

Medical University admission test: a confirmatory factor analysis of the results

Marion Luschin-Ebengreuth · Hans P. Dimai · Daniel Ithaler · Heide M. Neges · Gilbert Reibnegger

Received: 1 June 2015 / Accepted: 19 November 2015 / Published online: 10 December 2015
© Springer-Verlag Wien 2015

Summary

Background The *Graz Admission Test* has been applied since the academic year 2006/2007. The validity of the *Test* was demonstrated by a significant improvement of study success and a significant reduction of dropout rate. The purpose of this study was a detailed analysis of the internal correlation structure of the various components of the *Graz Admission Test*. In particular, the question investigated was whether or not the various test parts constitute a suitable construct which might be designated as “Basic Knowledge in Natural Science.”

Methods This study is an observational investigation, analyzing the results of the *Graz Admission Test* for the study of human medicine and dentistry. A total of 4741 applicants were included in the analysis. Principal component factor analysis (PCFA) as well as techniques from structural equation modeling, specifically confirmatory

factor analysis (CFA), were employed to detect potential underlying latent variables governing the behavior of the measured variables.

Results PCFA showed good clustering of the science test parts, including also text comprehension. A putative latent variable “Basic Knowledge in Natural Science,” investigated by CFA, was indeed shown to govern the response behavior of the applicants in biology, chemistry, physics, and mathematics as well as text comprehension. The analysis of the correlation structure of the various test parts confirmed that the science test parts together with text comprehension constitute a satisfactory instrument for measuring a latent construct variable “Basic Knowledge in Natural Science.”

Conclusions The present results suggest the fundamental importance of basic science knowledge for results obtained in the framework of the admission process for medical universities.

G. Reibnegger (✉)
Institute of Physiological Chemistry,
Center for Physiological Medicine, Medical University Graz,
Harrachgasse 21/2,
8010 Graz, Austria
e-mail: gilbert.reibnegger@medunigraz.at

M. Luschin-Ebengreuth · H. P. Dimai
Medical University Graz,
Auenbruggerplatz 2/4,
8010 Graz, Austria
e-mail: marion.habersack@medunigraz.at

H. P. Dimai
e-mail: hans.dimai@medunigraz.at

D. Ithaler · H. M. Neges
Medical University Graz,
Harrachgasse 21/6,
8010 Graz, Austria
e-mail: daniel.ithaler@medunigraz.at

H. M. Neges
e-mail: heide.neges@medunigraz.at

Keywords Medical university admission test · Confirmatory factor analysis · Latent construct variable · Basic Knowledge in Natural Science

Introduction

The scientific discussion on admission processes for medical universities addresses many different aspects. Among other things, construction, adaptation, and weighting of individual components of various admission processes as well as predictive validity (with regards to the future performance of the applicants) [1–5]; the necessity to test also interpersonal/personal skills—besides cognitive abilities [6–9], what results are obtained by whom, and of course, fairness in its broader sense [10, 11] are discussed. Depending on which countries the admission processes for medical college are performed in, this is, among other things, dominated by ethnicity

[12], gender equality [11, 13] or, generally, the necessity to educate a student population that corresponds to the societal requirements, and that is as diverse as possible [14].

While in many countries the development and application of medical college admission tests can look back on a long history, the three medical universities in Austria have been confronted with admission restrictions, the resulting selection of students and the conception of selection processes only since the academic year 2005/2006. As described earlier [15], at the Medical University of Graz an admission test (in the following designated as “Graz admission test”) was developed based mainly on secondary school knowledge in natural science disciplines (biology, chemistry, mathematics, and physics), amended by a text comprehension component and, since the academic year 2010/2011 by a situational judgment test (SJT) [16]. The validity of the Graz admission test was demonstrated by a significant improvement of study success [15] as well as by demonstration of a strongly significant reduction of the premature study dropout rate of the successful applicants [17, 18]. The purpose of this study is a detailed analysis of the internal correlation structure of the various components of the Graz admission test over three academic years (2010/2011 to 2012/2013). In particular, the question is investigated whether or not the various test parts constitute a suitable construct (a latent variable) which might be designated as “Basic Knowledge in Natural Science”. As a side topic, the study is intended to clarify to what extent the additional component SJT overlaps with the construct “Basic Knowledge in Natural Science” or constitutes a different test dimension in its own right. Clearly, the latter outcome appears much more probable in advance; nevertheless, it is interesting whether or not the achievements of the candidates in scientific test parts correlate with their achievements in the SJT part.

