Heuristic Constraint Management Methods in Multidimensional Adaptive Testing

Sebastian Born[a] and Andreas Frey[a,b]

[a]Friedrich Schiller University Jena, Germany

[b]Centre for Educational Measurement (CEMO) at the University of Oslo, Norway

Author Note

Correspondence concerning this article should be addressed to Sebastian Born, Institute of Educational Science, Department of Research Methods in Education, Friedrich Schiller University Jena, Am Planetarium 4, D-07743 Jena, Germany, e-mail: sebastian.born@uni-jena.de

Abstract

Although multidimensional adaptive testing has been proven to be highly advantageous with regards to measurement efficiency when several highly correlated dimensions are measured, there are few operational assessments that use MAT. This may be due to issues of constraint management, which is more complex in MAT than it is in unidimensional adaptive testing. Very few studies have examined the performance of existing constraint management methods (CMMs) in MAT. The present paper focuses on the effectiveness of two promising heuristic CMMs in MAT for varying levels of imposed constraints and for various correlations between the measured dimensions. Through a simulation study, the maximum priority index (MMPI) and multidimensional weighted penalty model (MWPM), as an extension of the weighted penalty model, are examined in regards to measurement precision and constraint violations. The results show that both CMMs are capable of addressing complex constraints in MAT. However, measurement precision losses were found to differ between the MMPI and MWPM. While the MMPI appears to be more suitable for use in assessment situations involving few to a moderate number of constraints, the MWPM should be used when numerous constraints are involved.

Keywords: item selection, computerized adaptive testing, constraint management, multidimensional adaptive testing

Heuristic Constraint Management Methods in Multidimensional Adaptive Testing

Computerized adaptive testing (CAT) is an approach that is used to measure person characteristics (van der Linden & Glas, 2010), whereby the item selection in CAT is based on the information acquired from responses to previously administered items. The approach has been proven to substantially increase measurement efficiency relative to linear tests involving a fixed number of items (Segall, 2005; Wang, Chen, & Cheng, 2004). For this reason, over the last decade, the relevance of CAT has increased considerably and it is now used in numerous fields (e.g., educational assessment and psychological testing). In order to assess multiple latent traits simultaneously, CAT has been generalized to multidimensional adaptive testing (MAT; e.g., Segall, 1996). In multiple simulation studies (Liu, 2007; Segall, 1996; Wang & Chen, 2004; Yao, 2010), MAT was found to be more efficient for correlated traits than unidimensional CAT. Nevertheless, only a few operational assessments have employed MAT (e.g., Mulcahey, Haley, Duffy, Pengsheng, & Betz, 2008). Even in the field of large-scale assessments, in which dimensions are often highly correlated, there is currently no form of operational assessment for which MAT is used. Possible reasons for this limited use of MAT include the fact that the management of test specifications in MAT is more complex than in unidimensional adaptive testing and the fact that only a few studies have addressed the management of test specifications in the multidimensional case (Frey, Cheng, & Seitz, 2011; Su, 2015; Su & Huang, 2015; Veldkamp & van der Linden, 2002; Yao, 2014).

Stocking and Swanson (1993) described test specifications as rules governing the assembly of tests, whereby these rules are related to one or more item or test properties. Test specifications can be, for example, the proportion of administered items on a specific topic, the test length or the total testing time. These specifications can be formulated as constraints or as objective functions that must be considered during the item selection process (van der Linden, 2005b). For standardized testing programs in particular, various test forms (which are typically needed due to the presence of large item pools) must be comparable in regards to a predefined set of test specifications. For linear tests, it is possible to assemble various test forms before administering tests. However, in adaptive testing, test specifications must be fulfilled over the course of a test. This challenging task can be addressed by modifying the item selection algorithm of an adaptive test to simultaneously consider statistical optimality criteria and the required test specifications. According to the literature, multiple constraint management methods (CMMs) can account for test specifications during the CAT item selection process.

He, Diao and Hauser (2014) gave a brief overview of the existing CMM and differentiated between two types. The first type of CMM, such as the constrained CAT method (Kingsbury & Zara, 1991) and modified multinomial model (Chen & Ankenman, 2004) can only address mutually exclusive constraints. The weighted deviation model (Stocking & Swanson, 1993), shadow test approach (STA; van der Linden & Reese, 1998), weighted penalty model (WPM; Shin, Chien, Way, & Swanson, 2009), and maximum priority index method (MPI; Cheng & Chang, 2009) belong to the second group of CMMs, which are also capable of addressing complex sets of constraints. The present study focuses on the second group of CMMs, as they are more flexible and can therefore be used to address a broad variety of constraint management problems. The main difference between approaches of this type pertains to the ways in which future item selection consequence projections are incorporated into the item selection process (He et al., 2014). The STA is a very flexible approach that has been proven to be successful in the management of multiple constraints for unidimensional (van der Linden & Reese, 1998) and multidimensional adaptive testing (Veldkamp & van der Linden, 2002). However, its use requires access to considerable knowledge on linear programming, and solver software must be available. For practitioners, solver software selection decisions can be challenging to make, as multiple issues must be considered in regards to specific test assembly problems (Donoghue, 2014). Such issues relate to the frequency of software program use, the size of the problem considered (e.g., the number of items and constraints), one's programming experience, and the financial resources available for purchasing licenses. Freeware such as lpSolveAPI could be an attractive alternative to commercial solver. Diao and van der Linden (2011) demonstrated its capacity to carry out CAT with STA for smaller number of constraints. However, the authors argue that the performance of this software must be evaluated on a case-by-case basis.

Based on this background information, heuristic CMMs (e.g., the WDM, WPM and MPI) are of particular interest to practitioners, as the requirements for their implementation are considerably low. Nonetheless, there are still considerable differences between the performance and maintenance of heuristic CMMs. In a study by He et al. (2014), the performance of the STA and that of the three heuristic CMMs (WDM, WPM, MPI) was compared. In regards to measurement precision, no significant differences were found between the heuristic CMMs. However, in regards to how well imposed constraints were met, the WPM outperformed the other heuristic methods. Furthermore, the MPI was described as the most “low maintenance,” and the WPM was described as the most “high maintenance” method. Unfortunately, few results have been recorded in regards to the performance of heuristic CMMs for the multidimensional case (Su, 2015; Yao, 2014).

The present study addresses this issue. According to He et al.'s (2014) results, the MPI and WPM are very promising candidates of constraint management in MAT. For the MPI a multidimensional extension already exists. It is named multidimensional maximum priority index (MMPI) and was presented by Frey et al. (2011). The WPM, however, has not yet been extended to the multidimensional case. Therefore, the first objective of the present study is to render the WPM applicable in MAT.

