Disattenuated Correlation and Unidimensionality

S. Alagumalai
J.P. Keeves
Njora
Flinders University of South Australia

Prior to using any multivariate or inferential statistics, items in an instrument have to be validated, i.e. with reference to the latent trait(s) it intends to measure. Particularly in the Rasch model, unidimensionality has to be ascertained, and precedes any item analysis. This article attempts to explore the techniques used to detect unidimensionality and provides a solution with Linear Structural equations through the use of disattenuated correlation. A complimentary IRT technique would also be explored through the Multi-Aspect Test Software.

Introduction

"In constructing an instrument for measuring psychological variables on an interval scale, - assumed that every item differentiates between individuals on one difficulty continuum, then the items must measure exactly the same trait but have different degrees of difficulty" (Magnusson, 1966, p.17). This concerns the dimensionality of the test. Measurement, thus, implies that one characteristic at a time is quantified (Lumsden, 1961, p. 268). A unidimensional test may be defined simply as a test in which all items are measuring the same thing (Lumsden, 1957).

Over the years, for both classical test theorists and current advocates of item response theory (IRT), the concept of unidimensionality (i.e. the existence of a single dimension in k-space in the measuring instrument) has been a focus of debate and refinement. Hambleton and Swaminathan (1985, p.17) indicate that, "it is commonly assumed that only one ability or trait is necessary to 'explain' or 'account' for examinees' test performance. IRT models that assume a single latent ability are referred to as unidimensional." The trait or ability describes the latent characteristic of the individual, which is estimated from the individual's responses to test items. Similar working definitions are advanced by Hulin, Drasgow and Parson (1983, p.10). They further add that under this definition of unidimensionality, all the items on a test or ability scale must measure a single latent trait of an individual and violation of this assumption would lead to seriously misleading results.
In order to make true sense of any model of item response theory (IRT), the dimension(s) underlying a test has to be ascertained and precedes further analysis of any kind. It is of the highest priority to investigate the 'psychological width' of items sets prior to any other forms of analysis. Thus, the concept of unidimensionality is of utmost importance in any attempts to make measurements (Gustafsson, 1977, p.94; McDonald, 1981).

This paper considers the importance of the assumption of unidimensionality in relation to IRT, reviews methods for ascertaining unidimensionality of an instrument and illustrates the use of linear structural equations analysis through the use of disattenuated correlations.

IRT and Unidimensionality

"There are a number of assumptions that must be fulfilled in order for any reasonable estimates of parameters to be achieved, and any sensible application to be made. The three most important assumptions are those pertaining to the dimensionality of the latent space, the principle of local statistical independence and the form of the item characteristic curve" (Gustafsson, 1977, p. 16). This is in contrast to classical test theory, as IRT makes strong demands in order to achieve its objectives (Weiss, 1983, p. 84). Thus, IRT explicitly acknowledges the requirement of unidimensionality, and other commonly used procedures (equipercentile equating) that transforms single scores are also unidimensional, even if they do not explicitly acknowledge this assumption (Hirsch, 1989).

One basic requirement of the Rasch model is the concept of unidimensionality (this follows earlier mentions of unidimensionality in test theories - See McNemar 1946, Solomon 1961, and Magnusson 1966). Each item in a test should measure the same unitary construct, that is the various items measure the same ability, achievement, attitude, etc., in addition to its specific and error components. "More precisely, a set of n tests is considered of dimension k if the residuals of the n variables about their regression on k further (hypothetical) variables, the common factors, are uncorrelated, i.e. if the partial correlations of the test scores are all zero if the common factors are partialled out" (McDonald, 1981).

