LUDWIG-
MAXIMILIANS-
UNIVERSITÄT
MÜNCHEN

MUNICH CENTER OF THE LEARNING SCIENCES
MCLS REPORTS

Rach, Stefanie; Sommerhoff, Daniel und Ufer, Stefan

KUM/MOAS

Technical Report - Knowledge for University Mathematics (KUM) and Mathematics Online Assessment System (MOAS)

MCLS Report No. 1, 2021

Munich Center of the Learning Sciences
University of Munich
http://www.en.mcls.lmu.de

2021

Technical Report
Knowledge for University Mathematics (KUM)
and
Mathematics Online Assessment System (MOAS)

Stefanie Rach

Daniel Sommerhoff

Stefan Ufer

Citation:

Rach, S., Sommerhoff, D., & Ufer, S. (2021). Technical Report - Knowledge for University Mathematics (KUM) and Mathematics Online Assessment System (MOAS). Magdeburg, Kiel, Munich.

Imprint:

Affiliation:

Mail: stefanie.rach@ovgu.de, sommerhoff@leibniz-ipn.de, ufer@math.lmu.de

ISSN/ISBN: xxxx

Magdeburg, Kiel, Munich, 2021

1 Introduction into KUM and MOAS – two connected projects 6
1.1 Background and motivation 6
1.2 KUM and MOAS as related projects 7
1.3 The KUM project 8
1.4 The MOAS project 8
1.5 KUM and MOAS studies up to now 9
2 Introduction to the presentation of the scales 11
2.1 Presentation of characteristic values 11
2.2 Data cleansing 11
2.3 Naming of the items and values 12
3 Sociodemographic data 13
3.1 Gender 13
3.2 School qualification grade 13
3.3 Grade in the last written exam in mathematics 14
3.4 Last oral grade in mathematics 15
3.5 Last grade in mathematics 16
3.6 Study program 16
4 Motivational and personal characteristics 17
4.1 Interest in mathematics 17
4.1.1 General interest in mathematics 17
4.1.2 Interest in calculation tasks 18
4.1.3 Interest in proving tasks 18
4.2 Self-concept in mathematics 19
4.2.1 General mathematical self-concept 19
4.2.2 Self-concept for calculating tasks 20
4.2.3 Self-concept for proving tasks 20
4.3 Study motives 21
4.3.1 Study motives: Perspective motives 21
4.3.2 Study motives: Application/job motives 22
4.3.3 Study motives: Intrinsic motives 22
4.3.4 Study motives: Scientific motives 23
4.4 Conscientiousness 23
5 Feedback 25
5.1 Review of prior research 25

5.1.1 Feedback content in the MOAS project 25
5.1.2 Determinants of feedback processing in the MOAS project 27
5.2 Students' expectations 29
5.3 Feedback perception 29
5.3.1 Reception of feedback 29
5.3.2 Usefulness of feedback 30
5.3.3 Fit of feedback and perceived performance 31
5.4 Information identified in the feedback 32
5.4.1 Information on expected performance 32
5.4.2 Criterial feedback information 34
5.4.3 Social feedback information 35
5.4.4 Dimensional feedback information 36
5.5 Consequences of feedback perception 38
5.5.1 Information on potential actions and consequences in the feedback 38
5.5.2 Planned actions and consequences based on the feedback 39
5.5.3 Reflection of study choice based on the feedback 40
5.5.4 Stability of study choice in view of feedback 41
6 The KUM scales 43
6.1 Review of prior research and test concept 43
6.1.1 Mathematics knowledge scales 43
6.1.2 Logic scale 45
6.2 Generating level models 49
6.3 Knowledge of analysis 49
6.3.1 Scaling results 49
6.3.2 Items, answer formats, and item parameters 50
6.3.3 Description of knowledge levels 51
6.4 Knowledge of linear algebra 52
6.4.1 Scaling results 52
6.4.2 Items, answer formats, and item parameters 52
6.4.3 Description of knowledge levels 54
6.5 Knowledge of calculus 56
6.5.1 Scaling results 56
6.5.2 Items, answer formats, and item parameters 56
6.5.3 Description of knowledge levels 58

6.6 Knowledge of analytical geometry 59
6.6.1 Scaling results 59
6.6.2 Items, answer formats, and item parameters 59
6.6.3 Description of knowledge levels 60
6.7 Knowledge of logic 61
6.7.1 Scaling results 61
6.7.2 Items, answer formats, and item parameters 62
6.7.3 Description of knowledge levels 64
7 Literature 67

1 Introduction into KUM and MOAS – two connected projects

1.1 Background and motivation

High drop-out rates, in particular in early phases of university mathematics programs, indicate that the transition from school to university poses challenges for beginning undergraduate mathematics students (Dieter, 2012; Heublein & Schmelzer, 2018). These high dropout rates are considered as a serious problem for individual students and for society more generally (Rasmussen & Ellis, 2013). The challenges causing these problems have been described internationally for more than a decade (e.g., Clark & Lovric, 2009; Gueudet, 2008; Tall, 2008; Ulriksen, Møller Madsen & Holmegaard, 2010). In this regard, a row of studies has indicated that the mathematical knowledge students bring from their school education is an important foundation for students' learning processes in their first semesters of university mathematics studies (Hailikari et al. 2008; Kosiol, Rach & Ufer, 2019; Rach & Heinze, 2017; Rach & Ufer, 2020; Ufer, 2015).

