

Chapter 18

Inventing Scales for a Multidimensional Model of Mathematics Self-efficacy to Analyse First-Year Students' Mathematical Self-assessment, Performance, and Beliefs Change



Boris Girnát

18.1 Introduction

Beliefs have become an important field of research in mathematics education (Philipp 2007). They are research objects of their own, but they are also used as background variables to explain psychological or behavioural aspects of the teaching and learning of mathematics—especially with regard to mathematics performance. One important question is the relationship between performance and self-assessment. Nearly every international large-scale study is combined with some scales to measure students' mathematical self-assessment mostly to predict and to explain their performance (OECD 2005; OECD 2013). Research has shown that the relationship among self-concept, self-efficacy, and mathematics performance is strong and that mathematics self-concept and self-efficacy are in general powerful predictors to students' mathematics performance (Multon et al. 1991). The aim of the study presented here is to extend the application of scales related to mathematics self-concept and self-efficacy from lower secondary schools to upper secondary schools and universities. New scales have been developed to cover the mathematical content and skills of the transition phase from school to university. This paper presents the results of a first application of these scales in the context of a two-week bridging course that was designed to repeat secondary school mathematics. All participants of this course had to take part in a mathematics pre-test and a post-test and both tests were combined with a context questionnaire including the new scales. The research questions are as follows: (1) Do the new scales possess satisfactory psychometric properties? (2) Can these scales reveal group differences concerning relevant co-variables? (3) How are these scales related to each other

B. Girnát (✉)

University of Hildesheim, Hildesheim, Germany

e-mail: girnát@imai.uni-hildesheim.de

© Springer Nature Switzerland AG 2020

C. Andrà et al. (eds.), *Theorizing and Measuring Affect in Mathematics Teaching and Learning*, https://doi.org/10.1007/978-3-030-50526-4_18

177

and to other attributes of the students (e.g. performance, mathematics self-concept, and mathematics anxiety)? (4) Are these scales good predictors for the students' performance? (5) Can the scales be used to detect belief changes between the pre-test and the post-test and—if so—how are these changes related to the change of the students' general mathematics self-concept and mathematics performance?

18.2 Theoretical Background

There are two different ways to conceptualise students' *mathematical self-assessment*. The first one is related to a person's *mathematics self-concept*, measured by general statements like "I have always believed that mathematics is one of my best subjects" (Marsh 1990). The second approach is called *mathematics self-efficacy*. It was introduced by Bandura's idea of measuring a person's self-assessment by his level of confidence about feeling able to solve specific problems that are relevant to the domain of interest (Bandura 1977, 1986). Bandura defined self-efficacy beliefs as "people's judgments of their capabilities to organise and execute courses of action required to attain designated types of performances" (Bandura 1986, p. 391). Related to this idea, Shavelson et al. (1976) invented a hierarchical model starting with the "general self-concept" on the most abstract level, going down on different steps like the "academic self-concept" and the "mathematical academic self-concept", and finally arriving at the "evaluation of behaviour in specific situations" that is very close to Bandura's concept of self-efficacy. Two theoretical claims of this hierarchical model are most relevant to our study: (1) the higher and more abstract the level, the more stable the associated concept and (2) the lower and less abstract the level is, the more the associated concept can be represented by several different dimensions (Shavelson et al. 1976, pp. 412–414). The second claim was the reason to use a multidimensional model consisting of five scales related to different aspects of mathematics instead of a one-dimensional scale. The first claim culminated in the fifth research question mentioned above, namely whether and how changes in mathematics self-efficacy are related to changes in mathematics self-concept and performance—in particular, whether a different degree of stability is recognisable concerning these three aspects. An examination of the students' mathematics anxiety was added to the questionnaire as the "emotional counterpart" of the mathematics self-concept: according to Hannula's distinction between mathematics affects as states and traits, mathematics anxiety as a trait is stable and typically highly negatively correlated to mathematics performance and self-concept (Hannula et al. 2019).