As the size of test assembly problems is a crucial issue, it is important to determine whether all methods are equally well suited to a particular number of constraints. While numerous studies have addressed the performance of CMMs (Cheng & Chang, 2009; Cheng, Chang, Douglas, & Guo, 2008; He et al., 2014; Shin et al., 2009; Su, 2015; van der Linden, 2005a), no existing results detail the relationship between performance and the number of constraints. For this reason, the second objective of this study is to compare multidimensional extensions of the MPI and WPM in regards to the relationship between their performance and number of constraints. From these results, we present recommendations on the use of the various approaches in MAT.

The remainder of the article is organized as follows. First, a brief introduction to MAT is given. Next, the two CMMs (MPI and MWPM) are introduced, and their extensions to the multidimensional case are is described. Finally, both approaches are evaluated through a simulation study, and recommendations for practitioners are presented.

Multidimensional Adaptive Testing

Multidimensional adaptive testing is proposed as a means of simultaneously measuring several traits. When employing MAT, two important issues must be addressed: the psychometric model and the item selection procedure. Multidimensional item response theory (Reckase, 2009) models are typically used as psychometric models for MAT. One general MIRT model is the multidimensional three-parameter logistic (M3PL) model, which specifies the probability that an examinee j will answer an item i correctly as a function of the ability vector $\boldsymbol{\theta}_j = (\theta_1, \theta_2, \dots, \theta_p)$ for p measured dimensions and for item parameters $\boldsymbol{a}_i'$, b_i, and c_i:

$$P\left(U_{ij} = 1 | \boldsymbol{\theta}_j, \boldsymbol{a}_i', b_i, c_i\right) = c_i + (1 - c_i) \frac{\exp\left(\boldsymbol{a}_i'(\boldsymbol{\theta}_j - b_i\mathbf{1})\right)}{1 + \exp\left(\boldsymbol{a}_i'(\boldsymbol{\theta}_j - b_i\mathbf{1})\right)} \quad . \tag{1}$$

The elements a_{ip} of the vector $\boldsymbol{a}_i'$ are the discrimination parameters, denoting the loadings of an item on the measured dimensions. The difficulty of item i is given by parameter b_i. Parameter c_i specifies the probability of a less capable examinee answering an item correctly by guessing (Hambleton & Swaminathan, 1985). The two-parameter logistic (M2PL) model and multidimensional Rasch (M1PL) model can be derived from the M3PL model shown in (*Equation 1*). The M2PL model is derived from the assumption that, for all test items, c_i is equal to zero. In addition to this assumption, in the M1PL model, elements of the vector $\boldsymbol{a}_i'$ are constrained to either one or zero, denoting that item i loads on dimension p with $a_{ip} = 1$ but that it does not for $a_{ip} = 0$. Due to the growing number of assessments that use the M2PL model, the present study focuses on this model.

The second important aspect in MAT pertains to the item selection method (see Yao, 2014 for an overview). Various approaches are used that differ with respect to the multivariable function that must be minimized or maximized (Yao, 2010): for example, maximizing the determinant of the Fisher information matrix (Segall, 1996), minimizing the trace of the inverse Fisher information matrix (van der Linden, 1999), maximizing the posterior expected Kullback-Leibler information (Veldkamp & van der Linden, 2002), and maximizing a simplified Kullback-Leibler information index

(Wang, Chang, & Boughton, 2011). One frequently investigated item selection method for MAT is Segall's Bayesian approach (1996), whereby the determinant of the Fisher information matrix is maximized. In regards to typical evaluation criteria (e.g., (conditional) bias and the measurement precision of ability estimates), this approach has been proven to be one of the best performing methods relative to other item selection methods (Mulder & van der Linden, 2009; Veldkamp & van der Linden, 2002; Wang & Chang, 2011; Wang et al., 2011; Yao, 2012, 2013, 2014). Nevertheless, according to some studies, other approaches perform slightly better (Wang & Chang, 2011). However, as Segall's item selection method has been shown to be robust in several studies and for various MAT specifications, it is used as the item selection procedure for the present study.

For Segall's Bayesian approach (1996), item selection is optimized by using the variance-covariance matrix $\mathbf{\Phi}$ of the measured latent traits as prior information. From the item pool, the item i^* that maximizes the determinant of the matrix $\mathbf{W}_{t+i^*}$ is selected.

$$|\mathbf{W}_{t+i^*}| = \left|\mathbf{I}\left(\boldsymbol{\theta}, \widehat{\boldsymbol{\theta}}_j\right) + \mathbf{I}(\boldsymbol{\theta}, u_{i^*}) + \mathbf{\Phi}^{-1}\right| \tag{2}$$

This matrix is determined by summing the information matrix of the previously t administered items $\mathbf{I}\left(\boldsymbol{\theta}, \widehat{\boldsymbol{\theta}}_j\right)$, the information matrix of the candidate item i^* $\mathbf{I}(\boldsymbol{\theta}, u_{i^*})$, and the inverse of the variance-covariance matrix of the prior distribution of the measured dimensions $\mathbf{\Phi}^{-\mathbf{1}}$. For estimating latent traits, Segall proposes using the multidimensional Bayes modal estimator in combination with the same prior information given by $\mathbf{\Phi}$.

Constraint Management Methods

In this section, the MPI and the WPM are described; their extensions to the multidimensional case–the MMPI and the MPWM–are introduced, and the similarities and differences between the two methods are outlined.

The Maximum Priority Index.

The MPI (Cheng & Chang, 2009) is based on the constraint relevancy matrix $\mathbf{C}$, where $\mathbf{C}$ is a matrix of size $I \times K$ with I representing the number of items in the pool and with K denoting the total number of constraints. Elements of $\mathbf{C}$ indicate that item i is relevant for the constraint k with $c_{ik} = 1$ and that it is not when $c_{ik} = 0$. An item is relevant for a constraint if it includes the property (e.g., multiple-choice format, a specific content area) that is associated with the constraint. Based on the constraint relevancy matrix, the MPI works via two major operations within each item selection step: first, the determination of the priority index (PI) for every eligible candidate item i^* in the item pool, and second, the selection of the item with the highest PI for administration.

