Psychological Test and Assessment Modeling, Volume 63, 2021 (4), 525-541

Chances and Psychometric Limits of Questionnaires for Field-Specific Interest: An Example from Mechanical Engineering

Lisbeth Weitensfelder [1]& Ilona Herbst[2]

Abstract

While interest questionnaires based on Holland's RIASEC-model face the challenge of finding an optimal congruence measure to match a person's interests with the environment, tailor-made specific interest questionnaires avoid this problem in a pragmatic way: Instead of resulting in a general interest profile, only the interests needed in the particular field are assessed. The particular field can then be assessed in more detail, while other areas and therefore a typological interest are not looked for. The article presents construction, implementation and psychometric analyses (i.e. according to the Rasch model) of a field-specific interest questionnaire for mechanical engineering, focussing on the struggle of how to create a uni-dimensional and therefore fair measurement.

Keywords: field-specific interests; interest assessment; questionnaire design; Rasch model

[1] Correspondence concerning this article should be addressed to: Lisbeth Weitensfelder, Department of Environmental Health, Center for Public Health, Medical University of Vienna, lisbeth.weitensfelder@meduniwien.ac.at

[2] Strategic Education Development Center, Vienna University of Technolgy

Introduction

There is a broad variety of useful and approved interest assessments for career counseling, so that at a first glimpse the invention of new inventories seems unnecessary. But as stated below, there might be cases where something other than the already existing inventories is needed to focus on more narrow aspects of interest. The present paper introduces an interest questionnaire that is tailor-made for a specific field of study (in the following: “field-specific” interest questionnaire) and discusses it from a psychometric point of view, focussing on Rasch model analysis and its benefits.

An uncommon approach to interest questionnaires

Though the application of interest assessments is indisputable in modern career assessments, no consistent definition of interest exists. Therefore it is not surprising that different research tendencies exist, investigating a broad variety of topics under the label “interest”. Applied psychology researches (longer-lasting) vocational interest, while educational psychology focuses on both the longer-lasting topic interest as well as on the situational perspective (e.g. Schraw & Lehman, 2001, for an educational overview). While some authors combine the two educational perspectives, e.g. by showing that situational interest might lead to longer-lasting topic interest (Hidi & Renninger, 2006; Krapp, 2002; Krapp, 2007), the two different psychological subfields with vocational vs. topic interest usually remain separated from each other. But when a person wants to study at the university, both subfields provide valuable information. Getting educated for a future profession might put university studies in a transitional field between educational and applied psychology, so that the two disciplines meet for field-specific interest questionnaires.

The Common Vocational Perspective

In general, the benefit of vocational interest questionnaires can be seen as a given, providing a key source for counselors (e.g. Sodano, 2015). Most vocational interest research is based on the empirically well-researched RIASEC-model of Holland (1959, 1997), which has been a “surpassing achievement in vocational psychology” (Savickas & Gottfredson, 1999). Holland’s theory comprises six different interest domains, with the initial letters naming the model: Realistic (R), Investigative (I), Artistic (A), Social (S), Enterprising (E) and Conventional (C) interests. Apart from one general factor (Tracey, 2012/ Prediger, 1982), it is assumed that Holland’s six domains are based on two factors (e.g. Prediger, 1982), usually called data/ideas and people/things, though other interpretations of underlying factors might also fit the data (e.g. sociability and conformity, see Hogan, 1983). Inventories following Holland’s approach offer the benefit that environments (like fields of study or professions) can also be classified regarding these six domains, making it possible to match people’s dispositions with environments (e.g. for reporting their "fit" with a specific environment or for finding suitable environments).

Limits of vocational interest questionnaires. Interest questionnaires are a useful source for career counseling, usually offering concrete examples for future career paths (e.g. the AIST-R, Bergmann & Eder, 2005) by matching the resulting interest type with environment codes. Nevertheless, when matching a dispositional interest type with an environment, several problems occur: First, finding an environment code can be problematic and the code might differ depending on the method used (Kaub, Stoll, Biermann, Spinath, & Brünken, 2014; Lent & Lopez, 1996; Rolfs & Schuler, 2002). But even if the environment code is already stated, another problem occurs when calculating the congruence of a testee's dispositional code with the environment code. Usually, the calculation of congruence happens in a typological manner, which means, the comparison between a person's dispositional code and the environment is done based on the person's intra-individually dominant interest areas. The absolute value of the testee's interest is not important – the rank order of the interest areas is the only criterion. Because of this, Rolfs and Schuler (2002) suggested a dimensional conception of congruence, where the absolute distinctions of interest areas are (inter-individually) aligned with the environment. Following this conception, it is important that a person is interested in an area that's requested from the environment, but it's not important whether the person would also be interested in something else. The authors could show in a sample of university students that a dimensional conception is superior to a typological conception in matters of well-being in academic studies. Moreover, the authors even assume that considering irrelevant interest areas might lead to a systematic underestimation of congruence and well-being. Another study of Kaub and her colleagues (2014) found similar results, leading them to speak out against the use of typological congruence indices: Testing teacher candidates, they found low to moderate correlations according to expectations between dimensional interest congruence and various criteria of study motivation and success. However, with a typological determination method, they partly even found correlations against their expectations, especially when it came to belief in one's own abilities.