18.3 Self-efficacy, Self-concepts, and Anxiety

The new scales are based on the PISA self-efficacy scale used in 2003 and 2012 (OECD 2005, pp. 291–294, OECD 2014, pp. 322–323). The PISA scale is short

and closely related to tasks typically used at lower secondary schools. But there were three obstacles to adopt this scale to university courses without changes: (1) the PISA scale was designed to measure mathematics self-efficacy only on lower secondary school level; (2) an exploratory factor analysis of this scale indicated that it cannot be considered as one-dimensional, but as multidimensional with respect to mathematical subdomains: algebra, elementary geometry, and applied mathematics/word problems (Girnat 2018); (3) Shavelson et al.'s model mentioned above is based on different subdomains on the more basic and less abstract levels. Insofar, the idea was to invent a bundle of self-efficacy scales that are (a) as short as possible, (b) related additionally to upper secondary school mathematics, and (c) multidimensional with respect to its different subdomains. To match the demands, it was necessary to define subdomains of upper secondary mathematics in a similar manner the re-analysis of the PISA scale suggested (Girnat 2018). To do so, some subdomains could remain, some had to be changed: algebra and applied mathematics are also relevant to upper secondary schools. These subdomains could remain and even some of the items of the PISA scale could be reused (indicated by "PISA"). Elementary geometry was replaced by analytic geometry; and calculus and probability theory were introduced as new subdomains. Since every subdomain should be represented by a scale that could stand for its own, four items per subdomain ought to be regarded as the minimum. Hence, the solution contains 20 items organised in five scales to represent a multidimensional model of mathematics self-efficacy (Table 18.1). The items were introduced with the following question: "How confident do you feel about having to do the following mathematics tasks (using a simple, non-graphing calculator)?"

The items on mathematics anxiety were directly reused from the PISA study in 2012 (OECD 2014, p. 323). The mathematics self-concept has to be adapted from the PISA scale (OECD 2014, p. 323), since some items with a direct reference to school contexts had to be removed or reformulated to be applicable at university level.

18.4 A Performance Test for a Bridging Course

The primary focus of this paper is the scales described above. Since they were used in the context of a bridging course that was evaluated by a mathematics performance test with a pre-/post-test design, it is also necessary to describe the bridging course and the performance tests. The teaching and learning of mathematics at university level have become a prominent part of mathematics education in recent years—especially the discrepancy between the requests of university mathematics and the skills students have achieved at school (Di Martino and Gregorio 2018). There are several proposals how to deal with this challenge. One of them is a bridging course to enable first-year students to university mathematics. In 2013, the University of Hildesheim decided to establish a voluntary bridging course for all courses of study including a substantial amount of mathematics (Hamann

Table 18.1 The items of the five mathematics self-efficacy scales

| Scale | Items |
|---|--|
| (a) Self-efficacy applied mathematics (self.app) | <ol style="list-style-type: none"> 1. Calculating how much cheaper a TV would be after a 30% discount. (PISA) 2. Calculating the petrol consumption rate of a car. (PISA) 3. Calculating how much interest is given on a savings plan within 10 years. 4. Calculating how long it takes to completely fill a swimming pool. |
| (b) Self-efficacy algebra (self.alg) | <ol style="list-style-type: none"> 1. Solving an equation like $3x + 5 = 17$. (PISA) 2. Solving an equation like $2(x + 3) = (x + 3)(x - 3)$. (PISA) 3. Multiply and simplify an algebraic expression like $2a(5a - 3b)^2$ 4. Solving an equation like $6x^2 + 5 = 29$ |
| (c) Self-efficacy calculus (self.calc) | <ol style="list-style-type: none"> 1. Determining the derivative of a function such as $f(x) = x \cdot e^x$. 2. Determining the maxima and minima of a function such as $f(x) = x^3 - 2x^2 + 1$. 3. Specifying the primitive of a function such as $f(x) = \frac{1}{2} \sin(x)$. 4. Calculating a definitive integral such as $\int_1^2 (x^2 - 2x) dx$. |
| (d) Self-efficacy analytical geometry (self.ageo) | <ol style="list-style-type: none"> 1. Calculating the length of a vector such as $v = (4, -2, 3)$. 2. Calculating the scalar product of two vectors such as $v = (-2, 1, 3)$ and $w = (2, -3, 2)$. 3. Calculating the intersection of a straight line such as $g : x = (1, -2, 0) + t \cdot (-1, 0, 1)$ with a plane. 4. Calculating the distance of a point such as $P(3 -1 4)$ from a plane (for example, $E : x - 3y + 2z = 3$). |
| (e) Self-efficacy probability theory (self.prob) | <ol style="list-style-type: none"> 1. Calculating the probability of throwing a 6 twice in a row. 2. Calculating the probability of winning the jackpot in the lottery. 3. Calculating how likely it is to draw two sweets of the same colour from a sweet jar. 4. Calculating how likely it is that two students in a class have their birthday on the same day. |