The PI for a candidate item i^* is computed using *Equation 3* where I_{i^*} represents Fisher information for the item i^* based on the provisional ability estimate $\hat{\theta}$, w_k as the weight of constraint k that can be used to control the relative importance of the various constraints, and f_k measures the scaled “quota left” (Cheng & Chang, 2009), which expresses how urgently a constraint k is needed at the current test stage.

$$\mathrm{PI}_{i^*} = I_{i^*} \prod_{k=1}^{K} (w_k f_k)^{c_{i^*k}} \tag{3}$$

The scaled “quota left” f_k is given by *Equation 4,* where b_k represents the number of items required for the test that are relevant for constraint k and where x_k is the number of relevant items that have been administered.

$$f_k = \frac{(b_k - x_k)}{b_k} \tag{4}$$

This ratio is equal to one if no item that is relevant for constraint k is presented $(x_k = 0)$. The value of f_k decreases as more relevant items are presented until it is equal to zero if the required number of items has been reached $(x_k = b_k)$. If a candidate item is not relevant for a constraint k $(c_{i^*k} = 0)$, the PI of this item is not affected by the term $w_k f_k$. Suppose constraint k refers to the number of items in the test that following a multiple-choice format, 10 items with this format are required (b_k) and five items of a multiple-choice format have already been selected (x_k). The resulting scaled quota left for this constraint k at this stage of the test is $f_k = 0.5$, but it only affects the PI of items with a multiple-choice format. The PI of items with an open-response format will be unaffected by the scaled quota left because such items are not relevant for this constraint k. In some CAT applications, the number of items required to fulfill a constraint is not fixed, but rather a minimum and maximum number is determined. In such cases, the MPI method can be used over a two-phase item selection procedure (Cheng, Chang, & Yi, 2007) by specifying a lower bound b_{kL} and upper bound b_{kU} for the constraints. To prevent a dysfunction of the MPI in some edge conditions where the PI of all eligible items becomes zero, He et al. (2014) developed two modifications of the MPI: M1_MPI and M2_MPI. As the M2_MPI was shown to perform slightly better than the M1_MPI, it is used in the present study.

The MPI was extended to the MMPI (Frey et al., 2011) by replacing Fisher item information I_{i^*} with Segall's Bayesian item selection criterion (*Equation 2*). Despite this modification, the underlying principle of the MMPI is analogous to that of the MPI. The PI of the candidate item i^* for the multidimensional case is thus given by:

$$\mathrm{PI}_{i^*} = |\mathbf{W}_{t+i^*}| \prod_{k=1}^{K} (w_k f_k)^{c_{i^*k}}. \qquad (5)$$

The Weighted Penalty Model.

Although not described explicitly in Shin et al.'s (2009) study, the WPM is also based on a constraint relevancy matrix $\mathbf{C}$ where the elements c_{ik} denote whether an item i is relevant for a constraint k or not. The Item selection with the WPM is conducted over three major operations: first, for every eligible item i^* in the pool, the weighted penalty value F_{i^*} is calculated; second, eligible items are assigned to various groups based on their desirability in regards to the specified constraints k (Shin et al., 2009), and third, the item belonging to the group of the highest priority and with the smallest weighted penalty value F_{i^*} within this group is administered.

Compared to the PI, the calculation of the weighted penalty value F_{i^*} is more complex and involves several sub-steps (Shin et al., 2009). For the sake of clarity, in the following paragraphs, only major steps are described. One step involves calculating the total content penalty value F_{i^*}''' for every eligible candidate item i^* using *Equation 6* where P_{i^*k} is the penalty value and w_k is the weight of constraint k that can be used to control the relative importance of the various constraints.

$$F_{i^*}''' = \sum_{k=1}^{K} P_{i^*k} \times w_k \qquad (6)$$

Small P_{i^*k} values denote that the item i^* is relevant for a constraint k that is needed at the present test stage. If an item is not relevant for a constraint k, the penalty value P_{i^*k} becomes zero. The total content penalty value F_{i^*}''' expresses the desirability of an item i^* with regards to all specified content constraints. As it is not limited to a specific range of values, the total content penalty value F_{i^*}''' is standardized using *Equation 7* where $min(F_{i^*}''')$ and $max(F_{i^*}''')$ are the minimum and maximum F_{i^*}''' over all eligible items i^*, respectively. In using this standardization, the resulting standardized total content penalty value F_{i^*}' is limited to values ranging from zero to one.

$$F'_{i^*} = \frac{F'''_{i^*} - \min(F''')}{\max(F'''_{i^*}) - min(F'''_{i^*})} \tag{7}$$

In an additional step, the standardized information penalty F''_{i^*} value is computed using *Equation 8,* where I_{i^*} is the Fisher item information for a candidate item i^* and I_{max} is the maximum Fisher item information across all eligible candidate items according to the provisional ability estimate $\hat{\theta}$.

$$F''_{i^*} = -\left(\frac{I_{i^*}}{I_{max}}\right)^2. \tag{8}$$

Finally, the weighted penalty value F_{i^*} is determined as the weighted sum of the two standardized penalty values (*Equation 9*). Weights w' and w'' are associated with the respective penalty values and can be used to determine the trade-off between content constraints and statistical information (Shin et al., 2009).

$$F_{i^*} = w'F'_{i^*} + w''F''_{i^*} \tag{9}$$

To extend the WPM to the multidimensional case (MWPM), only the calculation of the standardized information penalty F''_{i^*} needs to be modified. To ensure the comparability of MMPI and MWPM, the Fisher information I_{i^*} of the candidate item i^* and the maximum information I_{max} in *Equation* 8 are substituted with Segall's Bayesian item selection $\mathbf{W}_{t+i^*}$ and with the maximum Segall Bayesian item selection criterion across all eligible items $\mathbf{W}_{max}$. Hence, the information penalty F''_{i^*} for the multidimensional case can be determined using *Equation 10.* The standardized total content penalty value F'_{i^*} formula stays the same.

$$F''_{i^*} = -\left(\frac{|\mathbf{W}_{t+i^*}|}{|\mathbf{W}_{max}|}\right)^2. \tag{10}$$

Comparison between the MMPI and MWPM.

The MMPI and MWPM, in addition to their unidimensional ancestors, can be understood as penalty-based approaches. However, they differ in the ways in which they calculate the overall desirability of an item. For the MMPI, the statistical information of a candidate item i^* is multiplied by a term that denotes the suitability of this item in fulfilling a set of non-statistical constraints. Possible values for this term range from zero to one. Therefore, the penalty in the MMPI is expressed by low values for this term. In contrast to the MMPI, for the MWPM, a total penalty is determined as the sum of two separate penalty terms–one for statistical information and one for content constraints. Separate penalties used in the MWPM approach allow for a trade-off to be determined between the non-statistical constraints and statistical information. This procedure can be carried out easily with weights used for separate terms. For the MMPI, a means of determining such a trade-off has not yet been proposed. To make the two methods directly comparable, in the present study, the weights for separate terms used in the MWPM approach are set to a value of one.

