0683

DATA FEATURES AFFECTING THE DETECTION OF MULTIDIMENSIONALITY

BALAZS, K. and P. DE BOECK

Running head: DETECTION OF MULTIDIMENSIONALITY

Data Features Affecting the DETECTion of Multidimensionality

Katalin Balázs and Paul De Boeck

K.U. Leuven

Psychology Department
Tiensestraat 102.
B-3000 Leuven, Belgium

Data Features Affecting the DETECTion of Multidimensionality

Abstract

The dimensionality of binary data can be assessed by the DETECT procedure. Although DETECT has been used in several studies, the data features that affect the DETECT index are still relatively unknown. Previous studies were concentrated on the DETECT decision rules, so that the available evidence is only indirectly relevant for the index at the basis of the decision rules. Based on two simulation studies, the effects of several data features on the DETECT index are reported without applying a decision rule. The main data features in question are the correlation of the dimensions, and the relative importance of the dimensions both in terms of the distribution of the items over the dimensions, and in terms of the relative size of the dimensional variances. Further, also features, such as simple structure, the overall size of dimensional variance, variance of the item difficulties, and sample size are studied. The results provide both, a better understanding of the DETECT index and practical conclusions for its use.

Keywords: dimensionality, multidimensionality index, nonparametric, DETECT, binary data,

In psychological and educational measurement, an assessment of the dimensionality of the collected data is substantial, both for technical and theoretical reasons. As unidimensionality is a requirement for a number of standard statistical procedures, one may want to check unidimensionality before further analysis. Even if the data turn out to be multidimensional, any information on the underlying structure can be helpful for deciding on further analysis. From the substantive point of view, the researcher may want to test a theory about the dimensionality of individual differences, and/or one can be interested in the possible classification of the situations, stimuli or items at hand, based on the dimensions. Besides, exploratory dimensionality analysis can make sense in many cases, in order to obtain a better view on the data and the domain under investigation.

Hattie provided an empirical study (1984) and a profound review (1985) of statistical methods for assessing dimensionality. He pointed out that most of the available indices are not suitable for detecting dimensionality, because the underlying assumptions are not necessarily correct. Nevertheless, indices based on the residuals after fitting a two- or three-parameter logistic model are considered to be useful (Hattie, 1985).

Various methods are applied in practice for revealing the multidimensional structure of data, for example factor analysis (see e.g., De Ayala & Hertzog, 1991; Hattie, Krakowski, Rogers & Swaminathan, 1996; Reckrease, 1979), multidimensional nonlinear factor analyses (Wilson, Wood & Gibbons, 1991),Yen's indices for local independence (1984), DIMTEST (Hattie et al., 1996; Nanadakumar & Stout, 1993; Stout, Douglas, Junker, & Roussos, 1999), DETECT (Stout, Habing, Douglas, Kim, Roussos & Zhang, 1996). Several comparative studies are available on dimensionality assessment procedures (Bolt, 2001; Dimitriadou, Dolnicar & Weingessel, 2002; Hambleton & Rovinelli, 1986; Hattie, 1984; Roznowski, Tucker & Humphreys, 1991; Tate, 2003; van Abswoude, van der Ark & Sijtsma, 2004). The present paper concentrates on a particular and relatively new method, the DETECT procedure (Stout et al., 1996), and data features affecting its efficiency in revealing the dimensionality structure of data.

DETECT (Dimensionality evaluation to enumerate contributing traits), a prominent and popular method for the investigation of the underlying dimensional structure of data, is a promising recent method. The method was developed by Kim (1994) (Stout et al., 1996), based on the earlier work of Junker and Stout (1994). DETECT has been further investigated and improved since then (e.g., Zhang & Stout,1999). In case of a simple structure, DETECT discloses the number of underlying dimensions. The procedure also identifies the corresponding partitioning of the items and yields an index of multidimensionality. A simple structure is realized when non-overlapping item subgroups can be identified *and* the items of a subgroup measure the same ability (Zhang & Stout, 1999; Tate, 2003). The degree of simple structure is assessed by the R index, to be explained after the DETECT index is presented. When the data have approximate simple structure, the DETECT procedure finds the most prominent dimensions. However, in such cases, DETECT may merge minor dimensions (with small variance, a few items) with a dominant dimension (with large variance, many items) or it may partition items of one dimension, as if they formed more than one dimension.

DETECT uses a genetic algorithm for finding the optimal partitioning which provides the highest DETECT index. The theoretical DETECT index (D_α) is based on the sum of expected covariances of the item pairs conditional upon the test composite (see Equation 1).

$$D_\alpha(P) = \frac{2}{I(I-1)} \sum_{1 \le i_1 < i_2 \le I} \delta_{i_1 i_2}(P) E\left[Cov(X_{i_1}, X_{i_2} \middle| \theta_\alpha)\right], \quad (1)$$

where I is the number of the items, θ_α is the test composite, X_{i_1} and X_{i_2} are the scores on item pair i_1 and i_2, P refers to a given partition and $\delta_{i_1 i_2}(P)$ is an indicator having value +1 when the item pair (i_1, i_2,) belongs to the same cluster in the current partition, and –1 if not. $Cov(X_{i_1}, X_{i_2} | \theta_\alpha)$ is the conditional covariance given θ_α. The $D_\alpha(P)$ index reflects the amount of ignored multidimensionality when considering a general underlying trait θ_α. The bias corrected estimator of $E\left[Cov(X_{i_1}, X_{i_2} \middle| \theta_\alpha)\right]$ is $\frac{1}{2}\left[\hat{C}ov_{i1i2}(S) + \hat{C}ov_{i1i2}(T)\right]$, where $\hat{C}ov_{i1i2}(T)$ is the conditional covariance based on the total score, and $\hat{C}ov_{i1i2}(S)$ is the conditional covariance based on the rest score, which is the total score leaving out the item pair (Zhang & Stout, 1999). The DETECT procedure aims an optimal partition, such that the $D_\alpha(P)$ value is maximal. The corresponding estimate is denoted with $\hat{D}_{\max}$.

The DETECT software offers a cross-validation procedure, which provides the regular DETECT index estimate, called the *maximal* DETECT index ($\hat{D}_{\max}$), but only for a subset of the data ($\hat{D}'_{\max}$), and an additional cross-validated DETECT index estimate, called the *reference* DETECT index ($\hat{D}_{ref}$)(Zhang & Stout, 1999). In the DETECT cross-validation procedure, one half of the data is used for obtaining a $\hat{D}'_{\max}$ index and a partitioning of the items, which is further used for the second subset of the data, providing a $\hat{D}_{ref}$ index. Logically, the $\hat{D}_{ref}$ index is expected to approach the $\hat{D}'_{\max}$ index quite well when the partitioning fits the data well, but $\hat{D}_{ref}$ is much smaller than $\hat{D}'_{\max}$ when the partitioning is based upon chance.

Information about the degree of simple structure is provided by the R index. The R index is basically the DETECT index divided by a modified DETECT index, where the *absolute value* of the conditional covariances of the item pairs is used instead of the signed values:

$$R(P) = \frac{D_\alpha(P)}{\frac{2}{I(I-1)} \sum_{1 < i_1 < i_2 \le I} \left| E\left[Cov(X_{i_1}, X_{i_2} \middle| \theta_\alpha)\right] \right|} \qquad (2)$$

In case of cross-validation, beyond the $\hat{D}'_{\max}$ and $\hat{D}_{ref}$ indices, $\hat{R}$ and $\hat{R}_{ref}$ indices are also provided.

The DETECT index is a continuous measure of multidimensionality. For practical purposes, it is useful to have a decision rule in order to decide upon unidimensionality or multidimensionality. Ideally, one may want to have a criterion based on the distribution of the DETECT index under the null hypothesis of unidimensionality, but unfortunately this distribution is not known.

The DETECT manual (The William Stout Institute for Measurement, 2003), which also contains two published manuscripts (Stout et al., 1996; Zhang & Stout, 1999) and a research report (Douglas, Kim, Roussos, Stout & Zhang, 1999), offers more than one decision rule for assessing dimensionality. Unfortunately, this leads to different approaches in practice, as it will be illustrated in the following.

Decision rules

First, the regular DETECT index ($\hat{D}_{\max}$) calculated on the whole data set is often used for decision making. In the Law School Admission Council Statistical Report (Douglas et al., 1999), the data is classified based on this $\hat{D}_{\max}$ index, in agreement with Kim's work (1994). A $\hat{D}_{\max}$ (and similarly $\hat{D}'_{\max}$) index between 0 and .1 indicates unidimensionality, a $\hat{D}_{\max}$ index between .1 and .5, .5 and 1, 1 and 1.5, and 1.5 or higher value correspond to weak, moderate, strong and very strong multidimensionality, respectively (p. 7). In addition, it is emphasized that a $\hat{R}$ index higher than .8 is required for accepting a result as an indication of multidimensionality. In a simplified version for decision making (Stout et al., 1996), the five categories are reduced to three: unidimensionality ($\hat{D}_{\max} \leq 0.1$), moderate multidimensionality

($0.1 < \hat{D}_{\max} \leq 1$), and sizable multidimensionality ($1 \leq \hat{D}_{\max}$). In practice, this seems to be the common method (e.g., Bolt, 2001; Bouwmeester & Sijtsma, 2004, van Abswoude et al., 2004).

However, van Abswoude et al.(2004) proposed that the critical value for unidimensionality should be increased. Accordingly, Uribe-Zarain, Nandakumar and Yu (2004) applied a classification referring to a new DETECT manual in preparation (Moran, Roussos & Kim, 2005): unidimensionality if the $\hat{D}_{\max}$ index is smaller than 0.2; weak multidimensionality if the $\hat{D}_{\max}$ index is between 0.2 and 0.4; and moderate multidimensionality if the $\hat{D}_{\max}$ index is between 0.4 and 1; strong multidimensionality for $\hat{D}_{\max}$ indices higher than 1.