Another Approach as a Solution

The results above lead to the conclusion that interest is important – but the determination of a "classical" typological interest type might lead to several disadvantages. Since the main difference between dimensional and typological conception lies in the fact that a dimensional conception includes just the interests needed, the question arises, whether other interest areas should be determined in all cases. In other words: If known, that for a future profession only a person's value of the interest domains R (Realistic) and I (Investigative) would be important, why assess other interest areas at all, instead of focusing on the required domains more closely? If just required interests are assessed, the label of their underlying domain might not even be important. Based on this approach, the present study introduces an interest questionnaire that just focuses on what is needed.

While a general overview as in Holland's model is needed for vocational orientation, there might be several testees who can already narrow a future profession or field of study down – especially in those cases it could be more recommendable, economic, and reasonable to focus on just the topics needed.

So far, only very few questionnaires exist that focus on more narrow areas of interest (e.g. Bollschweiler & Bernath, 1998; Bollschweiler & Toggweiler, 2009), but there is not much research about narrow assessments. As an exception, a well-researched (e.g. Glavin, Richard & Porfeli, 2009; Porfeli, Richard, & Savickas, 2010; Sodano & Richard, 2009) specific inventory is the Medical Specialty Preference Inventory MSPI (Zimny, 1979) resp. its revision MSPI-R (Richard, 2011) – a questionnaire which aims to help American medical students in making the right choice of specialty. The MSPI-R shows satisfying predictive hit rates in medical specialty choice, while broad interest inventories reached their limits for such a specific question (Burns, 2016).

As a downside for narrow questionnaires, the existence of a general factor for interest has to be critically mentioned, which biases narrow interest-scales (Tracey, 2012). Given this, it might be especially important that narrow questionnaires prove to be valid and helpful for test-takers. Aside from validation, their construction should also fulfill psychometric requirements.

Construction of Questionnaires According to the Rasch Model

Out of all item response theory (IRT) models, especially the dichotomous logistic model of Rasch (1960) offers the benefit of empiric verifiability (e.g. Kubinger & Draxler, 2007). In essence, the so-called "Rasch model" indicates that the probability of solving an item (in the case of a questionnaire: the probability for answering in the affirmative) depends on the ability (in the case of a questionnaire: [personality] trait) of the testee and the difficulty (in the case of a questionnaire: challenge) of the item. Other parameters (e.g. other personality traits than the aimed one, guessing parameters, learning effects) should not systematically influence the score.

To calculate a single (overall) score without paying attention to which of the items were answered in the affirmative and which were not, the questionnaire has to measure uni-dimensionally regarding the Rasch model. The model allows testing whether the questionnaire score (sum of answers in the affirmative) establishes a sufficient statistic. If the Rasch model does not hold, the questionnaire score cannot be seen as a fair measure of a testee's test behaviour (e.g. Kubinger, 2005). If the Rasch model however stands the test, this confirms that all items are based on the same dimension.

In particular, the Rasch model makes it possible to test whether "differential item functioning" (DIF) applies for different groups – e.g. men and women. That is, certain items disclose different difficulties, even though the measured dimension is the same. Einarsdóttir and Rounds (2009) pointed out that two thirds of the famous Strong

Interest Inventory SII in its 1994-version (Harmon, Hansen, Borgen & Hammer, 1994) show DIF regarding the split criterion “gender”. DIF could also be found on each RIASEC-scale of the AIST-R (Bergmann & Eder, 2005), especially on the Realistic scale (Wetzel & Hell, 2013).

The application of the Rasch model to questionnaires is controversial and bears several measurement problems (e.g. Ortner, 2005, for an overview): Problems can derive from the items (e.g. bad formulation or poor discrimination), the testpersons (e.g. faking), the answer categories (in case of applying a Partial Credit Model [Masters, 1982]), or the assessed traits (e.g. some constructs could be more-dimensional or just not be a construct at all). Kubinger (2000) also criticizes that when applying the (dichotomous) Rasch model to questionnaires, the trait sought after could be hidden by another trait (to answer truly). Furthermore, the (dichotomous) Rasch model is only applicable for items with dichotomous answers. But Jansen and Roskam (1986) showed that a dichotomization of items “after the fact”, that means, after responses of testpersons are already given, is not essentially equal to dichotomization “before the fact”. When items are dichotomized “after the fact” the polychtomous Rasch model is usually not compatible with the dichotomized data (Jansen & Roskam, 1986). Ortner (2005) nevertheless showed, that it can indeed be possible to apply the Rasch model with dichotomized answer categories to a personality questionnaire, even if items or whole scales have to be removed. Other IRT models of the Rasch family have also be applied successfully to questionnaires: For example the General Partial Credit Model (Muraki, 1992, 1997) has repeatedly been applied to questionnaires to create computer-adaptive versions of single questionnaires (Walter, Becker, Fliege, et al., 2005; Walter & Holling, 2008).

