Multidimensional Item Response Theory Models in Vocational Interest Measurement: An Illustration Using the AIST-R

Eunike Wetzel1 and Benedikt Hell1

Abstract
Vocational interest inventories are commonly analyzed using a unidimensional approach, that is, each subscale is analyzed separately. However, the theories on which these inventories are based often postulate specific relationships between the interest traits. This article presents a multidimensional approach to the analysis of vocational interest data, which takes these relationships into account. Models in the framework of Multidimensional Item Response Theory (MIRT) are explained and applied to a widely used German vocational interest inventory based on the RIASEC model, the AIST-R. MIRT models were more appropriate to describe the data than unidimensional models. It follows that responses to some items were not only influenced by the interest type they were designed to measure but also by another dimension. The advantages of MIRT models are discussed.

Keywords
vocational interests, Multidimensional Item Response Theory (MIRT), RIASEC

Introduction
Theories of vocational interests posit differing numbers of interest dimensions. For example, according to Holland’s popular RIASEC model (Holland, 1959, 1997), six interest types describe vocational interests, namely, Realistic, Investigative, Artistic, Social, Enterprising, and Conventional. Numerous vocational interest inventories exist that assess these interest types (e.g., Self-directed Search; Holland, Fritzsche, & Powell, 1994). Analyses of data from vocational interest inventories usually treat the subscales as separate constructs with each item only assessing one construct. However, the theories on which these inventories are based often postulate specific relationships between the interest dimensions. For example, Holland (1997) theorizes that his RIASEC types can be represented by a hexagon in which spatial distances between interest domains indicate the degree to which they are related (see Figure 1). In the hexagon, adjacent interests (e.g., Realistic and Investigative) are more highly related than alternate interests (e.g., Realistic and Artistic), which are in turn more highly related than opposite interests.

1University of Konstanz, Baden-Württemberg, Germany

Corresponding Author:
Eunike Wetzel, Center for Educational Science and Psychology, Eberhard Karls University Tuebingen, Europastraße 6, 72072 Tübingen, Germany.
Email: eunike.wetzel@uni-tuebingen.de
Wetzel and Hell

The hexagonal (or circumplex; Guttman, 1954) structure of vocational interests based on the RIASEC model has been confirmed empirically (e.g., Rounds, Tracey, & Hubert, 1992). This raises the question whether relationships between interest domains should be taken into account in analyses of vocational interest data.

A promising development for the analysis of data from distinct, but related, latent traits such as the RIASEC factors, is Multidimensional Item Response Theory (MIRT). MIRT has been used extensively in the ability domain to analyze responses on items that require more than one ability. For example, Walker and Beretvas (2003) compared a one-dimensional model for general mathematical ability with a two-dimensional model (additionally including the ability to communicate in mathematics) for data from a large-scale mathematics test. Their results indicated that the two-dimensional model was more appropriate and, importantly, that inferences about student ability levels differed depending on which model was chosen. In one of the few applications of MIRT outside the ability domain, Osteen (2010) used MIRT to assess the latent factor structure of the Social Work Community of Practice Scale. As no applications of MIRT to the vocational interest domain could be found, it seems important to broaden knowledge here.

The general aim of this article is to demonstrate the application of MIRT to vocational interest measurement. As an example, we analyze a widely used German vocational interest inventory, namely, the General Interest Structure Test (AIST-R; Bergmann & Eder, 2005) using unidimensional and multidimensional models. This also allows us to investigate whether the unidimensional or multidimensional modeling of vocational interest data is more adequate for the AIST-R, although these results may not generalize to other interest inventories. The remainder of this article is structured as follows. First, an introduction to MIRT is given in which the differences between unidimensional and multidimensional Item Response Theory (IRT) will be depicted. Second, the specific IRT models used in this paper’s analyses will be explained briefly. Third, results from unidimensional and multidimensional models for the AIST-R will be compared. Finally, the results and their implications for vocational interest measurement will be discussed.