Second, Tate (2003) applies cross-validation in his study, and uses 0.1 and 1 as cut-off values (of $\hat{D}_{ref}$). These cut-off values are used for deciding upon unidimensionality or multidimensionality and upon essential or "sizeable" multidimensionality, respectively . Similarly, Zhang and Stout (1999) strongly recommend implementing cross-validation, and they introduce a new kind of criterion: when $(\hat{D}'_{\max} - \hat{D}_{ref})/\hat{D}_{ref} < 0.5$ or $\hat{D}_{ref} < 0.1$, unidimensionality is concluded (p. 247). In the following, $(\hat{D}'_{\max} - \hat{D}_{ref})/\hat{D}_{ref}$ as a single index, and $(\hat{D}'_{\max} - \hat{D}_{ref})/\hat{D}_{ref} < 0.5$ or $\hat{D}_{ref} < 0.1$ as a conjugate criterion will be referred to as the *discrepancy measure* and the *combined decision rule*, respectively. In another study (Balazs, Hidegkuti & De Boeck, submitted), these above described ad hoc cut-off values resulted in a rather large percentage of missers (type II errors, 22% of the data).

From this set of studies, it is not clear which set of decision rules is the best one and hence, should be used. As it is clear from the previous, there are basically four different indexes for deciding upon multidimensionality: $\hat{D}_{\max}$ for the total data set or $\hat{D}'_{\max}$ for a subset of the data (without cross-validation), $\hat{D}_{ref}$ (from the cross-validation), and the discrepancy measure. In the first simulation study to be reported, the basic indices will be compared on their validity, and then the best one will be used to investigate the effect of data features.

Factors affecting DETECT

Various factors or data features have been studied to find out the effects they have on DETECT: size factors such as sample size and number of items, as well as structure factors such as dimensionality and simple structure.

Size factors

Most of the time a large *sample size* is used, 2000 persons are common (Bolt, 2001; Tate, 2003). Van Abswoude et al. (2004) generated 2000 and 200 observations in their study. The sample size in Zhang and Stout's study (1999) was either 1600 or 800, and in Balazs, Hidegkuti and De Boeck's study (submitted) it was 200.

The results for the sample size were as follows: DETECT found the underlying structure of multidimensional data more often for data with 1600 examinees than for data with 800 examinees (Zhang & Stout, 1999). For a sample size as small as 200, DETECT performed less well than for 2000 examinees (van Abswoude et al. 2004). Although a small sample size seems to be problematic, in some domains, such as psychology, a sample size of 200 or 400 is rather common. In such a situation, a researcher who uses cross-validation would end up with 100/100 or 200/200 examinees per DETECT index, which may be too few for revealing the underlying structure. Unfortunately, based on the available literature, one cannot make an estimate of the smallest sample size for which DETECT can be used effectively.

The *number of items* was manipulated in a study by van Abswoude et al. (2004). Zhang and Stout (1999) used either 20 or 40 items, and in these latter simulations the distribution of the items over dimensions were not varied. The dimensional structure was recovered better for 21 items than for 7 items, in case of a five parameter acceleration model (van Abswoude et al., 2004), but the number of items did not have an effect in case of the two-parameter logistic model (van Abswoude et al., 2004).

In other studies, also the *distribution of the items over the dimensions* was manipulated. Tate (2003) generated two-dimensional data with a major dimension (50 items) and a minor dimension (10 items) and four-dimensional data with equally important dimensions. In the study of van Abswoude et al. (2004), there were two or four-dimensional data sets, with 7 or 21 items loading on the dimensions, in different combinations. Bolt (2001) assigned 25 items to two to four dimensions and manipulated the distribution of the items over the dimensions by grouping them in various ways.

In Tate’s study, DETECT was able to recover the dimensional structure of both the equally and the unequally distributed items. According to the result of van Abswoude et al. (2004), the recovery of the dimensional structure was better for data with equally distributed items over dimensions, than for data with unequally distributed items. This was explained by the location of the test composite that is closer to the item clusters with more items (van Abswoude et al., 2004). The same finding was reported in Bolt’s study (2001). When the correlation between the dimensions was not high (only .5), DETECT recovered the latent structure of data with two dimensions equally well independently of the distribution of the items. However, for the same correlation value, DETECT revealed only two dimensions out of three of a data set when 14, 7, and 4 items loaded on the dimensions. While DETECT found the three dimensions of data when 8, 8 and 9 items loaded on the dimensions. Similarly, the dimensional structure of a four-dimensional data set was revealed when 6, 6, 6 and 7 items belonged to the dimensions, and DETECT failed to reveal the dimensional structure in case of 10, 7, 5 and 3 items loading on the dimensions.

Structure factors

Different structures for data generation were used, most often the unidimensional and multidimensional (two to four dimensions) versions of the Rasch model (Tate, 2003), the two-parameter logistic model (Bolt, 2001; Tate, 2003; van Abswoude et al., 2004; Zhang & Stout, 1999) or the three-parameter logistic model (Tate, 2003; Zhang & Stout, 1999) were applied. The basic model for data generation does not seem to have an impact on the effectiveness of DETECT. Tate (2003) used the Rasch model, the 2PL and 3PL models, and the underlying structure was always found. It seems it is not the type of model that is vital, but other features of the model.

Besides, the *correlation* among the dimensions was often included as a factor in the simulation design (e.g., Bolt, 2001; Tate, 2003; van Abswoude et al. 2004). Van Abswoude et al. (2004) varied the correlation systematically form zero to one with steps of 0.2. It appears that a *correlation* from a value of .8 on is too high to differentiate between the dimensions based on the $D_{\max}$ index (Bolt, 2001; van Abswoude et al., 2004).

Some researchers used data with *a simple structure* (e.g., van Abswoude et al. 2004), others used data with an approximate simple structure (e.g., Zhang & Stout, 1999), or this data feature was included as a design factor in the study (e.g., Bolt, 2001; Tate, 2003). According to Zhang and Stout (1999), the effect of lack of simple structure is not strong, but in other studies, the precise structure could not be revealed in such cases. In Bolt's study

(2001), the dimensionality structure of data with lack of simple structure was not recovered. In his study, the data sets without simple structure contained items which measured various composites of two dimensions. Furthermore, in Tate's study (2003), DETECT was able to provide the dimensions of a data set with three underlying dimensions and approximate simple structure, and of a data set with five dimensions when one dimension correlated with the other four. But DETECT did not succeed in revealing the structure of a data set with one dominant dimension and a minor dimension, and of data with diffuse structure when the item discriminations were continuously varied between complete dominance of the first and complete dominance of the second dimension (Tate, 2003).

The extremity of the *discrimination parameters* (Tate, 2003; van Abswoude et al., 2004) and the extremity of the *item difficulties* (Tate, 2003) are further structural features that were explored for their effects.

The extremity of the item difficulties caused a problem for DETECT when the data were multidimensional (Tate, 2003). But the data sets with highly discriminating items were not problematic at all: the higher were the discrimination parameters, the better was the recovery of the structure of the data (Tate, 2003; van Abswoude et al., 2004).

Finally, the *variance of the underlying dimensions* has also been looked at (Balazs et al., submitted). The overall size of the dimensional variance was shown to have an effect on DETECT (Balazs et al., submitted). A variance of 0.2 seemed to be rather small for DETECT to detect multidimensionality. But in general, it is unknown which overall size of the dimensional variance is sufficient for DETECT to reveal multidimensionality. The size of the variance is an alternative way to approach discrimination strength, and therefore these results are in line with those of Tate (2003) and van Abswoude et at. (2004) on the extremity of the discrimination.

In sum, both the size of the data set and the structure of the data seem to have an effect on the DETECT index and hence, on the success of the procedure in revealing the underlying structure of data. However, since several decision rules are used in practice, overall conclusions about the conditions in which DETECT can be effectively used are difficult to draw. Because most studies report mainly the success or failure of a given criterion in revealing the underlying structure of the data, and another criterion may yield different results, it is not always possible to generalize from these studies.

The aim of the present studies is not to check the validity of the cut-off criteria, but to investigate the effects of various data features on the DETECT index. The validity of the cut-offs is a secondary issue because the cut-offs are defined on one or more DETECT indices. First the behaviour of the DETECT index should be explored. Considering several possibly influential data features, two simulation studies were conducted that will be described in detail after a discussion of the concept of dimensionality.

The concept of dimensionality

In order to interpret the findings of the studies to be described and of earlier studies, several data features will be discussed from the concept of dimensionality.

The concept of unidimensionality may not be confused with reliability, internal consistency and homogeneity referring to perfectly homogeneous intercorrelations (Hattie, 1985). The dimensionality of a test is independent from its reliability or internal consistency, and the desire for high intercorrelations of all items leads to tests with a rather narrow specific focus and should not be aimed at (Cattell 1964, 1978; Cattell & Tsujioka, 1964; Hattie, 1985).

Unidimensionality exists if there is only one underlying trait in the data. Classical test theory assumes that the items measure the same dimension (e.g., Nandakumar & Stout, 1993), although this assumption is often rather questionable (Hableton & Swaminathan, 1985; Humphreys, 1981,1985; Lumsden, 1961; Nandakumar, 1993; Nandakumar & Stout, 1993; Reckase, 1979, 1985; Stout, 1987; Traub 1983).

In traditional IRT, the dimensionality of a test is understood as the number of dimensions that results in a monotone and locally independency model (Tate, 2003). When one dimension is sufficient to fulfill these requirements, the data are *strictly unidimensional* (Hattie et al., 1996; Tate, 2003; van Abswoude et al., 2004). Nevertheless, since several examinee and item characteristics can influence dimensionality, this seems to be a very strict requirement (e.g., Hattie et al., 1996; Nandakumar & Stout, 1993).

McDonald (1979, 1981) suggested to revise the *strong local independence* principle, meaning that the probability of a given response pattern for a pair of items is a product of the probabilities of the separate responses given the latent trait (Lord & Novick, 1968). The revision implies the assumption of *weak local independence*, that is, the responses are mutually uncorrelated given the latent trait. In 1990, Stout proposed to use the *essential local independence* principle, namely that the conditional covariance of the items are small given the latent trait.

Based on this principle*, essential dimensionality* can be defined as the minimum number of dimensions needed to satisfy the assumptions of monotonicity and essential local independence. Consequently, the essential dimensionality of a test may be equal or lower than the strict and weak dimensionalities. In other words, items being *relatively* homogeneous and *mainly* reflecting the ability at the basis of the test can form *essentially unidimensional* test (Junker, 1991; Nandakumar, 1991, 1993; Stout, 1990; Stout et al., 1996; Tate, 2003; Zhang & Stout, 1999). The DETECT procedure is developed for investigating essential dimensionality (Zhang & Stout, 1999).