Green et al. (1984), content that unidimensionality is generally desirable with tests of ability, but is more critical for adaptive testing. Identification of the dimensionality of an item pool is crucial before fitting an IRT model in adaptive testing (De Ayala, 1992; Yen, 1985). Similar demands exist for vertical equating (Bogan & Yen, 1983, p. 3). Swaminathan et al. (1978), Lord (1980), Hattie
(1984), and Hambleton and Zaal (1991) all indicate that "testing unidimensionality takes precedence over other goodness of fit tests of a latent trait model, as violation of unidimensionality would imply violation of the basic assumption of the Rasch model." This is due to the invariance property of the item parameters, as they rely solely on the unidimensionality assumption, which lends indirect support to the assumption of local independence (Ree & Jensen, 1983, p. 84; Mellenbergh, 1994).

A study by Ackerman (1988, p.16) indicates that differential item functioning (DIF) can occur when there is a misspecification of the latent ability space. He indicates that multidimensional IRT should provide new methodology that would not only detect DIF, but would also provide some substantive support about why groups perform in the way that they do. However, Loyd and Hoover (1980, p.5) argue that the violation of the unidimensionality requirement has been a problem in estimating item parameters. Thus, given that the dimensionality requirement of unidimensional IRT subsumes the principle of local independence, violation of this requirement should affect the likelihood function used for parameter estimation (De Ayala, 1992).

Though numerous studies have argued about this, especially for the importance of a single dimension over small secondary dimensions, the current availability of multidimensional IRT model computer programs reiterates the need for the basic assumption of unidimensionality (See Andrich 1995). At this point it is important to note that unidimensionality does not imply that a single characteristics or process operates. If several characteristics or processes function in unison, then unidimensionality will hold. If the characteristics or processes do not function in concord, then it is not meaningful to assign numbers to a combination of the characteristics, unless another operation such as that of prediction provides the rule for combination. If a set of characteristics lacks this internal consistency and fidelity, or if it involves several identifiable dimensions, each must be measured separately and a profile of measures recorded (Bejar, 1983, p.31).

Stout (1987) outlines the importance of unidimensionality for a test and makes the following points:

It is important that a test that purports to measure the level of a certain ability is in reality not significantly contaminated by varying levels of one or more other abilities displayed by the examinees taking the test.

It is essential that a test designed to be used in the measurement of individual difference must in fact measure a 'unified trait'. In the use of Rasch models, it is a requirement to establish unidimensionality as a preliminary step prior to the use of any of the
standard (unidimensional) latent trait methodologies.

In order to operationally define undimensionality, researchers like Hulin, Drasgow and Parsons (1983), Green et al. (1984) and Hattie (1984), have attempted to look into the three-parameter IRT model. In the three-parameter IRT model, the logistic item response curve, indicating the probability that a person with ability \( \theta \) responds correctly to item \( j \), is

\[
P_j(\theta) = c_j + \frac{[1 - c_j]}{[1 + \exp(-1.7a_j (\theta - b_j)]}
\]

When \( \theta \) is very low, the probability is \( c_j \). As \( \theta \) increases, the probability rises from \( c_j \) to 1. The rise is sharp for high values of \( a \), gentle for low \( a \). The entire item response curve can be shifted laterally with respect to the \( \theta \) axis by changing \( b \), so \( a \) can be interpreted as item discriminability, \( b \) as item difficulty, and \( c \) as the pseudochance / guessing level (Lord 1980).

Green et al. (1984) indicate that items with high \( a \) values will tend to "ensure unidimensionality." However, Hattie (1984) has shown that a whole range of values for \( a \) could provide for unidimensionality. Furthermore, Hattie's (1985) study clearly indicates that almost all proposed criteria for unidimensionality do not serve their intended purpose. It should also be noted that \( c \), the guessing parameter, is not applicable for an attitude or view instrument. This leads us to the next section for testing unidimensionality.