### Unidimensional and Multidimensional IRT

In the framework of IRT, latent traits are related to responses using probabilistic models. These models describe the interaction between persons and items (Reckase, 1997). In the most basic IRT model, the one-parameter logistic model or Rasch model (Rasch, 1960), the probability of a
The response of 1 (correct response or agreement), is a function of the distance between a person’s trait level and the difficulty of the item. Many extensions of this basic IRT model exist (for an overview, see Embretson & Reise, 2000).

The main difference between unidimensional IRT (UIRT) and MIRT is that UIRT models posit that only one latent trait influences responses while MIRT models posit that several latent traits influence responses. Thus, UIRT models the probability of a certain response as a function of a single latent trait, while MIRT models the probability of a response depending on a vector of multiple latent trait dimensions. MIRT analyses enable researchers to investigate and take into account the underlying relationships in the data because MIRT models link multiple latent traits to item responses (Finch, 2010; Reckase, 2009). Besides viewing MIRT as an extension of UIRT, it can also be conceptualized as a special case of factor analysis. MIRT and factor analysis “define hypothetical scales that can be used to reproduce certain features of the data that are the focus of the analysis” (Reckase, 2009, p. 63). However, MIRT and factor analysis have different goals. While MIRT aims at modeling the interaction between persons’ latent trait levels and the responses to items, factor analysis usually aims at data reduction, that is, explaining relationships between items using a small number of factors. One further difference is that in MIRT, item characteristics (e.g., item difficulty and item discrimination) are an object of study, whereas in factor analysis, item characteristics are not of interest. For an extensive treatise of MIRT and its relationship to factor analysis, the interested reader is referred to Reckase (2009).

MIRT models, as described above, are sometimes called within-MIRT models to distinguish them from the so-called between-MIRT models. Between-MIRT models can be considered a special case of within-MIRT models because several dimensions are modeled simultaneously, but each item only measures one specific dimension. These two cases of multidimensionality are contrasted with the unidimensional case in Figures 2 to 4 for the first three RIASEC factors, namely, Realistic, Investigative, and Artistic. As can be seen in the comparison of Figures 2 and 3, between-MIRT models are similar to unidimensional models in that each dimension is assessed by a unique set of items. However, because between-MIRT models include several dimensions simultaneously, they allow the researcher to model the correlations between dimensions as well. In contrast, within-item multidimensionality assumes that responses to some items depend on the trait levels on more than one latent trait (see Figure 4).

**Unidimensional and Multidimensional Partial Credit Model**

In this article, unidimensional and multidimensional models based on the Partial Credit Model (PCM) will be used. The unidimensional PCM (uPCM) was formulated by Masters (1982) and forms an extension of the dichotomous Rasch Model (Rasch, 1960) to polytomous data with ordered response categories. Because the uPCM is a direct extension of the Rasch Model, all the desirable properties of the Rasch Model, such as the sum score being a sufficient statistic and the independence of item and person parameters (Embretson & Reise, 2000), are preserved. The multidimensional PCM (mPCM) was developed as a generalization of the uPCM to multidimensional data by Kelderman (1996). The main difference between the uPCM and the mPCM is that in the multidimensional case, more than one trait parameter exists. Furthermore, weights are included that indicate whether the response to an item depends on a certain trait. The uPCM and the mPCM assume that items only differ in their level of difficulty but not in their ability to differentiate between different trait levels. The Generalized Partial Credit Model (GPCM; Muraki, 1992) is an extension of the PCM that relaxes this assumption by additionally modeling discrimination parameters for each item. Model fit comparisons between the PCM and the GPCM indicate whether the assumption of equal item discriminations in the PCM holds. Detailed accounts of different versions of the PCM can be found in Masters (1982), Kelderman (1996), and Muraki (1992).
Most interest inventories (e.g., Self-directed Search; Holland et al., 1994) consist of subscales that are all purported to measure related, but distinct, latent traits. In these inventories, each item is assigned to one particular subscale and none of the items are common between subscales. The dominant approach is to analyze these subscales separately, ignoring the possibly existing multidimensionality between them. In this approach, relationships between latent traits are not modeled. According to Adams, Wilson, and Wang (1997), when subscales measure highly correlated latent traits, “a joint analysis with an appropriate model will lead to improved estimation of the item parameters and ability predictions” (p. 11). Thus, the unidimensionality assumption may be inappropriate for analyzing instruments assessing (at least in part) highly intercorrelated latent traits such as the RIASEC factors. Therefore, in this article UIRT and MIRT models will be applied. First, pre-analyses will be conducted to investigate which items may load on additional factors to the one they were designed to measure. Second, unidimensional and multidimensional models will be compared for each RIASEC factor to analyze whether multidimensional models provide a more adequate representation of the data. Third, a between-MIRT model and a within-MIRT model that incorporate all RIASEC factors simultaneously will be compared regarding their fit. Finally, correlations between the RIASEC factors will be contrasted between unidimensional models and a