Main Aims and Questions of the Study

Field-specific interest questionnaires offer an opportunity to avoid several of the disadvantages mentioned above: If a questionnaire just focusses on topics relevant in an aspired major, a typological calculation is neither possible nor necessary. However, it is not clear how such field-specific interest questionnaires fit in the existing interest theories. Unlike general vocational interest questionnaires, they do not give a broad overview over several domains, but it is probable that a whole field of study consists of more than just one topic: For example, the previously mentioned medicine-specific MSPI consists according to an exploratory factor analysis of 18 interest factors (Sodano & Richard, 2009), and a stepwise discriminant function analysis with the 18 interest factors as predictors for a specialty choice showed that the majority of the constructs underlying specialty choice could closely be modeled with 15 of the 18 factors (Porfeli, Richard & Savickas, 2010). The methodological approach is not based on the item response theory, but the result with the high amount of factors still raises the question whether it could be possible to assess a whole field of study with just one underlying dimension.

The article presents the construction and implementation of a field-specific interest questionnaire for mechanical engineering. While the aims and benefit of the questionnaire can already be deduced from the theoretical background, the subsequent research focuses on the question whether such a tailor-made construction meets psychometric requirements at all.

Method

Construction of a Field-Specific Interest Questionnaire for Mechanical Engineering

The interest questionnaire presented is part of an extensive self-assessment for study applicants at the Vienna University of Technology [Technische Universität Wien] (the Viennese Self-Assessments for technical majors are accessible at studienwahl.tuwien.ac.at , see also Kubinger, Frebort, Khorramdel & Weitensfelder, 2012, 2013). The self-assessment, where study applicants can test themselves via internet without the presence of a test proctor, is based on a specially conducted requirement analysis and includes a broad variety of personality questionnaires, ability tests and experimental-based behavior tasks. Being part of this test battery, a field-specific interest questionnaire is applied.

The interest questionnaire aims to ask (just) what topics are relevant for the field of study (bachelor's programme). Furthermore, the questionnaire aims to have an informative character about the study contents (see Weitensfelder, Undeutsch, Khorramdel & Useini, 2012). The item construction draws on the curriculum of the fields of study and contains real or at least potential teaching contents. First, the curriculum's course catalogue was used to formulate questionnaire items about what is done in the courses. After this first step based on the curriculum, the first draft of items contained many technical terms that appeared overly abstract and hard to understand, therefore they were revised and reformulated in consultation with advanced mechanical-engineering-students (e.g. via asking them how they would explain the content of the underlying course to a high-school graduate).

The resulting questionnaire consists of 35 questions (example: "I am looking forward to learn how processes in nature can be abstracted through mathematical models.") at a 4-point rating scale. In order to be able to use the Rasch model, the items had to be dichotomized: The answer categories "agree" and "rather agree" both were scored with a point, the answer categories "rather disagree" and "disagree" with no point.

Prior to knowing whether the interest for a whole field of study can be measured unidimensionally, testees are reported their overall score, but they also get additional information regarding their interest in the four broadest subjects reported as "strength" or "weakness". 30 of the 35 questions in the questionnaire are based on the four broadest subjects in an unequal distribution, five items assess other contents (see Tab. 1).

Study Questions

When implementing new questionnaires, numerous aspects can be looked upon. The present article focusses on psychometric aspects, including analyses regarding IRT, i.e. the Rasch model: Apart from fairness of the overall score, questionnaires should furthermore (not at least for sociopolitical reasons) be fair regarding the split criterion "gender" and not include items that handicap study applicants with another native language.

Hence the first questions (1., 2.) to be answered in this article are:

1. Is the calculation of an overall score, as it was assumed in the item construction process, justified?
2. Are the items fair regarding "gender" (2a) and "language" (2b)?

Once established, such fairness of the items does not necessarily mean that there are no distribution differences between any subsamples of examinees with respect to their scores – all above no differences of the mean scores. Hence a Welch test was applied. There are empirically well-confirmed differences between men and women in the people-things-dimension (Su, Rounds & Armstrong, 2009), though of course the present population may already be self-selected according to their interests (the sample consists of the target population: testees who want to find out whether they are fit to study *mechanical engineering*, see below). That might diminish gender differences, but there is no guarantee.

Participants

The sample consists of anonymous test takers of the *Viennese Self-Assessments* for *Mechanical Engineering* during a survey period in late 2011 and 2012.

Of course, there is no control over test conditions or the sincerity of the assessment process. The completion of the self-assessment (a test battery of several hours) itself serves as (the single) sincerity criterion.

Altogether, there are data from 320 applicants. The test battery is completed anonymously and demographic information is not requested. Only for a part of the survey period, was an additional questionnaire provided, where demographic data were requested from the testees. The additional questionnaire was administered to 219 testees. 212 of these testees gave demographic information that seemed to be plausible as well, as it was self-reported to be honest. 52 of the 212 testees reported to be women, 31 of the 212 had another mother-language than German; mean age of the sample was 21.45 years (SD: 4.184, MIN = 17, MAX = 42).