Multidimensionality can be differentiated in several ways. Multidimensional structures can be either compensatory or non-compensatory (see e.g., Bolt & Lall, 2003). In compensatory models (e.g., Reckase, 1985), the score on a dimension can compensate the scores on another dimensions unlike in non-compensatory multidimensional structure (e.g., Whitely, 1980). When between-item multidimensionality exists, none of the items are meant to measure more than one dimension, whereas in case of within-item multidimensionality they are (Wang, Wilson & Adams, 1997; Wang & Chen, 2004). Kirisci, Hsu and Yu (2001) differentiate between two types of multidimensionality generating techniques: equally dominant and perhaps correlated dimensions and dimensions differing in their dominance.

Furthermore, it is a basic idea that multidimensionality is gradual (e.g., Hattie, 1984, 1985) and that data features contribute to multidimensionality to a certain extent. Some features are directly relevant, others are less clearly relevant and a third category of data features is conceptually irrelevant. The results of the DETECT procedure will be explained keeping in mind this discussion of multidimensionality.

At one extreme, one finds the completely unidimensional structure. Moving away from this extreme to the opposite direction, the degree of the multidimensionality is increasing depending on the following two, directly relevant features and their combination.

First, the higher is the *correlation between the dimensions*, the more difficult it is to differentiate among them, and therefore the less multidimensional the data are.

Second, the relative importance of the dimensions plays a role. When the second (the third, etc.) dimension is less important than the first, the data are less multidimensional than when the dimensions are equally important. The importance of the dimension concerns the *relative size of the variance* and the *distribution of the items*.

Other features are less clearly relevant for multidimensionality, but are related to the way the DETECT index is defined. A first feature is the *overall size of dimensional variance*, given that there is more than one dimension. It is an issue whether a structure with two dimensional variances of 1 is more multidimensional than one with two dimensional variances of .2 instead. The way the DETECT index is constructed implies it is, because of the summation of conditional covariances in Equation 1, and the conditionality being based on the composite score.

A second feature in this category, is whether the *simple structure* is realized. Lack of simple structure does not necessarily imply less multidimensionality but less homogeneous dimensions. Again, the DETECT index depends on whether the simple structure is realized, because the way it is defined requires distinct item clusters.

Finally, other features such as the *sample size*, or the *variance of the item difficulties* do not play a role at all in multidimensionality, since there is no conceptual basis for a link. Other features may be considered to have an effect, but here only these two are investigated, among the broader set of features unrelated to multidimensionality.

From this conceptual analysis, it is desirable that DETECT is sensitive to the correlation of the dimensions, and to the relative importance of the additional dimensions as defined in terms of relative size of the dimensional variance and in terms of the distribution of items over the dimensions. It is expected, but not especially desirable, that DETECT is influenced by the overall size of the dimensional variance and by the degree of simple structure. Finally, it is undesirable that the sample size and the variance of the item difficulties affect the DETECT index. Apart from investigating these expectations, it is of interest to assess the size of the impact of these factors and their interactions.

Simulation studies

Two simulation studies were carried out. Most of the design features in the two studies are the same. The main differences concern the way in which multidimensionality was varied in the design.

In the first study, the correlation between dimensions and the relative importance in terms of items was varied but not the relative importance in terms of variance. There were always two dimensions with equal variances, but when the correlation between the two dimensions was one, the two dimensions collapse into one.

In the second study, the relative importance was varied both, in terms of relative size of the dimensional variance and in terms of the distribution of the items over the dimensions, but the correlation between the dimensions was always zero. The two dimensions reduce to one when the variance of the second dimension is zero.

In the following, the common features of the design structures are described, and next the specific factors of the designs are presented.

Common design factors

Data sets were generated based on a two-dimensional 2PL (Equation 3).

$$\text{logit}(P(y_{pi}=1)) = \alpha_{1i}\theta_{1p} + \alpha_{2i}\theta_{2p} - \beta_i, \tag{3}$$

where $P(y_{pi}=1)$ is the probability of success of person p on item i, α_{1i} is the discrimination of item i for ability $\theta_{1,}$ and β_i is the difficulty of item i. For an item of dimension one $\alpha_{1i}=1$ and $\alpha_{2i}=0$, and for an item of dimension 2 $\alpha_{1i}=0$ and $\alpha_{2i}=1$. In case of an approximate simple structure, $E(\alpha_{1i})=1$ and $E(\alpha_{2i})=0$ when the item belongs to the first dimension, and $E(\alpha_{1i})=0$ and $E(\alpha_{2i})=1$ when the items belongs to the second dimension.

The number of items was always 30. The *distribution of the items* over the dimensions was equal or unequal, 15/15 or 20/10, respectively.

In order to vary the degree of *simple structure*, the variances of the discrimination parameters within a dimension was either 0 or 0.2. When the variance was 0.2, the simple structure was not perfect, but only approximate. Since the definition of simple structure (see, e.g. Zhang and Stout, 1999) allows for correlated dimensions, the existence of simple structure was purely determined by the deviations of the item discriminations from the average discrimination for the items belonging to the given dimension.

The *sample size* was varied systematically: 100, 200, 400, 600, 800 and 1000 examinees were used. Note, that the sample size for the cross-validation procedure was only half of the above mentioned values. For the cross-validation, the data sets were randomly splitted into two halves.

The mean of the item difficulties was 1 in each data set, and the *variance of the item difficulty* parameter was 0.2, 0.6, 1.

Specific design factors of the studies

In Study one, two additional factors were varied. The *correlation between the dimensions* was 0, .2, .4, .6, .8, or 1. In the last case, the data were unidimensional, and not two-dimensional. DETECT is expected to be less efficient in revealing the multidimensionality of the data when the correlation is high.

The latent dimensions were generated with normal distribution and zero means. *The dimensional variances* were equal for the two dimensions, and the *overall size of the dimensional variance* was varied between 0.2 and 1 (for each dimension), with steps of 0.2.

All six design factors (distribution of the items over the dimension, simple structure, sample size, variance of item difficulty, correlation, overall size of the dimensional variance) were fully crossed, and one data set was generated for each cell in the design. In this way, 2160 data sets were generated for the first study.

Study two was planned in order to see how the *relative size of the dimensional variance* affects the DETECT index. The variance of the first dimension was always 1, whereas the variance of the second was varied between zero and one by steps of 0.2. When this latter variance is zero, the data are unidimensional. The correlation between the dimensions were always zero.

All five design factors (distribution of items over dimensions, simple structure, sample size, variance of item difficulty, relative size of the dimensional variance) were fully crossed, and one data set was generated for each cell. In this way, 432 data sets were generated for the second study.

Implementing the DETECT procedure

The DETECT procedure was applied both with and without cross-validation. For the cross-validation, data sets were randomly divided into two halves, the first half was used to obtain the $\hat{D}'_{\max}$ index and in the second subset of the data $\hat{D}_{ref}$ was determined. For analyses without cross-validation, the whole data sets were used to calculate $\hat{D}_{\max}$. Based on the instructions of the DETECT program (The William Stout Institute of Measurement, 2003), the number of vectors to be mutated should be between a fifth (6 in this study) and a tenth (3 in this study) of the number of the items, and therefore 5 vectors were mutated. As unidimensional and two-dimensional data were simulated, and the differentiation between the two is in the focus of this study, two was chosen as the maximum number of dimensions to run the program. For the analyses with cross-validation, the minimum number of examinees per total score cell was set to 10, 10, 7, 5, 3 and 1, in the case of sample size 1000, 800, 600, 400, 200 and 100, respectively. For a sample size of 100, minimum 72% and on average 83% of the observations belonged to a score group larger than 1 person. When the minimum number of examinees per total score cell was set to 2 for the data sets, 61% of the 360 data sets did not fulfill the requirement that 85% of the examinees should be used in the analysis. This problem did not occur for analyses without cross-validation, when the minimum number of examinees per total score was set to 20, 20, 15, 10, 5 and 2, for sample size of 1000, 800, 600, 400, 200 and 100, respectively.

Results

Study one

Since four DETECT indices are used for decision making in the literature, the validity of $\hat{D}_{\max}$, $\hat{D}^{'}_{\max}$, $\hat{D}_{ref}$ and the discrepancy measure were investigated with data from Study one in a first step, in order to select an index to concentrate on. This selection was made based on a logistic regression analysis using the four indices as predictors of multidimensionality. A binary variable was created with value 1 for all two-dimensional and value 0 for all unidimensional data sets, and was used as the dependent variable in an analysis with the four indices as predictors. The results showed that the discrepancy measure $(\hat{D}^{'}_{\max} - \hat{D}_{ref})/\hat{D}_{ref}$ was not a significant predictor of multidimensionality (p=.207), but $\hat{D}_{\max}$, $\hat{D}^{'}_{\max}$ and $\hat{D}_{ref}$ were (p<.001 for all).

The fit of the models with different sets of predictors is shown in Table 1. The R^2_{adj} statistics were calculated based on Cox and Snell (1989, pp. 208-209) and Nagelkerke (1991), as implemented in SAS (SAS Institute, 1999). It is clear from the results that the discrepancy measure does not contribute significantly to the prediction of multidimensionality. $\hat{D}'_{\max}$ can be used to predict multidimensionality, but not as effectively as $\hat{D}_{\max}$ or $\hat{D}_{ref}$. It is important to note that twice as much data is used for the calculation of $\hat{D}_{\max}$ than for $\hat{D}'_{\max}$; and $\hat{D}_{ref}$ is also calculated on only half of the data, although by implementing the partitioning from the analysis of the other half.

The goodness of fit of a model with $\hat{D}_{\max}$ as a single predictor of multidimensionality, was somewhat better than of a model with $\hat{D}_{ref}$ only. When the indices, $\hat{D}_{\max}$ and $\hat{D}_{ref}$, were combined, the fit further improved, but adding $\hat{D}'_{\max}$ has no effect. The weighted sum of $\hat{D}_{\max}$ and $\hat{D}_{ref}$ has a higher validity than each of the two separately. For reasons of simplicity and because the weights are empirically determined for this study, and cannot be generalized to other studies, a single index will be used to study the effect of features of the data. Based on the comparative results, the $\hat{D}_{\max}$ index was selected to concentrate on, although all analyses were repeated with $\hat{D}_{ref}$, and the results were very similar to those obtained with $\hat{D}_{\max}$, which is not surprising given the high correlation between $\hat{D}_{\max}$ and $\hat{D}_{ref}$ (r =.849). The only difference occurred with respect to the effect of the sample size, which will be discussed in detail later.