The two examined CMMs are also similar in the selective appropriateness for specific item pool structures. Numerous educational and psychological tests are based on item pools with between-item-multidimensionality structures, whereby each item in a pool assesses only one latent trait. However, for some multidimensional assessments, items measure multiple latent trait dimensions. The MMPI and MWPM in their current forms are better suited to between-item-multidimensionality structures. For assessments based on an item-pool with items measuring one latent trait or several traits, item selection based on the MMPI and MWPM tends to favor items with a single loading. This characteristic is attributable to the ways in which priority index and standardized total content penalty values are calculated, generating a smaller priority index or a higher penalty value for items that measure several traits. In reference to the MPI, Su and Huang (2015) recently described this problem and developed a modified MPI for item selection in cases of within-item multidimensionality. However, as we are focusing between-item multidimensionality, which is often used in operational tests, a modification of the MMPI and MWPM is not necessary in the present study.

Research Questions

As CMMs are designed to fulfill desired test specifications while optimizing statistical information of the presented items, CMM usage results in a more or less intense loss of measurement precision. The magnitude of this loss will depend on the CMM used (He et al., 2014), on the number of constraints imposed, and on the characteristics of the item pool. In addition to unidimensional adaptive testing, the correlation structure of the measured dimensions is crucial for measurement precision in MAT (Wang & Chen, 2004; Yoo, 2011). Although several studies have examined the performance of various CMMs, very few have been conducted in the context of MAT (Frey et al., 2011; Su, 2015; Su & Huang, 2015; Veldkamp & van der Linden, 2002; Yao, 2014). Furthermore, no previous study has systematically varied the number of constraints or has analyzed interactions between CMMs and the number of imposed constraints. In providing this information, which is essential for determining which CMMs should be used for MAT, the present study addresses four research questions.

1. What effect does the number of imposed constraints have on the extent to which the MMPI and MWPM are capable of fulfilling these constraints?

2. What effect does the number of imposed constraints have on the measurement precision of the MMPI and MWPM relative to that of an item selection procedure based solely on statistical optimality?

3. Are there specific assessment situations for which the MMPI or MWPM are recommended?

4. What effect does the correlation between the measured dimensions have on the performance of the MMPI and MWPM?

Method

Study Design.

To answer the four research questions presented above, a comprehensive study with simulated data was conducted. The study was based on a full factorial design with three independent variables (IVs). For all of the conditions, $p = 3$ latent trait dimensions were considered. The first IV *constraint management method* involved comparing an item selection based solely on a statistical optimality criterion (referred to as “none”) to the MMPI and MWPM. For the second IV *constraints*, the extent to which item selection was restricted was varied systematically. Levels of this IV stand for the number of imposed constraints (3, 8, 13, 18, 23, 28, 33, 38, 43, 48, and 53), which needed to be fulfilled over the course of the test. For the third IV *correlation*, correlation levels (.2, .5, and .8) between the three dimensions were specified. The fully crossed design used in the study generated 99 experimental conditions. For each design cell, the performance of the CMM was analyzed based on 100 replications with regards to various evaluation criteria.

Item Pools.

For each replication, an item pool with 600 items (200 items per dimension) was constructed. Each item measures exactly one of three dimensions (between-item-multidimensionality). Item discrimination parameters were drawn from a uniform distribution on the interval of real numbers (0.5, 1.5), and item difficulty parameters were drawn from a standard normal distribution, $b_i \sim \mathrm{N}(0,1)$.

Furthermore, for each replication a constraint relevancy matrix $\mathbf{C}$ was constructed to systematically vary the number of imposed constraints. The first three columns of this matrix denote an item loading on dimensions one to three for entries of zero or one. In addition, 50 dichotomous variables were generated, with each representing a fictitious categorical item property. Examples of such categorical item properties include the response format of an item (e.g., multiple choice, complex multiple choice, short answer), the answer key (e.g., first option, second option, third option, fourth option), and the cognitive level needed to solve an item (e.g., knowledge, comprehension, application) (van der Linden, 2005b). The item loadings on these item properties were randomly assigned with a probability of .5. Accordingly, approximately 50 percent of the items (ca. 300) in the pool were of relevance to a particular item property.

Data Generation.

For each replication, a sample of 1,000 simulees was generated. Ability parameters were randomly drawn from a multivariate normal distribution of $\boldsymbol{\theta} \sim \mathrm{MVN}(\boldsymbol{\mu}, \boldsymbol{\Phi})$ with $\boldsymbol{\mu} = (0, 0, 0)$ and

$$\boldsymbol{\Phi} = \begin{pmatrix} 1.00 & \mathrm{\rho} & \mathrm{\rho} \\ \rho & 1.00 & \rho \\ \rho & \rho & 1.00 \end{pmatrix}. \tag{11}$$

Three different levels of correlations ρ (.2, .5, .8) between the measured dimensions were used to study the effect of the correlation on the performance of the MMPI and MWPM. Binary responses on the items for the simulees were generated based on the M2PL model (*Equation 12*).

$$P\left(U_{ij} = 1 | \boldsymbol{\theta}_j, \boldsymbol{a}_i', b_i\right) = \frac{\exp\left(\boldsymbol{a}_i'(\boldsymbol{\theta}_j - b_i\mathbf{1})\right)}{1 + \exp\left(\boldsymbol{a}_i'(\boldsymbol{\theta}_j - b_i\mathbf{1})\right)} \tag{12}$$

MAT Specifications.

The simulations were performed using SAS® 9.4 for a fixed test length of 60 items. For all of the conditions, the ability vector $\boldsymbol{\theta}_j$ was estimated by the multidimensional Bayes modal estimator using the variance-covariance matrix $\boldsymbol{\Phi}$ from *Equation 11* as prior. The number of imposed constraints was reflected by a test blueprint in which the constraints, the associated weights and the lower and upper bounds of the constraints were specified. Table 1 shows an abridged version of an overall blueprint for all levels of the IV constraints (3, 8, 13, 18, 23, 28, 33, 38, 43, 48, and 53). The first level (constraints = 3) represents a test blueprint whereby only the number of administered items for each of the three dimensions is constrained. The blueprint of the next level includes the constraints of the previous level. For example, the blueprint of the second level (constraints = 8) contains the first three constraints of the first level and five additional constraints regarding the fictitious categorical properties.

--- Insert Table 1 about here ---

The weights for all constraints were set to a value of one. For the number of items per dimension, the lower bound was set to 18 and the upper bound was set to 22; the bounds for each categorical property were 28 and 32.

Evaluation Criteria.

As dependent variables (DV), the average mean squared error (MSE), the proportion of tests with at least one constraint violation (*%Viol*), and the average number of violations (*#Viol*) were used. MSE was calculated as the average squared difference between the ability estimates $\hat{\theta}$ and true ability θ across all p dimensions (*Equation 12*). Thus, a high degree of measurement precision is denoted by low values for the MSE.

$$MSE = \frac{1}{p \cdot N} \sum_{l=1}^{p} \sum_{j=1}^{N} \left(\hat{\theta}_{pj} - \theta_{pj} \right)^2 . \qquad (13)$$

The other two DVs were used to evaluate the extent to which the imposed constraints were fulfilled. %Viol was computed as the ratio of simulees taking a test with at least one constraint violation relative to all of the simulees multiplied by 100. #Viol was calculated as the average of constraint violations across all simulees N.