Measures

To answer the two study questions and hypothesis 3, the interest questionnaires described above are the only measures needed.

Procedure

Apart from Rasch model analyses, standard software (SPSS) was used.

Rasch model analyses. To test, whether interest for the study contents can be seen (and reported) as uni-dimensional, the program package *eRm* (Version 0.15-5, Mair, Hatzinger & Maier, 2015; see also Mair & Hatzinger, 2007) was used and for the statistical software *R* (Version 3.2.2, R Core Team, 2015). As a model test, Andersen's likelihood ratio-test (Andersen, 1973) was applied, testing whether the same item difficulty parameters are valid for different subgroups. Apart from splitting testees in two sub-samples based on their score, where testees with a higher reported degree of respective interest are opposed to testees with lower reported interest, the split criteria "gender" and "language" (German vs. non-German) are applied to ensure that the questionnaire does not disclose differential item functioning.

While no testees were excluded from the data set for the split criterion "score", for the split criteria "gender" and "language" only those testees remained in the sample, who plausibly completed the demographic declarations and who reported to have done so honestly.

Gender differences at scale level. As Student's t-test does not hold the type-I-risk if precondition tests are used in advance (Kubinger, Rasch, & Moder, 2009), the Welch test was applied. Type-I-risk was set to the $\alpha = .05$.

Results

Classical test theory analyses show a relatively high reliability coefficient for the overall score (Cronbach alpha: r = .828), but not for the subscales with exception of the largest one (see Tab. 1). Mean comparisons (Tab. 2) do not show differences with regard to "gender" and "language".

Rasch model analyses. To avoid an elevation of the type-I-error (e.g. Kubinger & Draxler, 2007), the level of significance was corrected (including the split criterion score, where the major part of the data set overlaps with the data set for the other split criteria). As suggested by Koller, Alexandrowicz, and Hatzinger (2012) the test-wise type-I-risk was set to α/q, that is .017. The results are as follows: With respect to the split criterion "language" the Rasch model is valid. Also regarding the split criterion "gender", the Rasch model is valid, but the graphical model check (GMC) shows two suspicious items (Fig. 1); hence their exclusion seemed advisable: as a matter of fact an *ex post* model fit of the Rasch model after exclusion (retested only for the split criterion "gender") occurred. The results of Andersen's likelihood ratio-tests are shown in Table 3.

Table 1

Scales and reliabilities

	Number of items	**Cronbach's Alpha**
Total	**35**	**.828**
Engineering science subjects	12	.708
Systems science subjects	8	.605
Construction science & production engineering subjects	7	.503
Mathematical and natural science subjects	3	.465
Other (diverse and economics)	5	-

Notes: All results based on the complete sample (n = 320). All translations of subjects by the author. Naming and translation based on the curriculum that was valid when the questionnaire was constructed.

Table 2

Welch-tests regarding gender and language differences

Gender					**Language**				
Women		Men			German		non-German		
M	*SD*	*M*	*SD*	*p*	*M*	*SD*	*M*	*SD*	*p*
28.60	5.948	28.17	5.015	.321	28.39	5.254	27.61	5.245	.226

Notes: All p-values one-sided.
n = 212: 52 women; 181 German-speaking

Table 3

Andersen's likelihood ratio-tests

Split Criterion	***n***	**Andersen χ^2**	***df***	$\chi^2_{\alpha=.017}$	***p***
Score	320	124.372	33	52.471	< 0.001
Language	212	42.918	34	53.731	0.14
Gender	212	51.561	34	53.731	0.027
Gender after item exclusion[a]	212	29.139	32	51.206	0.612

Note: [a]Item exclusion advisable due to the graphical model check

However, with regards to the split criterion “score”, the items proved not to measure uni-dimensionally. This might be due to the fact that the curriculum covers a broad variety of topics. Because of this assumption, separate analyses with the gathered subjects were calculated (α still was .017). The response patterns for most subjects show such a high endorsement, that only one subject group (systems science subjects) could be calculated - showing uni-dimensionality regarding the split criterion “score” (n = 320, p = 0.974, *Andersen* χ^2 = 1.71, crit. χ^2 = 17.062, df = 7). The other subjects had to be combined in order to test the model: Two subjects seem to be rather similar in their teaching contents (namely “engineering science subjects” with “construction science and production engineering subjects”), so their items were put together for an artificially combined “technical” subject group. Put together, the Rasch model could be calculated for most items, showing a model fit also for this artificially combined group (n = 320, p = 0.564, *Andersen* χ^2 = 13.491, crit. χ^2 = 28.813, df = 15). The Rasch model even holds, when “mathematical and natural science subjects” were added to that “technical” subject group (n = 320, p = .333, *Andersen* χ^2 = 21.072, crit. χ^2 = 34.286, df = 19). Due to the ex-post approach, the result has to be interpreted carefully, but it seems plausible that the questionnaire’s lack of uni-dimensionality might be caused by the broad variety of subjects in the curriculum. Figure 2 shows the GMCs for the split criterion “score” which enables you to see items from all subjects (above) as well as for just one subject group (system science subjects).