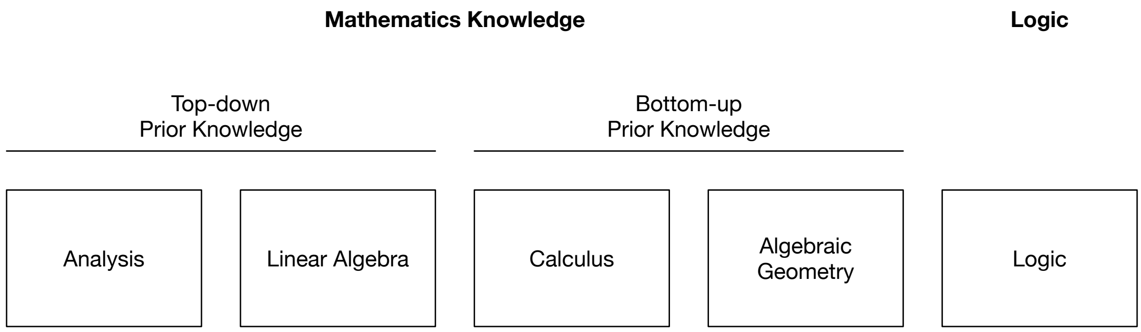
This line of research, focusing on the transition to university mathematics, has repeatedly shown that the *prior knowledge* about mathematics that students possess when entering a university mathematics program explains individual differences in students' success during the transition to university mathematics. Knowledge in this sense comprises individual representations of mathematical concepts and procedures. These mathematical concepts and procedures are defined in, described by, and used in mathematical practice in our society. The corresponding individual representations may be more or less in line with a social consensus of what forms "normatively correct" mathematical knowledge, and thus will lead individuals to provide more or less normatively correct solutions to problems that have to be solved with mathematical means (for this understanding of knowledge see also Greene, Sandoval & Bråten, 2016, p. 4).

However, even though research indicates that prior knowledge is a relevant factor when supporting or counselling future undergraduate mathematics students, further questions arise as to *which* knowledge is actually relevant for a successful transition:

Firstly, one might ask whether distinguishing *different facets* of mathematical knowledge is necessary to understand the relevance of knowledge for a successful transition. Corresponding approaches might structure different measures of prior knowledge based on models discerning different aspects of prior knowledge. For example, different dimensions for different knowledge types such as conceptual, procedural, and strategic knowledge might be differentiated (see Hailikari et al., 2007). However, prior research on the connection between conceptual and procedural knowledge point to a close entanglement of the two dimensions, making it hard to separate them empirically (Rittle-Johnson et al., 2015; Schneider, 2006). A more promising model might be to consider dimensions according to the mathematical content the knowledge refers to, for example knowledge of calculus and analysis (Rach & Ufer, 2020), knowledge of linear algebra (Dorier, J.-L. & Sierpinska, A. 2002; Stewart, 2017), or logical structures (e.g., Durand-Gurrier et al., 2012).

Secondly, based on the identification of (multiple) such knowledge facets, the question arises, which level along a coherent dimension of such a prior knowledge facet actually makes a difference between students coping with the transition to university mathematics successfully and those who do not. For example, Rach & Ufer (2020) have proposed four levels of prior knowledge for undergraduate analysis lectures based on IRT modelling of data from about 1500 future undergraduate mathematics students and identified a level of well-connected school-related knowledge as a soft threshold that differentiates between students who pass the analysis I exam and those who do not.

Thirdly, prior knowledge required during early undergraduate mathematics may be conceptualized from two perspectives: From a top-down perspective, the contents of undergraduate mathematics lectures, in particular mathematical concepts, principles, and procedures that are regarded as essential to comprehend and make sense of the contents introduced in these lectures, can be identified. Under the term *top-down prior knowledge,* we subsume "knowledge about mathematical concepts, that are being used, extended, or reconceptualized during university mathematics studies, and which has been acquired until end of secondary school. Based on cognitivist and constructivist perspectives on learning, learners reconstruct new information encountered in education individually, using their existing knowledge about concepts which are related to the new information. Thus, to study mathematics at university, learners most likely need appropriate prior knowledge to benefit from academic learning opportunities" (Rach & Ufer, 2020, p. 376). Corresponding measurement instruments of *prior knowledge* would embed these contents into items that are similar to situations in university lectures that require the use of this knowledge. From a bottom-up perspective, the contents of the secondary school curriculum may be surveyed for knowledge that arises in undergraduate mathematics and appears essential to cope with undergraduate mathematics. Corresponding measurement instruments of this *bottom-up prior knowledge w*ould embed these contents into items, which are typical for school context.



1.2 KUM and MOAS as related projects

The projects KUM and MOAS aim to contribute to our understanding of the role of prior knowledge in the transition to university mathematics and investigate a mechanism to support future undergraduate mathematics students by providing feedback based on a Mathematical Online Assessment System (MOAS). It is based on the Knowledge for Undergraduate Mathematics (KUM) project that aims at characterizing the relevant knowledge for a successful transition to undergraduate mathematics studies, proposing corresponding theoretical models and developing appropriate test instruments. The MOAS project implements and investigates an online assessment system based on the KUM measures and measures from other projects (e.g., SiSMa, Kosiol, Rach & Ufer, 2019; SEPP, Ufer, 2015).