Techniques for detecting Unidimensionality

The ascertainment of the indices of unidimensionality has been a challenge (See Hattie 1984; Stout 1987, 1990; Nandakumar 1994). Factor analysis has been a commonly used technique for demonstrating unidimensionality in classical test theory (Lumsden, 1957). Gurin, Gurin and Morrison (1978) noted that some earlier studies in this area, have apparently mistakenly interpreted the eigenvalue rule, and have applied the rule to eigenvalues resulting from rotated factors - without specifying the basis for selecting the number of factors to be initially rotated - and the common factor variance. A traditional procedure to test for unidimensionality has been suggested by Lord and Novick (1968, p. 381-382), who indicate that, "a principal axis analysis of the inter-item tetrachoric correlation matrix was performed after replacing the diagonal elements with the largest entry in their respective rows." An item set may be considered unidimensional if the first eigenvalue from this analysis is large compared to the second,
and all eigenvalues other than the first are about the same size (Lord & Novick, 1968; Holmes, 1982).

Slinde and Linn (1979) posit that in addition to the test for unidimensionality, the slopes of the items should be examined to determine if the items shared a common level of discrimination. The index of item discrimination employed should be a least squares estimate of the slope of the linearly transformed observed item characteristic curve for the items. Theoretically, this value should be unity for items that fit the Rasch model, although values within a pre-specified range might be considered similar enough to provide reasonable fit.

However, Jones (1980) and Nandakumar (1994), have suggested that the IRT model is not very sensitive to lack of unidimensionality. As a consequence, Stout (1987, 1990) and Nandakumar (1994) have attempted to use the DIMTEST computer program to determine unidimensionality. Linacre (1994), nevertheless, has indicated shortcomings of the DIMTEST program. Moreover, Hattie (1984) contends that indices based on answer patterns, reliability, component analysis, linear and nonlinear factor analysis, and on the one-parameter latent trait model were ineffective. However, this has been refuted by McDonald and Mok (1995).

Although numerous alternative procedures have been advanced, namely factor analysis, marginal maximum likelihood method, covariance structure analysis, to name a few, confirmatory factor analysis using LISREL is the most frequently used procedure (Marsh & Richards, 1987). It should be noted that confirmatory factor analysis, though a necessary initial procedure, is not the only test for unidimensionality (Hattie, 1984). This follows Lord and Novick's (1968, p. 382) findings that, "the dimensionality of the complete latent space does not depend on distributional assumptions, nor on a choice of a measure of inter-item correlation, nor on any choice of transformation of the latent variables. Thus the dimensionality of the complete latent space is a more basic concept than is the number of common factors."

In line with the 'basic concept' numerous fit indices have been suggested (Bentler & Bonett, 1981; Marsh & Hocevar, 1984; Bozdogan, 1987; McDonald & Marsh, 1990; Browne & Cudeck, 1993; McDonald & Mok, 1995) for structural equation models which discriminate effectively between unidimensional and higher dimensional item response models. One such index was advanced by Hattie (1984, 1985) who proposed using the average absolute residual covariance of the responses fitting the model through McDonald's ordinary least squares method or by Muthen's (1987) GLS program. However, this index is dependent on the functions of \( \phi \) and its distributed discrepancies. It should be noted that for a set of given data, confirmatory analysis can be performed with alternative models, until a model is found which is satisfactory both from the
point of view of goodness of fit and meaningfulness (Joreskog, 1971).

McDonald and Marsh (1990) suggest an unbiased relative fit index (RFI) relative to a null model.

\[ RFI = 1 - \left( \frac{dk}{do} \right) \]

where \( dk \) is the estimated noncentrality parameter for the k-dimensional model \([\text{Note: } dk = \frac{(2 - df)/N}{2}]\) and \( do \) is the same quantity computed for the model of zero dimension. The RFI expression is a monotone function of model complexity and requires a subjective evaluation of whether a statistically significant (\( \chi^2 \) is small enough to constitute an adequate fit.

McDonald and Mok (1995) indicate in their study that, "goodness-of-fit in linear structural models carries over to IRT, including all of the open questions as to how the proposed indices should be employed in application. They conclude that, "the treatments of goodness-of-fit that have been developed in the context of linear models are applicable to the problem of dimensionality in IRT." This includes the availability of a number of goodness-of-fit indices. However, they also indicate that, "further work is still needed before clear choices of indices and plain criteria for their applications can be offered." The next section considers how the concept of disattenuated correlation, through \( \chi^2 \) tests, and RFI differences could be used to ascertain unidimensionality with reasonable accuracy.