Methods

This study is an observational investigation analyzing the results of the Graz admission test for the study of human medicine and dentistry, obtained in the academic years 2010/2011, 2011/2012, and 2012/2013. Over the 3 years, there were 4741 applicants, all of whom are included in the study.

Context

The present study was conducted at the Medical University of Graz, one of three public medical universities in Austria. According to negotiations between the Austrian Federal Ministry of Science and the Medical University of Graz, the number of available study places for human medicine and dentistry for each academic year is limited to 360 (336 for human medicine and 24 for dentistry). These places are given to those applicants who rank high-

est according to the admission test. (Importantly, besides the general requirements for entering a university study such as having completed secondary education, ranking according to the admission test is the exclusive criterion for admission). If, in a given year, more than one applicant is ranked equally at the threshold of rank 360, all of them are admitted. Therefore, the yearly numbers of successful applicants are slightly higher than 360, ranging between 365 and 366 in the years under scrutiny.

Graz admission test

The Graz admission test, as applied in the academic years investigated, is divided into three sections containing in total six disciplines:

- i) Knowledge Test, based on secondary school level knowledge in
 - a. Biology
 - b. Chemistry
 - c. Physics
 - d. Mathematics
- ii) Text comprehension test
- iii) Situational Judgment Test (SJT).

In the following sections, these six different test disciplines (biology, chemistry, physics, mathematics, text comprehension, and SJT) are designated “test parts”. All test parts are uniformly done in the format of a written multiple choice test. Specifically, for each test item there are five distractors, one of which represents the correct answer. For correct answers, the applicants receive positive scores ranging from 2 to 5 in dependence on the test part; for wrong answers a negative score of -1 is counted. Leaving out an item without choosing one of the five distractors leads to a score of 0 for this item. This procedure was chosen for two reasons: first, in order to prevent pure guessing; second, because for future medical students the ability to realistically judge one’s own knowledge is deemed to be quite desirable.

For the determination of the ranking of the applicants—and hence, for the decision whether or not an applicant is admitted—the scores for each item are summed up to give a total score. Due to the different number of items in the various test parts, there is an implicit weight given to each of these parts.

Statistical analyses

For the present study, the total ranking and the admission decision are not of primary relevance; rather, here we are interested in the mutual correlations of the test parts as well as in revealing the latent structure determining the applicant’s answering behavior. Therefore, and to yield comparable results for the different test parts, we calculate “relative scores” for each test part using the following formula:

$$\text{relativescore} = \frac{\text{score} - \text{minimum}}{\text{maximum} - \text{minimum}}$$

Here, “score” is the absolute score of an applicant in a chosen test part, “minimum” represents the worst possible case of hypothetically answering all items of a test part wrongly, and “maximum” denotes the best possible case of hypothetically answering all items of a test part correctly. To give an example, suppose an applicant with a biology score of 45. In the respective admission test, suppose there are 90 biology items with possible scores of $-1/0/+2$, if the answer was false/no answer/correct. In this case, minimum = -90 and maximum = 180 . The applicant thus has a

$$\text{relative score} = \frac{45 - (-90)}{180 - (-90)} = \frac{135}{270} = 0.50.$$

Computing relative scores this way ensures that they can range from 0.0 (all items of a test part falsely answered) to 1.0 (all items of a test part correctly answered).

In order to detect potential underlying latent variables governing the behavior of the measured variables, principal component factor analyses (PCFA) as well as techniques from structural equation modeling (SEM) are employed. Specifically, we use confirmatory factor analysis (CFA). The advantage of CFA over PCFA is that each test part is allowed to have its own, unique variance which is not shared by the other test parts, and on the other hand, the shared variance(s) of the test parts included in CFA is(are) reflected by the putative latent variable(s). By isolating the shared variance from the unique variances of the test parts, a better measure of the latent variable is obtained.

For fitting a CFA model, there are several estimation techniques available; we use the maximum likelihood technique, because it is relatively robust even in the case of some violation of normality of the data. In order to be on the safer side regarding a possible violation of normality, we also compute the standard errors of the model by a bootstrap procedure: we resample our observations with replacement and fit the model 1000 times; then, the distributions of the parameter estimates over these replications are used to estimate the standard errors.