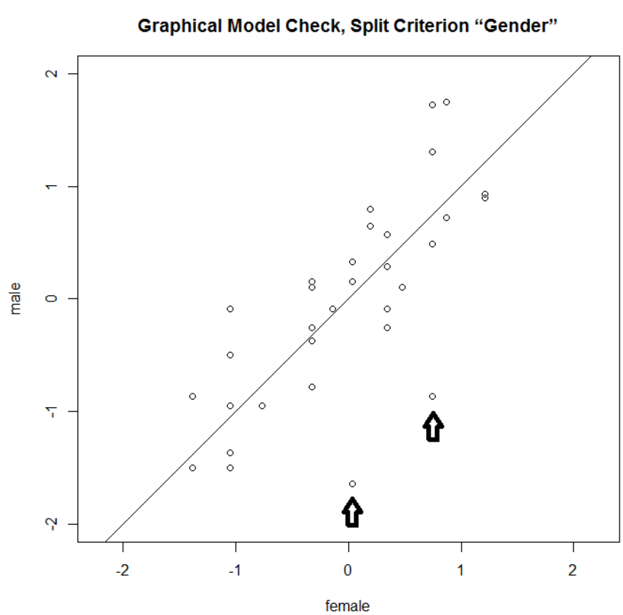

Figure 1. GMC for the split criterion “gender” with two graphically peculiar items (arrows)

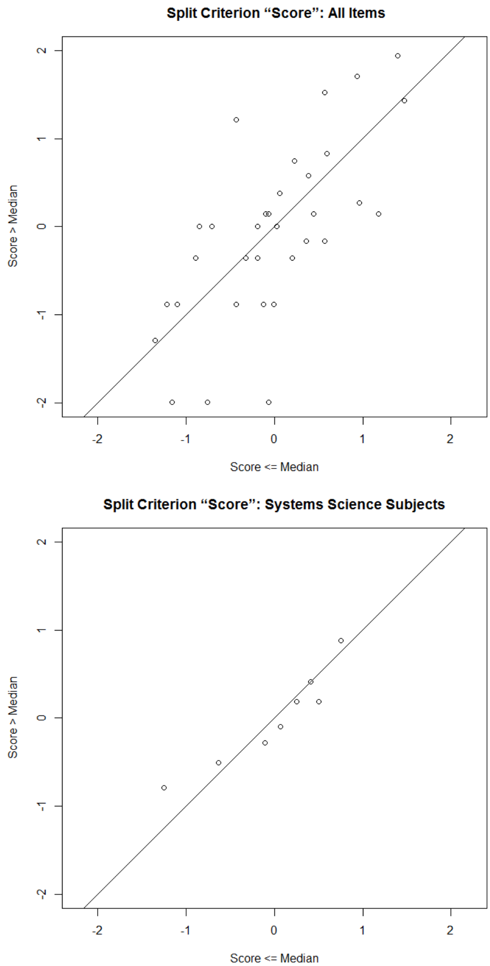

Figure 2. GMCs (split criterion "score") for all subjects combined (above) and "systems science subjects" separately

Discussion

Based on the theory that a certain interest pattern is not as important as a dimensional interest congruence with an aspired environment (Kaub et al., 2014; Rolfs & Schuler, 2002), field-specific interest questionnaires offer a pragmatic alternative for specific counseling: When only the topics needed are asked, the problematic (Lent & Lopez, 1996) determination of a congruence index is not necessary any more. Furthermore, tailor-made items can be constructed corresponding to very specific and detailed matters, making it possible to already give study applicants an idea about the curriculum. The item construction of the questionnaire presented is based on the curriculum and was revised with advanced students from that field of study, so the questionnaire can probably be called valid, though validation studies (which were not part of the present research) with outside criteria still need to be done and the questionnaire's benefit still needs to be proven. In a validation context, it would be especially interesting to investigate the incremental validity of field-specific questionnaires in combination with general interest questionnaires, since for counselling purposes they could perfectly complement each other: Where one type of questionnaire reaches its limits, the other one starts working. Applying a combination of the two different approaches, the current Viennese Self-Assessment reports the field-specific test result to testees, but in case testees score low it additionally describes the factors of the RIASEC model to present alternative ideas regarding a choice for a field of study.

When focusing on the psychometric criteria of the presented questionnaire, especially IRT results, the questionnaire shows potential, but also indicates the need for revision: A current disadvantage is that an overall score is reported, though the Rasch model does not hold for the split criterion "score", meaning that the items in the questionnaire measure more than one dimension. An explanation might be that the curriculum covers a wide variety of topics, which cannot be asked for on the same scale: studying mechanical engineering includes engineering and scientific knowledge as well as sound mathematical methods and analytical procedures. Since nowadays computer tools are used in all types of tasks connected with the development of innovative products and processes, an interest of sound knowledge in the field of information and communication technology is needed.

The analyses for the subjects seem to support the assumption, that the variety of study contents is too wide to be measured as a uni-dimensional construct. However, due to response patterns, only one of the subject groups (being uni-dimensional) could be calculated, and seemingly similar subjects showed a model fit after combining their items with an artificial "technical" subject group. This shows, how Rasch model analyses could help to create ex-post fair (sub-)scores of a questionnaire.