Disattenuated Correlation and unidimensionality

Within the framework of classical theory, but not item response theory, it has been argued that scales may be constructed to measure the relationship between latent traits. If the relationship between these scales is linear, then the correlation coefficient indicates the association between these scales. Lord and Novick (1968) also argue that the scales may contain error, and thus the correlation between the scales is less than the correlation between the traits. Hence, an attenuation formula, applied to an approximation, can be used to compute the actual correlation between the traits as the true scores which can be assumed to measure the latent traits in question (Lord & Novick, 1968, p. 69).

The correlation between true scores in terms of correlation between observed scores and the reliability of each measurement is given by

\[ \rho_{XY} = \frac{\rho_{\text{XY}}}{\sqrt{\rho_{XX} \rho_{YY}}} \]
where \((XY)\) is the population value of the correlation between test X and test Y, \((XX)\) and \((YY)\) are the population values of the reliability coefficients for the two tests, and \(((TX, TY))\) is the population value of the correlation between X and Y corrected for attenuation.

The above formula may be interpreted as the correlation between the latent traits being studied in terms of the observed correlation of the measures of these traits and the reliabilities of these measures. Under the assumption that a correlation exists between the observed score on one measurement and the true score on a second measurement, the above formula can be written as

\[ ((TX, TY)) = ((XY)) / ((YY')^{1/2}) \]

The above two equations are referred to as attenuation formulas (Lord, 1957). He termed \(((TX, TY))\) the disattenuated correlation. It follows the idea that the observed correlation between observed scores is less than the correlation between corresponding true scores because the former correlation is attenuated by the unreliability of the measurements. If the reliabilities of these measurements are known, then these formulas may be used to compute the disattenuated correlation, i.e. the correlation between the true scores [For a more detailed discussion, see Lord (1957) and Lord & Novick (1968)]. Thus, "if on the basis of a sample of examinees, the disattenuated correlation is near unity, than the experimenter concludes that the two tests are measuring the same trait" (Joreskog, 1971). This implies that when the disattenuated correlation approximates to unity, the instrument measuring the latent trait(s) is said to be unidimensional. Thus the disattenuated correlation answers a real and important question about the tests and traits under study (Lord, 1957).

This is in analogous to the concept of congeneric items in a test. A set of items are congeneric if they measure the same trait except for errors of measurement (Joreskog, 1971). Items are congeneric if the reliability of each item can be estimated directly from intercorrelations of the test scores without making any parallelity assumptions.

Here, the 'parallel tests' refers to test that have equal true score variances and equal error variances. Parallel tests can be used interchangeably as they have equal reliabilities and validities in predicting a given criterion. Furthermore, the variance-covariance matrix ( for such tests have equal variances and covariances(Joreskog, 1971). Thus, 'parallel tests' are special cases of congeneric tests, namely when any two true scores are identical (Lord and Novick, 1968).

Hence, with this concept of attenuation theory, which is central to classical test theory, simulated and real data studies would serve to provide evidence in the testing of the unidimensionality of an
instrument. It is important to note that this testing of the unidimensionality of an instrument is undertaken within the framework of classical test theory.

A simulated study

1) A general model (Instrument with 22 questions).

Dimension 1: Items x1, x2, ..... x11
Dimension 2: Items x12, x13, ..... x22.

In Joreskog's (1971) model for congeneric measurement applied to the item response, \( (ij = bj + aj ti \) where \( Var(Xij) = \sigma^2 \), which is also called the unidimensional latent trait model, the above disattenuated correlation between the latent traits can be identified. This is the linear regression of the item response or fixed values of the latent trait, where \( bj \) is the intercept of the regression and \( aj \) is the item discrimination parameter. A special case could be obtained by restricting \( bj, aj \) and \( \sigma^2 \). This leads to \( (ij = b + a ti \) where \( Var(Xij) = \sigma^2 \) which is the model for parallel items.

