



OPEN ACCESS

EDITED BY

Karen Stegers-Jager,
Erasmus Medical Center, Netherlands

REVIEWED BY

Fotios Milienos,
Panteion University, Greece
Suzanne Fikrat-Wevers,
The Institute of Medical Education Research
Rotterdam (iMERR), Netherlands

*CORRESPONDENCE

Daniel Weppert
✉ daniel.weppert@itb-consulting.de

SPECIALTY SECTION

This article was submitted to
Assessment, Testing and Applied Measurement,
a section of the journal
Frontiers in Education

RECEIVED 21 November 2022

ACCEPTED 06 March 2023

PUBLISHED 29 March 2023

CITATION

Weppert D, Amelung D, Escher M, Troll L,
Kadmon M, Listunova L and Montasser J
(2023) The impact of preparatory activities on
the largest clinical aptitude test
for prospective medical students in Germany.
Front. Educ. 8:1104464.
doi: 10.3389/educ.2023.1104464

COPYRIGHT

© 2023 Weppert, Amelung, Escher, Troll,
Kadmon, Listunova and Montasser. This is an
open-access article distributed under the terms
of the [Creative Commons Attribution License
\(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction
in other forums is permitted, provided the
original author(s) and the copyright owner(s)
are credited and that the original publication in
this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted which
does not comply with these terms.

The impact of preparatory activities on the largest clinical aptitude test for prospective medical students in Germany

Daniel Weppert^{1*}, Dorothee Amelung², Malvin Escher²,
Leander Troll¹, Martina Kadmon³, Lena Listunova² and
Jana Montasser¹

¹Institute for Test Development and Talent Research, ITB Consulting GmbH, Bonn, Germany, ²Faculty of Medicine, University of Heidelberg, Heidelberg, Germany, ³Deanery, Faculty of Medicine, University of Augsburg, Augsburg, Germany

Background: Aptitude tests are widely used for selecting medical students. Although their validity has been well documented, aptitude tests are sometimes suspected to create unequal opportunity for candidate groups with lower socioeconomic status due to limited resources (i.e., time, money, support) for preparatory activities. This study aims to explore how preparatory activities and money spent on preparation affect the results of the German aptitude Test for Medical Studies (TMS).

Methods: A standardized questionnaire was administered to all medical school applicants who sat the TMS in 2018. Participants were asked about the amount of time and money spent on different preparatory activities (i.e., information booklet, books, computer, study groups, and fee-based training courses) and their level of motivation during preparation. Univariate and multivariate multiple regressions were used to examine the influence of these variables on the TMS test score and its subtests.

Results: $N = 7903$ participants completed the questionnaire. Only preparation with books and training courses were significantly associated with an increase in the TMS total score. Self-reported motivation during preparation had a larger effect on test scores than money spent on preparation. However, all effect sizes were rather small. At the subtest level, preparation with books was the only activity which was significantly associated with an improvement in all subtests. The complex field-specific subtests were less associated with preparation than the less complex subtests.

Conclusion: The findings indicate that motivation may be a more important predictor for success in the TMS than money spent on preparation. As books were the most effective and cost-efficient way of preparation, financial

investments for test preparation do not appear to yield significant advantages, which is an important prerequisite for equal opportunity. Using more field-specific subtests and cost-free online training opportunities could be useful in further improving equal opportunity.

KEYWORDS

aptitude test, medical school selection, preparatory activities, motivation, admissions, equal opportunities

Introduction

In general, medical schools seek to identify the best and most motivated candidates to recruit their students (Turner and Nicholson, 2011) and frequently use aptitude test results as one important criterion to this aim. Despite their wide use and documented validity (e.g., Patterson et al., 2016; Schult et al., 2019), aptitude tests have been criticized to potentially disadvantage some candidate groups over others. For example, aptitude tests have been suspected to create unequal opportunity for candidate groups who might have limited access to preparatory activities, for example due to limited time for dedicated study or high cost (Laurence et al., 2013; Gliatto et al., 2016; Kumar et al., 2018). While the impact of preparatory activities on general aptitude or achievement tests has been studied for some time, there is significantly less research on tests used specifically for medical student selection. This is especially true for the selection of medical students in Germany. We therefore seek to determine the effects of various preparatory activities on the results of the widely used German Test for Medical Studies (TMS) to identify potential sources of unequal opportunity between candidate groups.

In the German medical school selection process, the TMS is one of the most important selection criteria in addition to school-leaving grades and predicts academic performance over and above school-leaving grades (Kadmon and Kadmon, 2016). Approximately two-thirds of all available study places are affected by its score and an applicant's TMS result may constitute the decisive factor for obtaining a study place as it can outweigh less-than-optimal school grades. Consequently, its use in the selection process has resulted in more diversified student cohorts (cf. Kadmon et al., 2012). However, more diversity is not synonymous with equal opportunity. For example, in Germany and other countries such as the UK, private-sector companies offer prospective applicants expensive preparatory training courses promising them better test results and thus higher chances to get into medical school. Whether this is ethical with regard to equal opportunity is one question, another question is whether the promise can indeed be kept - does paying more money for training courses have an actual effect on test results and, therefore, lead to higher chances of being admitted? As results of an aptitude selection test should be driven by the relevant or desirable characteristics of a medical student (i.e., their true aptitude), factors that are likely to introduce systematic bias - such as money spent on training courses - should be inconsequential.

Several studies and meta-analyses (e.g., Messick and Jungeblut, 1981; Powers and Rock, 1999; Briggs, 2004, 2009) have found

positive effects of training courses on the results of general university admission tests like the SAT. However, these effects are usually small and considerably lower than promised by the training course providers. For example, Briggs (2004) reported improvements of 0.11 to 0.31 standard deviations on the verbal and math subscales after training courses. For admission tests used specifically for medical school applications, there are mixed results. According to an Australian study, spending more money on training courses appeared to improve participants' confidence in their test results for the Australian Undergraduate Medicine and Health Sciences Admission Test (UMAT), while their test results did not exceed those of participants who did not attend training courses (Wilkinson and Wilkinson, 2013). Similar results were obtained for the UK (Lambe et al., 2012) and the US (McGaghie et al., 2004). However, in another Australian study, training courses yielded a positive effect on UMAT test results (Laurence et al., 2013). To our knowledge, similar and up-to-date results are currently not available for the TMS in Germany apart from some experiments in the 1980s pointing to only a small degree of trainability of certain subtest items (Deter, 1982).

Since courses are not the only type of preparation for the TMS that a prospective applicant can choose from, the general question is whether it is possible to achieve good test results with more affordable means of preparation such as books, the computer, or, in case of the TMS, the complimentary information booklet. However, a higher affordability of certain preparatory activities can only contribute to equal opportunity if these less costly alternatives do not require test takers to invest more time in them for comparable results, as time may be viewed as a resource that not everyone has equal access to. Dedicated time for study has already been discussed as one limiting factor in lower SES applicants (Girotti et al., 2020) and preparation time as well as test results are often positively associated, with both linear (Lambe et al., 2012) and logarithmic relationships (Messick and Jungeblut, 1981; Hausknecht et al., 2007) being found. However, a wealth of psychological research has found that too much of a positive virtue can have an inverse effect - which has been described as the "Inverted-U effect" (e.g., Yerkes-Dodson law, Grant and Schwartz, 2011). It is therefore also conceivable that too much time spent on preparation for the TMS might have adverse effects on results. While this study seeks to evaluate the relative impact of a range of intended and less intended factors that affect the TMS results, including ability (high school grades), type of preparatory activities, investments, and self-reported motivation during preparation, not all of the less intended effects necessarily introduce undesirable bias. In general, effects of preparation do not necessarily have to adversely affect the validity

of a test if they reflect actual gains in the measured ability (e.g., Briggs, 2009; Arendasy et al., 2016). However, strong effects of preparatory activities can be detrimental to equal opportunity if not all participants have the same access to effective preparatory activities or the same time resources to prepare for the test.