**Figure 2.** Unidimensional models.
between-MIRT model to investigate which model best represents the relationships posited by Holland’s model.

**Method**

**Instrument**

Participants filled out the *General Interest Structure Test* ("Allgemeiner Interessen-Struktur-Test"; AIST-R; Bergmann & Eder, 2005). The AIST-R is applied to assist occupational and career decisions (e.g., in the context of career counseling). Its 60 items (10 per dimension) describe occupational activities. For example, a Realistic item reads *to manufacture something according to a plan or a sketch*. Interests in the Investigative domain are, for example, assessed by the item *to examine the behavior of animals or plants*. The AIST-R’s rating scale is a 5-point Likert-type scale with the poles *I am not interested in this at all; I do not enjoy doing this at all and I am very interested in this; I enjoy doing this very much*. It takes about 10 to 15 min to complete. The AIST-R was developed according to Holland’s RIASEC model (Holland, 1959, 1997). A (three-letter) Holland code is derived from the test-taker’s subscale scores and can subsequently be matched with a list of occupations classified by their Holland codes. Thus, the test-taker is

![Between-multidimensional model](image-url)
provided with information on occupations that have identical or similar interest profiles as his or her own profile. Evidence supporting the validity of the AIST-R is reported in its manual (Bergmann & Eder, 2005) and includes relationships between the AIST-R’s RIASEC scales and different personality traits, as well as its satisfactory fit to the circumplex structure. Convergent and discriminant validity is supported by the relationships of the AIST-R’s RIASEC scales with the RIASEC scales in the German version of the Self-directed Search (Explorix; Jörin, Stoll, Bergmann, & Eder, 2003). IRT reliabilities for the RIASEC scales resulting from uPCMs and a between-MIRT PCM are reported in Table 1. These can be interpreted analogous to Cronbach’s alpha.

**Samples**

Two samples were used, one for the pre-analyses and another for the other (main) analyses. The pre-analyses sample comprised 797 students (62.4% women) who were enrolled either in college or vocational training. Participants were between 16 and 48 years old ($M = 21.42$, $SD = 2.89$). Data were collected online between October 2010 and March 2011.

The main analyses sample contained 3,997 students in their last year of school (56.7% women) with an age range from 17 to 30 years ($M = 19.21$, $SD = 1.10$). These data were collected between

---

**Figure 4.** Within-multidimensional model.
2003 and 2010 at the University of Linz in Austria. Participants were seeking vocational counseling and filled out a battery of tests (the AIST-R, personality inventories, and cognitive ability tests) described in Bergmann (2008). Participants received extensive feedback on their results. Raw mean scores and standard deviations for this sample on the AIST-R’s RIASEC scales are displayed in Table 1.

**Pre-Analyses**

Prior to estimating the IRT models, exploratory factor analyses (EFAs) using maximum likelihood estimation with Promax rotation were performed on the correlation matrix using SPSS (IBM, 2012). The purpose of these EFAs was to generate hypotheses concerning which items were measuring another interest domain besides the one they were intended to measure. Items that showed high loadings (> .30) on another factor in addition to the one they were designed to measure were modeled to assess two dimensions in the within-MIRT models: (a) the interest dimension they belonged to by instrument design and (b) the secondary dimension they had a high loading on.

