DOCUMENT RESUME

ED 343 959 TM 018 190

AUTHOR	Oort, Frans J.
TITLE	Using Restricted Factor Analysis in Test Construction.
PUB DATE	92
NOTE	6p.
PUB TYPE	Reports - Evaluative/Feasibility (142)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	*Attitude Measures; Educational Research; Elementary School Students; Elementary Secondary Education; *Evaluation Problems; Foreign Countries; *Item Bias; Mathematical Models; Measurement Techniques; Rating Scales; Secondary School Students; *Test Construction
IDENTIFIERS	Netherlands; *Restricted Factor Analysis; School Attitudes Questionnaire (Voorst); *Theory of Violators; Unidimensionality (Tests)

ABSTRACT

The use of restrictive factor analysis (RFA) is illustrated with data from the field of educational research. An attitude scale measuring the degree to which students feel they are accepted by their classmates is constructed in accordance with the Theory of Violators. The main idea of the Theory of Violators is that most problems associated with measuring psychological constructs can be reduced to the question of unidimensionality. The theory consists of three components: (1) a formal base of definitions of item purity and unidimensionality that introduces the concept of violator, a variable with respect to which the test item is biased; (2) a typology of violators; and (3) instruction on how to construct a scale that is free of item bias, unidimensional, and efficient. As a first step, test items were selected from an existing measure, the School Attitudes Questionnaire of H. C. M. Voorst (1990) administered in the Netherlands to 915 students aged nine to 12 years. Item bias detection was then performed through statistical analysis. After eliminating biased items, the remaining items were tested for unidimensionality through factor analysis. Inefficient items were then detected through RFA. This exercise demonstrates that RFA offers a single approach to various measurement problems. There is an eight-item list of references. (SLD)

USING RESTRICTED FACTOR ANALYSIS IN TEST CONSTRUCTION

Frans J. Oort
University of Amsterdam

Using Restricted Factor Analysis in Test Construction

Frans J. Oort
University of Amsterdam

In psychology (and education) there are all sorts of tests measuring abilities, aptitudes, personality traits, attitudes, moods, emotions, and so on. To be valid, each test scale should be unidimensional and free of item bias. Using the common factor model as an item response model, restrictive factor analysis (RFA) can be used to detect item bias, to test for unidimensionality, and to replace ordinary item analysis. In the present paper, the use of RFA is illustrated with some data from the field of educational research. An attitude scale, measuring the degree to which school pupils feel that they are socially accepted by their classmates, will be constructed in accordance with the Theory of Violators (Oort, 1989, 1991).

The main idea of the Theory of Violators is that most problems that are associated with measuring psychological constructs can be reduced to just one problem: The question of unidimensionality. The theory consists of three components. The first component is a formal base made up by definitions of item purity and unidimensionality. In these definitions, in order to link all measurement problems, the concept of a potential violator is introduced. A violator is a variable with respect to which a test item is biased. The second component is a typology of violators. The third component is an instruction on how to construct a scale that is free of item bias, unidimensional and efficient.

Item purity is defined as a conditional independence (Mellenbergh, 1985, 1989): An item X is pure (unbiased) with respect to potential violator V and given trait T if and only if

$$f (X \mid V = v, T = t) \quad = \quad g (X \mid T = t) \qquad (1)$$

for all values v and t of the variables V and T, where function f is the distribution of the item responses given v and t, and function g is the distribution of the item responses given t. Notice that the item purity definition does not depend on the measurement levels of variables X, V, and T. The unidimensionality definition simply states that a scale is unidimensional if each and every one of its constituting items is pure with respect to any potential violator that might be relevant.

A central idea in the Theory of Violators is that different measurement problems only differ in what is substituted for potential violator V. Five types of potential violators are distinguished: (1) the Item type, referring to the principle of local independence; (2) the Trait type, referring to the question of construct validity; (3) the Response Style type, referring to contaminating effects of response styles; (4) the Group type, referring to (in)stability of item parameters over groups (as in traditional item bias research); (5) the Time type, referring to (in)stability of item parameters over time.

The general goal in test construction is to arrive at a test of minimum length that will yield scores with the necessary degree of reliability and validity for the intended uses (e.g. Crocker & Algina, 1986). Beginning with quite a few items, a selection has to be made. To accomplish that this selection yields (sub-)scales that are free from item bias, unidimensional and efficient, three steps have to be taken: (1) choose and operationalize potential violators; (2) detect and remove biased items, and test for unidimensionality; (3) detect and remove inefficient items. These steps will be pursued when they are applied to the construction of the 'social acceptance' (SA) scale.