Insert Table 1 about here.

Following the manual (William Stout Institute for Measurement, 2003), for multidimensional data, an $\hat{R}$ index value of at least .8 is needed to interpret the DETECT index, which corresponds to approximate simple structure. However, the presence of a simple structure versus the presence of approximate simple structure was not significantly related to the $\hat{R}$ index (p = .836) in this study. At the same time, $\hat{R}$ was highly correlated with $\hat{D}_{\max}$ (r = .807). Neither was $\hat{R}_{ref}$ significantly related to simple structure (p=.369) but it was instead highly correlated with $\hat{D}_{ref}$ (r=.946). The simple structure concept is only valid for multidimensional data, so multidimensional data sets were also investigated separately. The conclusions are the same: the $\hat{R}$ index and the $\hat{R}_{ref}$ index were not significantly related to simple structure (p = .656 and p=.781, respectively), and were highly correlated with $\hat{D}_{\max}$ and $\hat{D}_{ref}$ (r=.787 and r=.939, respectively)

Furthermore, for most of the data sets the $\hat{R}$ and the $\hat{R}_{ref}$ indices were smaller than .8 (92.8% and 94.7%, respectively), which corresponds to the lack of approximate simple structure according to the DETECT manual. The ratio of the data sets with R indices smaller than .8 was not much smaller for multidimensional data (79.4% and 87%, considering $\hat{R}$ and $\hat{R}_{ref}$, respectively).

Since the $\hat{R}$ indices show only a weak relation with the simple structure factor, but a high correlation with the DETECT values, their use must be doubted in the present study. It is possible, however, that the R index is more sensitive to stronger deviations from the simple structure, so that it can fulfill its role to modify the interpretation.

The fitted models

First of all, $\hat{D}_{\max}$ was modeled with all main effects of the design: correlation, distribution of the items over the dimension, simple structure, overall size of the dimensional variance, sample size, variance of item difficulty (Model 1). Later, the model was completed with all pairwise interactions of the dummy coded variables (Model 2), and in a next step, also all possible three-way interactions were included in the model (Model 3). Because only those three-way interactions turned out to be significant which contained both the correlation and the overall size of the dimensional variance, another model was applied, in which only the three-way interactions containing the two-way interaction of correlation and variance were added to the elements of Model 2 (Model 4). These are also the two predictors with the highest conceptual relevance.

Second, these four models were restricted to the linear and quadratic components of the main effects and their interactions (from Model 2 on). Of course, this reduction is only relevant for factors with more than three levels.

Third, four analogous models were implemented replacing the corresponding design values with the real values. These real values deviate slightly from the design values for the variances and for the correlations because of the stochastic nature of the generated data.

In order to test whether the fit can be improved by including additional interactions in the models, regression tree analyses (see e.g., Breiman, Friedman, Olshen & Stone, 1984; Chaudhuri, Lo, Loh & Yang, 1995; Loh, 2002) were implemented on the $\hat{D}_{\max}$ index, using the design factors as predictors. The regression tree analyses were carried out by applying the GUIDE software (Loh, 2004). This procedure searches for meaningful splits of variables and interactions of those split variables. The first split was always forced between the unidimensional and multidimensional data sets, and two regression trees were grown on these trunks (see Dusseldorp & Meulman, 2004). The obtained leaves were included in a regression model for predicting the $\hat{D}_{\max}$ index. In a next step, the residuals of this regression analysis were used as dependent variable, and two residual regression trees were grown in the same way as before. Binary variables created from the obtained leaves of the two times two trees were used in new regression analyses. These variables accounted for 69.6% of the variance of

$\hat{D}_{\max}$. Seven of these variables were not redundant with Model 1 and could be included in this main effects model, and two of them were not redundant with the variables of Model 2. None of the tree variables could be added to the variables of Model 3 or Model 4, as they were all redundant.

Insert Table 2 about here.

Table 2 shows the adjusted R^2 of the Models 1 to 4 in their original form, and either complemented with a regression tree, or limited to the linear and quadratic trends, or finally, estimated with real values as predictors. The conclusions from the regression analyses and the regression tree analyses are as follows:

1. Model 4 with the reduced set of three-way interactions seems to be the best model. It performs almost as well as Model 3, and no regression tree variables were found that were not redundant with the factors of the model.
2. Restricting the effects only to linear and quadratic trends does not have a drastic effect, but a small effect instead.
3. Given the limitation to linear and quadratic effects, it does not really pay off to use real values instead of the true values.

Based on these conclusions, the effects of Model 4 with the design values instead of real values will be focused on but without the restriction to linear and quadratic trends (first column in Table 2).

The effects of the design factors

The significant (p<.05) effects from the selected regression analysis of Model 4, together with the corresponding probabilities and the effect sizes (η^2), are shown in Table 3. Based on these results, the findings are as follows. The three most important (effect sizes $\eta^2 > .1$) predictors of the cross-validated DETECT index were the correlation, the overall size of the dimensional variance and their interaction. Six additional predictors contributed sizably to the prediction of $\hat{D}_{\max}$ (effect sizes $\eta^2 > .01$), including the sample size and the distribution of the items over the dimensions. The simple structure and the variance of the item difficulties were among the ten variables with a significant effect, but their contribution to the prediction of the DETECT index was only minor.

Insert Table 3 about here.

In the following, the effects will be reported and discussed based on the three categories of the conceptual analysis of dimensionality (highly relevant; not directly relevant, but implied by DETECT; irrelevant).

First, as expected, the effect of *correlation* between the dimensions is strong, actually it is the strongest predictor of $\hat{D}_{\max}$ ($\eta^2 = .467$). Figure 1 shows the results. These results are a strong indication for the validity of the DETECT index, and they confirm the earlier observations (Bolt, 2001; van Abswoude et al., 2004), for example, that a correlation as high as .8, results in rather small average DETECT index value (mean $\hat{D}_{\max}$ = 0.319). This value turned out to be only slightly larger than the average DETECT index value for unidimensional data sets ($\hat{D}_{\max}$=0.282) . The differentiation of two underlying correlated dimensions seems to be very difficult.

Insert Figure 1 about here.

Similarly, the *distribution of the items over dimensions* affected the DETECT results in the expected way, but not as much ($\eta^2 = .017$) as could be expected from its conceptual importance. Also the interaction with the correlation is significant ($\eta^2 = .011$), see Table 2. Perhaps the inequality 10/20 is not sufficiently extreme to have a large effect. For clearly multidimensional data (r ≤ .6), the $\hat{D}_{\max}$ index is lower for unequally distributed items over the dimension than for equally distributed items over the dimensions, but for data with higher correlation values than .6, the means of $\hat{D}_{\max}$ are about the same (see Figure 2).

Insert Figure 2 about here.

Second, although it is not required, but implied by the DETECT procedure, the *overall size of the dimensional variance* affected the $\hat{D}_{\max}$ index substantially ($\eta^2 = .239$). The interaction between the dimensional variance and the correlation is also an important predictor of $\hat{D}_{\max}$ ($\eta^2 = .148$) as shown in Figure 3. It is natural that the higher the correlation is, the larger dimensional variance is needed, in order to indicate multidimensionality. A variance as small as 0.2 does not seem to suffice in order to differentiate between the dimensions, especially not for data with highly correlated dimensions. Similarly, a dimensional variance of 0.4 seems to differentiate between dimensions only with a correlation smaller than .6.

Insert Figure 3 about here.

In addition, an effect of *simple structure* is expected. The effect was significant indeed and had the expected sign, but the effect size was very small ($\eta^2 = .004$). This is perhaps not surprising, considering the fact that this effect refers to simple structure versus approximate simple structure (and not versus an extreme deviation from simple structure).

Third, although it is undesirable, the *variance of the item difficulties* had an impact on the $\hat{D}_{\max}$ index, but the effect size was negligible ($\eta^2 = .009$). In general, the larger the variance of the item difficulties was, the smaller the $\hat{D}_{\max}$ index value was.

The *sample size* had not only a moderately large main effect ($\eta^2 = .034$), but it also appears in several significant interactions. For sample size 100 and 200, the effect of the interdimensional correlation is seriously moderated, especially for high correlation values, as shown in Figure 4. In other words stated, there is no problem with small sample size if the correlation is low, but a moderately high DETECT index does not necessary indicate multidimensionality if the sample size is small.

Insert Figure 4 about here.

The results for $\hat{D}_{ref}$ are very similar to the results for $\hat{D}_{\max}$. The most important predictors ($\eta^2 > .1$) are the same: the correlation, the overall size of the dimensional variance and their interaction. The second group of predictors with $\eta^2 > .01$ consists of exactly the same variables as before, including the sample size and the distribution of the items. The effect sizes are very similar, except for the effect size of the variance of the item difficulties which is a bit smaller for $\hat{D}_{ref}$ ($\eta^2 = .002$) than for $\hat{D}_{\max}$ ($\eta^2 = .009$). All in all, the investigated data features have very similar effects on both DETECT indices.

In sum, the following conclusions can be drawn from a practical point of view. The DETECT procedure can be effectively used for investigating the dimensionality of data with 400 observations but less may be somewhat problematic. Also for data with weak dimensions (smaller dimensional variance than .4) or highly correlated dimensions (r ≥ .8), DETECT does less well. Since the R index is closely related to the DETECT index, but not to simple structure, the use of the R index is not really necessary for the kind of data as in Study one, but this conclusion may not be generalized to data structures with a larger deviation from simple structure.

Additional analyses related to the sample size

Based on the results, a larger sample size than 200 is to be recommended for using the DETECT indices. In order to check whether the fact that data sets with a smaller sample size than 400 were used affected the conclusions, a subset of the data with sample size lager than 200 was investigated. Also for this subset of the design, the $\hat{D}_{\max}$ index was a better predictor of multidimensionality than $\hat{D}_{ref}$ considering the deviance values of separate regression analyses (868.466 versus 924.2, respectively).