CFA also offers a suite of procedures which help in finding potentially better fitting models in which, for example, the unique variances of several of the variables in the model may be correlated with each other. On the one hand, there are several goodness-of-fit statistics. We use the root-mean-square error of approximation (RMSEA) and the comparative fit index (CFI). RMSEA penalizes over-parameterized models; it is recommended that RMSEA be less than 0.08 for a reasonable fit and less than 0.05 for a good fit. CFI lies between 0 and 100% and indicates in percent how much the model does better than a model with a baseline model assuming that there are no relationships between the variables of the model (the test parts in our case). On the other hand, so-

called modification indices can be computed estimating the extent of improvement of fit if we made the model somewhat more complex. Thus, by using these tools, the final CFA model is obtained.

All statistical analyses are performed using STATA 13 software (StataCorp. LP, College Station, TX, USA). Notably, since version STATA 12, this software package contains also a graphical user interface allowing drawing and estimating models with the STATA SEM builder. The software and a valuable guide for the usage of these new capabilities are described in a recent monograph [19].

Ethics statement

The authors gathered anonymized data from a data set that is routinely collected about medical students' admission, dropout, and graduation dates and examination history, as required by the Austrian Federal Ministry of Science and Research. Because the data were anonymous and no data beyond those required by law were collected for this study, the Medical University of Graz's ethical approval committee did not require approval for this study.

Results

Basic data

For the academic years 2010/2011 to 2012/2013, Table 1 shows basic data on the admission tests at the Medical University of Graz. As already described in an earlier publication, there are consistently more women than men among the applicants; however, among the successful candidates, men consistently outweigh women, mostly because of the better performance of men in the natural science parts of the test [20].

Principal component factor analysis

In order to get a first impression of the correlation structure underlying the relative scores in the various test parts, exploratory PCFA is performed on all 4741 applicants, and separately according to sex and year of admission test. Table 2 shows the cumulative results for all 4741 applicants. The results in the various subgroups according to sex and year of admission test are very similar and the table also shows the ranges observed in the six subgroup analyses. Obviously, the overall pattern of the correlation structure seen in the total analysis of 4741 participants is found in a very similar manner also in the six subgroup analyses; so in the following we discuss the total findings as representative also for the subgroup analyses.

There is one dominant factor with an eigenvalue of about 3.00 explaining about half of the variance. The loadings of this factor on test parts biology, chemistry,

Table 1 Distributions of applicants as well as of successful applicants according to sex and nationality in three consecutive academic years

	Applicants from	Total	Women		Men		Successful ap- plicants from	Total	Women		Men	
			Number	%	Number	%			Number	%		
2010	Austria	1029	576	55.98	453	44.02	Austria	274	122	44.53	152	55.47
	European Union	298	149	50.00	149	50.00	European Union	74	37	50.00	37	50.00
	Other nationalities ^a	26	7	26.92	19	73.08	Other nationalities	18	4	22.22	14	77.78
	<i>All nationalities</i>	<i>1353</i>	<i>732</i>	<i>54.10</i>	<i>621</i>	<i>45.90</i>	<i>All nationalities</i>	<i>366</i>	<i>163</i>	<i>44.54</i>	<i>203</i>	<i>55.46</i>
2011	Austria	1190	690	57.98	500	42.02	Austria	281	142	50.53	139	49.47
	European Union	493	268	54.36	225	45.64	European Union	76	34	44.74	42	55.26
	Other nationalities	19	10	52.63	9	47.37	Other nationalities	9	5	55.56	4	44.44
	<i>All nationalities</i>	<i>1702</i>	<i>968</i>	<i>56.87</i>	<i>734</i>	<i>43.13</i>	<i>All nationalities</i>	<i>366</i>	<i>181</i>	<i>49.45</i>	<i>196</i>	<i>50.55</i>
2012	Austria	1164	661	56.79	503	43.21	Austria	284	126	44.37	158	55.63
	European Union	510	288	56.47	222	43.53	European Union	76	32	42.11	44	57.89
	Other nationalities	12	5	41.67	7	58.33	Other nationalities	5	2	40.00	3	60.00
	<i>All nationalities</i>	<i>1686</i>	<i>954</i>	<i>56.58</i>	<i>732</i>	<i>43.42</i>	<i>All nationalities</i>	<i>365</i>	<i>160</i>	<i>43.84</i>	<i>205</i>	<i>56.16</i>
2010–2012	Austria	3383	1927	56.96	1456	43.04	Austria	839	390	46.48	449	53.52
	European Union	1301	705	54.19	596	45.81	European Union	226	103	45.58	123	54.42
	Other nationalities	57	22	38.60	35	61.40	Other nationalities	32	11	34.38	21	65.63
	<i>All nationalities</i>	<i>4741</i>	<i>2654</i>	<i>55.98</i>	<i>2087</i>	<i>44.02</i>	<i>All nationalities</i>	<i>1097</i>	<i>504</i>	<i>45.94</i>	<i>593</i>	<i>54.06</i>