Regarding the validity of the Rasch model for the split criterion "language", the questionnaire is uni-dimensional. Also regarding the split criterion "gender", the items can be seen as gender-fair, since the Rasch model is valid. At an overall level, no gender differences could be found for users who wanted to test their fit for mechanical engineering studies.

Summing up the Rasch model evaluations (and answering study question 1), the calculation of the overall score is not justified. When implementing assessments that should meet both scientific and practical requirements, compromises are often inevitable. However, the findings show how it could be possible to make adjustments via just reporting (aggregated) subscales instead of an overall score to meet psychometric requirements. Of course, such a solution (when supported by a later cross-validation) might have negative effects on reliability. Therefore it needs to be weighted whether the requirement that subscales might represent a certain curriculum order or psychometric requirements (validity of the Rasch model, reliability) are seen as more important. The correspondence of subscales with curriculum contents seems promising for counseling at a first glimpse, but apart from reliability concerns, frequent curricula changes or re-namings represent a downside of this approach. Orientation towards study content that is especially crucial in the curriculum and hence valid for a long-lasting period of time (e.g. mechanics in mechanical engineering) could be a possible solution, but even contents that are set as core contents might vary over time or new contents might amend: If, as in the example of mechanical engineering, a new focus is set in the bachelor's degree (like mechanical engineering management), an adaptation of the underlying interest questionnaires is recommended: Study applicants interested in studying mechanical engineering with a focus on economics should also find their (additional) interests in an adapted or maybe separate interest questionnaire.

Answering study question 2, the questionnaire can be seen as fair regarding both gender (2a) and language (2b).

As a limitation to the previous results, it has to be mentioned that all results are based on a preselected sample: Namely the (anonymous) data of (probable future) study applicants for a technical major, who completed an online Self-Assessment for their aspired field of study. Though this sample is exactly the population that the questionnaire was constructed for, no conclusions can be made for people who are not interested in the field of study. It is possible that overall gender effects might occur when the questionnaire is applied to other than that target population.

Several questions still remain open for future examinations: When still focusing on the psychometric point of view, it might be worth questioning, whether the dichotomization of single items - though being the basis for calculating the Rasch model – might also lead to disadvantages. The famous research article of MacCallum, Zhang, Preacher and Rucker (2002), which criticizes the technique of variable dichotomization, does not target the dichotomization of single items, but one of the authors' major concerns – the loss of information – might also apply to item level, when items are dichotomized. And Jansen and Roskam (1986) showed that also from a methodological point of view an item dichotomization after testpersons have already given their answers, is highly problematic. It could be a matter of future examination, whether the questionnaire could benefit from a polytomous data model, like a Partial Credit Model (Masters, 1982). Especially, since the overall score is very skewed, showing a high endorsement (see average scores in Tab. 2), a detailed discrimination in the different answer categories it could be a benefit to discriminate a bit more regarding the four different answer categories rather than also giving testees with "rather agree" full points for an item.

Concluding, the given findings can give some ideas or guidelines for the construction of field-specific interest questionnaires. The results underline the necessity of Rasch model analyses, since even narrow fields might be too versatile to be uni-dimensional, causing overall scores not to be a fair measure. In such a case, a fragmentation to subscales as discussed above could be a solution.

Acknowledgements

All data refers to the “Viennese Self-Assessment” project (Kubinger, Frebort, Khorramdel & Weitensfelder, 2012, 2013). The project was a collaboration between the *Centre of Testing and Consulting, Division of Psychological Assessment and Applied Psychometrics, Faculty of Psychology, University of Vienna* and the *TU Wien*, represented by then vice chancellor Univ.-Prof. Dr. Adalbert Prechtl.

References

Andersen, E. B. (1973). A goodness of fit test for the Rasch model. *Psychometrika, 38,* 123-140. https://doi.org/10.1007/BF02291180

Bergmann, C., & Eder, F. (2005). *AIST-R. Allgemeiner Interessen-Struktur-Test mit Umwelt-Struktur-Test (UST-R) – Revision* [GIST-R. General Interest Structure Test with Environment Structure Test (EST-R) – Revision]. Göttingen: Beltz Test GmbH.

Bollschweiler, R., & Bernath, W. (1998). *SIT: Sozial-Interessentest* [SIT: Social Interest Test]. Dübendorf: SVB.

Bollschweiler, R., & Toggweiler, S. (2009). *KV-Interessentest R* [KV Interest Test R] *(2. Ed.).* Bern: SDBB.

Burns, S. T. (2016). History of medical specialty interest assessment. *Cogent education, 3(1),* 1131062. https://doi.org/10.1080/2331186X.2015.1131062

Einarsdóttir, S., & Rounds, J. (2009). Gender bias and construct validity in vocational interest measurement: Differential item functioning in the Strong Interest Inventory. *Journal of Vocational Behavior, 74(3),* 295-307. https://doi.org/10.1016/j.jvb.2009.01.003

Glavin, K. W., Richard, G. V., & Porfeli, E. J. (2009). Predictive validity of the medical speciality preference inventory. *Journal of Vocational Behavior, 74(1),* 128-133. https://doi.org/10.1016/j.jvb.2008.11.004

Harmon, L. W., Hansen, J. C., Borgen, F. H., & Hammer, A. L. (1994). *Strong Interest Inventory: Applications and technical guide.* Palo Alto, CA: Consulting Psychologists Press, Inc.