et al. 2014): computer science, business informatics, and elementary and secondary education with mathematics as a major. Although there are different views on which mathematical skills first-year students should possess, it is common sense that the skills students should have learnt at secondary schools are regarded as crucial for being successful at university (Nicholas et al. 2015). The department of mathematics decided to restrict the content to the following topics, since the bridging course had to be limited to 2 weeks and a preceding study had shown that the students demanded a special need of algebra, arithmetic, functions, graphs, and calculus and their applications in real-world situations (Kreuzkam 2013). These five topics have become the core area of the bridging course and the items of the test had to match these circumstances. The bridging course took place in October 2018. 312 students attended the course; 271 of them took part in the pre-test, 224 in the post-test, and 194 in both tests. The participants were distributed as shown in Table 18.2.

Table 18.2 Properties and sub-groups of the sample

| Gender | Study course | Mathematics at school | Course left before end |
|---------------------|---------------------------|-----------------------|------------------------|
| Female: 136 | Primary education: 138 | Basic level: 152 | Not left earlier: 224 |
| Male: 168 | Secondary education: 73 | Advanced level: 125 | Left earlier: 88 |
| | Computer science: 48 | | |
| | Business informatics: 41 | | |
| Others/no answer: 8 | Others: 12 | No answer: 35 | |

Table 18.3 Reliability and fit indices of the scales in the pre-test

| Scale | Cronbach's alpha | RMSEA | CFI | SRMR |
|-----------|------------------|-------|-------|-------|
| Self.app | 0.75 | 0.029 | 0.998 | 0.031 |
| Self.alg | 0.76 | 0.000 | 1.000 | 0.024 |
| Self.calc | 0.78 | 0.043 | 0.997 | 0.034 |
| Self.ageo | 0.80 | 0.046 | 0.997 | 0.038 |
| Self.prob | 0.83 | 0.000 | 1.000 | 0.024 |
| Matcon | 0.74 | 0.000 | 1.000 | 0.007 |
| Matanx | 0.75 | 0.000 | 1.000 | 0.012 |

The labels “basic level” and “advanced level” refer to a characteristic of the German school system: students have to choose in grade 10 if they want to be taught in mathematics on a basic or on an advanced level during grade 11 and 12.

18.5 Psychometric Properties of the Scales and the Test

The psychometric properties of the scales and the performance test are reported now. The reliabilities of the scales were estimated by Cronbach's alpha (Cronbach 1951). Since we will later use confirmatory factor analyses (CFA) and structural equation modelling (SEM), the typical fit indices used within this paradigm are reported additionally (Beaujean 2014, pp. 153–166). All calculations were done using R (R Core Team 2018) and the R package “lavaan” (Rosseel 2012) with a DWLS estimator (Beaujean 2014, pp. 92–113).

Table 18.3 contains the values of the pre-test (the values of the post-test are quite similar and are omitted to save space). According to the usual criteria, all scales have good properties. For the performance test, a unidimensional Rasch model was used to obtain a scale that expresses the “overall” mathematics performance of the participants (Linden 2016). The analysis was done using the R package “TAM” (Robitzsch et al. 2018). The Rasch model had an excellent EAP reliability (0.904).

18.6 Results

The results are presented in four steps: (1) correlations between all scales; (2) group differences with respect to background variables mentioned in Table 18.2; (3) Table 18.3 shows the changes between the pre-test and the post-test; (4) linear models to explain students' mathematics performance and beliefs. The asterisks stand for the usual significance levels: * for $p < 0.05$, ** for $p < 0.01$, and *** for $p < 0.001$.

Correlations

Table 18.4 shows the latent correlations (Beaujean 2014, pp. 100–103) between the performance test, the scales of the context questionnaire, and the final school exam mark with respect to the data of the pre-test—estimated using a structural equation model with good fit indices (RMSEA 0.013, CFI 1.000, SRMR 0.052). The correlations are mostly as expected: there are substantial correlations between the five scales of the multidimensional self-efficacy model, but they are not that high that they could not be empirically differentiated. This is different in case of the mathematics self-concept (matcon) and mathematics anxiety (matanx). This finding supports the hypothesis that both scales are (positive and negative) indicators of the same underlying concept (Hannula et al. 2019, mentioned above).