1.3 The KUM project

The KUM project initially focused on undergraduate analysis courses and strongly built on existing works on the role of knowledge of calculus concepts for learning in undergraduate analysis lectures (Hailikari et al., 2008; Rach & Heinze, 2017). In particular, the tests used in these publications and the four-level model derived by Rach & Ufer (2020) were adapted to a test KUM-A (Analysis) that addresses relevant prior knowledge for analysis lectures on the four levels. However, based on these efforts and achievements, the KUM project now focuses on the transition to university mathematics more broadly, also including mathematical topics other than analysis and also including knowledge regarding other aspects such as logic.

The goals of the KUM project are:

(1) **Developing theoretical level models** to describe potentially relevant top-down prior knowledge for linear algebra lectures (KUM-LA), relevant bottom-up prior knowledge of calculus (KUM-CA) and algebraic geometry (KUM-AG), as well as prior knowledge of logic (KUM-LO).
(2) Developing **reliable and valid instruments** to measure these knowledge facets that can be used in an online testing environment.
(3) Investigating the **validity of the assumed theoretical models in terms of dimensionality** of the five scales, as well as the proposed knowledge levels.
(4) Investigating to which extent the proposed knowledge facets **explain individual differences in learning outcomes** in undergraduate mathematics programs beyond other measures, such as school grades, individual interest, and individual self-concept.

At point of writing, a first final version of the theoretical models and test instruments have been developed. This report primarily presents results on the evaluation of the developed instruments.

1.4 The MOAS project

The MOAS project builds on KUM in the sense that the KUM instruments are embedded into a Mathematics Online Assessment System (MOAS). Based on feedback models (Hattie & Timperley, 2007) and results on the role of formative feedback in learning (Harks et al., 2014), the system not only measures participants' knowledge regarding different facets, but also provides individual feedback based on the measured performance. In particular, the level models developed within KUM for each knowledge facet allow to provide criterion-oriented feedback, connecting students' scores to their performance on specific items.

The main goals of the MOAS project are:

(1) Developing feedback messages for each level of each knowledge facet, which provide students with **formative feedback** on their current level as diagnosed by the assessment system, as well as directions for further learning.
(2) Implementing the KUM scales and the provision of the feedback messages into an **adaptive online assessment system** that requires minimal testing by adaptively selecting tasks based on real-time predictions of students' knowledge levels.
(3) **Investigating students' expectations towards the system**, as well as their processing of the feedback and their (intended) actions based on the feedback.

(4) Investigating the **effects and possible added value of criterion-oriented and social-comparative feedback**.

At the time of writing, the feedback messages have been developed, and the implementation of the knowledge scales in an adaptive testing system is finalized. A first study on questions (1) to (4) has been conducted.

1.5 KUM and MOAS studies up to now

(1) KUM pilot studies

In spring 2018, $N = 26$ secondary school students in their final year as well as university students in their first two years from LMU Munich participated in a study to pilot the bottom-up prior knowledge scales KUM-CA and KUM-AG. The main goal of the study was to investigate the psychometric properties of the items and revise the items and scales for the later inclusion in MOAS. For validating items of the scale KUM-LA, two Bachelor students conducted an interview study with 34 secondary school students in February 2020. The participants were asked to work on eight of the multiple-choice items and explain their answers. The results indicate that the items are valid in the sense that the students' reported reasons for choosing an attractor or a distractor fitted to the chosen distractor and our assumptions made during item design.

(2) KUM Scaling study

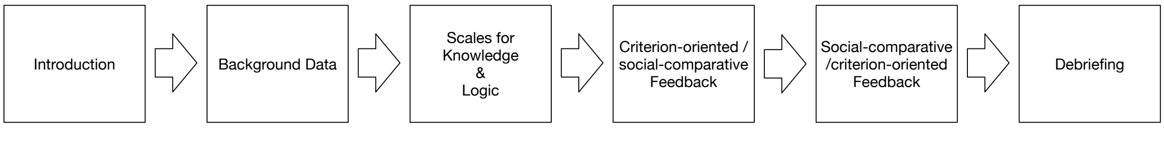
In autumn 2018, a total of $N = 182$ future undergraduate mathematics students from LMU Munich participated in the KUM scaling study during a preparatory course for university mathematics before starting the first semester.

The participants of the course completed the five knowledge scales in three sessions spread over three days to reduce fatigue effects. For each scale, two item booklets, which differed by reverse item sequences, were used. The data was used to validate the level models for the five scales, to calculate item parameters, and to derive cut-off values on the corresponding IRT scales for the level boundaries.

(3) MOAS study I

In autumn 2020, future undergraduate mathematics students from LMU Munich, University of Regensburg, and Otto-von-Guericke-University Magdeburg ($N = 188$) participated in a first study on the MOAS system.

The study consisted of the following six phases:



1. **Introduction** to the MOAS system and the goals of the study (excluding the comparison of the two feedback forms), consent to participation in the study.
2. Providing relevant **background data**, and reporting expectations on the feedback provided by the MOAS system (questionnaires).

3. Working on each of the **five knowledge scales** in the adaptive online assessment system, for a maximum of 10 minutes per instrument.
4. Receiving either the **criterion-oriented or the social-comparative feedback** (random assignment) and answering questions on the individual processing of the feedback and their (intended) actions based on the feedback.
5. Receiving the **other feedback type** and answering the same questions with a focus on this feedback, again.
6. **Debriefing** on the goals of the study.

2 Introduction to the presentation of the scales

For each scale, a short overview about the scale is given, followed by a more detailed description (similar structure to Carstensen et al., 2020).