Hidi, S., & Renninger, K. A. (2006). The Four-Phase Model of Interest Development. *Educational Psychologist, 41(2),* 111-127. https://doi.org/10.1207/s15326985ep4102_4

Hogan, R. (1983). A socioanalytic theory of personality. In M. M. Page (Ed.), *Nebraska symposium on motivation 1982. Personality: Current theory and research* (pp. 55–89). Lincoln, NE: University of Nebraska Press.

Holland, J. L. (1959). A theory of vocational choice. *Journal of Counseling Psychology, 6,* 35-45.

Holland, J. L. (1997). *Making vocational choices: A theory of vocational personalities and work environments* (3rd ed.). Odessa, FL: Psychological Assessment Resources.

Kaub, K., Stoll, G., Biermann, A., Spinath, F. M., & Brünken, R. (2014). Interessenkongruenz, Belastungserleben und motivationale Orientierung bei Einsteigern im Lehramtsstudium [Congruence, stress experience, and motivational orientation of first-year teacher candidates]. *Zeitschrift für Arbeits- u. Organisationspsychologie, 58(3),* 125-139. https://doi.org/10.1026/0932-4089/a000149

Koller, I., Alexandrowicz, R., & Hatzinger, R. (2012). *Das Rasch Modell in der Praxis: Eine Einfhrung mit eRm* [The Rasch model in practical applications: An introduction with eRm]. Vienna: facultas.wuv, UTB.

Krapp, A. (2002). Structural and dynamic aspects of interest development: theoretical considerations from an ontogenetic perspective. *Learning and Instruction, 12(4),* 383-409. https://doi.org/10.1016/S0959-4752(01)00011-1

Krapp, A. (2007). An educational-psychological conceptualization of interest. *International Journal of Educational and Vocational Guidance, 7,* 5-21. https://doi.org/10.1007/s10775-007-9113-9

Kubinger, K. D. (2000). Und für die Psychologische Diagnostik hat es doch revolutionäre Bedeutung [And yet there is revolutionary importance for Psychological Assessment]. *Psychologische Rundschau, 51(1),* 33-34. https://doi.org/10.1026//0033-3042.51.1.33

Kubinger, K. D. (2005). Psychological Test Calibration Using the Rasch Model – Some Critical Suggestions on Traditional Approaches. *International Journal of Testing, 5(4),* 377-394. https://doi.org/10.1207/s15327574ijt0504_3

Kubinger, K. D., & Draxler, C. (2007). Probleme bei der Testkonstruktion nach dem *Rasch*-Modell [Some problems in calibrating an item pool according to the Rasch model]. *Diagnostica, 53(3),* 131-143. https://doi.org/10.1026/0012-1924.53.3.131

Kubinger, K. D., Frebort, M., Khorramdel, L., & Weitensfelder, L. (Eds.) „Wiener Autorenkollektiv Studienberatungstests“ (2012). *Self-Assessment: Theorie und Konzepte* [Self-Assessment: Theory and concepts]. Lengerich: Pabst.

Kubinger, K.D., Frebort, M., Khorramdel, L. & Weitensfelder, L. (2013). Prinzipien und Verfahren der Self-Assessments vom „Wiener Autorenkollektiv Studienberatungstests“ [Principles and Inventories of the Self-Assessments from the „Viennese Author Collective for Study Aptitude Testing”]. *Wirtschaftspsychologie, 1,* 17-24.

Kubinger, K. D., Rasch, D., & Moder, K. (2009). Zur Legende der Voraussetzungen des *t*-Tests für unabhängige Stichproben [About the legend of t-test requirements for independent samples]. *Psychologische Rundschau, 60(1),* 26-27. https://doi.org/10.1026/0033-3042.60.1.26

Lent, E. B., & Lopez, F. G. (1996). Congruence from Many Angles: Relations of Multiple Congruence Indices to Job Satisfaction among Adult Workers. *Journal of Vocational Behavior, 49(1),* 24-37. https://doi.org/10.1006/jvbe.1996.0031

MacCallum, R. C., Zhang, S., Preacher, K. J., & Rucker, D. D. (2002). On the Practice of Dichotomization of Quantitative Variables. *Psychological Methods, 7(1),* 19-40. https://doi.org/10.1037//1082-989X.7.1.19

Mair, P., & Hatzinger, R. (2007). Extended Rasch modeling: The eRm package for the application of IRT models in R. *Journal of Statistical Software, 20(9),* 1-20. https://doi.org/10.18637/jss.v020.i09

Mair, P., Hatzinger, R., & Maier M. J. (2015). *eRm: Extended Rasch Modeling (Version 0.15-5).* Retrieved from http://erm.r-forge.r-project.org/

Masters, G. N. (1982). A Rasch Model for Partial Credit Scoring. *Psychometrika, 47,* 149-74. https://doi.org/10.1007/BF02296272

Muraki, E. (1992). A generalized partial credit model: Application of an EM algorithm. *Applied Psychological Measurement, 16(2),* 159–176.