Moreover, motivation is a factor that an aptitude test is not intended to measure directly, but usually has positive effects on test results, and may be assumed to be a desirable characteristic in a medical student (Kusurkar et al., 2013; Orsini et al., 2016). According to the expectancy-value theory of motivation (Wigfield and Eccles, 2000) an individual's persistence and success with a task results from the individual both ascribing high value to the task (intrinsic motivation) and being confident in his or her own ability to master it (which should be correlated with actual ability). Both have been found to positively predict learning choices and academic performance in longitudinal studies (Durik et al., 2006; Denissen et al., 2007). Additionally, motivation, especially if it can be considered intrinsic or autonomous (i.e., not induced by outside sources or incentives, but rather coming from the person itself), is increasingly acknowledged as an important factor to consider in the selection and teaching of medical students (Kusurkar et al., 2013; Orsini et al., 2016). Autonomous motivation has also been found to predict academic performance in and adherence to medical studies (Sobral, 2004). Therefore, even though aptitude tests in medical selection may not be intended to measure motivation, motivation may affect the results without limiting the predictive validity of the test.

In sum, this study seeks to explore the roles and relative impacts of preparation time and types, motivation, and financial investment on overall TMS performance. Moreover, as the TMS-subtests differ in the degree of complexity and abilities measured, another aim of this study is to examine whether the subtests are affected by these factors in different ways. As different aptitude tests vary internationally with regard to the specific cognitive abilities they are designed to measure (e.g., Mathew and Thomas, 2018), this might shed some light on the generalizability of our results to different kinds of cognitive aptitude tests. Thus, our research questions are as follows:

1. What is the relative impact of motivation, time, and financial investment into certain preparatory activities on the TMS results?
2. Are certain preparatory activities more advantageous than others with regard to the TMS test results?
3. Are certain TMS-subtests more robust to preparatory activities than others?

Materials and methods

Sample and procedure

We administered a paper-based questionnaire (Supplementary Figure 1) to all 10,433 participants of the 2018 TMS in all 48 test centers. Participation in the evaluation study was optional and took place before the actual test. TMS scores of all participants were gathered on the same day.

To match the TMS result with the answers in the voluntary evaluation questionnaire, we only included participants who provided their TMS-participant ID in the questionnaire and gave answers to all relevant questions. The TMS-participant ID was then used to match the TMS result as well as the participants' sex and age with the results of the voluntary evaluation. The final sample consisted of 7903 participants (75.8% of the overall cohort), of which 30.2% were male. On average, the participants were 19.20 years old ($SD = 1.97$) and had an average high school GPA of 1.68 ($SD = 0.46$).

Measures

Questions and response scales were based on previous studies (e.g., Deter, 1982; Bartussek et al., 1984; Trost et al., 1998; Briggs, 2009) and were developed in a consensus process between expert groups comprised of test developers and members of the medical faculty of the University of Heidelberg involved in the selection process.

The questionnaire was kept as short as possible due to limited time and to reduce participant burden. Questions were tested in two pilot studies before being used in the current study. Piloting was aimed at resolving issues with understandability and scaling. Questions and response scales are described in detail in the following paragraphs.

Predictors

Age: The participants' date of birth was collected as part of the TMS-registration process and was used to calculate the participants' age at the time of testing.

Sex: The participants' sex was collected as part of the TMS-registration process.

GPA: Participants were asked to enter their high school grade point average in a two-digit open response field (from 0.7 = "highest possible result" to 6.0 = "insufficient").

Motivation: We asked participants to indicate their degree of motivation during preparatory activities on a five-point rating scale ("My motivation during preparations was. . ." ranging from "very low" to "very high").

Time spent on preparation: Participants were asked to indicate the respective number of hours spent on a scale ranging from 0 to 80 hours in 10-hour-increments (i.e., 0, 1-10, 11-20, 21-30, 31-40, 41-50, 51-60, 61-70, 71-80) for each of the following preparatory activities:

- Cost-free complimentary information booklet provided by the TMS provider upon registration including an in-depth explanation of the type of tasks including some sample items.
- Books (e.g., books with original earlier versions of the TMS published for training purposes by the TMS provider).
- Commercially organized courses by third-party providers.
- Computer-based opportunities (e.g., forums, social media, exchange platforms).
- Learning in self-organized study groups with peers.

For data analyses, the mean value of the respective response option was used. For example, the option 31-40 h was treated as

35.5 h in the analyses. Responses were given within a three-digit open response field in case the number of hours exceeded 80 for any given preparatory activity. If there was a number of hours specified in the three-digit open response field, that amount of time was used for the analyses. In previous analyses of preparation activities for the TMS (Deter, 1982; Bartussek et al., 1984; Trost et al., 1998) participants were asked whether they had used the information booklet, books, or training courses for preparation. Evaluation questions for a similar test used in Switzerland, the EMS (Eignungstest für das Medizinstudium), additionally included time spent in study groups (e.g., Zentrum für Testentwicklung und Diagnostik, 2022). We included all of these preparation activities and additionally introduced computers as an option.

Preparation costs: We asked participants to indicate the amount of money spent on preparatory activities on a scale ranging from EUR 0 to 200 in EUR 20-increments (0, 1-20, 21-40, 41-60, 61-80, 81-100, etc.) supplemented by a four-digit open response field in case the amount of money spent exceeded EUR 200. Again, the mean value of the respective response option was used for the analyses. If there was an amount of EUR specified in the four-digit open response field, that amount was used for the analyses.

Outcome

Test results: The TMS in 2018 comprised nine subtests: Pattern Assignment (Muster Zuordnen; MUZ), Basic Medical and Scientific Understanding (Medizinisch-naturwissenschaftliches Grundverständnis; MNGV), Tube Figures (Schlauchfiguren; SF), Quantitative and Formal Problems (Quantitative und formale Probleme; QFP), Concentrated and Precise Working (Konzentriertes und sorgfältiges Arbeiten; KONZ), Text Comprehension (Textverständnis, TV), Learning of Figures (Figuren lernen; FIL), Learning of Facts (Fakten lernen; FAL), Charts and Tables (Diagramme und Tabellen; DUT). Previous confirmatory factor analyses have shown that a three-factor structure (i.e., reasoning, perceptual speed and memory) is most suitable to describe the structure of the TMS (Trost et al., 1998). According to Trost et al. (1998) the subtests can additionally be divided by complexity (i.e., complex and less complex subtests). The four complex deductive reasoning subtests have an especially

strong focus on medicine-related content and require rule inference and problem solving based on different information from complex verbal and/or numerical material (Trost et al., 1998). The five remaining subtests are referred to as less complex, because they capture individual, more narrowly defined abilities (e.g., Trost et al., 1998). An overview of all TMS-subtests together with the respective tasks, the associated factors in the three-factor model, and the complexity of each subtest are depicted in Table 1. We used raw scores for all analyses. For all subtests except KONZ the raw score was derived by summing up the number of correctly solved items. For KONZ the scoring consists of three steps. First, the number of correctly evaluated symbols is determined for each participant, which results in a value of -800 to 1600. Second, the top and bottom 2.5 percent of all participants receive a score of 0 and 20, respectively. Third, the values in between are divided into equal intervals and are assigned a score from 1 to 19 accordingly.