Using the LISREL measurement model,

\[
\begin{align*}
C1 \\
C2 \\
& \vdots \\
C10 \\
C11 \\
& \vdots \\
C12 \\
C13 \\
& \vdots \\
C21 \\
C22 \\
\end{align*}
\]

In matrix form,
In this model, \( x_1 \ldots x_{11} \) are congeneric measures of latent trait (1), and \( x_{12} \ldots x_{22} \) are congeneric measures of latent trait (2). The disattenuated correlation (12) is the correlation between (1) and (2). In order to analyse the data, and to answer the question whether the disattenuated correlation coefficient is consistent with the hypothesis that sub-sections are really measuring the same ability or trait, four hypotheses were set up:

H1: \( (1) = (2) = (3) = \ldots = (11), \quad (12) = (13) = (14) = \ldots = (22), \quad (1) = (2) = (3) = \ldots = (11), \quad (12) = (13) = (14) = \ldots = (22), \quad (1) = 1. \)

H2: \( (1) = (2) = (3) = \ldots = (11), \quad (12) = (13) = (14) = \ldots = (22), \quad (1) = (2) = (3) = \ldots = (11), \quad (12) = (13) = (14) = \ldots = (22). \)

H3: \( (1) = 1. \)

H4: \( (1), \ldots, (11), (12), \ldots, (22), (1), \ldots, (11), (12), \ldots, (22), \text{ and } \) \( (1) \) are unconstrained.

(2) Simulated data

The Stout's Data Generator (1995) was used to generate unidimensional and 2-dimensional data set for the simulated study. The program allows the user to input both the sample size needed and the number of items required. As the real data set had 2511 students and required both unidimensional and 2-dimensional data, the above data were generated and tested.

(2a) 1 dimensional data (Item number = 6; Examinees = 2511)

Dimension 1: Items  \( x_1, x_2, \ldots, x_6 \)
To test the parallel/congeneric hypotheses, items were grouped as follows:

Group 1 ((1): x1, x2, x3  
Group 2 ((2): x4, x5, x6

Table 1. Lisrel Output for 1-dimensional Simulated Data-Set

Each hypothesis in the Table 1 is tested against the general alternative that ( (the variance-covariance matrix) is unconstrained. To consider various hypotheses that can be tested, the four $\chi^2$ values of the above table are recorded as indicated below.

Table 2. Parallel and Congeneric Test of Hypotheses

With reference to Table 2, test of H1 against H2 gives $\chi^2 = 1.29$ with one degree of freedom. An alternative test of H3 against H4, which gives $\chi^2 = 0.66$ with one degree of freedom. Thus, the hypothesis $\chi = 1$ is not rejected. There is strong evidence that the two sets of items do measure the same trait, i.e. the items x1 ...... x6 are unidimensional.

(2b) 2 dimensional data (Item number = 6 + 6; Examinees = 2511)

Dimension 1: Items x1, x2, ...... x6  
Dimension 2: Items x7, x8, ...... x12

To test the parallel/congeneric hypotheses, items were grouped as follows:

Group 1 ((1): x1, x2, ...... x6  
Group 2 ((2): x7, x8, ...... x12

Table 3. Lisrel Output for 2-dimensional Simulated Data-Set

As above, each hypothesis in the above table is tested against the
general alternative that $\psi$ is unconstrained. To consider various hypotheses that can be tested, the four $\chi^2$ values of the above Table 3 are recorded as indicated below.

Table 4. Parallel and Congeneric Test of Hypotheses

With reference to Table 4, test of H1 against H2 gives $\chi^2 = 3577.25$ with one degree of freedom. An alternative test of H3 against H4, which gives $\chi^2 = 4550.93$ with one degree of freedom. Thus, the hypothesis $\psi = 1$ is rejected. There is strong evidence that the two sets of items do not measure the same trait, i.e. the items $x_1$ ...... $x_{12}$ are not unidimensional.

