA of Standardised Residuals

Items	W6	W14	W17	W3	W5	W7	W10	T12	T9	T1
Loadings	0.50	0.48	0.42	0.42	0.42	0.32	0.22	0.02	0.02	0.01
Items	E8	E18	E16	T13	E15	E11	E9	E2	W4	E10
Loadings	-0.46	-0.45	-0.44	-0.44	-0.38	-0.37	-0.28	-0.27	-0.14	-0.06

The two items with the highest positive loadings are W6: “The harder I work at taking a test, the more confused I get” and W14: “I seem to defeat myself while working on important tests”. The two items with the highest negative loadings are E8: “I feel very jittery when taking an important test” and E18: “I feel my heart beating very fast during important tests”. It is clear that all four statements describe consequences of test anxiety giving further support of the unidimensional structure of the scale.

One further point that should be made is the very clear separation of the worry from the emotionality items. All the worry items (except W4) have positive loadings on the first factor and all emotionality items negative. This good separation can be seen on Figure 4.

Figure 4: Rasch First Factor Loadings



Taken separately, for diagnostic purposes, the two groups of items seem to be measuring two different factors. However the disattenuated correlation between the two sets of person measures (one set from the items with the positive factor loadings and one from the ones with the negative loadings) was found to be $r_d = 0.907$ (corresponding to $r = 0.752$). The evidence collected in the Rasch approach includes:

- Variance explained by measures is 45.1% of total variance.
- Variance explained by first factor has an eigenvalue of just 2.4.
- Variance explained by first factor is 12.2% of the unexplained variance and just 6.7% of the total variance.
- The ratio of the variance explained by measures: variance explained by first factor was 7:1.
- The content of the two items with the highest positive loadings on the first factor extracted is no different than that of the two items with the highest negative loadings.
- The disattenuated correlation between person measures estimated with the two distinct groups of items was very high at 0.907.

All the evidence collected indicates that the scale is unidimensional, despite the clear separation between the two factors in the PCA of the standardised residuals.

4. Discussion

The aim of this study was to use the three methods (EFA, CFA and Rasch measurement) on a set of high school students' responses to a well-known scale, which has a unidimensional structure. The scale used was the maths TAI, the main construct of which, test anxiety, comprised of two distinct components: worry and emotionality. The main objective of this study was therefore to compare the three methods in terms of how well they can identify the two components and the unidimensional structure of the scale.

4.1 EFA and CFA

EFA demonstrated a satisfactory separation of the two components. At the same time, the disattenuated correlation between these was significant, implying that the two components do work in unison, thus measuring the construct of maths test anxiety.

CFA confirmed the 2-factor model for maths test anxiety. At the same time there was good separation of the two components and a high disattenuated correlation between them, (significantly higher than in the EFA case) implying again, as with EFA, that the scale can be considered unidimensional.

An important criticism of CFA is that when a model is confirmed, that only means that the particular model fits the data. It does not mean that there are no other models fitting the data, or even fitting better [14]. This study put Kline's criticism to the test and indeed the 1-factor, 3-factor and 4-factor models also fit the data equally well. In all cases unidimensionality was confirmed but the factor structure varied from a one-component to a four-component model. This raises an important question as to which model to choose. In accordance with many authors, the formulation, specification and evaluation of the factor model require strong theoretical or empirical foundation [8,13,16-19]. This means that the choice between the different models must be based on previously reported theory or results. This answer however gives rise to other questions: What if the theory was originally flawed? What if previously reported results were flawed? And perhaps more importantly would such an approach (confirming previous theories or results) not bring us to a standstill?

Finally, these statistical methods, as with IRT models, are driven by the desire to model or accommodate all of the characteristics in the data, regardless of whether they contribute to the measurement process. These methods correspond to modeling and not measurement.

4.2 Rasch Measurement

The Rasch approach showed the best separation between the worry and emotionality components of maths test anxiety. Worry items had positive loadings on the first factor extracted by PCA of the standardised residuals and emotionality items negative. At the same time they showed the strongest evidence of the unidimensional structure of the scale through a significantly higher disattenuated correlation between the two components and strong support from the analyses of PCA of the standardised residuals.