To calculate the TMS total score, all subtest scores were added up to obtain the resulting sum ranging from 0 to 178 points. The complete scoring of the TMS is explained in more detail in the information booklet for participants (ITB Consulting GmbH, 2022). The TMS results were then matched with the evaluation questionnaire data and with data from the participant database provided by participants during the registration process for the TMS. Table 2 summarizes the means and standard deviations of all predictors and outcomes.

Data analysis

To determine the relationship between time and money spent on preparatory activities and the total TMS score, we used a hierarchical multiple regression analysis. For analyses of the TMS-subtests, we conducted a hierarchical multivariate multiple regression to determine the global effect of all predictors (motivation, expenses, and preparatory activities) on all nine TMS-subtests simultaneously. Subsequent univariate multiple regression analyses were performed to determine the unique relationships between each of the predictor and criterion variables. To account

TABLE 1 Subtests of the TMS.

Subtest (abbreviation)	Task	Factor ^a	Complexity
MUZ	Detection of the correct image section from a complex image	Perceptual speed	Low
MNGV	Comprehension of short texts with medical or natural scientific content	Reasoning	High
SF	Mental rotation of three-dimensional objects	Perceptual speed/Reasoning	Low
QFP	Dealing with numbers, variables, and formulas within a medical or natural scientific context	Reasoning	High
KONZ	Detection of certain letter or symbol sequences in a row of 1600 letters or symbols	Perceptual speed	Low
TV	Comprehension and analysis of longer schoolbook-like texts	Reasoning	High
FIL	Memorizing figural information in a short period of time	Memory/Perceptual speed	Low
FAL	Memorizing verbal information in a short period of time	Memory	Low
DUT	Interpretation of graphs and tables within a medical or natural scientific context	Reasoning	High

^aFactor structure according to Trost et al. (1998); some subtests are allowed to load on more than one factor in the three-factor model.

for multiple testing in the subsequent univariate analyses, the alpha level was adjusted to 0.005 via Bonferroni correction. In the first step of every regression, we controlled for age, sex and high school GPA. Because participants could not be randomly assigned to the different conditions, this approach was chosen in accordance with Briggs (2004), as these variables have been demonstrated to be associated with performance on aptitude tests (e.g., Briggs, 2004; Buchmann et al., 2010) and therefore are potentially confounding variables. The multivariate multiple regression was performed with R (R Core Team, 2022) and all other analyses were conducted with SPSS 24.0.

Our results are reported as standardized and unstandardized regression coefficients with 95% confidence intervals. Standardized regression coefficients are used to make comparisons between predictors which are on different scales (e.g., preparation time and motivation), whereas unstandardized regression coefficients are used for comparisons of the preparatory activities. We report corrected R^2 determinants to adjust for potential overestimation bias due to a high number of predictors.

TABLE 2 Descriptive summary for all variables.

Variable	<i>n</i>	%	
Sex			
Female	5516	69.8	
Male	2387	30.2	
Variable	<i>M</i>	<i>SD</i>	Scale
Age	19.20	1.97	–
High school GPA	1.68	0.46	–
Motivation	3.62	0.84	1-5
Preparation duration (in hours)			0-80 (10-h increments) plus 3-digit input field
Preparation with information booklet	6.08	7.65	
Preparation with books	35.78	33.27	
Preparation with computer	5.64	11.46	
Preparation in study groups	3.05	8.65	
Training courses	5.39	13.61	
Expenses (in EUR)	139.11	275.74	0-200 (20 Euro increments) plus 4-digit input field
Test scores			
TMS score	96.66	21.29	0-178
MUZ score	9.29	2.82	0-20
MNGV score	10.32	3.58	0-20
SF score	14.43	3.33	0-20
QFP score	8.91	3.98	0-20
KONZ score	10.41	4.49	0-20
TV score	10.15	3.48	0-18
FIL score	10.45	3.52	0-20
FAL score	12.82	4.11	0-20
DUT score	9.89	3.65	0-20

N = 7,903. *n*, sample size; *M*, mean; *SD*, standard deviation.

Requirements for regression analysis

Regarding multicollinearity, all variables were found to have a tolerance value above 0.67, well above the minimum threshold of 0.25 suggested by Urban and Mayerl (2006), p. 232). In addition, the VIF value for each predictor is close to 1, as such being far below the threshold of 5.

To check for correlation between residuals, we ran the Durbin-Watson test with a resulting value of 1.88 which is well within a tolerable range of 1.5-2.5. Hence, we assume no problem related to residual autocorrelation. Graphical analyses (p-p-plots) show that residuals follow a normal distribution (Supplementary Figures 2-4).

To test for the heteroscedasticity of residuals, we constructed a chart of residuals (*y*-axis) and regression values (*x*-axis). As values scatter randomly in a dot-like shape (Supplementary Figures 5-7), we assume no problem of heteroscedasticity. Taken together, the requirements for regression analyses were met.

Check for non-linear effects

To consider possible non-linear (in particular quadratic and logistic) effects, especially for the time participants spent on preparing for the TMS, we performed graphical analyses (i.e., residual and scatter plots) in addition to the CURVEFIT procedure in SPSS, neither of which provided strong enough evidence of a non-linear relationship. Entering quadratic terms for the predictors in the multiple regression analyses yielded negligible amounts of additional explained variance at best (ranging from $\Delta R^2 \leq 0.001$ to $\Delta R^2 = 0.014$). Therefore, we report the more parsimonious linear models.

Results

The associations of preparatory activities, motivation, and expenses with the TMS total score

The results of our regression model testing the associations of motivation, preparatory activities, and expenses in addition to age, sex and high school GPA with the TMS total score are depicted in Table 3. 79.3% of all participants used the information booklet for preparation, 92.9% used books, 49.0% used the computer, 23.8% used learning groups, and 18.9% used training courses. 2.6% of all participants indicated that they had invested more than 80 hours in at least one of the preparatory activities.

The corresponding correlation matrix of all predictors and the outcome variable can be found in Table 4. We also calculated correlations between all predictors with the relative times spent on each preparatory activity (absolute time spent on preparatory activity *x* divided by the sum total of times spent on all preparatory activities) to facilitate interpretation. The results are depicted in Table 5.

In total, the model explains 20% of variance, where 9% of variance is explained by the control variables (high school GPA, age, and sex) and 11% is explained by the predictors motivation, time, and expenses incrementally over GPA, age, and sex.

Our results indicate that with an increase in self-reported motivation by one point on the respective scale, such as from

low to medium or from medium to medium-high, the TMS score increases by 3.90 points ($b = 3.90$, CI [3.35, 4.46]) on a scale from 0 to 178 points, holding all other predictors constant.

Regarding the time spent on different preparatory activities, significant but small associations can be detected: their average magnitude equals less than a one point change of the 178 points in the TMS score.