Results

In this section, the four research questions of the present study are answered. First, the results regarding the constraint violations are presented. Second, the measurement precision of the MMPI and MWPM is compared. Then, the performance of the CMM in the various assessment situations is evaluated. Finally, the performance of the CMM for the various correlation levels between the measured dimensions is analyzed.

Constraint Violations.

The first research question focuses on the CMM's capacity to fulfill the desired test specifications. To answer this question, we evaluated the proportion of tests with at least one constraint violation (*%Viol*) and the average number of violations (*#Viol*). Tables 2 and 3 show the results of the various correlation levels.

--- Insert Table 2 about here ---

It can be concluded that the MMPI and MWPM perfectly met all imposed constraints for all of the conditions. Accordingly, no test was conducted with at least one violation, and the average number of violations was zero. When item selection was solely based on issues of statistical optimality (CMM = "none"), for almost all of the conditions, the proportion of tests with at least one violation was higher than 97 percent. However, when constraints were only imposed on the number of administered items per dimension (constraints = 3), the percentage of tests with at least one violation was considerably lower. When no CMM was used, the average number of violations increased with a growing number of imposed constraints.

--- Insert Table 3 about here ---

Measurement Precision.

The second research question focuses on the effect the number of imposed constraints has on the measurement precision of the MMPI and MWPM. Table 4 presents the average mean squared error (MSE) of the various correlation levels.

--- Insert Table 4 about here ---

We designated the condition wherein item selection was based solely on the statistical optimality criterion (‘none’) as the baseline condition. As expected, the MSE of this condition was independent of the number of imposed constraints. For the MMPI and MWPM, the MSE increased and, accordingly, the measurement precision decreased when more constraints needed to be considered. When constraints only referred to the number of administered items per dimension, there was no loss in the measurement precision of the MMPI and MWPM in relation to the baseline condition. However, when numerous constraints were involved, the measurement precision of the two heuristic CMM decreased significantly relative to the baseline.

Performance in Various Assessment Situations.

In addressing the third research question, we aim to make recommendations in regards to the two CMMs for specific assessment situations. We thus examined interactions between the CMM, the number of test specifications and the performance of the CMM used. Performance was assessed based on constraint violations and measurement precision. As stated above, both CMMs perfectly met all of the imposed constraints for all of the conditions. However, a closer examination of this result shows that the loss in measurement precision resulting from an increasing number of imposed constraints was different for the two CMMs. For low to moderate numbers of imposed constraints, the measurement precision of the MMPI was higher than that of the MWPM. This changed when numerous constraints were involved. Here, the MWPM outperformed the MMPI. This interaction between the constraint management methods and the number of imposed constraints is shown in Figure 1. The graph intersection point for all of the correlation levels occurs within 23 and 28 constraints. Thus, for this number of constraints and above, the MWPM performs better than the MMPI.

--- Insert Figure 1 about here ---

Performance of Various Correlation Levels.

The fourth research question concerns the effect that the correlation between the measured dimensions has on the performance of the MMPI and MWPM. Tables 2, 3, and 4 present the criteria for evaluating performance depending on the correlation levels between the measured dimensions. The two CMMs do not differ in regards to constraint violations ($\%Viol$ and $\#Viol$) for various correlation levels. By contrast, measurement precision is affected by correlations between the measured dimensions. For all of the conditions, a higher correlation between the measured dimensions resulted in lower MSE values and thus in higher degrees of measurement precision. Furthermore, the loss in measurement precision derived from an increasing number of imposed constraints was related to correlations between the measured dimensions. In Table 4, it is clearly shown that the least correlated dimensions were associated with a greater loss of measurement precision.

Discussion

The present study focused on the effectiveness of two promising heuristic CMMs in MAT for varying levels of imposed constraints and for various correlations between the measured dimensions. The multidimensional extension of the WPM was introduced, and its performance was compared to the MMPI through a simulation study. The performance of the two CMMs was evaluated based on constraint violation and measurement precision.

The study shows that both the MMPI and MWPM are capable of addressing complex sets of constraints in MAT without causing any violations. By contrast, when item selection is based only on a statistical optimality criterion, the number of tests with constraint violations is quite high. The small proportion of tests with at least one constraint violation found for the condition where no CMM was used and where only the number of administered items per dimension was constrained may erroneously support the conclusion that CMMs may not be needed to balance the proportion of items per dimension. However, this result is rather attributable to the built-in “minimax mechanism” of the D-optimal method (Mulder & van der Linden, 2009), which tends to select items belonging to the dimension with the least information. Consequently, the requested proportions of items per dimension are automatically balanced when (as in the present study) the item pool is balanced. Though we expected the two CMMs examined to perform better in regards to the fulfillment of test specifications, it is not a trivial finding that all of the constraints were met perfectly in each condition. Thus, this study highlights the capability of the MMPI and MWPM in meeting test specifications when numerous constraints are involved.

To fulfill imposed constraints, statistical information must be sacrificed to a particular degree. Accordingly, measurement precision decreased with an increasing number of constraints for both CMMs. This result is not surprising, because with an increasing number of constraints and a constant item pool size, the proportion of items with a specific combination of properties decreases. In turn, the penalty for non-statistical constraints becomes more critical to the item selection process, and the relevance of statistical information decreases. In particular, when numerous constraints are involved, the loss in measurement precision is significant compared to that found for MAT without CMM (“none”).

This study shows that the performance of the two heuristic CMMs for various correlation levels does not differ in regards to constraint violations. In accordance with the results of Wang and Chen (2004) and Yoo (2011), measurement precision was affected by correlation levels among the dimensions. The loss in measurement precision that resulted from an increasing number of imposed constraints was related to the correlation level and to CMM used. For tests with highly correlated dimensions, the resulting loss was considerably smaller than that for lowly correlated dimensions, and especially when numerous constraints were imposed. In regards to the CMMs used, the MMPI performed slightly better than the MWPM for low to moderate numbers of constraints. Although the difference in measurement precision between the MMPI and MWPM was rather small, we recommend using the MMPI for assessment situations involving few constraints, as this is a low maintenance method (He et al., 2014). When numerous constraints are involved, the MWPM appears to be more suitable. These findings seem to contradict results presented by He et al. (2014), who found the WPM method to perform considerably better for a moderate number of constraints in the unidimensional case. However, these results are not directly comparable, as the two studies differ in some major respects (e.g., the number of items in the pool per dimension, the item selection criterion, the use of constraint weights).