**Model Fit Comparisons**

Model fit comparisons between unidimensional and multidimensional models were conducted first for each RIASEC factor separately and second for models that incorporated all RIASEC factors simultaneously. Two types of unidimensional models were estimated for each RIASEC factor: (a) uPCMs that only assume differences in item difficulty and (b) uGPCMs that additionally allow differences in item discrimination. For RIASEC factors that the EFA indicated contained items that loaded on a second dimension, within-MIRT models were estimated (mPCM and mGPCM). Thus, this model fit comparison allowed us to test the dimensionality of the data (uPCM vs. mPCM and uGPCM vs. mGPCM) and whether equal item discriminations could be assumed (uPCM vs. uGPCM and mPCM vs. mGPCM). These models were analyzed using the mirt package (Chalmers, 2012) in R (R Core Team, 2013). mirt uses the Metropolis-Hastings Robbins-Monro algorithm for model estimation.

Then, a between-MIRT model and a within-MIRT model based on the mPCM that contained six dimensions (one for each RIASEC factor) were compared regarding their fit. In the between-MIRT model, each RIASEC factor was modeled to be assessed by the items it was designed to measure, with no items overlapping between dimensions. The within-MIRT model also contained six dimensions, one for each RIASEC factor, but here, items with substantial

---

**Table 1.** Means, Standard Deviations, and Item Response Theory Reliability Estimates for the AIST-R Scales.

<table>
<thead>
<tr>
<th>Interest Type</th>
<th>M (SD)</th>
<th>uPCM</th>
<th>mPCM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Realistic</td>
<td>24.51 (8.59)</td>
<td>0.86</td>
<td>0.85</td>
</tr>
<tr>
<td>Investigative</td>
<td>30.37 (7.93)</td>
<td>0.84</td>
<td>0.83</td>
</tr>
<tr>
<td>Artistic</td>
<td>30.28 (8.55)</td>
<td>0.84</td>
<td>0.83</td>
</tr>
<tr>
<td>Social</td>
<td>32.77 (7.91)</td>
<td>0.88</td>
<td>0.87</td>
</tr>
<tr>
<td>Enterprising</td>
<td>37.02 (7.64)</td>
<td>0.86</td>
<td>0.86</td>
</tr>
<tr>
<td>Conventional</td>
<td>30.86 (7.15)</td>
<td>0.84</td>
<td>0.84</td>
</tr>
</tbody>
</table>

Note. uPCM = unidimensional Partial Credit Model; mPCM = multidimensional Partial Credit Model (between-MIRT model).
cross-loadings in the EFA were modeled to measure more than one dimension. These two models were estimated using ConQuest (Wu, Adams, Wilson, & Haldane, 2007) because latent correlations (see below) could only be obtained in ConQuest. ConQuest applies marginal maximum likelihood estimation. The identification constraint was set on the latent trait parameters (i.e., they have a normal distribution), whereas item parameters were estimated freely.

Two information criteria, namely, Akaike’s Information Criterion (AIC; Akaike, 1973) and the Bayesian Information Criterion (BIC; Schwarz, 1978), were consulted to evaluate model fit. For the AIC and the BIC, the better-fitting model is the one with a lower AIC or BIC value, respectively.

**Comparison of Correlations**

Correlations between the RIASEC types were compared between the unidimensional models and the between-MIRT model. For the unidimensional models, correlations between latent traits could not be estimated directly as in the between-MIRT model. Thus, the two-step procedure described in Wang, Chen, and Cheng (2004) was applied. First, the Pearson product–moment correlation between the weighted likelihood estimates (WLEs) for two latent traits were computed. WLEs (Warm, 1989) are estimates for the person parameters in the framework of IRT, that is, they indicate a person’s standing on the latent trait. Second, the disattenuation formula proposed by Spearman (1904) was applied to correct the correlations for the unreliability of the measures. The reliability coefficients used here were the WLE person separation reliabilities computed in ConQuest for each model. This two-step procedure was used for all bivariate correlations between the RIASEC interest domains. In the between-MIRT model, correlations can be estimated directly and therefore do not contain measurement error due to unreliability as in the unidimensional approach. These latent correlations between the factors were compared with the attenuated and disattenuated correlations obtained from the UIRT models.