When studying one hour longer and holding all other factors constant, using the information booklet is associated with a decrease of 0.211 ($b = -0.211$, CI [-0.267, -0.155]) points in the TMS score, using books with an increase of 0.118 ($b = 0.118$, CI [0.104, 0.132]) points. With an average across all participants of about 36 hours spent preparing with books, this factor accounts for an increase of about 4.2 points or 0.19 standard deviations on average (note that we used the standard deviation of 21.56 points from the total sample as reference). In contrast, using computer programs was associated with a decrease of 0.10 ($b = -0.10$, CI [-0.137, 0.063]) points and using training courses with an increase of 0.076 ($b = 0.076$, CI [0.038, 0.114]) per hour. With the average preparation time of roughly 5.5 h, preparation with training courses

would account for an increase of 0.4 points or 0.02 standard deviations on the TMS total score. Spending more time learning in study groups is not significantly associated with the TMS score ($b = 0.06$, CI [0.01, 0.111]).

Expenses for preparatory activities show a significant association with the TMS score. Each Euro spent on preparatory activities is associated with an increase of 0.004 ($b = 0.004$, CI [0.002, 0.006]) points in the TMS-score. This means that – according to our model – an investment of 250 EUR would be associated with a TMS total score increase of one point out of 178 points. 136 EUR – the average amount spent on preparation – on the other hand, would result in a predicted TMS total score increase of about 0.6 points or 0.03 standard deviations. Additionally, test-takers who spend more money overall on preparation are significantly more likely to spend most of their time preparing with training courses suggesting that training courses are the main driver of preparation cost ($r(7901) = 0.51$, $p < 0.001$). Comparing the standardized regression coefficients, the association of motivation with the TMS score is 2.5 times stronger than the association of expenses with the TMS score.

TABLE 3 Hierarchical multiple regression analysis predicting the TMS-score.

	<i>b</i>	SE (<i>b</i>)	β	<i>t</i>	<i>p</i> -value	95% CI LL for <i>b</i>	95% CI UL for <i>b</i>	<i>R</i> ²	ΔR^2
Step 1								0.09	
Control variables ^a									
Step 2								0.20	0.11
Constant	105.676	2.36	/	44.78	<0.001	101.050	110.302		
Motivation	3.904	0.28	0.15	13.79	<0.001	3.349	4.459		
Preparatory activities:									
Information booklet	-0.211	0.03	-0.08	-7.41	<0.001	-0.267	-0.155		
Books	0.118	0.01	0.18	16.45	<0.001	0.104	0.132		
Computer	-0.100	0.02	-0.05	-5.28	<0.001	-0.137	-0.063		
Study groups	0.060	0.03	0.02	2.34	0.020	0.010	0.111		
Training courses	0.076	0.02	0.05	3.95	<0.001	0.038	0.114		
Expenses	0.004	0.00	0.06	4.60	<0.001	0.002	0.006		

N = 7,903. *b*, unstandardized regression coefficient; SE, standard error; β , standardized regression coefficient; *t*, *t*-test; *p*-value, probability; CI, confidence interval; LL, lower limit; UL, upper limit; *R*², adjusted coefficient of determination; ΔR^2 , change in the corrected coefficient of determination; ^aControl variables included age, sex, and high school GPA.

TABLE 4 Pearson correlations of all predictors in the final regression model.

	1	2	3	4	5	6	7	8	9	10
(1) Information booklet										
(2) Books	0.04**									
(3) Computer	0.12**	0.00								
(4) Study groups	0.04**	0.14**	0.04**							
(5) Training courses	0.01	0.13**	0.03**	0.24**						
(6) GPA	0.06**	0.07**	0.07**	0.06**	0.07**					
(7) Motivation	0.11**	0.40**	0.10**	0.13**	0.15**	0.02*				
(8) Expenses	-0.03*	0.18**	0.03**	0.17**	0.55**	0.05**	0.16**			
(9) Age	0.06**	0.08**	0.06**	0.01	-0.02*	0.33**	0.00	0.01		
(10) Sex ^d	0.06**	0.09**	-0.02*	0.00	-0.01	-0.07**	0.09**	0.02*	-0.06**	
(11) TMS score	-0.08**	0.23**	-0.06**	0.07**	0.11**	-0.33**	0.22**	0.13**	-0.08**	-0.06**

N = 7,903; * $p < 0.05$, ** $p < 0.01$; ^a0, male, 1, female.

TABLE 5 Pearson correlations of predictors with relative time spent on each preparatory activity.

Predictor	1	2	3	4	5
(1) Information booklet					
(2) Books	-0.52**				
(3) Computer	0.06*	-0.56**			
(4) Study groups	-0.13**	-0.22**	-0.10**		
(5) Training courses	-0.21**	-0.32**	-0.15**		
(6) GPA	-0.05**	-0.16	0.02	0.03**	0.05**
(7) Motivation	-0.22**	0.21**	-0.11**	0.02*	0.09**
(8) Expenses	-0.22**	-0.07*	-0.13**	0.07**	0.51**
(9) Age	-0.04**	0.05**	0.01	-0.02**	-0.05**
(10) Sex ^a	-0.03*	0.10**	-0.07**	-0.03*	-0.03*
(11) TMS score	-0.20**	0.19**	-0.17**	0.02	0.11**

N = 7,903; relative preparation time was calculated by dividing the absolute time spent on the respective preparatory activity by the sum total of times spent on all preparatory activities; *p < 0.05, **p < 0.01; ^a0, male, 1, female; The correlations for the variables 6 to 11 can be found in Table 4 and are thus not displayed here.

The association of preparatory activities, motivation, and expenses with TMS subtest results

Because the multivariate multiple regression yielded significant results for all predictors (Table 6), univariate multiple regressions were performed for each subtest to determine the unique relationships with each predictor (Tables 7, 8). The univariate regression analyses showed differences with respect to the incremental variance explained by motivation, preparatory activities, and expenses. When controlled for age, sex, and GPA the complex field-specific subtests TV ($\Delta R^2 = 0.012$), QFP ($\Delta R^2 = 0.013$), MNGV ($\Delta R^2 = 0.014$), DUT ($\Delta R^2 = 0.008$) are less affected by preparatory activities, expenses and motivation than the remaining subtests MUZ ($\Delta R^2 = 0.075$), SF ($\Delta R^2 = 0.121$), FIL ($\Delta R^2 = 0.097$), FAL ($\Delta R^2 = 0.118$), KONZ ($\Delta R^2 = 0.05$).

Self-reported motivation showed a significant association with all subtest-scores. Regarding the complex subtests, a one-point increase in self-reported motivation was associated with a score increase of 0.211 points (DUT) to 0.306 points (MNGV). For the less complex subtests, a one-point increase in self-reported motivation was associated with a score increase of 0.445 (KONZ) to 0.759 (FAL) points.

Concerning the different preparatory activities, studying with the information booklet and books showed a significant association with all subtests. Preparation with the information booklet was associated with a decrease of points in all subtests, ranging from -0.02 (TV) to -0.026 (QFP) points per self-reported hour for the complex subtests and from -0.016 (MUZ) to -0.036 (FAL) points for the less complex subtests. For preparatory activities with books, one self-reported hour was associated with an increase in points ranging from 0.004 (DUT) to 0.007 (MNGV and QFP) for the complex subtests and from 0.011 (MUZ) to 0.025 (FAL) points for the less complex subtests. Preparatory activities on the computer were not associated with the complex subtests MNGV and QFP. For DUT and TV preparation on the computer was

TABLE 6 Hierarchical multivariate multiple regression analysis predicting the nine TMS-subtests.