Name of the variable:	Name of the scale and stem of the single item names in the dataset
Measurement point:	When we used the scale
Prompt in the tool (or booklet or questionnaire or test):	Prompt for the participants
Scaling:	Description of the response format in the tool
Reversed Items:	Items that were reversed before computing the scale
Source:	Author and year of the publication of the scale respectively items. "KUM – own development" means that the scale was developed by members of the project
Notes:	Important information concerning special features of the items, the scale, or the analysis of data

2.1 Presentation of characteristic values

For measuring control variables, we mainly used published questionnaires and adopted them for this project. The presented scale values are the results of calculating the mean values of the single items. For the calculation of the mean values, we only included those participants that had dealt with more than half of the items. The descriptive statistics of the items (mean values, standard deviations, corrected item-total-correlations) are presented for the reversed items. Moreover, there is information concerning the reliability (Cronbachs α) and the descriptive statistics of the scales.

2.2 Data cleansing

In the main study, N_{inv} = 441 students were invited to participate. Repeated participation of students was prevented by using individual codes that were required to participate in MOAS and that were distributed to the students prior to the studies. The students were informed about the study and data protection and data use regulations. They could only participate in the study after explicit consent. There were N_{raw} = 244 cases in the raw data set. All of them completed the background questionnaires and started at least one knowledge test item. *N* = 229 completed all five knowledge tests. *N = 215* viewed the first feedback and completed the questionnaire on this feedback. *N = 188* viewed the second feedback and completed the questionnaire on this feedback. All further analyses are based on this sample of *N = 188* complete datasets. 125 of these students were enrolled at the LMU Munich, 47 at the University of Regensburg, and 16 at the OvGU Magdeburg.

2.3 Naming of the items and values

Abbreviations for scales and items	
R	Reversed (needs to be reversed for the calculation of the overall scale)
Abbreviations of statistical parameters	
M	Mean values, rounded to two decimal places
SD	Standard deviations, rounded to two decimal places
N	Number of participants who worked on the item
$r_{it\text{-}i}$	Corrected item-total-correlations, rounded to two decimal places
α or WLE	Cronbachs α or WLE of the scale, rounded to two decimal places

All items were implemented in German and translated in English for this manual. For research projects, the items of the questionnaires can be made available upon request.

3 Sociodemographic data

3.1 Gender

Name of the variable:	demo_gender
Prompt in the tool	Which gender do you assign yourself to?
Scaling:	Multiple choice
Reversed Items:	0
Source:	“KUM – own development”
Notes:	none

Answers	*N*
female	94
male	93
divers	1

3.2 School qualification grade

Name of the variable:	demo_abitur
Prompt in the tool	Overall qualification grade (1.0-4.0)
Scaling:	open
Reversed Items:	0
Source:	“KUM – own development”
Notes:	In German upper secondary school, grades for individual tests or oral grades are given on a scale from 0 (worst) to 15 (best). These grades are aggregated and rescaled to grades from 1.0 (best) to 4.0 (worst) when calculating overall grades.

Answers	*N*
1.0	18
1.1	12
1.2	11
1.3	8
1.4	15
1.5	11
1.6	8
1.7	9
1.8	7
1.9	17
2.0	14
2.1	6
2.2	7
2.3	8
2.4	5
2.5	5
2.6	5

2.7	3
2.8	5
2.9	5
3.0	3
3.1	2
3.2	2
3.3	0
3.4	2
3.5	0
3.6	0
3.7	0
3.8	0
3.9	0
4.0	0

3.3 Grade in the last written exam in mathematics

Name of the variable:	demo_Mschriftlich
Prompt in the tool	Last written grade in mathematics (0-15 points)
Scaling:	open
Reversed Items:	0
Source:	"KUM – own development"
Notes:	In German upper secondary school, grades for individual tests or oral grades are given on a scale from 0 (worst) to 15 (best).

Answers	*N*
0	0
1	1
2	2
3	2
4	1
5	4
6	4
7	2
8	6
9	12
10	18
11	12
12	18
13	26
14	32
15	43

3.4 Last oral grade in mathematics

Name of the variable:	demo_Mmuendlich
Prompt in the tool	Last oral grade in mathematics (0-15 points)
Scaling:	open
Reversed Items:	0
Source:	“KUM – own development”
Notes:	In German upper secondary school, grades for individual tests or oral grades are given on a scale from 0 (worst) to 15 (best).

Answers	*N*
0	4
1	2
2	2
3	0
4	0
5	0
6	0
7	0
8	0
9	5
10	7
11	10
12	14
13	26
14	38
15	65

3.5 Last grade in mathematics

Name of the variable:	demo_Mzeugnis
Prompt in the tool	Last grade in mathematics (0-15 points)
Scaling:	open
Reversed Items:	0
Source:	"KUM – own development"
Notes:	In German upper secondary school, grades for individual tests or oral grades are given on a scale from 0 (worst) to 15 (best).