**Results**

First, results from the pre-analyses using EFA will be reported. Second, model fit comparisons between the unidimensional models and multidimensional models will be described. Third, correlations obtained from unidimensional models and the between-MIRT model will be compared.

**Pre-Analyses**

The EFA conducted using the pre-analyses sample indicated that all RIASEC factors except for Social contained at least one item that loaded substantially (> .30) on a second factor besides the one it was designed to measure. These items were modeled to measure a second dimension in addition to the one they were designed to measure in the within-MIRT models (see Table 2).

**Model Fit Comparisons**

Table 2 shows the model fit comparisons for each RIASEC factor. Model comparisons between the uPCM, uGPCM, mPCM, and mGPCM yielded the best fit for the mGPCM according to AIC and BIC for Realistic, Investigative, Artistic, and Enterprising. Thus, for these interest types, a multidimensional model that accounts for differing item discriminations was better able to represent the data than a unidimensional model. Note, however, that for Enterprising the difference in AIC and BIC between the uGPCM and the mGPCM was miniscule. For Social, no MIRT models were estimated because none of the items showed substantial cross-loadings on a second factor.
Here, the comparison between the uPCM and the uGPCM resulted in a better fit for the uGPCM. The uGPCM was also the best-fitting model for Conventional although the difference between uGPCM and mGPCM was again very small. Thus, the GPCM consistently yielded a better fit than the PCM for all interest types, indicating that the assumption of equal item discriminations made in the PCM did not hold. Concerning dimensionality, Realistic, Investigative, and Artistic were best represented using multiple dimensions, whereas Social appeared to be unidimensional.

Drawing on these results, the within-MIRT model including all six interest types was modeled with items loading on another dimensions besides the one they were designed to measure for Realistic, Investigative, Artistic, and Enterprising. The comparison of model fit between the between-MIRT model and the within-MIRT model yielded an AIC (BIC) of 625,728 (627,370) for the former and 458,374 (459,639) for the latter. Thus, the within-MIRT model showed a better fit than the between-MIRT model, indicating that the within-MIRT model was more adequate to describe the data.

In addition, WLE person separation reliabilities for the RIASEC scales were compared between the unidimensional models and the between-MIRT model. As can be seen in Table 1, these were practically identical. Furthermore, fit analyses were also conducted at the item level using the weighted mean square computed by ConQuest. Five items exceeded the criteria for low-stake tests according to which MNSQ values between 0.70 and 1.30 are acceptable (Wright & Linacre, 1994). For the between-MIRT and within-MIRT models, fewer items (three and two,

Table 2. Model-Fit Comparison Between Unidimensional and Multidimensional Models.