The results of the present study make a substantial contribution to the management of test specifications in several respects. First, through the extension of the WPM to the MWPM, a new heuristic CMM that can address complex sets of constraints was made available for MAT. Second, the results underline that the number of imposed constraints constitutes a crucial factor affecting the management of test specifications (Donoghue, 2014). Third, the study pointed out that for some assessment situations, one specific CMM is more suitable to use than another. In selecting a heuristic CMM to address complex constraints, the MMPI should be used when a low to moderate number of constraints is involved, and the MWPM should be employed when numerous constraints are involved.

The findings of the present study are limited to assessment situations in which the item pool is constructed so desired test specifications can be fulfilled. This assumption is appropriate, as test specifications are typically known prior to item pool construction. Furthermore, we used Segall's Bayesian item selection method (1996), a powerful item selection method that is likely the most frequently studied and applied item selection procedure for MAT. Nevertheless, as other MAT item selection methods may have specific effects on CMM performance, a comparison between the presented CMM and various item selection procedures should be conducted. As the CMMs used in this study are better suited to cases of between-item-multidimensionality, the extension of methods to within-item-multidimensionality contexts represents another area for future inquiry.

In conclusion, the MWPM and MMPI are two heuristic CMMs that can manage complex sets of constraints in MAT. In applying these two methods, which can be selected depending on the number of constraints involved, the management of test specifications in MAT becomes much easier, and more operational applications of this powerful method may be generated as a result.

References

Chen, S.-Y., & Ankenman, R. D. (2004). Effects of practical constraints on item selection rules at the early stages of computerized adaptive testing. *Journal of Educational Measurement*, *41*, 149–174. doi:10.1111/j.1745-3984.2004.tb01112.x

Cheng, Y., & Chang, H.-H. (2009). The maximum priority index method for severely constrained item selection in computerized adaptive testing. *British Journal of Mathematical and Statistical Psychology*, *62*, 369–383. doi:10.1348/000711008X304376

Cheng, Y., Chang, H.-H., Douglas, J., & Guo, F. (2008). Constraint-Weighted a-Stratification for Computerized Adaptive Testing With Nonstatistical Constraints: Balancing Measurement Efficiency and Exposure Control. *Educational and Psychological Measurement*, *69*(1), 35–49. doi:10.1177/0013164408322030

Cheng, Y., Chang, H.-H., & Yi, Q. (2007). Two-phase item selection procedure for flexible content balancing in CAT. *Applied Psychological Measurement*, *31*, 467–482. doi:10.1177/0146621606292933

Diao, Q., & van der Linden, W. J. (2011). Automated test assembly using lp_Solve version 5.5 in R. *Applied Psychological Measurement*, *35*, 398–409. doi:10.1177/0146621610392211

Donoghue, J. R. (2014). *Comparison of integer programming (IP) solvers for automated test assembly (ATA)* (ETS Research Report No. RR-15-05). Princeton, NJ.

Frey, A., Cheng, Y., & Seitz, N. N. (2011, April). *Content balancing with the maximum priority index method in multidimensional adaptive testing.* Paper presented at the 2011 meeting of the National Council on Measurement in Education (NCME), New Orleans, LA.

Frey, A., & Seitz, N.-N. (2009). Multidimensional adaptive testing in educational and psychological measurement: Current state and future challenges. *Studies in Educational Evaluation, 35*(2-3), 89–94. doi:10.1016/j.stueduc.2009.10.007

Hambleton, R. K., & Swaminathan, H. (1985). *Item response theory: Principles and applications*. *Evaluation in education and human services*. Boston: Kluwer.

He, W., Diao, Q., & Hauser, C. (2014). A comparison of four item-selection methods for severely constrained CATs. *Educational and Psychological Measurement.* doi:10.1177/0013164413517503

Kingsbury, C. G., & Zara, A. R. (1991). A comparison of procedures for content-sensitive item selection in computerized adaptive tests. *Applied Measurement in Education*, *4*, 241–261. doi:10.1207/s15324818ame0403_4

Liu, J. (2007). *Comparing multi-dimensional and uni-dimensional computer adaptive strategies in psychological and health assessment* (Doctoral dissertation). University of Pittsburgh, Pittsburgh, PA.

Mulcahey, M. J., Haley, S. M., Duffy, T., Pengsheng, N., & Betz, R. R. (2008). Measuring physical functioning in children with spinal impairments with computerized adaptive testing. *Journal of Pediatric Orthopaedics*, *28*, 330–335. doi:10.1097/BPO.0b013e318168c792

Mulder, J., & van der Linden, W. J. (2009). Multidimensional adaptive testing with optimal design criteria for item selection. *Psychometrika*, *74*, 273–296. doi:10.1007/S11336-008-9097-5

Reckase, M. D. (2009). Computerized adaptive testing using MIRT. In M. D. Reckase (Ed.), *Statistics for Social and Behavioral Sciences. Multidimensional item response theory* (pp. 311–339). New York: Springer.

Segall, D. O. (1996). Multidimensional adaptive testing. *Psychometrika*, *61*, 331–354.

Segall, D. O. (2005). Computerized adaptive testing. In K. Kempf-Leonard (Ed.), *The encyclopedia of social measurement* (pp. 429–438). Boston: Elsevier/Academic.

Shin, C. D., Chien, Y., Way, W. D., & Swanson, L. (2009). Weighted Penalty Model for content balancing in CATs. Pearson. Retrieved from http://images.pearsonassessments.com/images/tmrs/tmrs_rg/WeightedPenaltyModel.pdf?WT.mc_id=TMRS_Weighted_Penalty_Model_for_Content_Balancing

Stocking, M. L., & Swanson, L. (1993). A method for severely constrained item selection in adaptive testing. *Applied Psychological Measurement*, *17*, 277–292. doi:10.1177/014662169301700308

Su, Y.-H. (2015). The Performance of the Modified Multidimensional Priority Index for Item Selection in Variable-Length MCAT. In van der Ark, Andries L, M. D. Bolt, W.-C. Wang, A. J. Douglas, & S.-M. Chow (Eds.), *Quantitative Psychology Research: The 79th Annual Meeting of the Psychometric Society, Madison, Wisconsin, 2014* (pp. 89–97). Cham: Springer International Publishing.

Su, Y.-H., & Huang, Y.-L. (2015). Using a modified Multidimensional Priority Index for item selection under within-item multidimensional computerized adaptive testing. In R. E. Millsap, D. M. Bolt, van der Ark, L. Andries, & W.-C. Wang (Eds.), *Springer Proceedings in Mathematics & Statistics. Quantitative psychology research* (Vol. 89, pp. 227–242). Springer International Publishing.