The same models were fitted to the restricted data set as before. Based on a further inspection of the results (for N≥400), it can be concluded, that both indices ($\hat{D}_{\max}$ and $\hat{D}_{ref}$) are still significantly related to the sample size (p=.002, p<.001, respectively), although the effect sizes were negligible ($\eta^2 = .003$, $\eta^2 = .005$, respectively). For this set of data, the interaction of the sample size with the correlation was slightly related to the DETECT indices ($\eta^2 = .002$ for $\hat{D}_{\max}$ and $\eta^2 = .003$ for $\hat{D}_{ref}$). When also the data sets with a sample size of 400 were omitted, the $\hat{D}_{\max}$ index was no longer significantly affected by the sample size (p=.109), whereas $\hat{D}_{ref}$ index sill was (p<001, $\eta^2 = .003$), which is probably because $\hat{D}_{ref}$ is calculated on half of the data. However, none of the indices were affected by the interaction of the sample size with the correlation in this subset of the data.

The most important predictors (for N≥600) were the correlation ($\eta^2 = .524$), the overall size of the dimensional variance ($\eta^2 = .280$), the interaction of the two ($\eta^2 = .143$), the distribution of the items ($\eta^2 = .017$), and the interaction of the correlation with the distribution of the items ($\eta^2 = .013$).

In sum, based on the above described additional analyses, the DETECT procedure is less vulnerable when one uses data with a *larger* sample size than 400. The most important effects are similar for smaller sample sizes than for larger ones.

Study two

In Study one, the inequality of the importance of the dimensions was restricted to the aspect of the distribution of the items over the dimensions. In Study two, the relative importance was manipulated in two ways: through the distribution of the items over the dimensions *and* through the relative size of the variances. The variance of the second dimension was varied in order to manipulate the relative importance of the dimensions in terms of the variance. An additional difference between the designs was that the correlation between the dimensions was not manipulated in the second study, but fixed to zero.

The R indices

The R indices were again highly correlated with DETECT indices, the R and $\hat{R}_{ref}$ indices were not significantly related to simple structure (p=.399 and p=.345, respectively) The correlation between R and $\hat{D}_{\max}$ was .751, and the correlation between the $\hat{R}_{ref}$ value and $\hat{D}_{ref}$ was .893. The omission of the unidimensional data sets from the analyses did not change the results.

Finally, although all data sets had at least approximate simple structure, 75.9% of the $\hat{R}$ index values and 78.7% of the $\hat{R}_{ref}$ index values were smaller than .8. Considering only the multidimensional data sets, the corresponding ratios were 71.1% and 74.4%, respectively. Based on this analysis of the R indices, similar conclusions can be drawn as before, that is the R indices have only limited utility, when the deviation from simple structure is not too large.

The fitted models

The analyses of the data were carried out as before. First, a model with only main effects was fitted to the data (Model 1), also models with all pairwise effects (Model 2) and with all three-way interactions (Model 3). The significant three-way interactions were added to Model 2, in order to construct Model 4 (the lower-order effects are always included if higher-order effects are). In a second step, these analyses were repeated for the linear and quadratic trends, and in a final, third step, the linear and quadratic trends of the *realize values* of the design factors were used in the models. In addition, regression tree analyses were carried out, in the same way as in Study one. The regression tree resulted in three interaction variables. Also the residual tree was fitted, which detected three additional threshold interactions. The resulting tree variables were used in a model as predictors of $\hat{D}_{\max}$, and accounted for 55.7% of the variance of $\hat{D}_{\max}$. One of the variables was not redundant with the variables of Model 1, but all of them were redundant with the predictors of Model 2, Model 3 and Model 4. As it can be seen in Table 4, based on the adjusted R-square values, the model in the first column with all three-way interactions had the best performance in terms of adjusted explained variance ($R_{adj}^2 = .834$). Since Model 4 contains all significant effects from Model 3, and has a similarly good performance ($R_{adj}^2 = .808$), this model was chosen to concentrate on. .

Insert Table 4 about here.

The effects of the design factors

In Table 5, the significant predictors of this model are shown in the order of the effect sizes (expressed as η^2).

Insert Table 5 about here.

According to the results, the relative importance of the dimensions in terms of the dimensional variance and the distribution of the items were the most important predictors, these are the two predictors with the highest conceptual validity. In addition, eight other predictors had sizable effects ($\eta^2 > .01$), including the variance of the item difficulties and the sample size. The simple structure of the data had only a minor effect on the $\hat{D}_{\max}$ index (η^2 =.004). The results will be discussed in detail for all three predictor categories based on the multidimensional concept.

First, as the correlation of the dimensions was not manipulated, the effect of correlation could not be observed. However, the fact, that the two dimensions were not correlated at all has implications, because Study one shows that the correlation interacts with other factors, so that their effect is different when the correlation is zero.

Based on the multidimensionality concept, it was expected that the *relative importance of the dimensions in terms of variance* was one of the most meaningful predictors. This predictor itself explained 70.2% of the variance of the data. Also, the *relative importance of the dimensions in terms of the distribution of the items* over the dimensions seems to have a substantial effect, but a clearly smaller one ($\eta^2 = .093$). The interaction of these two relative importance measures was also significant, but it is only minor ($\eta^2 = .012$). The means of $\hat{D}_{\max}$ indices are shown for each combination of these two predictors in Figure 5.

Insert Figure 5 about here.

The $\hat{D}_{\max}$ index increased with the relative importance of the second dimension in terms of variance. For unequally distributed items, the $\hat{D}_{\max}$ index values were always lower than for data with equally distributed items. The larger the relative importance of the second dimension was in terms of variance, the larger effect the distribution of the items had on $\hat{D}_{\max}$. The DETECT procedure was sensitive to both type of inequalities of the dimensions, which is a desirable characteristic. The effect of an unequal distribution of items showed somewhat better in terms of explained variance when the dimensional correlation is zero, in comparison with the results of Study one.

Second, although it is not necessarily implied by the concept of multidimensionality, the *simple structure* had a minor effect on the $\hat{D}_{\max}$ index ($\eta^2 = .004$). This effect was slightly stronger in interaction with the variance of the second dimension ($\eta^2 = .006$). When the variance of the second dimension was smaller in comparison to the variance of the first dimension, the $\hat{D}_{\max}$ index values were smaller for data with simple structure, but when the variance of both dimensions was high, simple structure leads to higher $\hat{D}_{\max}$ index values than an approximate simple structure did, as it can be seen in Figure 6. Simple structure seemed to have the expected kind of role when the dimensions are balanced.

Insert Figure 6 about here.

Third, the *sample size* appeared only in interactions and did not have such a large effect on $\hat{D}_{\max}$ as in Study one. This result is actually in agreement with the findings of Study one, since it is shown in Figure 4, that the sample size does not have a sizable effect for small correlation values. The results for $\hat{D}_{ref}$ differ from the results for $\hat{D}_{\max}$ only with respect to the sample size. The sample size in the cross-validation, for $\hat{D}_{ref}$, is only half of the sample size for $\hat{D}_{\max}$ in the regular procedure. This may explain that the sample size had a strong effect on $\hat{D}_{ref}$ in Study two ($\eta^2 = .140$).

The *variance of the item difficulties* affected the DETECT procedure to some extent, in the same way as in Study one.

The results of Study two can be summarized briefly in the following way. Both the relative size of the dimensional variances and the distribution of the items over the dimensions affected the DETECT index in the expected way. The less multidimensional the data were, the smaller was the $\hat{D}_{\max}$ index. Furthermore, the effect of the sample size on $\hat{D}_{\max}$ could not be seen in Study two, because of the uncorrelated dimensions, whereas it did have an effect on $\hat{D}_{ref}$. Study two confirmed the results of Study one that simple structure (versus approximate simple structure) is at best slightly related to the DETECT indices. Besides, the variance of the item difficulties again affected the DETECT procedure in a minor way. Finally, the R indices did not seem to function well as a moderator in the interpretation of the DETECT index when the data have a simple or an approximate simple structure.

Discussion and conclusion

The effects of several important data features on DETECT have been investigated. Not the effectiveness of a decision rule was studied, but the sensitivity of the regular DETECT value for the whole data ($\hat{D}_{\max}$). The sensitivity of the cross-validated DETECT index ($\hat{D}_{ref}$) seems to very similar, although $\hat{D}_{ref}$ is more influenced by the sample size than $\hat{D}_{\max}$. This is not surprising, since for the calculation of $\hat{D}_{ref}$, only half as much data is used effectively than for the calculation of $\hat{D}_{\max}$. Consequently, $\hat{D}_{\max}$ seems to be a better indicator than $\hat{D}_{ref}$.

According to the results, all data features that are conceptually connected with the multidimensionality affect DETECT as desired, such as the correlation of the dimensions, and the inequality of the dimensional importance. Other data features are less naturally related to multidimensionality, but still have an expected effect on DETECT, based on the way DETECT indices are constructed, such as simple structure and the overall size of the dimensional variance. A third group of effects is undesirable because they are unrelated to multidimensionality, such as the sample size, and the variance of the item difficulties. Especially the sample size is shown to have some effect on DETECT, if the sample size is 100 or 200. This finding is relevant in that large sample sizes are common in studies in educational measurement, but not in studies in psychology. It should require further investigation to find out how large sample size should be in relation to the number of items.

The higher the correlation of the dimensions is, the smaller the DETECT indices are. In the same way, the more similar the relative importance of the dimensions is, the more multidimensional the data are, and the higher the DETECT indices are. This applies to importance in terms of variance and in terms of the number of the corresponding items. Besides, the larger the overall size of the dimensional variance, the more sensitive the DETECT index is. Finally, it is important to note that the DETECT procedure indicates only sizeable multidimensionality. Dimensions with relatively small importance, containing only a few items or having small variance may be overlooked by the DETECT procedure. In conclusion, the DETECT indices, especially the $\hat{D}_{\max}$ index can be seen as a relatively good measure of multidimensionality, but one should be aware of some limitations.

Author Note

Katalin Balázs and Paul De Boeck, Department of Psychology, K. U. Leuven, Belgium. We acknoledge the financial support from K.U. Leuven the COE grant to Katalin Balázs and from the IAP/5 network grant to Paul De Boeck.

Correspondence concerning this article should be addressed to Katalin Balázs, Department of Psychology, K. U. Leuven, Tiensestraat 102, B-3001, Leuven, Belgium, Katalin.Balazs@psy.kuleuven.be.

References

Balazs, K., Hidegkuti, I., & De Boeck, P. (submitted). Detecting heterogeneity in logistic regression models

Bolt, D. (2001). Conditional covariance-based representation of multidimensional test structure. *Applied Psychological Measurement, 25*, 244-257.