Answers	*N*
0	1
1	1
2	3
3	0
4	1
5	1
6	1
7	3
8	8
9	7
10	10
11	20
12	20
13	26
14	47
15	35

3.6 Study program

Name of the variable:	demo_study
Prompt in the tool	What degree program are you enrolled in?
Scaling:	open
Reversed Items:	0
Source:	"KUM – own development"
Notes:	none

Answers	*N*
Bachelor mathematics	60
Bachelor business mathematics	33
Teacher education program, primary level	4
Teacher education program, lower secondary level	6
Teacher education program, upper secondary level	60
Teacher education program, vocational school	1
Others (mainly STEM-related programs)	24

4 Motivational and personal characteristics

4.1 Interest in mathematics

4.1.1 General interest in mathematics

Name of the variable:	bg_ial
Prompt in the tool	Your attitudes concerning mathematics. Please rate the following statements on a scale from disagree to agree.
Scaling:	0 = disagree, 1 = somewhat disagree, 2 = somewhat agree, 3 = agree
Reversed Items:	0
Source:	Pekrun, Goetz, Titz, and Perry (2002)
Notes:	None

Name of the item	*M*	*SD*	r_{it-i}
bg_ial1	2.77	0.48	.46
bg_ial2	2.42	0.62	.65
bg_ial3	2.45	0.65	.56
bg_ial4	1.50	0.82	.30
bg_ial5	2.28	0.63	.59

Scale			
N	*M*	*SD*	α
188	2.29	0.45	.73

Sample Item	
bg_ial1	Mathematics is fun to me.

4.1.2 Interest in calculation tasks

Name of the variable:	bg_iakr
Prompt in the tool	see 4.1.1
Scaling:	0 = disagree, 1 = somewhat disagree, 2 = somewhat agree, 3 = agree
Reversed Items:	0
Source:	SISMa – own development (see Ufer, Rach & Kosiol, 2017)
Notes:	none

Name of the item	*M*	*SD*	$r_{it\text{-}i}$
bg_iakr1	2.38	0.66	.47
bg_iakr2	2.07	0.82	.51
bg_iakr3	2.05	0.88	.49
bg_iakr4	2.32	0.67	.74
bg_iakr5	2.61	0.55	.33

Scale			
N	*M*	*SD*	α
188	2.29	0.51	.74

Sample Item	
bg_iakr2	I like to deal with complicated calculations.

4.1.3 Interest in proving tasks

Name of the variable:	bg_iakb
Prompt in the tool	see 4.1.1
Scaling:	0 = disagree, 1 = somewhat disagree, 2 = somewhat agree, 3 = agree
Reversed Items:	0
Source:	SISMa – own development (see Ufer et al., 2017)
Notes:	none

Name of the item	*M*	*SD*	$r_{it\text{-}i}$
bg_iakb1	1.78	0.85	.65
bg_iakb2	2.14	0.80	.73
bg_iakb3	1.72	0.80	.60
bg_iakb4	2.29	0.74	.66

Scale			
N	*M*	*SD*	α
188	1.98	0.65	.83

Sample item	
bg_iakb3	Reading mathematical proofs is fun to me.

4.2 Self-concept in mathematics

4.2.1 General mathematical self-concept

Name of the variable:	bg_ska
Prompt in the tool	Your attitudes concerning mathematics. Please rank the following statements on a scale from disagree to agree.
Scaling:	0 = disagree, 1 = somewhat disagree, 2 = somewhat agree, 3 = agree
Reversed Items:	0
Source:	Kauper, Retelsdorf, Bauer, Rösler, Möller, & Prenzel (2012)
Notes:	none

Name of the item	*M*	*SD*	r_{it-i}
bg_ska1	2.12	0.65	.69
bg_ska2	2.04	0.73	.75
bg_ska3	2.16	0.74	.69
bg_ska4	2.03	0.61	.60

Scale			
N	*M*	*SD*	α
188	2.09	0.57	.84

Sample item	
bg_ska2	I am very good in mathematics.

4.2.2 *Self-concept for calculating tasks*

Name of the variable:	bg_skr
Prompt in the tool	see 4.2.1
Scaling:	0 = disagree, 1 = somewhat disagree, 2 = somewhat agree, 3 = agree
Reversed Items:	1
Source:	SISMa – own development (see Rach, Kosiol & Ufer, 2019)
Notes:	none

Name of the item	*M*	*SD*	$r_{it\text{-}i}$
bg_skr1 (reversed)	1.95	0.70	.55
bg_skr2	1.74	0.68	.61
bg_skr3	2.11	0.64	.42
bg_skr4	1.89	0.71	.63
bg_skr5	1.66	0.66	.43

Scale			
N	*M*	*SD*	α
188	1.87	0.48	.76

Sample item	
bg_skr1 (reversed)	I often miscalculate when dealing with complicated terms or equations.

4.2.3 *Self-concept for proving tasks*

Name of the variable:	bg_skb
Prompt in the tool	see 4.2.1
Scaling:	0 = disagree, 1 = somewhat disagree, 2 = somewhat agree, 3 = agree
Reversed Items:	1
Source:	SISMa – own development (see Rach et al., 2019)
Notes:	none

Name of the item	*M*	*SD*	$r_{it\text{-}i}$
bg_skb1	1.65	0.72	.57
bg_skb2 (reversed)	1.77	0.77	.66
bg_skb3	1.63	0.67	.74
bg_skb4	1.43	0.64	.63

Scale			
N	*M*	*SD*	α
188	1.62	0.57	.82

Sample item	
bg_skb3	It is easy for me to understand mathematical proofs.

4.3 Study motives

4.3.1 Study motives: Perspective motives

Name of the variable:	bg_swm_e
Prompt in the tool	Your study choice. Please rank the following statements on a scale from disagree to agree. I chose to study mathematics because …
Scaling:	0 = disagree, 1 = somewhat disagree, 2 = somewhat agree, 3 = agree
Reversed Items:	0
Source:	Ufer (2015)
Notes:	none

Name of the item	*M*	*SD*	r_{it-i}
bg_swm_e1	1.01	0.94	.60
bg_swm_e2	1.12	0.99	.77
bg_swm_e3	1.53	0.99	.63

Scale			
N	*M*	*SD*	α
188	1.22	0.83	.82

Sample Item	
bg_swm_e2	… I will earn a lot of money as a mathematician.