<table>
<thead>
<tr>
<th>Interest Type</th>
<th>Model</th>
<th>Items on Second Dimension</th>
<th>Log-Likelihood</th>
<th>AIC</th>
<th>BIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Realistic</td>
<td>uPCM</td>
<td></td>
<td>−49896.66</td>
<td>99881.32</td>
<td>100158.23</td>
</tr>
<tr>
<td></td>
<td>uGPCM</td>
<td></td>
<td>−49145.62</td>
<td>98399.24</td>
<td>98739.08</td>
</tr>
<tr>
<td></td>
<td>mPCM</td>
<td>7, 19, 49, 55</td>
<td>−50570.91</td>
<td>101227.82</td>
<td>101498.43</td>
</tr>
<tr>
<td></td>
<td>mGPCM</td>
<td>7, 19, 49, 55</td>
<td>−4903.61</td>
<td>98177.22</td>
<td>98535.94</td>
</tr>
<tr>
<td>Investigative</td>
<td>uPCM</td>
<td></td>
<td>−54223.90</td>
<td>108543.80</td>
<td>108845.90</td>
</tr>
<tr>
<td></td>
<td>uGPCM</td>
<td></td>
<td>−53161.61</td>
<td>106439.20</td>
<td>106804.20</td>
</tr>
<tr>
<td></td>
<td>mPCM</td>
<td>38, 44</td>
<td>−54794.32</td>
<td>109682.60</td>
<td>109978.40</td>
</tr>
<tr>
<td></td>
<td>mGPCM</td>
<td>38, 44</td>
<td>−53126.35</td>
<td>106370.70</td>
<td>106742.00</td>
</tr>
<tr>
<td>Artistic</td>
<td>uPCM</td>
<td></td>
<td>−56746.00</td>
<td>113588.00</td>
<td>113890.10</td>
</tr>
<tr>
<td></td>
<td>uGPCM</td>
<td></td>
<td>−55610.48</td>
<td>111337.00</td>
<td>111702.00</td>
</tr>
<tr>
<td></td>
<td>mPCM</td>
<td>9, 15, 27, 39, 51</td>
<td>−57240.79</td>
<td>114575.60</td>
<td>114871.40</td>
</tr>
<tr>
<td></td>
<td>mGPCM</td>
<td>9, 15, 27, 39, 51</td>
<td>−54754.22</td>
<td>109632.40</td>
<td>110022.60</td>
</tr>
<tr>
<td>Social</td>
<td>uPCM</td>
<td></td>
<td>−50305.09</td>
<td>100714.18</td>
<td>101041.43</td>
</tr>
<tr>
<td></td>
<td>uGPCM</td>
<td></td>
<td>−49459.88</td>
<td>99043.76</td>
<td>99433.95</td>
</tr>
<tr>
<td>Enterprising</td>
<td>uPCM</td>
<td></td>
<td>−48812.71</td>
<td>97717.42</td>
<td>98006.91</td>
</tr>
<tr>
<td></td>
<td>uGPCM</td>
<td></td>
<td>−48336.67</td>
<td>96785.35</td>
<td>97137.77</td>
</tr>
<tr>
<td></td>
<td>mPCM</td>
<td>23</td>
<td>−48743.69</td>
<td>97577.37</td>
<td>97860.57</td>
</tr>
<tr>
<td></td>
<td>mGPCM</td>
<td>23</td>
<td>−48336.40</td>
<td>96784.80</td>
<td>97137.22</td>
</tr>
<tr>
<td>Conventional</td>
<td>uPCM</td>
<td></td>
<td>−53561.42</td>
<td>107216.80</td>
<td>107512.60</td>
</tr>
<tr>
<td></td>
<td>uGPCM</td>
<td></td>
<td>−53038.84</td>
<td>106191.70</td>
<td>106550.40</td>
</tr>
<tr>
<td></td>
<td>mPCM</td>
<td>48</td>
<td>−53641.42</td>
<td>107374.80</td>
<td>107664.30</td>
</tr>
<tr>
<td></td>
<td>mGPCM</td>
<td>48</td>
<td>−53041.46</td>
<td>106196.90</td>
<td>106555.60</td>
</tr>
</tbody>
</table>

Note. AIC and BIC values of the best-fitting model are depicted in italics. N for all models was 3,997. AIC = Akaike’s Information Criterion; BIC = Bayesian Information Criterion; uPCM = unidimensional Partial Credit Model; mPCM = multidimensional Partial Credit Model.
respectively) demonstrated an MNSQ below 0.70 or above 1.30. Detailed results on the item fit analyses can be obtained from the first author on request.

**Comparison of Correlations**

Attenuated and disattenuated correlations between latent traits using WLEs from the uPCMs1 are reported in Table 3. As would be expected, disattenuated correlations were generally higher than attenuated correlations. Overall, correlations confirm the RIASEC circumplex structure, for example, the adjacent traits Realistic and Investigative correlated higher than the opposite traits Realistic and Social ($r = .78$ and $r = −.24$ for disattenuated uPCM correlations).

Correlations between latent traits were estimated directly in the between-MIRT model. Table 3 contrasts these latent correlations with the attenuated and disattenuated correlations resulting from the separate uPCMs. Correlations obtained in the between-MIRT model were higher than uPCM correlations in most cases. However, for disattenuated correlations, this difference is reduced compared with the attenuated correlations. For example, the latent correlation between Realistic and Investigative is .83 (between-MIRT model), while it is .66 (disattenuated .78) for the uPCM.