	Wilks' λ	Df	F
Step 1			
Control variables ^a			
Step 2			
Motivation	0.96	9, 7884	37.19**
Preparatory activities:			
Information booklet	0.99	9, 7884	7.59**
Books	0.93	9, 7884	61.97**
Computer	0.99	9, 7884	4.50**
Study groups	1.00	9, 7884	2.94**
Training courses	0.99	9, 7884	8.627**
Expenses	0.99	9, 7884	7.467**

N = 7,903. TMS-subtests are MUZ, MNGV, SF, QFP, KONZ, TV, FIL, FAL and DUT. Df = degrees of freedom; ^aControl variables included age, sex, and high school GPA; *p < 0.05, **p < 0.01.

associated with a decrease of 0.009 and 0.01 points per self-reported hour respectively. For all less complex subtests preparation on the computer was associated with a decrease in points, ranging from 0.011 (SF) to 0.014 (MUZ and FAL) per self-reported hour.

Spending time preparing in study groups only showed significant associations with the less complex subtests MUZ and KONZ with an increase of 0.012 to 0.016 points per each self-reported hour, respectively. Time spent in training courses also had no significant association with the complex subtests. In contrast, the time spent in training courses was associated with a point increase on all less complex subtests. The increase of points per self-reported hour ranges from 0.011 (FIL) to 0.022 (KONZ) points.

Expenses for preparatory activities showed significant associations with the less complex subtests except for KONZ. The increase of points per self-reported hour was 0.001 for FIL, MUZ, SF, and FAL respectively. No associations of expenses were found with the complex subtests.

Discussion

The association of motivation, time, and financial investment with TMS results

In the current study, we assessed the association of self-reported motivation, time spent on various preparatory activities, and financial investment with TMS results of a cohort of medical school applicants. Most notably, preparatory activities and money spent for preparation had only small associations with the TMS result, while, for example, self-reported motivation during preparatory activities had a 2.5-fold stronger association with the TMS score than the money spent on it.

As we controlled for preparation time, the aforementioned association of motivation with the TMS result cannot be explained by a mere difference in duration, but rather appears to point to a difference in the quality of time spent. Motivation might play a role in how a person “uses” the preparation time and also how he

or she works on the TMS during the actual test-taking procedure. Highly motivated persons might take the TMS more seriously. Additionally, persons with higher motivation may perceive the TMS tasks as more joyful than others and have a higher interest in them, because motivation is associated with positive emotions such as interest and enjoyment (Deci and Ryan, 1985; Brandstätter et al., 2018).

Interestingly, motivation also appears to be a factor in the differences we found in the effectiveness of time spent on each preparatory activity. The higher the motivation, the more absolute time test-takers spent on either of the preparatory activities. It is not surprising that participants who are more motivated invest more time in preparation, but it is also noteworthy that they appear to invest more time in preparatory activities that are efficient (especially books). This may be the result of a combination of two different pathways: motivated test-takers may be drawn to the modes of preparation which are 'per se' more effective modes of preparation, and it is also conceivable that due to higher motivation, the overall quality of the time they spent on the more effective activities positively affected results.

On the other hand, we investigated the association of money spent on preparation with the TMS result. Previous studies on whether financial resources are a significant predictor of high-stakes admission test results have yielded contradictory findings (e. g. Lambe et al., 2012; Laurence et al., 2013; Wilkinson and Wilkinson, 2013). Our results show, that financial investments in preparatory activities play a statistically significant but a practically small role. Therefore, our findings are in line with studies in the context of other study admission tests showing the direct effects of money spent on preparation to be very small and, therefore, either non-significant or overrated (e.g., Powers and Rock, 1999; Sackett et al., 2008; Wilkinson and Wilkinson, 2013). All other factors (including motivation) being equal, mere time spent on the different preparatory activities – the information booklet, books, computer, study groups, and training – yielded only small positive and sometimes even small negative associations with the total TMS score and its subtests, respectively. This suggests that the mere quantity of time invested in preparation is not the decisive factor in achieving good test results. However, within these small associations, some relevant differences were detected.

Are certain preparatory activities more advantageous than others?

Our results indicate, that of all preparatory activities, preparation with books descriptively yielded the highest positive association with the TMS score. Additionally, it is the only activity that showed a positive association with the performance in all subtests. Interestingly, the magnitude of these associations seems to be more in line with effects found for training courses and test retaking in previous meta-analyses (e.g., Messick and Jungeblut, 1981; Powers and Rock, 1999; Briggs, 2004, 2009; Hausknecht et al., 2007) than with usually smaller effects reported for preparation with books. For example, Briggs (2004) reported effects of 0.11 to 0.31 standard deviations for training courses, but no significant effects for preparation with books. Hausknecht et al. (2007) found an adjusted effect of $d = 0.26$ for retesting of different admission tests (i.e. taking a parallel or the same test form a second time).

However, on a closer look, the relative effectiveness of preparation with books does not surprise in the context of the TMS. Books typically used for preparation contain original earlier TMS test versions, which means that participants who spend time preparing with those books would practice with original earlier tests. Thus, for the TMS, preparation with books can be seen as a form of retesting, which would explain why the associations found in our study are more comparable to the effects of retesting reported by Hausknecht et al. (2007). Practicing with books, therefore, is highly encouraged by the test administrators as the most effective and cost-efficient way of preparation.

Preparing with the information booklet does not appear to yield any additional training benefit beyond mere informational purposes and general advice about the style, format, and general requirements associated with the different subtests of the TMS. In addition, only a small number of practice items are provided, which may cause participants who spend too much time on the booklet to become too focused on this narrow subset of items, potentially providing them with a false sense of security.

Accordingly, time spent in study groups was found to have no association with the overall TMS score. One possible reason for this finding is, that study groups may vary considerably with regard to the structure and content discussed and therefore may not yield reliable positive results overall. Concerning training courses, we found a small positive association with the TMS total score. For 20 h of preparation with training courses, our model would predict a score increase of 0.07 standard deviations – considerably less than the training effects of 0.12 or 0.20 standard deviations reported by Messick and Jungeblut (1981) or the training effects reported by Briggs (2004) and Hausknecht et al. (2007). To achieve a similar effect for the TMS it would require about 34 to 80 h of preparation.

Are certain subtests more robust to preparation than others?

Analyses suggest, that the complex subtests may be more robust to preparation than the less complex subtests. At the subtests level, preparation with books appears to be the only activity that has a positive association with the performance in all subtests, while the score increases were higher for the less complex subtests. Training courses, on the other hand, appear to yield no association with the complex subtests. In contrast, training courses had positive associations with all less complex subtests, while the score gain per hour was somewhat higher for MUZ and KONZ compared to preparation with books. For all other less complex subtests, preparation with books seems to be more effective than preparation with training courses. On the subtest level, preparation on the computer and preparation in training courses do not appear to be effective types of preparation. Both activities yielded negative associations with most of the complex and all less complex subtests. Again, associations were predominantly higher for the less complex subtests than for the complex subtests. Even though preparation with study groups was not associated with the overall TMS score, at the subtest level, small positive associations were found for MUZ and SF. Suggesting that, at least for some of the less complex subtests, positive results can be obtained with this mode of preparation.