Group Differences

The focus of the analysis is now set to mean differences related to different sub-groups of the sample. Here, we omit the co-variate “choice of the study course”, since there are no significant differences observable. This is very surprising, since the most relevant study to this topic (Betz and Hackett 1983) indicates that mathematics self-efficacy is a good predictor for the students' choices of their study courses. The reason may be the fact that the possibilities to choose a study course is rather limited at the University of Hildesheim and, therefore, remarkable differences cannot occur. To make the differences on different scales comparable, we report the differences in terms of Cohen's d (Cohen 1988), i.e. the mean of one group (the “reference group”) is set to zero and the mean of the other group is given as the difference to zero on a standardised metric. Cohen's d is usually interpreted as follows (Cohen 1988): $d = 0.2$ indicates a small effect, $d = 0.5$ a medium effect, and $d = 0.8$ a strong effect (Table 18.5).

There is no significant gender difference concerning the performance test ($d = -0.120$), but the situation is quite diverse with respect to the different self-efficacy scales. There is a small to medium difference ($d = 0.339^*$) in favour to the female group in case of analytic geometry; and there is nearly a large ($d = -0.678^{***}$) difference to the detriment of the female group in case of applied mathematics. Both cases are remarkable aberration in perception compared to the measured (insignificant) performance difference.

Table 18.5 Mean differences between sub-groups of the sample (pre-test)

| Variable | Gender (ref. gr.: male) | Level in mathematics (ref. gr.: basic level) | Course left before end (ref. gr.: not left) |
|-----------|----------------------------|---|--|
| Pre-test | $d = -0.120$ | $d = 1.041^{***}$ | $d = -0.661^{***}$ |
| Seff.app | $d = -0.678^{***}$ | $d = 0.087$ | $d = 0.168$ |
| Seff.alg | $d = 0.106$ | $d = 0.953^{***}$ | $d = -0.319^*$ |
| Seff.calc | $d = 0.134$ | $d = 1.020^{***}$ | $d = -0.180$ |
| Seff.ageo | $d = 0.339^*$ | $d = 0.763^{***}$ | $d = -0.337^*$ |
| Seff.prob | $d = -0.209$ | $d = 0.414^{**}$ | $d = -0.090$ |
| Matcon | $d = 0.074$ | $d = 0.501^{***}$ | $d = -0.243$ |
| Matanx | $d = 0.340^*$ | $d = -0.370^*$ | $d = -0.043$ |

Table 18.6 Mean differences between pre-test and post-test

| Variable | Difference |
|------------------|-------------------|
| Performance test | $d = 0.978^{***}$ |
| Seff.app | $d = 0.581^{***}$ |
| Seff.alg | $d = 0.848^{***}$ |
| Seff.calc | $d = 0.887^{***}$ |
| Seff.ageo | $d = 0.210^*$ |
| Seff.prob | $d = 0.026$ |
| Matcon | $d = 0.144$ |
| Matanx | $d = -0.105$ |

Differences Between Pre-test and Post-test

Now, we compare the results of the pre-test and the post-test. The two tests were connected using the linking method according to Stocking and Lord (1983). Again, the differences are expressed in terms of Cohen's d (Table 18.6).

The difference concerning the performance tests is huge ($d = 0.978^{***}$). However, it is remarkable that this successful development is reflected very diversely with respect to the beliefs and emotions measured by the scales of the context questionnaires: both mathematics self-concept ($d = 0.144$) and mathematics anxiety ($d = -0.105$) did not change significantly. That may be an evidence for the conjecture that these beliefs (or emotions, respectively) belong to the set of central beliefs within a beliefs system that do not change rapidly—especially not on the basis of a relatively short experience of a two-week bridging course (Philipp 2007, p. 260). Therefore, they seem to be no good indicators for detecting short-term changes. It is exactly the opposite concerning the scales of mathematics self-efficacy: these scales reflect the achievement change almost in the same size that is observed in the performance tests.