**Discussion**

In this article, the unidimensional approach to analyzing interest inventories was extended by analyses in the framework of MIRT. Model fit comparisons indicated that for Realistic, Investigative, Artistic, and Enterprising, item responses were influenced by more than one interest type. Enterprising and Conventional only contained one item loading on a second dimension, whereas for Realistic, Investigative, and Artistic, two or more items loaded on a second dimension. This may have led to the very similar AIC and BIC values between the uGPCM and the mGPCM for Enterprising and Conventional, indicating that without this one item, unidimensionality may also exist for Enterprising. This illustrates that MIRT can be used as a tool to explore the underlying dimensionality of the data (Reckase, 1997; Yao & Schwarz, 2006). Despite the promising developments in the area of MIRT, it should not be overlooked that having items that measure more than one trait can still be considered problematic from a measurement perspective, especially if the intention was to construct unidimensional scales. If scales are shown to be unidimensional (as in our study for Social and Conventional), analyzing each scale separately is justified. Nevertheless, between-MIRT models can still be useful in this case because they provide latent estimates of the correlations between scales.

The model fit comparison between the between-MIRT model and the within-MIRT model showed that the within-MIRT model, which allowed several items on Realistic, Investigative, Artistic, and Enterprising to load on secondary dimensions, was more adequate to describe the

---

**Table 3. Correlations Between RIASEC Factors for Separate uPCMs and Between-MIRT Model.**

<table>
<thead>
<tr>
<th></th>
<th>Realistic</th>
<th>Investigative</th>
<th>Artistic</th>
<th>Social</th>
<th>Enterprising</th>
<th>Conventional</th>
</tr>
</thead>
<tbody>
<tr>
<td>Realistic</td>
<td>—</td>
<td>.83</td>
<td>−.22</td>
<td>−.28</td>
<td>−.25</td>
<td>.05</td>
</tr>
<tr>
<td>Investigative</td>
<td>.66; .78</td>
<td>—</td>
<td>−.04</td>
<td>−.14</td>
<td>−.20</td>
<td>.00</td>
</tr>
<tr>
<td>Artistic</td>
<td>−.16; −.19</td>
<td>−.00; −.00</td>
<td>—</td>
<td>.56</td>
<td>.24</td>
<td>−.00</td>
</tr>
<tr>
<td>Social</td>
<td>−.21; −.24</td>
<td>−.11; −.13</td>
<td>.45; .52</td>
<td>—</td>
<td>.47</td>
<td>.17</td>
</tr>
<tr>
<td>Enterprising</td>
<td>−.19; −.22</td>
<td>−.15; −.18</td>
<td>.20; .24</td>
<td>.37; .43</td>
<td>—</td>
<td>.53</td>
</tr>
<tr>
<td>Conventional</td>
<td>.05; .06</td>
<td>.01; .02</td>
<td>.00; .00</td>
<td>.14; .16</td>
<td>.44; .52</td>
<td>—</td>
</tr>
</tbody>
</table>

Note. Above diagonal: latent correlations for the between-MIRT model, below diagonal: attenuated and disattenuated correlations from uPCMs. uPCM = unidimensional Partial Credit Model.
data. This confirms the results obtained from separate analyses of the interest types that responses to the AIST-R’s items on these subscales are influenced not only by the interest dimension they were designed to measure but also by a second or third interest dimension. These results also coincide with a recent analysis of the AIST-R regarding differential item functioning (DIF; Wetzel & Hell, 2013), which found that, for some items showing gender-DIF, the differential functioning of the items appears to be related to a second interest dimension.

WLE person separation reliabilities were practically identical between the unidimensional and multidimensional (between) approach. Thus, in our study, no differences in measurement precision were found between these two approaches. However, other studies (e.g., Cheng, Wang, & Ho, 2009) showed that MIRT models usually yield higher measurement precision. Item fit was problematic for several of the AIST-R’s items, especially in the unidimensional PCMs. In the MIRT models, fewer of these items demonstrated problematic item fit.