Limitations and directions for further research

Due to time constraints during test-taking, we only assessed motivation with a single item directly and retrospectively asking for participants' subjective degree of motivation during preparation, instead of a validated questionnaire designed to assess the entirety of the construct. This is associated with a number of additional sources of error and provides room for alternative explanations of our findings. For example, participants' answers could have been affected by their degree of confidence with regard to the test material, which in turn might have been affected by their actual or perceived ability and their expectation to succeed in the test. Therefore, participants who believe that they will succeed in the test might indicate higher levels of motivation during preparation in retrospect than they actually experienced at the time. Also, participants who are more optimistic about their general chances to obtain a study place – for example due to very good GPA grades, a generally high self-efficacy, or

high academic confidence – might indicate higher motivation at the time of assessment regardless of their actual motivation during preparation. All of these pathways are conceivable, and we cannot rule out the possibility that our measure of motivation is affected by other constructs such as self-efficacy, optimism, actual ability, or academic self-concept rather than measuring intrinsic motivation at the time of preparation. Therefore, results regarding the role of motivation should be further investigated with validated questionnaires and in a longitudinal design. Nevertheless, we do believe that our interpretation of the results is still valid, as motivation and constructs such as self-efficacy and ability itself are generally correlated (Goldberg, 1994; Richardson et al., 2012). We therefore believe that our results can provide a first indication that motivation may not only be relevant during university training and for a successful career in medicine (e.g., Lawler and Hall, 1970; Zapata-Phelan et al., 2009; Latham, 2012) but also already at the stage of application and admission.

Since the complex and less complex subtests seem to be differentially associated with preparatory activities, it should be mentioned that preparation effects do not necessarily harm the

TABLE 7 Hierarchical univariate multiple regression analyses predicting the complex TMS-subtests.

	MNGV			QFP			TV			DUT		
	<i>b</i>	β	R^2	<i>b</i>	β	R^2	<i>b</i>	β	R^2	<i>b</i>	β	R^2
Step 1			0.09			0.17			0.08			0.15
Control variables ^a												
Step 2			0.10			0.18			0.10			0.15
Constant	12.436**			14.748**			11.511**			15.907**		
Motivation	0.306**	0.07		0.239**	0.05		0.296**	0.07		0.211**	0.05	
Preparatory activities:												
Information booklet	-0.021**	-0.05		-0.026**	-0.05		-0.02**	-0.05		-0.025**	-0.05	
Books	0.007**	0.07		0.007**	0.06		0.006**	0.06		0.004*	0.04	
Computer	-0.008	-0.02		-0.009	-0.03		-0.010*	-0.03		-0.009*	-0.03	
Study groups	0.005	0.01		-0.005	-0.01		0.002	0.01		-0.002	-0.01	
Training courses	-0.004	-0.02		0.009	0.03		-0.006	-0.03		0.002	0.01	
Expenses	0.000	0.01		0.000	0.02		0.000	0.02		0.000	0.01	

N = 7,903; *b*, unstandardized regression coefficient; β , standardized regression coefficient; R^2 , corrected coefficient of determination; ^aControl variables included age, sex, and high school GPA; Alpha level was adjusted for multiple tests via Bonferroni correction, **p* < 0.005, ***p* < 0.001.

TABLE 8 Hierarchical univariate multiple regression analyses predicting the less complex TMS-subtests.

	MUZ			SF			KONZ			FIL			FAL		
	<i>b</i>	β	R^2	<i>b</i>	β	R^2	<i>b</i>	β	R^2	<i>b</i>	β	R^2	<i>b</i>	β	R^2
Step 1			0.01			0.02			0.02			0.02			0.03
Control variables ^a															
Step 2			0.09			0.14			0.07			0.12			0.15
Constant	8.647**			13.793**			9.329**			9.478**			9.819**		
Motivation	0.450**	0.13		0.588**	0.15		0.445**	0.08		0.611**	0.15		0.759**	0.16	
Preparatory activities:															
Information booklet	-0.016**	-0.04		-0.029**	-0.07		-0.019*	-0.03		-0.019**	-0.04		-0.036**	-0.07	
Books	0.011**	0.13		0.019**	0.20		0.018**	0.13		0.02**	0.19		0.025**	0.20	
Computer	-0.014**	-0.06		-0.011**	-0.04		-0.013*	-0.03		-0.012**	-0.04		-0.014**	-0.04	
Study groups	0.012*	0.04		0.012*	0.03		0.016	0.03		0.008	0.02		0.014	0.03	
Training courses	0.013**	0.06		0.016**	0.07		0.022**	0.07		0.011*	0.04		0.014**	0.06	
Expenses	0.001**	0.05		0.001**	0.08		0.000	0.03		0.001**	0.05		0.001**	0.07	

N = 7,903; *b*, unstandardized regression coefficient; β , standardized regression coefficient; R^2 , corrected coefficient of determination; ^aControl variables included age, sex, and high school GPA; Alpha level was adjusted for multiple tests via Bonferroni correction, **p* < 0.005, ***p* < 0.001.

validity of a test. Preparation can even increase the validity by reducing the construct-irrelevant variance in participants' test scores attributable to differences in participants' familiarity with the test (cf. Briggs, 2009; Arendasy et al., 2016). Also, it may be possible that the underlying skills measured are trainable to a different extent and that the differing effects found for the complex and less complex subtests reflect differing actual gains in ability (cf. Arendasy et al., 2016). In this case, the training effects would not affect validity. To be able to assess the meaning and the differences of preparation effects in the context of the TMS and other admission tests, it would therefore also be important to investigate the influence of preparation on validity.

In selecting the different preparatory activities listed in our questionnaire, we followed prior relevant studies (e.g., Deter, 1982; Bartussek et al., 1984; Trost et al., 1998; Briggs, 2004, 2009). Nevertheless, it is possible that we did not include other relevant preparatory activities. Therefore, one possibility for future research would be to allow participants to specify additional preparatory activities by use of an open response format. However, based on the previous studies mentioned above, we assume that at least the most essential preparatory activities are included in our study.

Another important limitation is that the data was collected cross-sectionally. Therefore, no conclusive statements can be made about the causality of the associations found. However, it is also important to note that the questionnaire on preparatory activities was conducted shortly before the actual testing. Thus, the participants did not yet have an accurate impression of their test performance, which might then have affected the statements regarding their preparatory activities. However, a longitudinal study would be an important approach for future research to clearly answer the question of causal direction.

We also had no information on the extent to which the participants already had connections within the medical field (e.g., through parents, siblings, or friends). For example, it is conceivable that well-connected individuals have better knowledge of and access to the different types of preparation, which may moderate the observed effects to some extent and should therefore be taken into account in future studies.

Previous studies have also shown that SES can be associated with performance on high school aptitude tests (e.g., Briggs, 2004). Since we had no information on the participants' SES, we could not control for this potentially confounding variable. For example, it may be the case that some effective preparatory activities were more available to applicants with higher SES. Around 93% of the participants in our study reported using books for preparation, which indicates that almost all participants had knowledge of and access to this type of preparation. Thus, the likelihood that access to preparation books is related to the participants' background, seems to be quite low. Nevertheless, this possibility cannot be ruled out completely at this moment. The situation is different for training courses, however: only 19% of the participants stated that they used this type of preparation. Since training courses are usually expensive, they are most likely not available to all subgroups of participants. For future studies, it would be important to investigate this relationship in more detail. Nevertheless, since books are significantly less costly and more effective, applicants who attended costly training courses are not likely to have an advantage over participants that cannot afford this kind of preparation. As noted before, time can also be seen as a resource not everyone may have equal access to. Therefore, it may be the case that SES

is related to the time spent for preparation. Since all preparation effects are small, this study provides an indication that the effects of such a relationship on the selection process would likely also be small. However, in order to answer the question conclusively and to rule out the possibility that some groups of participants are systematically disadvantaged, the aforementioned limitation must be taken into account in future studies.