Bolt, D. M., & Lall, V. F. (2001). Estimation of compensatory and noncompensatory multidimensional item response models using Markov chain Monte Carlo. *Applied Psychological Measurement, 27,* 395-414.

Bouwmeester, S., & Sijtsma, K. (2004). Measuring the ability of transitive reasoning, using product and strategy information. *Psychometrika, 69*, 123-146.

Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J.(1984). *Classification and regression trees*, New York: Chapman & Hall..

Cattell, R. B. (1964). Validity and reliability: a proposed more basic set of concepts. *Journal of Educational psychology, 55,* 1-22.

Cattell, R. B. (1978). *The scientific use of factor analysis in behavioral and life sciences.* New York: Plentum.

Cattell, R. B., & Tsujioka, B. (1964). The importance of factor-trueness and validity, versus homogeneity and orthogonality, in test scales. *Educational and Psychological Measurement, 24,* 3-30.

Chaudhuri, P., Lo, W.-D., Loh, W.-Y., & Yang, C.-C. (1995). Generalized regression trees, *Statistica Sinica, 5,* 641-666.

Cox, D. R., & Snell, E. J. (1989). *Analysis of binary data.* Second edition, London: Chapman and Hall.

De Ayala, R. J., & Hertzog, M. A. (1991). The assessment of dimensionality for use in item response theory. *Multivariate Behavioral research, 26,* 765-792.

Dimitriadou, E., Dolnicar, S., & Weingessel, A. (2002). An examination of indexes for determining the number of clusters in binary data sets. *Psychometrika, 67,* 137-160.

Douglas, J., Kim, H.-R., Roussos, L., Stout, W., & Zhang, J. (1999). LSAT Dimensionality analysis for the December 1991, June 1992, and October 1992 Administrations [Law School Admission Council Statistical Report 95-05]

Dusseldorp, E., & Meulman, J. J. (2004). The regression trunk approach to discover treatment covariate interaction. *Psychometrika, 69,* 355-374.

Hambleton, R. K., & Rovinelli R. J. (1986). Assessing the dimensionality of a set of test items. *Applied Psychological Measurement, 10,* 287-302.

Hambleton, R. K., & Swaminathan, H.(1985). *Item response theory: Principles and applications.* Boston: Kluwer-Nijhoff Publishing.

Hattie, J. (1984). An empirical study of various indices for determining unidimensionality. *Multivariate Behavioral Research, 19,* 49-78.

Hattie, J. (1985). Methodology review: Assessing unidimensionality of tests and items. *Applied Psychological Measurement, 9,*139-164.

Hattie, J., Krakowski, K., Rogers, H. J., & Swaminathan, H. (1996). An assessment of Stout's index of essential unidimensionality. *Applied Psychological Measurement, 20,* 1-14.

Humphreys, L. G. (1981). The primary ability. In M. P. Friedman, J. P. Das & N O'Connor (Eds.) *Intelligence and learning,* (pp. 87-102). New York: Plenum.

Humphreys, L. G. (1985). General intelligence: An integration of factor, test, and simplex theory. In B. B. Wolman (Ed.), *Handbook of intelligence* (pp. 201-224). New York: Wiley.

Junker, B. (1991). Essential independence and likelihood-based ability estimation for polytomous items. *Psychometrika, 56,* 255-278.

Junker, B., & Stout, W. (1994). Robustness of ability estimation when multiple traits are present with one trait dominant. In D. Laveault, B. Zumbo, M. Gessaroli, & M. Boss (Eds.), *Modern theories of measurement: Problems and issues* (pp.31-61), Ottawa, Canada: Edumetrics Research Group, University Ottawa.

Kim, H. R. (1994). New techniques for the dimensionality assessment of standardized test data. (Doctoral dissertation, University Illinois at Urbana Campaign). *Dissertation Abstracts International, 55-12B,* 5598

Kirisci, L., Hsu, T.-c., & Y., L. (2001). Robustness of item parameter estimation programs to assumptions of unidimensionality and normality. *Applied Psychological Measurement, 25,* 146-162.

Loh, W-Y. (2002). Regression trees with unbiased variable selection and Interaction detection. *Statistica sinica, 12,* 361-386.

Loh, W-Y. (2004). GUIDE (ver. 3) User Manual. Retrieved April 25, 2005, from http://www.stat.wisc.edu/~loh/treeprogs/guide/guideman.pdf

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores.* Menlo Park, CA: Addison-Wesley.

Lumsden, J. (1961). The construction of unidimensional tests. *Psychological Bulletin, 58,* 122-131.

McDonald, R. P. (1979). The structural analysis of multivariate data: A sketch of general theory. *Multivariate Behavioral Research, 14,* 21-38.

McDonald, R. P. (1981). The dimensionality of tests and items. *British Journal of Mathematical and Statistical Psychology, 34,* 100-117.

Moran, J., Roussos, L., & Kim, H.-R. (2005). DETECT manual.

Nagelkerke, N. J. D. (1991). A note on the general definition of the coefficient of determination, *Biometrika, 78,* 691-692.

Nandakumar, R.(1991). Traditional dimensionality versus essential dimensionality. *Journal of Educational Measurement, 28*, 99-117.

Nandakumar, R.(1993). Assessing essential unidimensionality of real data (1993), *Applied Psychological Measurement, 17,* 29-38.

Nandakumar, R., & Stout, W. (1993). Refinements of Stout's procedure for assessing latent trait unidimensionality. *Journal of Educational measurement, 18,* 41-68.

Reckase, M. R. (1985). The difficulty of test items that measure more than one dimension. *Applied Psychological Measurement, 9,* 401-412.

Reckase, M. D. (1997). The past and the future of multidimensional Item Response Theory. *Applied Psychological Measurement, 21,* 25-36.

Roznowski, M., Tucker, L. R., & Humphreys, L. G. (1991). Three approaches to determining the dimensionality of binary items. *Applied Psychological Measurement, 15,* 109-127.

SAS Institute, Inc. (1999). SAS online doc (Version 8) [Software manual on CD-ROM]. Cary, NC: SAS Institute, Inc.

Stout, W. F. (1987). A nonparametric approach for assessing latent trait dimensionality. *Psychometrika, 52,* 79-98.

Stout, W. F. (1990). A new item response theory modeling approach with applications to unidimensionality assessment and ability estimation. *Psychometrika, 55,* 293-325.

Stout, W. F., Douglas, J., Junker, B., & Roussos, L. (1993). DIMTEST manual. University of Illinois, Urbana-Champaign.

Stout, W., Habing, B., Douglas, J., Kim, H. R., Roussos, L., & Zhang, J. (1996). Conditional covariance-based nonparametric multidimensionality assessment. *Applied Psychological Measurement, 20,* 331-354.

Tate, R. (2003). A comparison of selected empirical methods for assessing the structure of responses to test items. *Applied Psychological measurement*. 27, 159-203.

The William Stout Institute for Measurement. (2003). DETECT manual.

Traub, R. E. (1983). A priori considerations in choosing an item response model. In R. K. Hambleton (Ed.) *Applications of item response theory* (pp. 57-70). Vancouver: Educational Research Institute of British Columbia.

Uribe-Zarain, X., Nandakumar, R., & Yu, F. (2005). Determining the dimensional structure of real test data. Paper presented at the annual meeting of the National Council on Measurement in Education, Montreal

van Abswoude, A. A. H., van der Ark, L. A., & Sijtsma, K. (2004). A comparative study of test data dimensionality assessment procedures under nonparametric IRT models. *Applied Psychological Measurement, 28,* 3-24.

Wang, W.-C., & Chen, P.-H. (2004). Implementation and measurement efficiency of multidimensional computerized adaptive testing. *Applied Psychological Measurement, 28,* 295-316.

Wang, W.-C., Wilson, M., & Adams, R. J. (1997). Rasch models for multidimensionality between items and within items. In M. Wilson, G. Engelhard, Jr., & K. Draney (Eds.) *Objective measurement: Theory into practice* (vol. 4, pp. 139-155). Norwood, NJ: Ablex.

Whitely, S. E. (1980). Multicomponent latent trait models for ability tests. *Psychometrika, 45,* 479-494.

Wilson, D. T., Wood, R., & Gibbons, R. (1991). TESTFACT: Test scoring, item statistics, and item factor analysis [Computer Software]. Chicago: Scientific Software International.

Yen, W. M. (1984). Effects of local item dependence on the fit and equating performance of the three-parameter logistic model. *Applied Psychological Measurement, 8,* 125 -145.

Zhang, J., & Stout, W. (1999). The theoretical DETECT index of dimensionality and its application to approximate simple structure. *Psychometrika, 64,* 213-249.

Table 1

The fit of different models for predicting multidimensionality in Study one

Predictor(s)	Deviance	R^2_{adj}
$(\hat{D}_{\max} - \hat{D}_{ref})/\hat{D}_{ref}$	1943.949	.002
$\hat{D}_{\max}$	1528.502	.296
$\hat{D}'_{\max}$	1652.943	.214
$\hat{D}_{ref}$	1541.976	.288
$\hat{D}_{\max}, \hat{D}_{ref}$	1412.587	.369
$\hat{D}_{\max}, \hat{D}_{ref}, \hat{D}'_{\max}$	1412.475	.369

Table 2

Adjusted R-square statistics of the models in Study one

	Analysis with theoretical values (dummy coding)	Analysis with theoretical values extended with variables from regression tree	Analysis with theoretical values (linear and quadratic trends)	Analysis with real values (linear and quadratic trends)
Model 1	.679	.781	.663	.669
Model 2	.856	.857	.828	.849
Model 3	.866	-	.834	.856
Model 4	.866	-	.834	.855

Table 3

The predictors with significant contributions in Model 4 in Study one

variable	p>F	η^2	variable	p>F	η^2
r	<.001	.467	r* σ_θ^2 *Eq	<.001	.005
σ_θ^2	<.001	.239	σ_θ^2 *Eq	<.001	.005
r* σ_θ^2	<.001	.148	Ss	<.001	.004
N	<.001	.034	r* σ_θ^2 * σ_β^2	.052	.004
r*N	<.001	.020	r* σ_β^2	<.001	.003
Eq	<.001	.017	σ_θ^2 * σ_β^2	<.001	.003
σ_θ^2 *N	<.001	.013	Eq*Ss	<.001	.001
r* σ_θ^2 *N	<.001	.012	σ_θ^2 * σ_β^2 *Ss	.016	.001
r*Eq	<.001	.011	r*Eq*Ss	.021	.001
σ_β^2	<.001	.009			