4.3.2 Study motives: Application/job motives

Name of the variable:	bg_swm_a
Prompt in the tool	I chose to study mathematics because …
Scaling:	0 = disagree, 1 = somewhat disagree, 2 = somewhat agree, 3 = agree
Reversed Items:	0
Source:	Ufer (2015)
Notes:	none

Name of the item	*M*	*SD*	r_{it-i}
bg_swm_a1	2.20	0.79	.18
bg_swm_a2	1.55	1.03	.58
bg_swm_a3	2.09	0.93	.58

Scale			
N	*M*	*SD*	α
188	1.95	0.69	.62

Notes: After 2020, we replace the item bg_swm_a1, which doesn't fit to the other two items of the scale, by the item: "… I will learn many things which will be important in my future job."

Sample item	
bg_swm_a3	… I will be well prepared for my future job.

4.3.3 Study motives: Intrinsic motives

Name of the variable:	bg_swm_i
Prompt in the tool	I chose to study mathematics because …
Scaling:	0 = disagree, 1 = somewhat disagree, 2 = somewhat agree, 3 = agree
Reversed Items:	0
Source:	Ufer (2015)
Notes:	none

Name of the item	*M*	*SD*	r_{it-i}
bg_swm_i1	2.13	0.66	.33
bg_swm_i2	2.51	0.63	.55
bg_swm_i3	2.49	0.67	.49

Scale			
N	*M*	*SD*	α
188	2.38	0.50	.64

Sample item	
bg_swm_i2	… I like to deal with questions in mathematics.

4.3.4 Study motives: Scientific motives

Name of the variable:	bg_swm_w
Prompt in the tool	I chose to study mathematics because …
Scaling:	0 = disagree, 1 = somewhat disagree, 2 = somewhat agree, 3 = agree
Reversed Items:	0
Source:	Ufer (2015)
Notes:	none

Name of the item	*M*	*SD*	r_{it-i}
bg_swm_w1	1.99	0.92	.62
bg_swm_w2	1.78	0.95	.68
bg_swm_w3	2.19	0.85	.33

Scale			
N	*M*	*SD*	α
188	1.99	0.72	.71

Sample item	
bg_swm_w2	… I want to learn about current research in mathematics.

4.4 Conscientiousness

Name of the variable:	bg_gewi
Prompt in the tool	How do you work? Please rank the following statements on a scale from disagree to agree.
Scaling:	0 = disagree, 1 = somewhat disagree, 2 = somewhat agree, 3 = agree
Reversed Items:	1
Source:	Dehne & Schupp (2007).
Notes:	none

Name of the item	*M*	*SD*	r_{it-i}
bg_gewi1	2.41	0.67	.47
bg_gewi2 (reversed)	1.84	0.88	.41
bg_gewi3	2.24	0.60	.22

Scale			
N	*M*	*SD*	α
188	2.16	0.52	.54

Sample item	
bg_gewi1	I am someone who works thoroughly.

5 Feedback

After assessing multiple knowledge facets, MOAS provides students with individual feedback for each of these knowledge facets based on their scores. In particular, students receive criterial and social feedback. To evaluate the effectiveness of either type of feedback and improve feedback in future versions of MOAS, students were asked to rate the quality of either type of feedback regarding several aspects. In this manual, we first give a review of prior research that the feedback within MOAS as well as the scales to evaluate the feedback were based upon. Afterwards, we report on the quality of these scales separately for every type of feedback (criterial and social).

5.1 Review of prior research

Feedback describes instructional interventions that provide learners with information and an assessment of their current or prior performance, often meant to stimulate future learning processes that reduce the gap between current and expected performance (Hattie & Timperley, 2007). To describe feedback content and feedback processing in the MOAS project, we build on the following models.

5.1.1 Feedback content in the MOAS project

Hattie and Timperley (2007) describe the content of feedback messages. In particular, they point out that feedback should provide answers to three major questions:

(1) *Where am I going?*
This refers to information about goals, that is the results of learning expected in the respective learning setting.

(2) *How am I going?*
This refers to information about students' current performance and progress towards these goals.

(3) *Where to next?*
This refers to information about activities that need to be undertaken to make further progress towards these goals.

Beyond this, Hattie and Timperley (2007) distinguish between feedback on different levels that relate to the goals resp. expected learning results:

a. *Task level*
Information about how well tasks are performed or understood.

b. *Process level*
Information about the main processes required to perform or understand the task.

c. *Self-regulation level*
Information that addresses self-monitoring and self-regulation of the learners' learning actions.

d. *Self level*
Information that contains general evaluations of the learner or affective messages towards the learner.

In their review of feedback studies, Hattie and Timperley (2007) mention that feedback on the self level has little effect on learning outcomes. Feedback on the task level is considered most promising for simple tasks, for which the learners have sufficient knowledge to find a correct solution strategy by themselves. However, it is considered less effective for more complex tasks, unless it directs learners' attention towards the solution process. Feedback on the task level has little transfer effects. Feedback on the process level may, for example, trigger learners to search for errors, to activate conceptual knowledge about the task, and to initiate a restructuring of conceptual knowledge. Feedback on the self-regulation level is also considered effective if the task has substantial self-regulatory demands and if the feedback directs the learners towards sustained self-regulated work on the learning tasks. However, feedback on the self-regulation level requires that learners have favorable attitudes towards the task and self-regulated learning techniques.