^aFor applicants from other nationalities, the percentages are given just for completeness despite the very small absolute numbers

Table 2 Results of a principle component factor analysis of the relative scores of the various test parts. Upper subtable: factor eigenvalues and proportion of explained variance. Lower subtable: factor loadings on the variables (i.e., the test parts)

Factor	Eigenvalue	Proportion	Cumulative
Factor 1	3.18 (2.80–3.39) ^a	0.53 (0.47–0.56)	0.53 (0.47–0.56)
Factor 2	1.07 (0.98–1.16)	0.18 (0.16–0.19)	0.71 (0.65–0.74)
Factor 3	0.65 (0.55–0.85)	0.11 (0.09–0.14)	0.82 (0.80–0.83)
Factor 4	0.54 (0.51–0.57)	0.09 (0.08–0.10)	0.91 (0.89–0.92)
Factor 5	0.33 (0.30–0.41)	0.06 (0.05–0.07)	0.96 (0.96–0.97)
Factor 6	0.23 (0.20–0.27)	0.04 (0.03–0.05)	1.00
Test part	Factor 1	Factor 2 ^b	Alpha ^c
Biology	0.83 (0.75–0.86)	–0.14 (–0.40–0.14)	0.76 (0.71–0.79)
Chemistry	0.86 (0.79–0.89)	–0.19 (–0.46–0.21)	0.75 (0.70–0.78)
Physics	0.83 (0.76–0.84)	–0.23 (–0.29–0.13)	0.76 (0.71–0.80)
Mathematics	0.73 (0.59–0.77)	–0.07 (–0.13–0.41)	0.79 (0.76–0.82)
Text comprehension	0.68 (0.64–0.71)	0.40 (0.33–0.48)	0.79 (0.72–0.84)
SJT	0.27 (0.19–0.43)	0.89 (0.68–0.88)	0.85 (0.78–0.87)

^aShown are the results of an analysis of all 4741 applicants as well as, in parentheses, the minimum and maximum values obtained by subgroup analyses regarding sex and year of admission (there are $2 \times 3 = 6$ subgroup analyses)

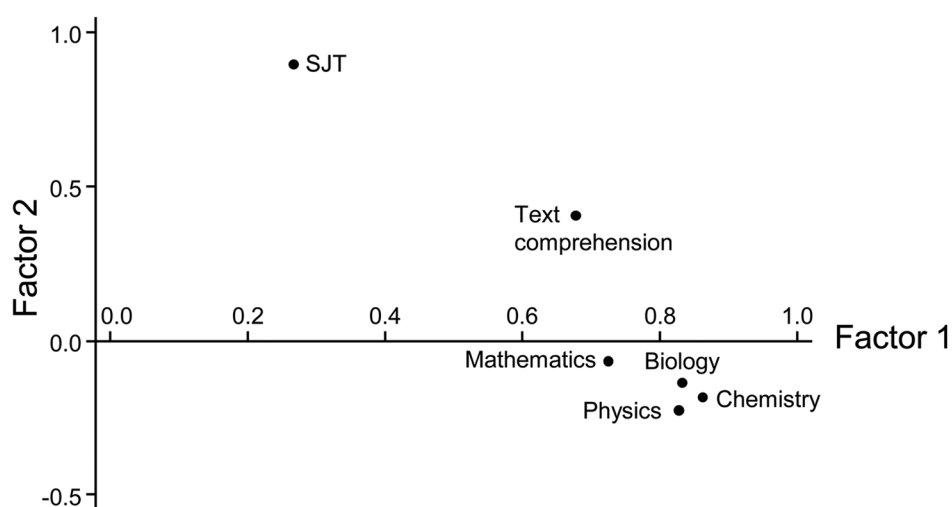
^bIn one subgroup (2012, men) only the first factor was above 1.00; the second factor was below 1.00 and hence, no loadings were computed for a second factor in this subgroup analysis. Therefore, the loading values of the case with all 4741 subjects do not necessarily fall into the ranges observed in the subgroup analyses