Ss=simple structure, Eq=equally distributed items over dimensions, N=sample size, r= correlation of the dimensions

Table 4

Adjusted R-square statistics of the models in Study two

	Analysis with theoretical values (dummy coding)	Analysis with theoretical values extended with variables from regression tree	Analysis with theoretical values (linear and quadratic trends)	Analysis with real values (linear and quadratic trends)
Model 1	.732	.736	.734	.740
Model 2	.783	-	.771	.774
Model 3	.834	-	.776	.776
Model 4	.808	-	.771	.778

Table 5

The predictors with significant contributions in Model 4 in Study two

variable	p>F	η^2	variable	p>F	η^2
$\sigma^2_{\theta_2}$	<.001	.702	Eq*N*σ^2_{β}	.006	.013
Eq	<.001	.093	N*σ^2_{β}	.007	.013
$\sigma^2_{\theta_2}$ *N*σ^2_{β}	.021	.038	Eq*$\sigma^2_{\theta_2}$	<.001	.012
σ^2_{β}	<.001	.032	$\sigma^2_{\theta_2}$ *Ss	.030	.006
$\sigma^2_{\theta_2}$ *σ^2_{β}	<.001	.028	Eq*σ^2_{β}	.021	.004
$\sigma^2_{\theta_2}$ *N	.011	.023	Ss	.008	.004
$\sigma^2_{\theta_2}$ *Eq*N	.021	.022	Eq*Ss	.023	.003

Ss=simple structure, Eq=equally distributed items over dimensions, N=sample size

Figure captions

Figure1. The $\hat{D}_{\max}$ index values as a function of the correlation in Study one

Figure 2. The $\hat{D}_{\max}$ means per correlation value, for equally and unequally distributed items over dimensions in Study one

Figure 3. The $\hat{D}_{\max}$ as a function of the overall size of the dimensional variance and the correlation of the dimensions in Study one

Figure 4. The $\hat{D}_{\max}$ as a function of the sample size and the correlation of the dimensions in Study one

Figure 5. The $\hat{D}_{\max}$ as a function of the relative size of the variance of the second dimension and the distribution of the items over dimensions in Study two

Figure 6. The $\hat{D}_{\max}$ as a function of the variance of the second dimension in case of simple structure and approximate simple structure in Study two

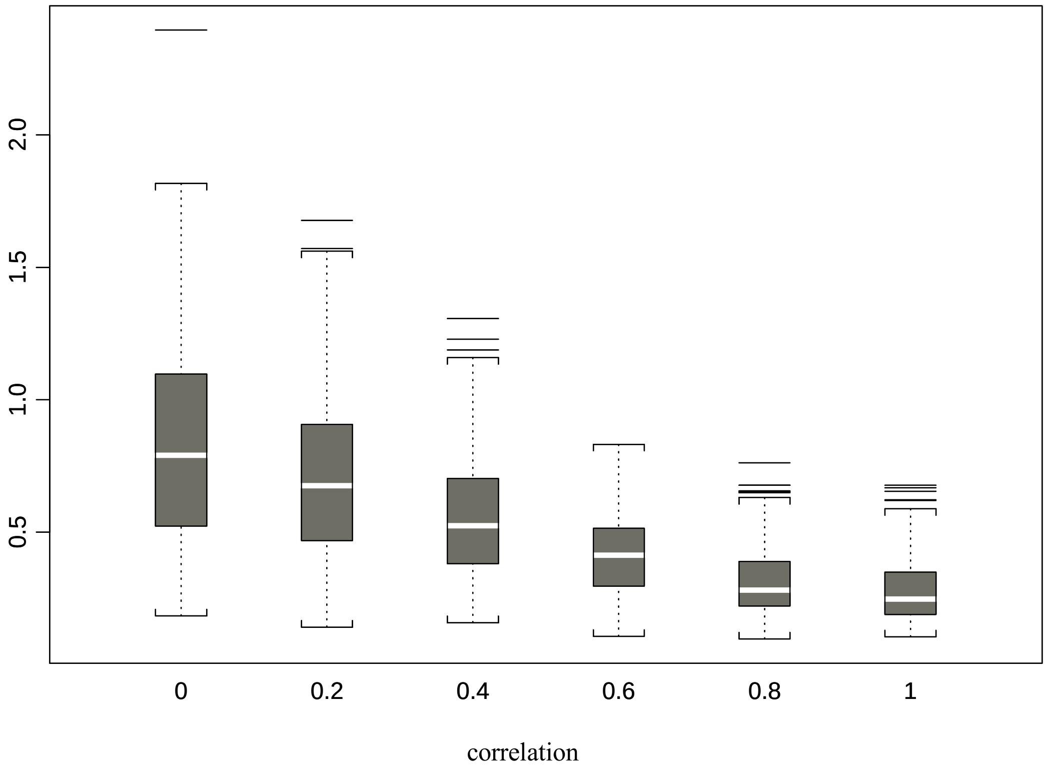

Figure 1

The $\hat{D}_{\max}$ index values as a function of the correlation in Study one

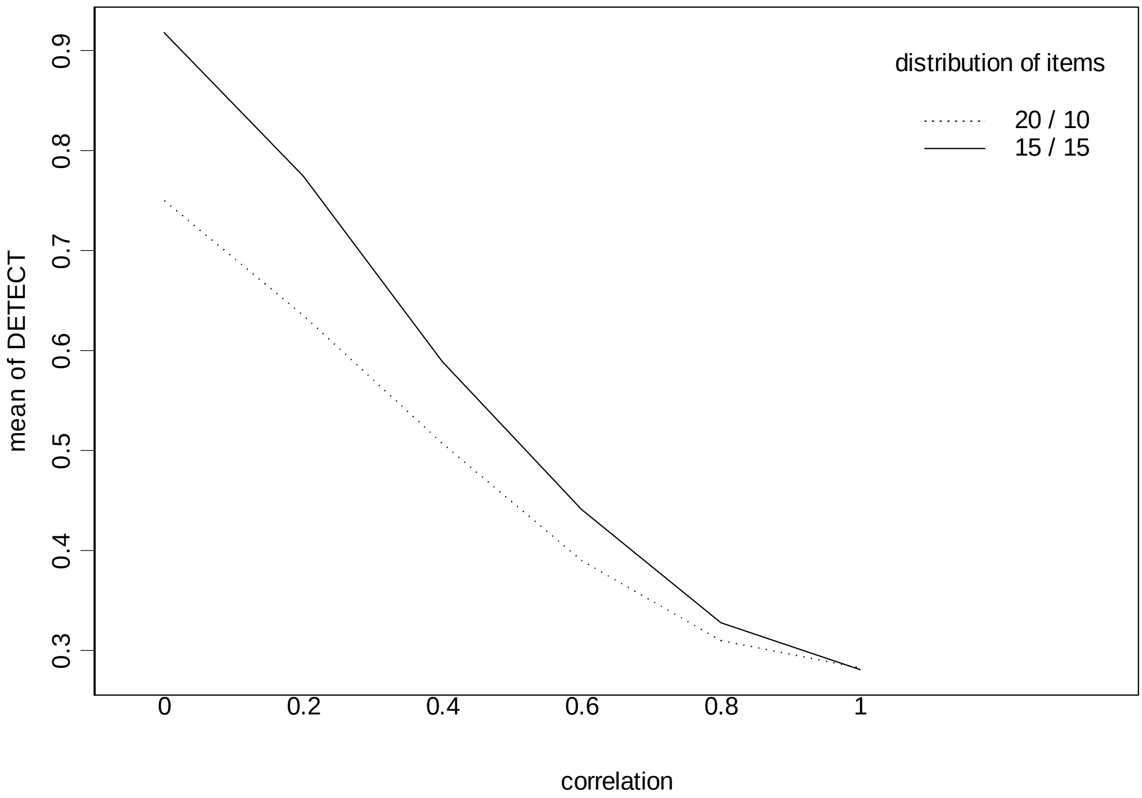

Figure 2

The $\hat{D}_{\max}$ means per correlation value, for equally and unequally distributed items over dimensions in Study one

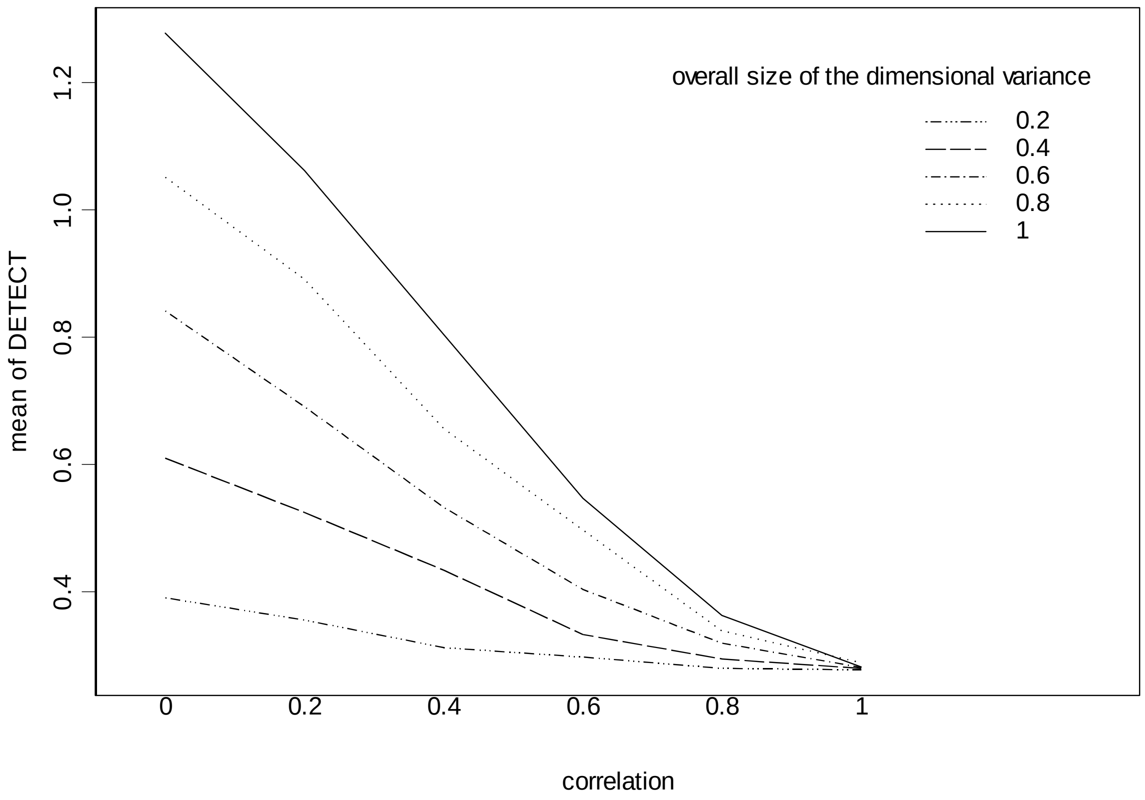

Figure 3

The $\hat{D}_{\max}$ as a function of the overall size of the dimensional variance and the correlation of the dimensions in Study one