In view of the Hattie and Timperley model, MOAS intends to provide feedback on each of the three feedback questions: (1) What is the level of knowledge that is considered sufficient to productively engage in a university mathematics program (e.g., Rach & Ufer, 2020)? (2) In relation to this level of knowledge, how can the current prior knowledge of the learner be described? (3) Which actions might be helpful to develop the learners' prior knowledge towards the next higher knowledge level? The main type of feedback in MOAS is criterial feedback (resp. "sachliche Normen") that compares individual performance to a reference performance (Hattie and Wollenschläger, 2014; Kopp and Mandl, 2014). For each prior knowledge facet, the criterial feedback in MOAS first provides a process-level feedback of the students' skills. This information is based on the knowledge level for the respective facet, to which the student has been assigned based on her or his test performance. It primarily comprises information about the knowledge that seems to be rightly available and the kinds of tasks the student can already solve (question 2). Moreover, the student receives information about problems and knowledge facets that go beyond her or his current knowledge level (questions 1 and 2). This information is based on the knowledge level above the one the student was assigned to. Additionally, the feedback contains an evaluation of the student's current knowledge level in relation to the level that is considered necessary for a successful entry into a mathematics program (based on Rach & Ufer, 2020 for top-down prior knowledge of analysis, and parallel considerations for the other facets, question 1). Finally, question 3 is addressed by providing hints for further study, and two exemplary tasks that might be worth studying. In line with typical implementations of social feedback, this feedback type primarily provided information on students' achievement in relation to a comparison group (question 2).

Hattie and Wollenschläger (2014) as well as Kopp and Mandl (2014) point out that feedback content may also draw on other norms to assess learners' current performance. Beyond criterial feedback, comparing individual performance to a reference performance (see above), they differentiate a social norm that compares individual performance to the distribution of performance in a group of learners, and an individual norm that compares the current individual performance with prior individual performance. In MOAS, we investigated a social norm feedback besides the criterial norm feedback to find out which norm is more useful for students. The social feedback in MOAS reports students' individual knowledge level for each facet as a grade on an equivalent scale to German school grades (from 1="very good" to 5="inadequate"; the worst level 6="insufficient" was not used in this feedback) and presents a table with the frequency of each grade that would be expected for a typical classroom with 30 students. The frequency table data was based on the MOAS pilot studies.

In the MOAS main study, all participants received feedback with a social norm as well as feedback with a criterial norm separately and one after the other. The sequence of the two types (first criterial then social or vice versa) was randomized.

5.1.2 Determinants of feedback processing in the MOAS project

Even though feedback research has strongly focused on how feedback is delivered and what it should contain, the call to consider how learners actually *process* feedback has increased over the last years (Strijbos & Müller, 2014; Narciss, 2013). In particular, Narciss's (2013) model for interactive tutoring feedback proposes to differentiate between an internal feedback loop and external feedback. As part of the internal feedback loop, the learner evaluates her or his perceived performance (individual feedback) against her or his individual goals. This process may be influenced by external feedback if it stimulates the comparison of the external feedback with either the perceived performance or the individual goals. Furthermore, it is plausible that feedback may also trigger comparisons between individual goals and goals communicated in the feedback message (Where am I going?).

In particular, feedback is intended to lead to self-evaluations of the learner based on the feedback message that in turn should trigger further learning processes (e.g., Narciss, 2013). Depending on which norm (social, criterial, individual) is used to assess students' performance in the feedback message, different internal comparisons may be triggered. In the context of self-concept development, the internal/external frame of reference model (e.g., Marsh et al., 2015) differentiates *social comparisons* (corresponding to a social norm), *criterial comparisons* (corresponding to a criterial norm), *temporal comparisons* (corresponding to an individual norm), and *dimensional comparisons*. *Dimensional comparisons* also correspond to comparisons within individual learners and thus a kind of individual norm but refer to comparisons between performance in different fields or tasks (and not regarding prior individual performance). For example, a learner might perceive his or her performance in science as higher than his or her performance in languages, and consequently arrive at a positive self-evaluation in terms of science performance.

Such comparison processes as well as the further actions taken based on the feedback processing may be influenced not only by characteristics of the feedback message and its presentation but also by a number of personal characteristics. Strijbos and Müller (2014) highlight attributional processes, in which learners explain potential differences between individual performance and individual or external goals. The authors point out that these attributional processes are closely connected to learners' self-efficacy expectations regarding the task resp. their individual self-concept, which both are strongly connected to their past performance. Summarizing, they argue that learners with high self-efficacy perceive negative feedback as less threatening and are likely to increase their efforts to reach the goal. Learners with low self-efficacy, in contrast, tend to attribute negative feedback to internal and stable factors, which may decrease efforts towards the intended goal – in particular if the goal appears to be out of reach for the learner. Finally, focusing on the current knowledge about the relation of person characteristics and the use of feedback, the authors conclude that, for example, higher agreeableness may decrease the relation between emotions and the individual processing of feedback, while higher conscientiousness is generally assumed to have a positive impact on feedback processing. Based on these prior works, the MOAS project investigates plausible predictors of feedback processing, including individual performance, self-concept and interest, and conscientiousness.