^cReliabilities: Total reliability = 0.82 (0.77–0.84). If a test part is removed from the model, the reliability of the new model is changed to the respective value in the Alpha column. In the total analysis as well as in every single subgroup analysis, SJT is the only test part the removal of which leads to an increase in reliability; removal of any of the other test parts leads to decrease in total reliability

and physics are uniformly high; for mathematics and text comprehension the loadings are somewhat smaller but still quite substantial. None of these test parts, on the other hand, shows substantial loadings on the second factor, which, with an eigenvalue of slightly above 1.00, is only marginally greater than unity and uniformly loads predominantly on the test part SJT. All other factors are

associated with eigenvalues below 1.00. Figure 1 shows a representative factor loading plot for the total analysis of 4741 subjects: from this plot it is obvious that the four science test parts group closely together with high loadings on the first factor and very small loadings on the second; SJT loads highly only on the second factor, and text comprehension shows moderate loadings on both factors but

Fig. 1 Factor loadings on the different test parts. (Results of a principal component factor analysis of the relative scores of the various test parts)



resembles more closely the science test parts group than SJT.

From this analysis we arrive at the picture that there is one significant latent variable controlling the answer behavior of the applicants in the four science disciplines and—albeit somewhat weaker—in text comprehension. As this latent variable apparently governs the response pattern to the natural science test parts, for convenience we name it “Basic Knowledge in Natural Science.” The SJT part obviously does not reflect this putative latent variable; rather, SJT measures a completely different quality of the applicants. This impression is additionally supported by an analysis of the reliabilities of the total model and the contributions of the various test parts to this overall reliability: as shown in the lower part of Table 2, reliability of the two factor-PCFA model would increase when SJT would be removed (in fact, only one significant factor would be retained in that case), but decrease on removal of any of the other test parts.

Confirmatory factor analysis

In order to get an even better view of the structure governing the response behavior of the applicants, CFA is employed. For this analysis, we are interested in how strongly the putative latent factor “Basic Knowledge in Natural Science” accounts for the results in the test parts biology, chemistry, physics, mathematics, and text comprehension. Based on the results of the PCFA, the test part SJT is not included into the CFA model.

The CFA model resulting after few steps of iterative improvement by using the tools of goodness-of-fit statistics and modification indices is depicted in Table 3 and Fig. 2. While in Fig. 2 the variable loadings, the variances and the covariance are shown, Table 3 also contains the standard errors and the 95% confidence intervals, as obtained by a bootstrapping procedure using 1000 replicate computer experiments. (Notably, standard errors obtained without bootstrapping deviate less than 0.5% from the bootstrapping values; thus, the deviation from

Table 3 Results of confirmatory factor analysis on the latent variable “Basic Knowledge in Natural Science.” “Basic Knowledge in Natural Science” (shortly denoted as “Science” in the figure). A bootstrap analysis is done estimating the standard errors from 1000 replicate model estimations^a. All applicants ($N=4741$) are included

Variable loadings	Coefficient	SE	<i>P</i>	95% confidence interval
Biology	0.681	0.010	<0.0001	0.660–0.701
Chemistry	0.756	0.010	<0.0001	0.737–0.775
Physics	0.896	0.009	<0.0001	0.879–0.913
Mathematics	0.646	0.010	<0.0001	0.626–0.666
Text comprehension	0.669	0.013	<0.0001	0.643–0.694
<i>Variances^d</i>				
ϵ .biol.	0.537	0.014	– ^b	0.510–0.565
ϵ .chem.	0.428	0.015	–	0.400–0.458
ϵ .phys.	0.197	0.015	–	0.169–0.229
ϵ .math.	0.583	0.013	–	0.557–0.610
ϵ .text comp.	0.553	0.017	–	0.520–0.588
Latent factor ^c	1.000			
<i>Covariances^d</i>				
ϵ .biol.— ϵ .chem.	0.520	0.015	<0.0001	0.490–0.549
ϵ .chem.— ϵ .text comp.	–0.150	0.020	<0.0001	–0.189–0.112
ϵ .phys.— ϵ .text comp.	–0.620	0.057	<0.0001	–0.733–0.508

^aIn fact, the bootstrap results deviate only minimally (in all cases less than 0.5%) from the calculation assuming normality

^bBecause of a boundary problem (a variance can never be smaller than zero), STATA does not report *P* values for variances of the error terms. However, the narrow 95% confidence intervals demonstrate that the error variances all are highly significant

^cThe variance of the latent factor “Basic Knowledge in Natural Science” is by definition 1.000 because of using standardized variables

^dError terms are denoted by “ ϵ ”

normality of our data can be neglected.) In Table 3 as well as in Fig. 2, standardized results are reported. In addition to Table 3 reporting detailed results on the analysis including all applicants, Fig. 2 also shows results for separate analyses on female as well as male applicants.