(3) Real data

The data-set used for this section was obtained from the Australian Mathematics Study (1964). For the 1-dimensional data, 6 items were selected from the 'Views of Mathematics Teaching' section and for the 2-dimensional item, the former was used together with 6 other questions from the 'Views About School and School Learning' section.

(3a) 1 dimensional data (Item number = 6; Examinees = 2511) [6 items from 'Views of Mathematics Teaching' Section of the 1964 data-set.]

Dimension 1: Items $x_1$, $x_2$, ...... $x_6$

To test the parallel/congeneric hypotheses, items were grouped as follows:

Group 1 ((1) :$x_1$, $x_2$, $x_3$
Group 2 ((2):$x_4$, $x_5$, $x_6$

Table 5. Lisrel Output for 1-dimensional-Real Data-Set

Each hypothesis in the above table is tested against the general alternative that $\psi$ is unconstrained. To consider various hypotheses that can be tested, the four $\chi^2$ values of the above Table 5 are recorded as indicated below.
Table 6. Parallel and Congeneric Test of Hypotheses

With reference to Table 6, test of H1 against H2 gives $\chi^2 = 1.02$ with one degree of freedom. An alternative test of H3 against H4, which gives $\chi^2 = 3.46$ with one degree of freedom. Thus, the hypothesis $\gamma = 1$ is not rejected. There is strong evidence that the two sets of items do measure the same trait, i.e. the items $x_1 \ldots x_6$ are unidimensional.

(3b) 2 dimensional data (Item number = 6 + 6; Examinees = 2511) [6 items from 'Views of Mathematics Teaching' Section and 6 items from the 'Views About School and School Learning' of the 1964 data-set.]

Dimension 1: Items $x_1, x_2, \ldots, x_6$
Dimension 2: Items $x_7, x_8, \ldots, x_{12}$

To test the parallel/congeneric hypotheses, items were grouped as follows:

Group 1 ((1): $x_1, x_2, \ldots, x_6$
Group 2 ((2): $x_7, x_8, \ldots, x_{12}$

Table 7. Lisrel Output for 2-dimensional Real Data-Set

As above, each hypothesis in the above table is tested against the general alternative that $\gamma$ is unconstrained. To consider various hypotheses that can be tested, the four $\chi^2$ values of the above Table 7 are recorded as indicated below.

Table 8. Parallel and Congeneric Test of Hypotheses

With reference to Table 8, test of H1 against H2 gives $\chi^2 = 149.00$ with one degree of freedom. An alternative test of H3 against H4, which gives $\chi^2 = 219.73$ with one degree of freedom. Thus, the hypothesis $\gamma = 1$ is rejected. There is strong evidence that the two sets of items do
not measure the same trait, i.e. the items x1 ...... x12 are not unidimensional.

Comparison of above procedure with the IRT Techniques

As indicated above, the test of unidimensionality through the use of 'disattenuated correlation' procedure adheres to the realm of classical test theory. To further validate and compliment the above procedure, MATS (Wu, Adams and Wilson, 1996) which employs the IRT models would be explored. As indicated by Bentler & Houck (1996), "programs like LISREL and EQS perform the equivalent of a 2-parameter IRT model analysis." Since, the above test of unidimensionality was attempted using LISREL, it would be fair to argue that the 'results' of the dimensionality test would also apply.

As MATS (Wu, Adams and Wilson, 1996) enables the testing of dimensionality of an instrument, the above data-sets were subjected to further analysis. Table 9 summarises the output from MATS (Wu, Adams and Wilson, 1996).

Table 9. Output of MATS (1996) analysis of data-sets

It is evident that the correlation coefficient for unidimensional data is relatively higher than for 2-dimensional data.