Linear Models

Finally, we come back to the pre-test. We suspected that the scales of the context questionnaire may be good predictors for the students' performance. We analyse this hypothesis using linear models (Searle and Gruber 2016) that include scales of the questionnaire as predictors (independent variables) and the results of the pre-test as

Table 18.7 Linear models explaining students' mathematics performance (pre-test)

| Predictors | b_i | $SE(b_i)$ | β_i | Total R^2 |
|------------|----------|-----------|-----------|---------------|
| Seff.app | -0.117 | 0.232 | -0.031 | $R^2 = 0.357$ |
| Seff.alg | 1.532*** | 0.267 | 0.417*** | |
| Seff.calc | 0.188 | 0.172 | 0.103 | |
| Seff.ageo | 0.508 | 0.270 | 0.159 | |
| Seff.prob | 0.069 | 0.112 | 0.033 | |
| Matcon | 0.076 | 0.121 | 0.022 | |
| Matanx | -0.035 | 0.080 | -0.014 | |
| Seff.alg | 2.087*** | 0.307 | 0.579*** | $R^2 = 0.335$ |

Table 18.8 Linear models explaining students' mathematics self-concept (pre-test)

| Predictors | b_i | $SE(b_i)$ | β_i | Total R^2 |
|------------|---------|-----------|-----------|---------------|
| Seff.app | 0.248* | 0.112 | 0.215* | $R^2 = 0.369$ |
| Seff.alg | 0.159 | 0.098 | 0.137 | |
| Seff.calc | 0.178* | 0.072 | 0.307* | |
| Seff.ageo | 0.012 | 0.094 | 0.013 | |
| Seff.prob | 0.092 | 0.054 | 0.140 | |
| Seff.app | 0.261** | 0.054 | 0.422*** | $R^2 = 0.353$ |
| Seff.calc | 0.397** | 0.119 | 0.311** | |

the dependent variable to explain. The model was defined in two steps (Table 18.7): firstly, all variables were used as predictors (model above the line, RMSEA 0.36, CFI 0.963, SRMR 0.53); then, all insignificant predictors were removed leading to the result that only one predictor has a significant and substantial explorative value: the self-efficacy concerning algebra (model below the line, RMSEA 0.32, CFI 0.976, SRMR 0.43). We used b_i to denote the unstandardised regression coefficients, $SE(b_i)$ for their standard errors, and β_i for the standardised regression coefficients.

Since we already observed that the mathematics self-concept scale has different properties than the self-efficacy scales, we analyse the relationship between these concepts, defining linear models in the same way as above, but now using the mathematics self-concept as depending variable (Table 18.8).

The result is remarkable. The self-concept is explained by only two significant predictors: the self-efficacy concerning applied mathematics and calculus. That may be an explanation why the mathematics self-concept acts in different way than the self-efficacy scales: it is mostly determined by two self-efficacy scales that are irrelevant for explaining the students' performance (Table 18.7).

18.7 Conclusion

The first application of the new scales on mathematics self-efficacy in the context of a bridging course shows the following results: firstly, the five scales of this model (related to algebra, applied mathematics, calculus, analytic geometry, and

probability) are short and easy to apply, have good statistical properties, and are substantially, but not too highly correlated, so that they describe different facets of interrelated concepts. Secondly, they allow analysing group differences in details. For example, the self-concept did not allow detecting gender differences; the self-efficacy scales, on the contrary, showed a large difference specifically related to applied mathematics and—not having a similar large counterpart in mathematics performance—suggesting that this difference is “exaggerated”. Thirdly, the scales are useful tools to describe and to explain changes related to beliefs and to mathematics performance. Whereas the mathematics self-concept is more stable and obviously a part of a student’s central beliefs system, the scales of self-efficacy are more peripheral, so that they are appropriate to detect short-term changes as shown by the analyse of the two-week bridging course. Lastly, the self-efficacy scales can explain a considerable amount of both students’ mathematics performance and self-concept, revealing that the self-concept is mostly determined by the students’ *self-efficacy* related to *applied mathematics* and to *calculus*, whereas their *performance* could be explained best by their *algebraic self-efficacy*. Insofar, the scales of self-efficacy can clarify relationships between different mathematically relevant beliefs and concepts.

All results have to be understood against to the background that the two-week bridging course was the first and very limited opportunity to apply and to test the scales. This fact induces some limitations: until now, there is no information about the marks and the drop out of the participants during their upcoming academic courses; there are also no data about the long-term effects on the participants’ performance, beliefs, and emotions; and there was no possibility to control the tasks given during the bridging course. Especially the latter would be desirable to check if the predominant role of the algebraic self-efficacy as a predictor of the students’ performance is a general fact or if this role is limited to the specific content of the bridging course at Hildesheim.