Swanson, L., & Stocking, M. L. (1993). A model and heuristic for solving very large item selection problems. *Applied Psychological Measurement*, *17*, 151–166. doi:10.1177/014662169301700205 van der Linden, W. J. (1999). Multidimensional adaptive testing with a minimum error-variance criterion. *Journal of Educational and Behavioral Statistics*, *24*, 398–412. doi:10.3102/10769986024004398

van der Linden, W. J. (2005a). A comparison of item-selection methods for adaptive tests with content constraints. *Journal of Educational Measurement*, *42*, 283–302. doi:10.1111/j.1745-3984.2005.00015.x

van der Linden, W. J. (2005b). *Linear models for optimal test design*. New York: Springer.

van der Linden, W. J., & Glas, C. A. W. (Eds.). (2010). *Statistics for Social and Behavioral Sciences*. *Elements of adaptive testing*. New York: Springer.

van der Linden, W. J., & Reese, L. M. (1998). A model for optimal constrained adaptive testing. *Applied Psychological Measurement*, *22*, 259–270. doi:10.1177/01466216980223006

Veldkamp, B. P., & van der Linden, W. J. (2002). Multidimensional adaptive testing with constraints on test content. *Psychometrika*, *67*, 575–588. doi:10.1007/BF02295132

Wang, C., & Chang, H.-H. (2011). Item selection in multidimensional computerized adaptive testing—Gaining information from different angles. *Psychometrika*, *76*, 363–384. doi:10.1007/S11336-011-9215-7

Wang, C., Chang, H.-H., & Boughton, K. A. (2011). Kullback-Leibler Information and its applications in multidimensional adaptive testing. *Psychometrika*, *76*, 13–39. doi:10.1007/s11336-010-9186-0

Wang, W.-C., & Chen, P.-H. (2004). Implementation and measurement efficiency of multidimensional computerized adaptive testing. *Applied Psychological Measurement*, *28*, 295–316. doi:10.1177/0146621604265938

Wang, W.-C., Chen, P.-H., & Cheng, Y.-Y. (2004). Improving measurement precision of test batteries using multidimensional item response models. *Psychological Methods*, *9*(1), 116–136. doi:10.1037/1082-989X.9.1.116

Yao, L. (2010). Reporting valid and reliable overall scores and domain scores. *Journal of Educational Measurement*, *47*(3), 339–360. doi:10.1111/j.1745-3984.2010.00117.x

Yao, L. (2012). Multidimensional CAT item selection methods for domain scores and composite scores: Theory and applications. *Psychometrika*, *77*, 495–523. doi:10.1007/s11336-012-9265-5

Yao, L. (2013). Comparing the performance of five multidimensional CAT selection procedures with different stopping rules. *Applied Psychological Measurement*, *37*, 3–23. doi:10.1177/0146621612455687

Yao, L. (2014). Multidimensional CAT item selection methods for domain scores and composite scores with item exposure control and content constraints. *Journal of Educational Measurement*, *51*, 18–38. doi:10.1111/jedm.12032

Yoo, H. (2011). *Evaluating several multidimensional adaptive testing procedures for diagnostic assessment* (Doctoral Dissertations). Retrieved from http://scholarworks.umass.edu/dissertations/AAI3465252

Table 1

Overall Test Blueprint for all Research Conditions

Constraints	*Weight*	*Lower bound*	*Upper bound*
Dim 1	1	18	22
Dim 2	1	18	22
Dim 3	1	18	22
C4	1	28	32
C5	1	28	32
C6	1	28	32
C7	1	28	32
C8	1	28	32
C9	1	28	32
C10	1	28	32
C11	1	28	32
C12	1	28	32
C13	1	28	32
C14	1	28	32
.	.	.	.
.	.	.	.
.	.	.	.
C48	1	28	32
C49	1	28	32
C50	1	28	32
C51	1	28	32
C52	1	28	32
C53	1	28	32

Table 2

Percentage of Tests with at least one Violation (%Viol) and Standard Error of different Constraint Management Method for Multidimensional Tests with differently Correlated Dimensions

No. of Constraints	None			MMPI			MWPM		
	low ρ = .2	moderate ρ = .5	high ρ = .8	low ρ = .2	moderate ρ = .5	high ρ = .8	low ρ = .2	moderate ρ = .5	high ρ = .8
3	15.69 (3.75)	15.26 (3.38)	13.10 (3.51)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
8	97.27 (2.41)	97.17 (2.64)	97.05 (3.14)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
13	99.94 (0.14)	99.94 (0.18)	99.96 (0.18)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
18	100.00 (0.01)	100.00 (0.01)	100.00 (0.01)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
23	100.00 (0.00)	100.00 (0.00)	100.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
28	100.00 (0.00)	100.00 (0.00)	100.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
33	100.00 (0.00)	100.00 (0.00)	100.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
38	100.00 (0.00)	100.00 (0.00)	100.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
43	100.00 (0.00)	100.00 (0.00)	100.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
48	100.00 (0.00)	100.00 (0.00)	100.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
53	100.00 (0.00)	100.00 (0.00)	100.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)