Conclusion

Within the framework of classical test theory, it is evident from the above findings that the disattenuated correlation is a powerful concept in ascertaining the unidimensionality of an instrument. The comparison of (2 is in line with Joreskog's (1971) contention that it is the differences between (2 values that matter rather than the absolute (2 values themselves. Furthermore, if the change in (2 is large compared to the difference in degree of freedom, then the changes made to the model represents an improvement. However, if the change in (2 is close to the difference in the number of degrees of freedom, it indicates that the improvement in fit is obtained by 'chance' and the added parameters may not have any real significance or meaning. Furthermore, the complimentary technique of using MATS (Wu, Adams and Wilson, 1996), further lends support to Bentler & Houck's (1996) argument that LISREL could perform the equivalent of IRT model analysis. This would indicate that the concept of 'unidimensionality' is 'unique' and should be testable regardless whether it is from the classical test theory perspective or the IRT perspective.
The above methodology also provides a further advantage of being able to choose individual items meaningfully and conceptually and testing its dimensionality, unlike confirmatory factor analysis where 'faulty items' may be 'shielded' during 'batch' analysis. This is in line with Gustafsson's (1977, p. 95) argument that, "to construct a test intended to measure an attribute we of course need a conception of the attribute at once when the work is begun. But this conception is likely to be vague and there will be little basis for deciding whether an item or an item type does reflect the attribute. But through a continuing process of revision of the conception of the attribute and revision of items used to measure the attribute we are likely to obtain a better understanding both of the attribute and the measuring device."

Apart from using the concept of disattenuated correlation and testing the various hypotheses under different constraints, the difference in RFI between the model is another indication of 'lack of unidimensionality'. The 'unidimensional RFI difference' is \( \pm 0.01 \) if \( \text{RFIH}_1 - \text{RFIH}_2 \ (\text{RFIH}_3 - \text{RFIH}_4 \ < \ 0.01) \) while the 'non-unidimensional RFI difference' is \( > \pm 0.1 \) if \( \text{RFIH}_1 - \text{RFIH}_2 \ (\text{RFIH}_3 - \text{RFIH}_4 \ > \ 0.1) \]. These differences could be studied further with instruments with a large number of items. Furthermore, the psychometric properties of achievement test data can be compared to those of aptitude tests. Birenbaum and Tatsuoka (1983), contend that the main difference between achievement and aptitude tests lies in their dimensionalities.

Since, "no actual psychological measurement is likely to be exactly unidimensional," (Hulin, Drasgow and Parsons, 1983, p. 40) it is better to attempt to identify the items through the above process to 'qualify' the instrument to allow application of IRT. In real situation, there is the possibility of less than ideal unidimensional data existing (De Ayala, 1992). Thus, in line with the argument advanced by Reckase, Ackerman and Carlson (1988) that the unidimensionality requirement of IRT does not necessarily demand that test items measure a simple ability, but rather the unidimensionality assumption requires that the test items measure the same composite of abilities, lends credibility to the above methodology of identifying unidimensionality. This is parallel to Bejar's (1983, p.31) argument that unidimensionality does not imply that a single characteristics or process operates. If several characteristics or processes function in unison, then unidimensionality will hold. If the characteristics or processes do not function in concord, then it is not meaningful to assign numbers to a combination of the characteristics, unless another operation such as that of prediction provides the rule for combination. If a set of characteristics lacks this internal consistency and fidelity, or if it involves several identifiable dimensions, each must be measured separately and a profile of measures recorded.
Reference


Linacre, J.: 1994, 'A DIMTEST Diminuendo', Rasch Measurement Transactions, 8(3). (Internet article - gopher_root_eric_ae:[rasch]rm83.txt;1)


1 If the test scores $x_1$, $x_2$, ..., $x_m$ are congeneric, then any pair of true scores $t_i$ and $t_j$ correlate unity and hence there exists a random variable $\zeta$ such that each $t_i$ is linearly related to $\zeta$.

Disattenuated Correlation and Unidimensionality

1