The Rasch analyses provide more reliable results for another reason; they transform first the ordinal scores into interval-level data. Thus all analyses thereafter are performed, as they should be on interval-level data. The Rasch models construct a one-dimensional measurement system regardless of the dimensionality of the data, and the fit of the data to this measurement system, and not the other way around, can be tested. Apart from the methods used in this study, fit statistics, investigation into whether invariance holds and point-measure correlations can also help in the assessment of unidimensionality.

5. Conclusions

Of the three methods used in this exemplification, with data from the psychological measurement setting, the Rasch approach demonstrated better assessment of the unidimensional structure of the scale and at the same time superior identification-separation of the two components of the maths test anxiety construct. The researcher suggests that setting aside all the other measurement advantages of the Rasch models, they offer a more robust investigation of the dimensionality of any test or scale.

As this study used just one instrument in comparing the three methods, it is premature to make generalizations based on it. Notwithstanding, the aim of this study was to generate interest in further investigations of the three methods in this and other settings such as educational or medical measurement

References

[1] Bond, T. G. & Fox, C. M. (2007). *Applying the Rasch model: Fundamental measurement in the social sciences*. New Jersey: Lawrence Erlbaum Associates.

[2] Hambleton, R. K. (1993). Principles and selected applications of Item Response Theory. In Linn, R. L. (Eds.), *Educational Measurement*. (3rd *ed.)*, (pp 13-104), Phoenix: Oryx Press.

[3] Hambleton, R. K. Swaminathan, H. & Rogers, H. J. (1991). *Fundamentals of Item Response Theory*. California: SAGE Publications, Inc.

[4] Smith, Jr., E. V. (2004). Detecting and evaluating the impact of multidimensionality using item fit statistics and principal components analysis of residuals. In E. V Smith Jr, & R. M Smith (Eds), *Introduction to Rasch Measurement* (pp. 575-599). Minnesota: JAM Press.

[5] Wright, B. D. & Linacre, J. M. (1989). Observations are always ordinal; measurements, however, must be interval. *Archives of Physical Medicine and Rehabilitation*, *70*, 857-860.

[6] Linacre, J. M. (1998). Detecting multidimensionality: Which residual data-type works best? *Journal of Outcome Measurement*, *2*(3), 266-283.

[7] Nunnally, J. C. & Bernstein, I. H. (1994) *Psychometric Theory*, (3rd ed.). USA: McGraw-Hill, Inc.

[8] Stapleton, C. D. (1997). *Basic concepts and procedures of confirmatory factor analysis*. Paper presented at the annual meeting of the Southwest Educational Research Association, Austin, Texas, January, 1977.

[9] Panayides, P. & Walker, M. J. (2012). Evaluation of the Psychometric Properties of the Internet Addiction Test (IAT) in a sample of Cypriot High School Students: The Rasch measurement perspective. *Europe's Journal of Psychology*, *8*(3), 327- 351.

[10] Wright, B. D. & Masters, G.N. (1982). *Rating scale analysis*. Chicago: MESA Press.

[11] Wright, B. D. & Mok, M. (2000). Rasch Models Overview. *Journal of Applied Measurement, 1*(1), 83-106.

[12] Panayides, P. (2009). Exploring the reasons for aberrant response patterns in classroom maths tests. (Unpublished doctoral dissertation). Durham University, UK.

[13] Brown, T. A. (2006). *Confirmatory Factor Analysis for applied research*. New York: The Guilford Press.

[14] Kline, P. (2000). *A psychometrics primer*. London: Free Association Books.

[15] Linacre, J. M. (2005). *WINSTEPS Rasch measurement computer program*. Chicago: Winsteps.com

[16] Joreskog, K. & Sorbom, D. (1996). *LISREL 8: Structural equation modeling with the SIMPLIS command language*. Chicago: Scientific Software International.

[17] Kline, P. (1994). *An Easy Guide to Factor Analysis*. London: Routledge

[18] Nunnally, J. C. (1978). *Psychometric Theory* (2nd ed.). New York: McGraw Hill.

[19] Stevens, J. (1996). *Applied multivariate statistics for the social sciences (3rd ed.)*. Mahwah, NJ: Lawrence Elbaum.