Table 3

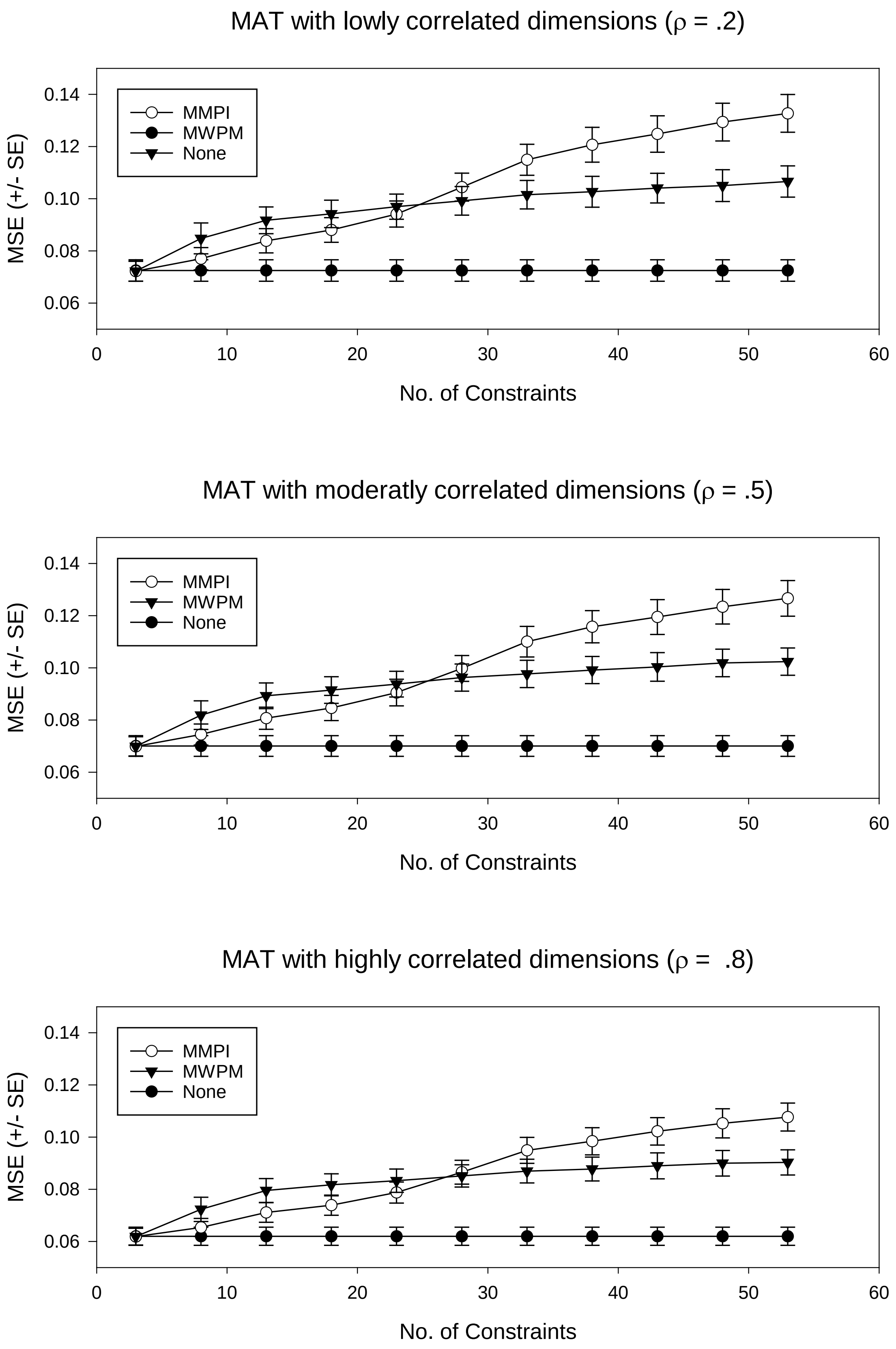
Average Number of Violations (#Viol) and Standard Error of different Constraint Management Method for Multidimensional Tests with differently Correlated Dimensions

	None			MMPI			MWPM		
No. of Constraints	low ρ = .2	moderate ρ = .5	high ρ = .8	low ρ = .2	moderate ρ = .5	high ρ = .8	low ρ = .2	moderate ρ = .5	high ρ = .8
3	0.21 (0.05)	0.21 (0.05)	0.18 (0.05)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
8	2.77 (0.38)	2.77 (0.38)	2.73 (0.41)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
13	5.45 (0.50)	5.44 (0.50)	5.43 (0.52)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
18	8.00 (0.56)	8.01 (0.58)	8.00 (0.64)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
23	10.58 (0.65)	10.59 (0.67)	10.60 (0.71)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
28	13.20 (0.73)	13.20 (0.73)	13.21 (0.76)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
33	15.83 (0.76)	15.84 (0.76)	15.83 (0.78)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
38	18.50 (0.80)	18.50 (0.81)	18.49 (0.83)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
43	21.11 (0.84)	21.11 (0.84)	21.10 (0.89)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
48	23.74 (0.94)	23.74 (0.93)	23.76 (0.97)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
53	26.33 (1.08)	26.32 (1.05)	26.34 (1.06)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)

Table 4

Average Mean Squared Error (MSE) and Standard Error of different Constraint Management Method for Multidimensional Tests with differently Correlated Dimensions

	None			MMPI			MWPM		
No. of Constraints	low ρ = .2	moderate ρ = .5	high ρ = .8	low ρ = .2	moderate ρ = .5	high ρ = .8	low ρ = .2	moderate ρ = .5	high ρ = .8
3	0.072 (0.004)	0.070 (0.004)	0.062 (0.003)	0.072 (0.004)	0.070 (0.004)	0.062 (0.003)	0.072 (0.004)	0.070 (0.004)	0.062 (0.003)
8	0.072 (0.004)	0.070 (0.004)	0.062 (0.003)	0.077 (0.004)	0.074 (0.004)	0.065 (0.003)	0.085 (0.006)	0.082 (0.005)	0.072 (0.005)
13	0.072 (0.004)	0.070 (0.004)	0.062 (0.003)	0.084 (0.005)	0.081 (0.004)	0.071 (0.004)	0.092 (0.005)	0.089 (0.005)	0.080 (0.005)
18	0.072 (0.004)	0.070 (0.004)	0.062 (0.003)	0.088 (0.005)	0.085 (0.005)	0.074 (0.004)	0.094 (0.005)	0.091 (0.005)	0.082 (0.004)
23	0.072 (0.004)	0.070 (0.004)	0.062 (0.003)	0.094 (0.005)	0.091 (0.005)	0.079 (0.004)	0.097 (0.005)	0.094 (0.005)	0.083 (0.004)
28	0.072 (0.004)	0.070 (0.004)	0.062 (0.003)	0.104 (0.006)	0.100 (0.005)	0.087 (0.005)	0.099 (0.005)	0.096 (0.005)	0.085 (0.004)
33	0.072 (0.004)	0.070 (0.004)	0.062 (0.003)	0.115 (0.007)	0.110 (0.006)	0.095 (0.005)	0.102 (0.005)	0.098 (0.005)	0.087 (0.005)
38	0.072 (0.004)	0.070 (0.004)	0.062 (0.003)	0.121 (0.007)	0.116 (0.006)	0.098 (0.005)	0.103 (0.006)	0.099 (0.005)	0.088 (0.005)
43	0.072 (0.004)	0.070 (0.004)	0.062 (0.003)	0.125 (0.007)	0.119 (0.007)	0.102 (0.005)	0.104 (0.006)	0.100 (0.005)	0.089 (0.005)
48	0.072 (0.004)	0.070 (0.004)	0.062 (0.003)	0.129 (0.007)	0.123 (0.007)	0.105 (0.006)	0.105 (0.006)	0.102 (0.005)	0.090 (0.005)
53	0.072 (0.004)	0.070 (0.004)	0.062 (0.003)	0.133 (0.007)	0.127 (0.007)	0.108 (0.005)	0.107 (0.006)	0.102 (0.005)	0.090 (0.005)


MAT with lowly correlated dimensions (ρ = .2)

MAT with moderatly correlated dimensions (ρ = .5)

MAT with highly correlated dimensions (ρ = .8)

Figure 1 Average Mean Squared Error (*MSE*) on Number of Imposed Constraints for the Different Constraint Management Methods and Different Correlation Levels