Muraki, E. (1997). A generalized partial credit model. In W.J. van der Linden, & R.K. Hambleton (Eds.), *Handbook of modern item response theory* (pp. 153–164). Berlin: Springer.

Ortner, T. M. (2005). *Möglichkeiten und Grenzen adaptiver Persönlichkeitsfragebogen. Ein Beispiel.* Lengerich: Pabst Science Publishers.

Porfeli, E. J., Richard, G. V., & Savickas, M. L. (2010). Development of specialization scales for the MSPI: A comparison of empirical and inductive strategies. *Journal of Vocational Behavior, 77(2),* 227-237. https://doi.org/10.1016/j.jvb.2010.04.011

Prediger, D. J. (1982). Dimensions underlying Holland's hexagon: Missing link between interests and occupations? *Journal of Vocational Behavior, 21(3),* 259-287. https://doi.org/10.1016/0001-8791(82)90036-7

R Core Team (2015). *R: A language and environment for statistical computing* [Computer software]. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from https://www.R-project.org/

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests.* Copenhagen: The Danish Institute of Education Research. (Expanded Edition, 1980. Chicago: University of Chicago Press.)

Richard, G. (2011). *Medical Speciality Preference Inventory, Revised Edition.* Washington DC: Association of American Medical Colleges.

Rolfs, H., & Schuler, H. (2002). Berufliche Interessenkongruenz und das Erleben im Studium. [Interest congruence and the experience of academic studies]. *Zeitschrift für Arbeits- und Organisationspsychologie, 46(3),* 137-149. https://doi.org/10.1026//0932-4089.46.3.137

Savickas, M. L., & Gottfredson, G. D. (1999): Holland's Theory (1959-1999): 40 Years of Research and Application. *Journal of Vocational Behavior, 55(1),* 1-4. https://doi.org/10.1006/jvbe.1999.1694

Schraw, G., & Lehman, S. (2001). Situational interest: A Review of the Literature and Directions for Future Research. *Educational Psychology Review, 13,* 23-52. https://doi.org/10.1023/A:1009004801455

Sodano, S. M. (2015). Meaning, measurement, and assessment of vocational interests for career intervention. In Paul J. Hartung, Mark L. Savickas, Bruce W. Walsh (Eds.), *APA handbook of career intervention, Volume 1: Foundations. APA handbooks in psychology* (pp. 281-301). Washington, DC: American Psychological Association.

Sodano, S. M., & Richard, G. V. (2009). Construct validity of the medical specialty preference inventory: A critical analysis. *Journal of Vocational Behavior, 74(1),* 30-37. https://doi.org/10.1016/j.jvb.2008.10.001

Su, R., Rounds, J., & Armstrong, P. I. (2009). Men and Things, Women and People: A Meta-Analysis of Sex Differences in Interests. *Psychological Bulletin, 135(6),* 859-884. https://doi.org/10.1037/a0017364

Tracey, T. J. G. (2012). Problems with single interest scales: Implications of the general factor. *Journal of Vocational Behavior, 81(3), 378-384.* https://doi.org/10.1016/j.jvb.2012.10.001

Walter, O. B., Becker, J., Fliege, H., Bjorner, J., Kosinski, M., Walter, M., Klapp, B. F., & Rose, M. (2005). Entwicklungsschritte für einen computeradaptiven Test zur Erfassung von Angst (A-CAT1). *Diagnostica, 51(2),* 88-100. https://doi.org/10.1026/0012-1924.51.2.88

Walter, O. B., & Holling, H. (2008). Transitioning from Fixed-Length Questionnaires to Computer-Adaptive Versions. *Zeitschrift für Psychologie, 216(1),* 22-28. https://doi.org/10.1027/0044-3409.216.1.22

Weitensfelder, L., Undeutsch, N., Khorramdel, L., & Useini, C. (2012). Intrinsische Studienmotivation: Interesse, Erwartungen und Selbstkonzept eigener Fähigkeiten [Intrinsic study motivation: Interest, expectations and self-concept of own abilities]. In K. D. Kubinger, M. Frebort, L. Khorramdel & L. Weitensfelder („Wiener Autorenkollektiv Studienberatungstests“) (Eds.), *Self-Assessment: Theorie und Konzepte* [Self-assessment: Theory and concepts] (pp. 119-142). Lengerich: Pabst.

Wetzel, E., & Hell, B. (2013). Gender-Related Differential Item Functioning in Vocational Interest Measurement. An Analysis of the AIST-R. *Journal of Individual Differences, 34(3),* 170-183. https://doi.org/10.1027/1614-0001/a000112

Zimny, G. H. (1979). *Manual for the medical specialty preference inventory.* St. Louis, MO: St. Louis University School of Medicine.