Fig. 2 Results of confirmatory factor analysis on the latent variable “Basic Knowledge in Natural Science.” (Results of a confirmatory factor analysis of the relative scores of the science test parts. The latent factor “Basic Knowledge in Natural Science” is abbreviated to “Science.” More details and the meaning of the numbers are explained in Table 3. In addition to Table 3 which shows only the detailed results for the analysis including all applicants, in the figure there are also shown the results of separate analyses using either females or males. Numbers in the figure are reported in the order “all applicants/females only/males only (abbreviated as “all/f/m”))”

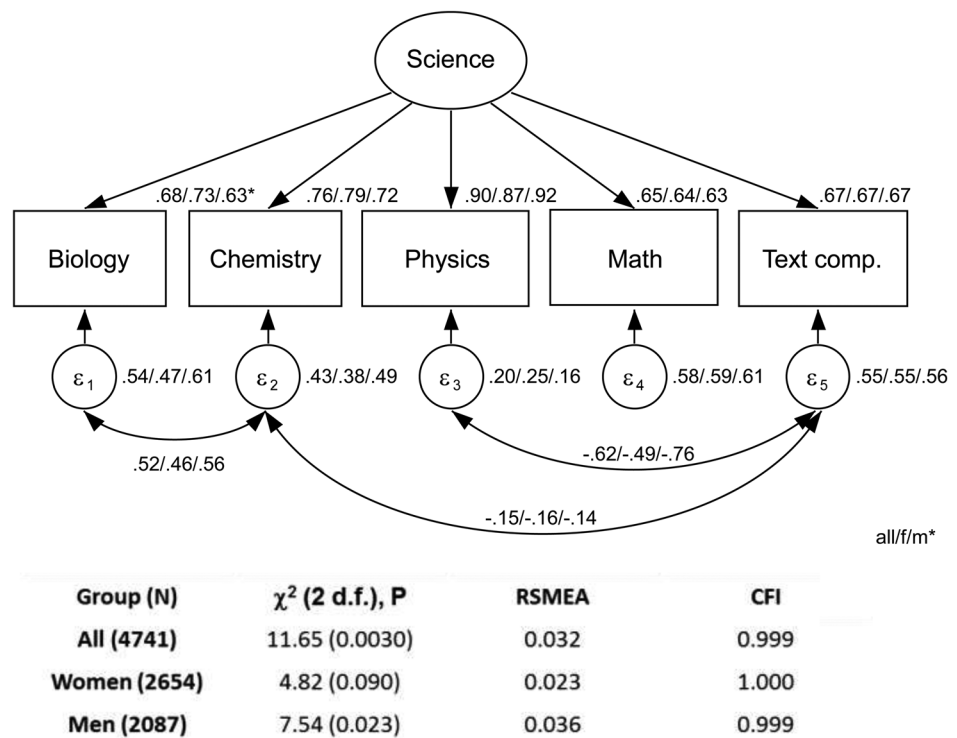


Figure 2 also contains the goodness-of-fit statistics of the model: while the chi-squares are still slightly significant for all applicants and for men, both RSMEA and CFI values are excellent, indicating that the model is well-suited to describe the structure underlying our data. In fact, the chi-square statistics are strongly influenced by the high number of cases investigated, so even minute model deviations are significant.

The results indicate that the putative latent variable “Basic Knowledge in Natural Science” indeed governs the response behavior of the applicants in biology, chemistry, physics, and mathematics as well as text comprehension: all factor loadings (i.e., regression coefficients) are substantial and differ highly significantly from zero (Table 3). The strongest loading is observed for physics, followed by chemistry. Biology, text comprehension, and mathematics show somewhat weaker but still substantial loadings. Moreover, each of the five variables still retains some unique variance which is not represented by the common latent variable. Also these unique variances are significantly different from zero. Finally, in order to get this excellently fitting model, following the respective modification indices we allow covariance between the error terms of biology and chemistry, chemistry and text comprehension, and physics and text comprehension: these additional flexibilities have great impacts on improvement of the goodness-of-fit statistics and also strongly improved the overall fit which is indicated by the chi-square statistics.