To have some items to start off with, SA items are simply taken from an existing standard test, the School Attitudes Questionnaire (SAQ) of Vorst (1990). The SAQ is a Dutch questionnaire consisting of 160 items with three-point response scales measuring children's educational attitudes: motivation for school tasks, satisfaction with school life, and self-confidence about one's scholastic capabilities. There are 16 items measuring SA. This SA scale will be shortened to become free of item bias, unidimensional, and efficient. As mentioned above, the first step in test construction is to choose and operationalize potential violators. For the assertion of unidimensionality the choice of potential violators is crucial. In this case, potential violators are taken from the SAQ as well. Variables operating as potential violators of the Trait type are: Motivation and Self-confidence, both operationalized by adding 48 SAQ items, and Social Desirability, operationalized by adding 16 SAQ items. Yea-saying is a potential violator of the Response Style type and is operationalized as the number of times that a confirmative answer is given, irrespective of item content. Two demographic variables, Sex and Age, serve as potential violators of the Group type. The SAQ was administered to 915 school pupils, 467 males and 448 females, aged nine to twelve.

The second step in test construction, item bias detection, is done through statistical analysis. Many item bias detection methods are available. The choice of a particular item bias detection method is dependent on the levels of measurement of test items and potential violators. Here, RFA is applied. By accommodating all items and all potential violators in one factor model, it is possible to test for direct effects of potential violators on items (Oort, in press). Three SA items appear to be severely biased, one with respect to Yea-saying, and two with respect to other SA items, thereby violating the principle of local independence. Three more SA items appear to be moderately biased with respect to Social Desirability, Age, and Yea-saying, respectively. After the removal of all biased items, the set of remaining items should be unidimensional (by definition). Using the common factor model as an item response model, unidimensionality may be tested for by fitting an one-factor model to the data.

The third step in test construction is to detect inefficient items. Thus far, the focus was on item bias, i.e., on items that measure something else besides what they are supposed to measure. Another possible flaw of test items is that they measure the trait of interest insufficiently. Remembering the general goal of test construction, an item should be efficient. An efficient item contributes to the reliability and validity of a scale. To detect inefficient items, a traditional item analysis can be done. Alternatively, one can make use of the information that one gets from fitting an item-response model. Again, the choice of model is partly determined by the measurement level of the test items. Here, the common factor model is used as an item response model and RFA is applied once more. Inefficient items can be detected by looking at the item reliability index or the information value. The item reliability index equals the quotient of common variance and total variance, and the item information value equals the quotient of common variance and residual variance (Mellenbergh, 1991). A rigid look at the SA items, reduced to ten already, yields two inefficient items. Removing them has only a small effect on the reliability of the SA scale as a whole.

This exercise in using RFA to construct a scale in accordance with the Theory of Violators shows that there is more to item bias research than just bias in dichotomous items with respect to nominal violators. This exercise also shows that RFA offers a single approach to various measurement problems. And, what is more, by investigating item bias with respect to several potential violators at once, item bias often is better understood.

However, RFA has its disadvantages too. Firstly, RFA really cannot be applied to dichotomous items, though there are ways round the problem of factor analyzing non-continuous variables. Secondly, by using RFA, the notion of statistical independence in the item purity definition is relaxed to the notion of linear independence. Therefore, non-uniform bias cannot be investigated.

Finally, there is one caveat that has to be addressed. When working with RFA in the above way, one has to be careful not to over-emphasize unidimensionality at the cost of, for example, content validity. A test constructor has to compromise between unidimensionality, efficiency, reliability and validity.

References

Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. New York: Holt, Rinehart and Winston.

Mellenbergh, G.J. (1985). Vraag-onzuiverheid: definitie, detectie en onderzoek. [Item bias: Definition, detection, and research.] *Nederlands tijdschrift voor de psychologie, 40*, 425-435.

Mellenbergh, G.J. (1989). Item bias and item response theory. In: Hambleton, R.K. (Ed.), Applications of item response theory [Special issue]. *International Journal of Educational Research, 13*, 127-143.

Mellenbergh, G.J. (1991, July). *The unidimensional latent trait model for continuous item responses.* Paper presented at the 7th European Meeting of the Psychometric Society, Trier, Germany.

Oort, F.J. (1989). *Potentiële schenders van de eendimensionaliteit van psychologische meetinstrumenten.* Unpublished manuscript. Amsterdam: Psychologisch Laboratorium, University of Amsterdam.

Oort, F.J. (1991, July). *Theory of Violators: Assessing Unidimensionality of Psychological Measures.* Paper presented at the 7th European Meeting of the Psychometric Society, Trier, Germany.

Oort, F.J. (in press). Using restricted factor analysis to detect item bias. *Methodika.*

Vorst, H.C.M. (1990). *Handleiding en verantwoording bij de SchoolVragenLijst.* [*School Attitudes Questionnaire Test Manual*]. Nijmegen: Berkhout Nijmegen BV.