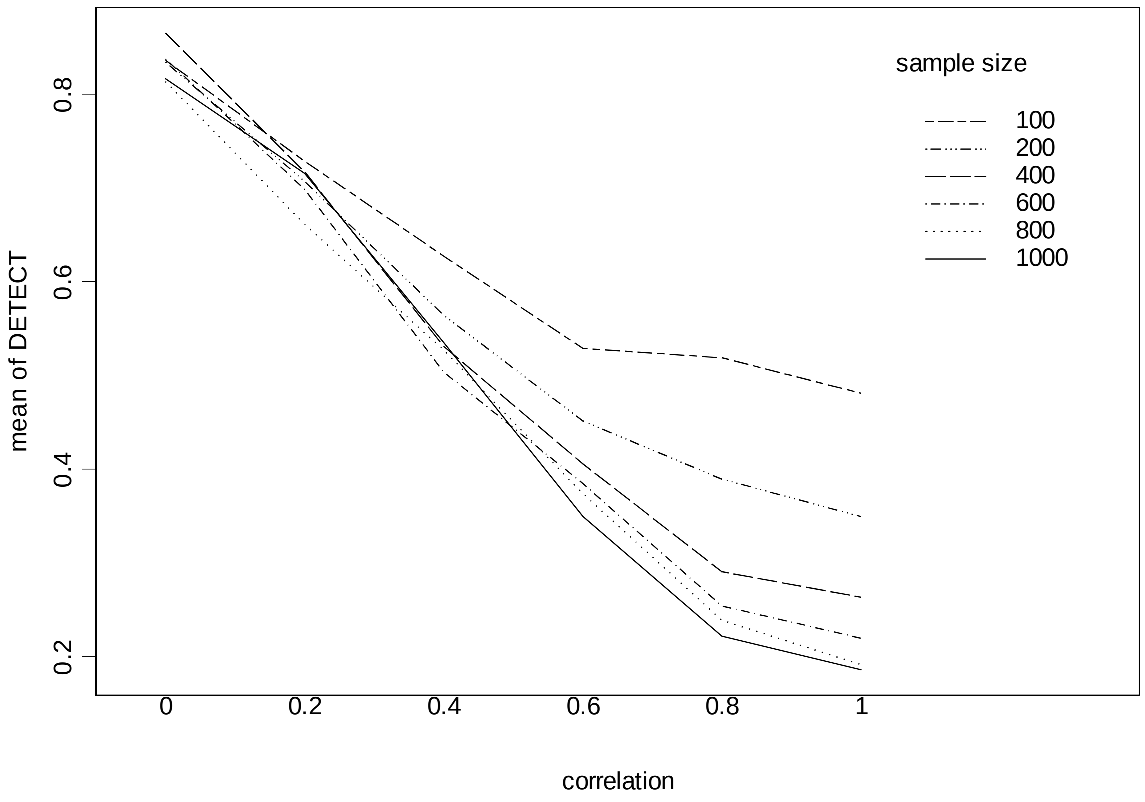

Figure 4

The $\hat{D}_{\max}$ as a function of the sample size and the correlation of the dimensions in Study one

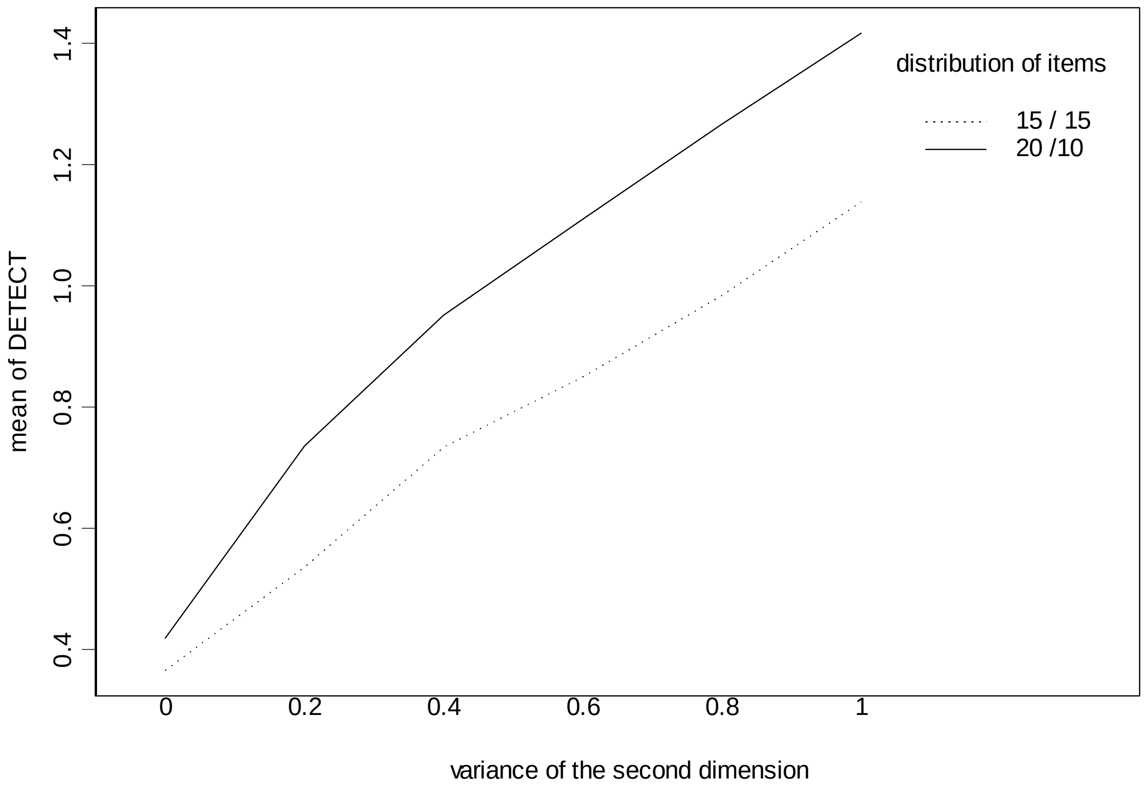

Figure 5

The $\hat{D}_{\max}$ as a function of the relative size of the variance of the second dimension and the distribution of the items over dimensions in Study two

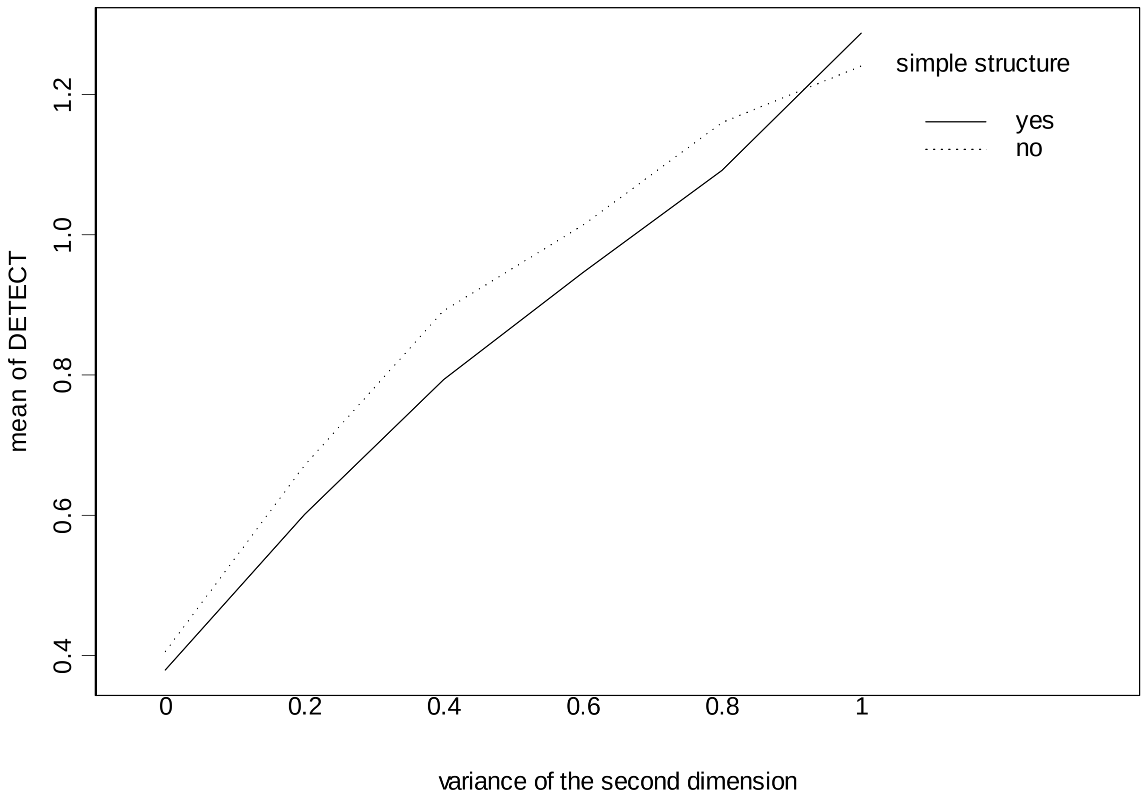

Figure 6

The $\hat{D}_{\max}$ as a function of the variance of the second dimension in case of simple structure and approximate simple structure in Study two