Inspection of Fig. 2 shows that there are small, however notable, deviations between the results observed for all applicants versus females and males separately: while for mathematics and text comprehension the factor loadings are practically the same for women and men, we find

higher loadings for biology and chemistry for women but the opposite for physics. Consequently, in women test results in biology and chemistry are more strongly governed by the latent variable, and hence, less individual variances remain for these test parts in women, while the opposite is true in men. In physics, the results are just reversed.

Discussion

In this study, the focus is directed onto the correlation structure of the test response data: a preliminary PCFA shows that the responses to the natural science disciplines biology, chemistry, physics, and mathematics exhibit a strong mutual correlation; text comprehension shows a slightly smaller degree of correlation, and SJT is, not unexpectedly, essentially unrelated to the science test parts.

The PCFA results stimulate the question whether one could construct a latent variable, designated “Basic Knowledge in Natural Science,” which might govern the response behavior of applicants with respect to the science disciplines as well as to text comprehension. SJT is deliberately not included into this analysis because of its lacking correlation. A CFA confirms this expectation and in fact, a corresponding model with one such latent variable fits the data sample very well (CFI=0.999; RSMEA=0.032). The influence of the latent variable on the response behavior in text comprehension can, among other things, be ascribed to the fact that the texts used aim at the comprehension of biological, chemical, physical, and mathematical processes. It is remarkable that the observed (measured) variables are linked in a

certain way: the natural science-oriented subsections as well as the subsection text comprehension, evaluate the cognitive abilities of the applicants in the natural science subjects. Similar results are noted by Collin et al. [21]. The authors conclude that, due to the focus on cognition-based aptitude in the framework of admission processes for medical colleges, the underlying structure likewise has an intensive reference to cognitive ability. Situational Judgment Inventories, as Oswald et al. [22] remark, measure constructs that are not exclusively identical with cognitive ability. Especially in the framework of a “typical performance test” (among other things, SJT with behavioral tendency response instructions), in contrast to “maximal performance tests” (among other things, knowledge test), lower cognitive correlates are to be expected [23–25].

Interestingly, in the CFA analyses we observed slight but notable differences in the variable loadings as well as in the individual variances of the variables (i.e., the test parts) between women and men, whereby the results were qualitatively different for biology and chemistry versus physics (for mathematics and text comprehension the sex effect was negligible). Whether or not these differences might contribute to an explanation for the consistently observed “gender gap” (men always achieved somewhat better results in the basic science test parts; see also [20]) remains unclear. At least we know [20] that in physics the gender gap is particularly strong; so the deviating behavior of the variable “physics” in our CFA models might be partly understandable.

Conclusions

The findings of the present study identify or confirm, respectively, a latent variable, “Basic Knowledge in Natural Science,” as an essential influencing factor for the response behavior of applicants in the natural science test parts of the Graz admission test. Finally, it has to be noted that the present results not only suggest the fundamental importance of basic science knowledge for results obtained in the framework of the admission process for medical college, but that furthermore, a basic science knowledge gap between men and women can be assumed—when the less good test results in female applicants are considered.

Authors’ contributions

Marion Luschin-Ebengreuth has made substantial contributions to conception and design, drafted the article, and revised it critically.

Hans P. Dimai participated in the acquisition of data and helped to draft the manuscript.

Daniel Ithaler has made substantial contributions to the analysis of data and revised the manuscript critically.

Heide M. Neges participated in the acquisition of data and revised the manuscript critically.

Gilbert Reibnegger has made substantial contributions to conception and design and performed the statistical analysis.

All authors read and approved the final manuscript.

Conflict of interest

M. Luschin-Ebengreuth, H. P. Dimai, D. Ithaler, H. M. Neges, and G. Reibnegger declare that there are no actual or potential conflicts of interest in relation to this article.