Practical implications

An important implication for TMS participants is that the test scores are only associated to a comparatively small extent with any given preparatory activity, indicating little influence of preparatory activities on the test result. However, the results also suggest that a certain amount of preparation and familiarization with the test material is sensible. Based on our findings, preparation with books can be recommended for this purpose. In contrast, it is neither necessary nor advisable to spend a lot of money on expensive courses, especially because more effective and at the same time less expensive options such as books are available. For comparison, books with original TMS versions are available for the cost of approximately 13 euros, whereas training courses often range in cost from 150 to 500 euros.

The findings also have important implications in terms of equal opportunity. First, money spent on preparatory activities is only marginally associated with the test result indicating that participants with more money available to spend on preparation are not likely to be advantaged over other participants. Since there still is a small positive association, further research is needed to address the practical significance of this association regarding the allocation of study places. However, as already mentioned, the results so far indicate that money spent on preparation is not the decisive factor in context of the TMS. Second, of all preparatory activities included in the study, books seem to be the most effective and – because of the relatively low cost – most accessible type of preparation. This is also reflected in the fact that most participants prepared using books, which in turn minimizes the likelihood that certain subgroups of applicants are disadvantaged because they lack access to this type of preparation. However, we are aware that books still cost money and may not be equally accessible to everyone. To address this issue, the test developers have invited all test takers of 2021 to practice free of charge with original test items through an online-preparation platform as a first pilot. This project was continued in 2022 and is planned to be further expanded and evaluated in the future. Third, although preparation with books and training courses is associated with test results, the effects of these preparatory activities are modest. This again limits the likelihood that subgroups of applicants with more time to spend for preparation have a systematic advantage over other applicants. As mentioned earlier, this topic needs to be explored further in future research because the effects of other important variables, such as participants' SES or any pre-existing connections to the medical field (such as having parents who are doctors), could not be addressed in this study. In summary, we are still confident that the results provide initial indications that important prerequisites for equal opportunities are met. However, as mentioned before, further research is necessary in order to be able to treat this topic conclusively.

Since the results suggest a linear relationship between preparatory activities and test results, this finding should not be misinterpreted in a way that the test score can be increased to any degree by excessive preparation. This is because our results apply primarily to preparation times up to a maximum of 80 hours, as only a small proportion of participants reported preparation times above 80 hours (mainly for books). Thus, for preparation times exceeding 80 hours, other associations (e.g., logarithmic or quadratic relationships) with the test results are still conceivable (e.g., Messick and Jungeblut, 1981; Hausknecht et al., 2007).

In addition, our results provide evidence, that the TMS-subtests are differently affected by preparatory activities. Although the associations of preparatory activities with the TMS were small for all subtests, mostly the complex and more field-specific TMS-subtests appear to be resistant to preparatory activities. On the other hand, performance on more unidimensional subtests such as concentration, mental rotation, and memory seems more likely to improve with preparation. This also has important implications for test construction. For example, a promising approach might be using a substantial number of more complex field-specific subtests to further minimize the trainability of admission tests and thus the impact of resources such as time and money.

Conclusion

Overall, our results provide evidence relevant to the topic of equal opportunity with regard to the choice of career and access to one's preferred field of study in the context of scholastic aptitude tests: Motivation – as assessed in our study – appears to be a predictor for success in the context of the TMS, while more money spent on preparation is only marginally associated with better test results. Given that most training courses include preparation with original test items derived from books, in addition to our result that books appear to be the most time- and cost-efficient ways of preparation, we conclude that financial investments for test preparation likely do not yield substantial advantages, which is an important prerequisite for equal opportunity. However, further research is needed on the role of SES in terms of equal opportunity and future research should also take into account the impact of preparatory activities on the validity of admission tests.

Data availability statement

The datasets presented in this article are not readily available because of confidentiality agreements. Requests to access the datasets should be directed to DW, daniel.weppert@itb-consulting.de.

References

Arendasy, M. E., Sommer, M., Gutiérrez-Lobos, K., and Punter, J. F. (2016). Do individual differences in test preparation compromise the measurement fairness of admission tests? *Intelligence* 55, 44–56. doi: 10.1016/j.intell.2016.01.004

Ethics statement

Ethical approval was not provided for this study because data collection was initially conducted as part of annual quality assurance measures. Within this context, an ethical approval was not required in accordance with the institutional requirements. General written informed consent for participation in the TMS (as part of which the survey was administered) was obtained from all participants or legal guardians. Therefore, no separate informed consent was obtained for participation in the study.

Author contributions

DW, DA, ME, MK, and JM contributed to conception and design of the study. DW and JM organized the database and provided the initial draft of the manuscript. DW, LT, and JM performed statistical analyses. DW, DA, and ME planned the concept for statistical analyses and interpreted the results. DA, ME, LT, and LL wrote sections of the manuscript. All authors contributed to manuscript revision, read, and approved the submitted version.

Conflict of interest

DW, LT, and JM are working for the ITB Consulting GmbH, the organization that is developing the Test for Medical Studies (TMS) evaluated in this article.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/feduc.2023.1104464/full#supplementary-material>

Bartussek, D., Raatz, U., Stapf, K. H., and Schneider, B. (1984). *Die evaluation des tests für medizinische studiengänge. Erster zwischenbericht*. Bonn: Ständige Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland.