When item content overlaps between traits, multidimensional models may be more appropriate than unidimensional models to analyze response data because they explicitly take these underlying relationships into account. Thus, by using MIRT models, we can model the theoretical assumptions made by Holland concerning the circumplex structure of the RIASEC factors. This is also reflected by the correlations between the interest traits that confirmed Holland’s (1997) predictions. In MIRT models, these can be estimated directly and are not attenuated with measurement error due to unreliability. Consequently, correlations were higher compared to the attenuated correlations obtained in the unidimensional models. However, they were often also higher than the disattenuated unidimensional correlations, which indicates that the multidimensional models have a stronger capability of capturing the real relationship between the traits than the unidimensional models.

As noted in the introduction, MIRT and factor analysis models share a number of properties although they follow different goals. One of the benefits of MIRT over factor analysis is that it addresses the complexity in the interaction between persons and items by modeling variation in item properties such as difficulty and discrimination. On the other hand, factor analysis usually analyzes a correlation matrix and thus ignores other item characteristics. Thus, MIRT can provide a better understanding of the items.

**Implications of MIRT Models to Vocational Interest Measurement**

There are several reasons why we think vocational interest measurement could benefit from analyses in the framework of MIRT. From a psychometric perspective, it is important to consider multidimensionality in the data. If the measured latent trait is multidimensional, the fundamental assumption of local independence made by UIRT models may be violated, resulting in biased item parameter estimates (Finch, 2010) and person trait estimates (Ackerman, 1992). In contrast, MIRT makes use of the relationships between dimensions, which enables it to produce more accurate item and person parameter estimates (Cheng et al., 2009). Furthermore, multidimensionality can cause DIF (Ackerman, 1992). The multidimensional approach allows a direct estimation of the correlations between latent traits (without measurement error; Wang et al., 2004), which is preferable to correlating trait estimates derived from unidimensional models. Kirisci, Hsu, and Yu (2001) recommend the use of MIRT models when correlations between latent traits vary as is the case in many RIASEC-based interest inventories.

From a substantive perspective, MIRT analyses may deepen the understanding of vocational interest data compared with traditional methods. In MIRT, items provide information about multiple latent traits simultaneously. This additional information that results from linking item responses to multiple traits can have diagnostic value. For example, diagnostic profiles of participants’ latent trait levels can be estimated more precisely when relationships between traits are taken into account (Yao & Schwarz, 2006). This is especially important in the context of
counseling because using unidimensional models to model multidimensional data can lead to different inferences concerning trait levels compared to when the appropriate multidimensional model is applied (Walker & Beretvas, 2003). MIRT allows modeling RIASEC data in a way that is more consistent with the theoretical assumptions of the model because relationships between dimensions are taken into account in the estimation of correlations and trait levels. However, more research on the application of MIRT to vocational interest measurement is clearly needed.

Limitations
In this study, both samples consisted of students. It is not clear to what extent the results can be generalized to persons from other populations. Moreover, MIRT analyses were only conducted on one interest inventory, namely, the AIST-R. Future research could analyze other widely used interest inventories regarding the multidimensionality of their data. Furthermore, this study mainly addressed two aspects, namely, model fit comparisons and comparisons of correlations. Further research could compare the trait estimates obtained from different models, although a simulation study would be needed to be able to judge which trait estimates are more accurate. MIRT models can also be extended to include background variables (e.g., gender, cognitive abilities), which may yield further information on relationships between trait variables and background characteristics. This was not implemented in this study as it would have added complexity to the models and their presentation.

Conclusion
This article attempted to demonstrate the benefits of MIRT analyses to vocational interest measurement. MIRT analyses (when applied to multidimensional data) have several advantages, for example, greater precision in the estimation of item and person parameters, direct estimation of the correlations between dimensions, and a more accurate representation of theoretical assumptions. From a practical perspective, it is especially important to apply multidimensional models to multidimensional data to draw correct inferences concerning test-takers’ trait levels.

Declaration of Conflicting Interests
The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding
The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was supported by the German Federal Ministry of Education and Research and by the European Social Fund of the European Union (grant agreement no.: 01FP0930).

Note
1. For the correlations to be comparable between the unidimensional models and the between-Multi-dimensional Item Response Theory models, all estimations were based on the Partial Credit Model, despite the Generalized Partial Credit Model showing a better fit for four interest types.

References