References

1. Shulruf B, Poole P, Wang GY, Rudland J, Wilkinson T. How well do selection tools predict performance later in a medical programme? *Adv Health Sci Educ Theory Pract*. 2012;17(5):615–26.
2. Lumsden MA, Bore M, Millar K, Jack R, Powis D. Assessment of personal qualities in relation to admission to medical school. *Med Educ*. 2005;39(3):258–65.
3. Hissbach JC, Klusmann D, Hampe W. Dimensionality and predictive validity of the HAM-Nat, a test of natural sciences for medical school admission. *BMC Med Educ*. 2011;11:83.
4. Siu E, Reiter HI. Overview: what’s worked and what hasn’t as a guide towards predictive admissions tool development. *Adv Health Sci Educ Theory Pract*. 2009;14(5):759–75.
5. Hurwitz S, Kelly B, Powis D, Smyth R, Lewin T. The desirable qualities of future doctors—a study of medical student perceptions. *Med Teach*. 2013;35(7):e1332–9.
6. Albanese MA, Snow MH, Skochelak SE, Huggett KN, Farrell PM. Assessing personal qualities in medical school admissions. *Acad Med*. 2003;78(3):313–21.
7. Wilson IG, Roberts C, Flynn EM, Griffin B. Only the best: medical student selection in Australia. *Med J Aust*. 2012;196(5):357.
8. Lievens F. Adjusting medical school admission: assessing interpersonal skills using situational judgement tests. *Med Educ*. 2013;47(2):182–9.
9. Oates K, Goulston K. How to select the doctors of the future. *Intern Med J*. 2012;42(4):364–9.
10. Emery JL, Bell JF, Vidal Rodeiro CL. The BioMedical admissions test for medical student selection: issues of fairness and bias. *Med Teach*. 2011;33(1):62–71.
11. Cuddy MM, Swanson DB, Clauser BE. A multilevel analysis of examinee gender and USMLE step 1 performance. *Acad Med*. 2008;83(10 Suppl):S58–62.
12. Tiffin PA, Dowell JS, McLachlan JC. Widening access to UK medical education for under-represented socioeconomic groups: modelling the impact of the UKCAT in the 2009 cohort. *BMJ*. 2012;344:e1805.
13. Case SM, Becker DF, Swanson DB. Performance of men and women on NBME Part I and Part II: the more things change.... *Acad Med*. 1993;68(10 suppl):S25–7.
14. Reiter HI, Lockyer J, Ziola B, Courneya CA, Eva K. Should efforts in favor of medical student diversity be focused during admissions or farther upstream? *Acad Med*. 2012;87(4):443–8.
15. Reibnegger G, Caluba HC, Ithaler D, Manhal S, Neges HM, Smolle J. Progress of medical students after open admission or admission based on knowledge tests. *Med Educ*. 2010;44(2):205–14.

16. Luschin-Ebengreuth M, Dimai HP, Ithaler D, Neges HM, Reibnegger G. Situational judgment test as an additional tool in a medical admission test: an observational investigation. *BMC Res Notes*. 2015;8:81.
17. Reibnegger G, Manhal S. Study dropout in medical education. In: Walsh K, editor *Oxford textbook of medical education*. Oxford; 2013. p. 398–496.
18. Reibnegger G, Caluba HC, Ithaler D, Manhal S, Neges HM, Smolle J. Dropout rates in medical students at one school before and after the installation of admission tests in Austria. *Acad Med*. 2011;86(8):1040–8.
19. Acock AC. *Discovering structural equation modeling using stata, revised edition*. Texas: Stata Press; 2013.
20. Habersack M, Dimai HP, Ithaler D, Reibnegger G. Time: an underestimated variable in minimizing the gender gap in medical college admission scores. *Wien Klin Wochenschr*. 2015;127(7–8):241–9.
21. Collin VT, Violato C, Hecker K. Aptitude, achievement and competence in medicine: a latent variable path model. *Adv Health Sci Educ Theory Pract*. 2009;14(3):355–66.
22. Oswald FL, Schmitt N, Kim BH, Ramsay LJ, Gillespie MA. Developing a biodata measure and situational judgment inventory as predictors of college student performance. *J Appl Psychol*. 2004;89(2):187–207.
23. McDaniel MA, Morgeson FP, Finnegan EB, Campion MA, Braverman EP. Use of situational judgment tests to predict job performance: a clarification of the literature. *J Appl Psychol*. 2001;86(4):730–40.
24. McDaniel MA, Hartman NS, Whetzel DL, Grubb WL. Situational judgment tests, response instructions, and validity: a meta-analysis. *Pers Psychol*. 2007;60(1):63–91.
25. Whetzel DL, McDaniel MA, Nguyen NT. Subgroup differences in situational judgment test performance: a meta-analysis. *Hum Perform*. 2008;21(3):291–309.