- Brandstätter, V., Schüler, J., Puca, R. M., and Lozo, L. (2018). *Motivation und emotion*. Berlin: Springer Berlin Heidelberg. doi: 10.1007/978-3-662-56685-5
- Briggs, D. C. (2004). "Evaluating SAT coaching: Gains, effects, and self-selection," in *Rethinking the SAT: The future of standardized testing in university admissions*, ed. R. Zwick (New York, NY: Routledge Falmer).
- Briggs, D. C. (2009). *Preparation for college admission exams (2009 NACAC discussion paper)*. Alexandria, VA: National Association for College Admission Counseling.
- Buchmann, C., Condron, D. J., and Roscigno, V. J. (2010). Shadow education, American style: Test preparation, the SAT and college enrollment. *Soc. Forces* 89, 435–461. doi: 10.1353/sof.2010.0105
- Deci, E. L., and Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behavior*. New York, NY: Plenum Press. doi: 10.1007/978-1-4899-2271-7
- Denissen, J. J., Zarrett, N. R., and Eccles, J. S. (2007). I like to do it, i'm able, and i know i am: Longitudinal couplings between domain-specific achievement, self-concept, and interest. *Child Dev.* 78, 430–447. doi: 10.1111/j.1467-8624.2007.01007.x
- Deter, B. (1982). *Zum einfluss von übung und training auf die leistung im "test für medizinische studiengänge" (TMS)*. Hanover: Agentur Petersen.
- Durik, A. M., Vida, M., and Eccles, J. S. (2006). Task values and ability beliefs as predictors of high school literacy choices: A developmental analysis. *J. Educ. Psychol.* 98:382. doi: 10.1037/0022-0663.98.2.382
- Girotti, J. A., Chanatry, J. A., Clinchot, D. M., McClure, S. C., Swan Sein, A., Walker, I. W., et al. (2020). Investigating group differences in examinees' preparation for and performance on the new MCAT exam. *Acad. Med.* 95, 365–374. doi: 10.1097/ACM.0000000000002940
- Gliatto, P., Leitman, I. M., and Muller, D. (2016). Scylla and charybdis: The MCAT, USMLE, and degrees of freedom in undergraduate medical education. *Acad. Med.* 91, 1498–1500. doi: 10.1097/ACM.0000000000001247
- Goldberg, M. D. (1994). *A developmental investigation of intrinsic motivation: Correlates, causes, and consequences in high ability students*. Charlottesville, VA: University of Virginia.
- Grant, A. M., and Schwartz, B. (2011). Too much of a good thing: The challenge and opportunity of the inverted U. *Perspect. Psychol. Sci.* 6, 61–76. doi: 10.1177/1745691610393523
- Hausknecht, J. P., Halpert, J. A., Di Paolo, N. T., and Moriarty Gerrard, M. O. (2007). Retesting in selection: A meta-analysis of coaching and practice effects for tests of cognitive ability. *J. Appl. Psychol.* 92, 373–385. doi: 10.1037/0021-9010.92.2.373
- ITB Consulting GmbH (2022). *Informationsbroschüre für den testdurchgang im november 2022*. Bonn: ITB Consulting GmbH. Available at: https://www.tms-info.org/wp-content/uploads/informationsbroschuere_tms.pdf (accessed October 26, 2022).
- Kadmon, G., and Kadmon, M. (2016). Academic performance of students with the highest and mediocre school-leaving grades: Does the aptitude test for medical studies (TMS) balance their prognoses? *GMS J. Med. Educ.* 33:Doc7.
- Kadmon, G., Kirchner, A., Duelli, R., Resch, F., and Kadmon, M. (2012). Warum der test für medizinische studiengänge (TMS)? *Z. Evid. Fortbild. Qual. Gesundheits.* 106, 125–130. doi: 10.1016/j.zefq.2011.07.022
- Kumar, K., Roberts, C., Bartle, E., and Eley, D. S. (2018). Testing for medical school selection: What are prospective doctors' experiences and perceptions of the GAMSAT and what are the consequences of testing? *Adv. Health Sci. Educ.* 23, 533–546. doi: 10.1007/s10459-018-9811-8
- Kusurkar, R. A., Ten Cate, T. J., Vos, C. M. P., Westers, P., and Croiset, G. (2013). How motivation affects academic performance: A structural equation modelling analysis. *Adv. Health Sci. Educ.* 18, 57–69. doi: 10.1007/s10459-012-9354-3
- Lambe, P., Waters, C., and Bristow, D. (2012). The UK clinical aptitude test: Is it a fair test for selecting medical students? *Med. Teach.* 34, e557–e565. doi: 10.3109/1042159X.2012.687482
- Latham, G. P. (2012). *Work motivation: History, theory, research, and practice*. Thousand Oaks, CA: Sage. doi: 10.4135/9781506335520
- Laurence, C. O., Zajac, I. T., Lorimer, M., Turnbull, D. A., and Sumner, K. E. (2013). The impact of preparatory activities on medical school selection outcomes: A cross-sectional survey of applicants to the university of Adelaide medical school in 2007. *BMC Med. Educ.* 13:159. doi: 10.1186/1472-6920-13-159
- Lawler, E. E., and Hall, D. T. (1970). Relationship of job characteristics to job involvement, satisfaction, and intrinsic motivation. *J. Appl. Psychol.* 54, 305–312. doi: 10.1037/h0029692
- Mathew, M., and Thomas, K. (2018). Medical aptitude and its assessment. *Nat. Med. J. India* 31, 356–363. doi: 10.4103/0970-258X.262905
- McGaghie, W. C., Downing, S. M., and Kubilius, R. (2004). What is the impact of commercial test preparation courses on medical examination performance? *Teach. Learn. Med.* 16, 202–211. doi: 10.1207/s15328015tlm1602_14
- Messick, S., and Jungeblut, A. (1981). Time and method in coaching for the SAT. *Psychol. Bull.* 89, 191–216. doi: 10.1037/0033-2909.89.2.191
- Orsini, C., Binnie, V. I., and Wilson, S. L. (2016). Determinants and outcomes of motivation in health professions education: A systematic review based on self-determination theory. *J. Educ. Eval. Health Prof.* 13:19. doi: 10.3352/jeehp.2016.13.19
- Patterson, F., Knight, A., Dowell, J., Nicholson, S., Cousans, F., and Cleland, J. (2016). How effective are selection methods in medical education? A systematic review. *Med. Educ.* 50, 36–60. doi: 10.1111/medu.12817
- Powers, D. E., and Rock, D. A. (1999). Effects of coaching on SAT I: Reasoning test scores. *J. Educ. Meas.* 36, 93–118. doi: 10.1111/j.1745-3984.1999.tb00549.x
- R Core Team (2022). *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing. Available online at: <https://www.R-project.org/> (accessed October 3, 2022).
- Richardson, M., Abraham, C., and Bond, R. (2012). Psychological correlates of university students' academic performance: A systematic review and meta-analysis. *Psychol. Bull.* 138, 353–387. doi: 10.1037/a0026838
- Sackett, P. R., Borneman, M. J., and Connelly, B. S. (2008). High stakes testing in higher education and employment: Appraising the evidence for validity and fairness. *Am. Psychol.* 63:215. doi: 10.1037/0003-066X.63.4.215
- Schult, J., Hofmann, A., and Stegt, S. J. (2019). Leisten fachspezifische studierfähigkeitstests im deutschsprachigen raum eine valide studienerefolgsprognose? Ein metaanalytisches update. *Z. Entwicklungspsychol. Pädagog. Psychol.* 51, 16–30. doi: 10.1026/0049-8637/a000204
- Sobral, D. T. (2004). What kind of motivation drives medical students' learning quests? *Med. Educ.* 38, 950–957. doi: 10.1111/j.1365-2929.2004.01913.x
- Trost, G., Blum, F., Fay, E., Klieme, E., Maichle, U., Meyer, M., et al. (1998). *Evaluation des tests für medizinische studiengänge (TMS). Synopse der ergebnisse*. Bonn: ITB.
- Turner, R., and Nicholson, S. (2011). Reasons selectors give for accepting and rejecting medical applicants before interview. *Med. Educ.* 45, 298–307.
- Urban, D., and Mayerl, J. (2006). *Regressionsanalyse: Theorie, technik und anwendung (2. Aufl.)*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Wigfield, A., and Eccles, J. S. (2000). Expectancy-value theory of achievement motivation. *Contemp. Educ. Psychol.* 25, 68–81. doi: 10.1006/ceps.1999.1015
- Wilkinson, T. M., and Wilkinson, T. J. (2013). Preparation courses for a medical admissions test: Effectiveness contrasts with opinion. *Med. Educ.* 47, 417–424. doi: 10.1111/medu.12124
- Zapata-Phelan, C. P., Colquitt, J. A., Scott, B. A., and Livingston, B. (2009). Procedural justice, interactional justice, and task performance: The mediating role of intrinsic motivation. *Organ. Behav. Hum. Decis. Processes* 108, 93–105. doi: 10.1016/j.obhdp.2008.08.001
- Zentrum für Testentwicklung und Diagnostik (2022). *Vorbereitungsreport*. Available online at: <http://www.ztd.ch/w/index.php?title=Vorbereitungsreport> (accessed October 26, 2022).