References

- Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, *84*, 191–215.
- Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Upper Saddle River: Prentice Hall.
- Beaujean, A. (2014). *Latent variable modeling using R – A step-by-step guide*. New York: Routledge.
- Betz, N. E., & Hackett, G. (1983). The relationship of mathematics self-efficacy expectations to the selection of science-based college majors. *Journal of Vocational Psychology*, *23*(3), 329–345.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.) Mahwah: Lawrence Erlbaum Associates.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*, 297–334.
- Di Martino, P., & Gregorio, F. (2018). The first-time phenomenon: Successful students’ mathematical crisis in secondary-tertiary transition. In E. Bergqvist, M. Österholm, C. Granberg,

- & L. Sumpter (Eds.), *Proceedings of the 42nd Conference of the International Group for the Psychology of Mathematics Education* (Vol. 2, pp. 339–346). Taipei: PME.
- Girnat, B. (2018). The PISA mathematics self-efficacy scale: Questions of dimensionality and a latent class concerning algebra. In H. Palmér & J. Skott (Eds.) *Students' and teachers' values, attitudes, feelings and beliefs in mathematics classrooms* (pp. 89–99). New York: Springer. https://doi.org/10.1007/978-3-319-70244-5_9
- Hamann, T., Kreuzkam, S., Sander, J., Schmidt-Thieme, B., & de Wiljes, J.-H. (2014). The first academic year – Steps on the way to mathematics. In B. Ubuz, C. Haser, & M. A. Mariotti (Eds.), *Proceedings Cerme 8* (pp. 2292–2293). Ankara: Middle East Technical University.
- Hannula, M. S., Leder, G., Morselli, F., Vollstedt, M., & Zhang, Q. (Eds.) (2019). *Affect and mathematics education: Fresh perspectives on motivation, engagement, and identity*. (ICME-13 Monographs). Springer. <https://doi.org/10.1007/978-3-030-13761-8>.
- Kreuzkam, S. (2013). Mangel an mathematischen Routinefähigkeiten – Basiswissen Mathematik (Lack of mathematical routine skills – Basic mathematics). In G. Greefrath, F. Käpnick, & M. Stein (Eds.), *Beiträge zum Mathematikunterricht* (Vol. 1, pp. 564–567). Münster: WTM Verlag.
- Linden, W. (2016). Unidimensional logistic response models. In W. Linden (Ed.), *Handbook of item response theory. Vol. 1: Models* (pp. 13–30). Boca Raton: CRC Press.
- Marsh, H. W. (1990). *Self-description questionnaire (SDQ) II: A theoretical and empirical basis for the measurement of multiple dimensions of adolescent self-concept: An interim test manual and a research monograph*. New York: The Psychological Corporation.
- Multon, K. D., Brown, S. D., & Lent, R. W. (1991). Relation of self-efficacy beliefs to academic outcomes: A meta-analytic investigation. *Journal of Counselling Psychology*, 38, 30–38. American Psychological Association
- Nicholas, J., Poladian, L., Mack, J., & Wilson, R. (2015). Mathematics preparation for university: Entry, pathways and impact on performance in first year science and mathematics subjects. *International Journal of Innovation in Science and Mathematics Education*, 23(1), 37–51.
- OECD (2005). *PISA 2003 – Technical report*. Paris: OECD Publishing.
- OECD (2013). *PISA 2012 Results: Ready to Learn: Students' Engagement, Drive and Self-Beliefs* (Vol. III). Paris: OECD Publishing.
- OECD (2014). *PISA 2012 – Technical report*. Paris: OECD Publishing.
- Philipp, R. A. (2007). Mathematics teachers' beliefs and affect. In F. K. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 257–315). Charlotte: Information Age Publishing.
- R Core Team (2018). *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Robitzsch, A., Kiefer, T., & Wu, M. (2018). *TAM: Test analysis modules*. R package version 3.0–21.
- Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, 48(2), 1–36. <http://www.jstatsoft.org/v48/i02/>.
- Searle, S. R., & Gruber, M. H. J. (2016). *Linear models* (2nd ed.). Hoboken: Wiley.
- Shavelson, R. J., Hubner, J. J., & Stanton, G. C. (1976). Self-concept – Validation of construct interpretations. *Review of Educational Research*, 46(3), 407–441.
- Stocking, M., & Lord, F. M. (1983). Developing a common metric in item response theory. *Applied Psychological Measurement*, 7, 207–210.