A possible factor influencing the processing of the feedback message (including comparison, generation, and selection of control actions; Narciss, 2013) is the perception of the feedback, in particular its *perceived usefulness* (e.g., Harks et al., 2014a). Harks et al. (2014b) find that criterial feedback was perceived as more useful than social feedback (called process-oriented resp. grade-oriented feedback in Harks' works), and the usefulness totally mediated the relation between feedback type and achievement resp. interest. The effect of perceived usefulness on intrinsic learning motivation was stronger for learners with higher interest and weaker for learners with higher self-concept (Harks et al., 2014b). Considering the comparisons assumed in Narciss's (2013) model, it seems plausible that the *fit between individually perceived performance and external assessment* of the performance might increase feedback acceptance and processing.

A more direct measure of feedback processing might be the extent to which students actually report to be /are able to excerpt and draw on information that is included (resp. that is not included) in the feedback message, such as information on the relevant learning or performance *goals* in the corresponding context (Where am I going?), on *assessments* of the own performance based on criterial, social, and dimensional norms (How am I going?), or on *potential actions* that might be taken based on the feedback (Where to next?).

However, being able to take up the relevant information included in feedback does not guarantee that a learners' future actions actually reflect that information and may thus still be a too distal measure for feedback. The actual effects of feedback (beyond the processing of the information therein and the subsequent comparison processes) might thus be measured best based on the degree to which *concrete actions* are planned based on the feedback, whether the feedback stimulates a *reflection* of individual plans, for example the choice of a study program, or if the feedback affects the *stability of the decision* for a concrete study program. That is, focusing on whether the learners actually plan to change their behavior based on the feedback may be the most proximal measure of feedback effectiveness.

The MOAS draws on a set of instruments to measure students' feedback perception and processing. These instruments mainly comprise self-developed questionnaire scales.

5.2 Students' expectations

At the very beginning of the system, the students state their expectations concerning the feedback using single items.

Name of the variable:	stu_exp
Prompt in the tool	What do you expect from MOAS? I expect to …
Scaling:	0 = disagree, 1 = somewhat disagree, 2 = somewhat agree, 3 = agree
Source:	KUM – own development
Notes:	none

Name of the item	*M*	*SD*
exp_criterial1	2.68	0.56
exp_criterial2	2.71	0.57
exp_dimensional	2.21	0.80
exp_social	1.62	1.02

	Items
exp_criterial1	… know which tasks I can already solve in each topic.
exp_criterial2	… know what I still have to learn about every topic.
exp_dimensional	… be able to compare my performances between different topics.
exp_social	… know from the feedback, how well I am doing in each topic compared to the other students.

5.3 Feedback perception

5.3.1 Reception of feedback

Name of the variable:	stu_exp_c
Prompt in the tool	How do you experience the feedback?
Scaling:	0 = disagree, 1 = somewhat disagree, 2 = somewhat agree, 3 = agree
Source:	KUM – own development
Notes:	none

Name of the item	*M*	*SD*
fb_reze_read_c	2.63	0.61
fb_reze_comp_c	2.52	0.61

Items	
fb_reze_read_c	I carefully read the feedback.
fb_reze_comp_c	I understood the feedback.

Name of the variable:	stu_exp_s
Prompt in the tool	How do you experience the feedback?
Scaling:	0 = disagree, 1 = somewhat disagree, 2 = somewhat agree, 3 = agree
Source:	KUM – own development
Notes:	none

Name of the item	*M*	*SD*
fb_reze_read_s	2.54	0.69
fb_reze_comp_s	2.62	0.57

Items	
fb_reze_read_s	I carefully read the feedback.
fb_reze_comp_s	I understood the feedback.

5.3.2 Usefulness of feedback

Name of the variable:	feed_useful_c
Prompt in the tool	see 5.3.1
Scaling:	0 = disagree, 1 = somewhat disagree, 2 = somewhat agree, 3 = agree
Reversed Items:	1
Source:	adapted from different sources, e.g. Rakoczy et al., 2019
Notes:	none

Name of the item	*M*	*SD*	r_{it-i}
fb_qual_hful_c	2.37	0.71	.81
fb_qual_help_c	2.22	0.79	.77
fb_qual_nuse_c (reversed)	2.38	0.83	.48
fb_qual_usef_c	2.35	0.70	.80
fb_qual_valu_c	2.09	0.78	.61

Scale			
N	*M*	*SD*	α
188	2.28	0.61	.86

Sample item	
fb_qual_help_c	The feedback is helpful.

Name of the variable:	feed_useful_s
Prompt in the tool	see 5.3.1
Scaling:	0 = disagree, 1 = somewhat disagree, 2 = somewhat agree, 3 = agree
Reversed Items:	1
Source:	adapted from different sources, e.g. Rakoczy et al., 2019
Notes:	none

Name of the item	*M*	*SD*	r_{it-i}
fb_qual_hful_s	2.26	0.76	.77
fb_qual_help_s	2.10	0.81	.76
fb_qual_nuse_s (reversed)	2.18	0.96	.40
fb_qual_usef_s	2.16	0.77	.75
fb_qual_valu_s	1.97	0.80	.68

Scale			
N	*M*	*SD*	α
188	2.13	0.65	.85

Sample item	
fb_qual_help_s	The feedback is helpful.

5.3.3 Fit of feedback and perceived performance

Name of the variable:	feed_fit_c
Prompt in the tool	see 5.3.1
Scaling:	0 = disagree, 1 = somewhat disagree, 2 = somewhat agree, 3 = agree
Reversed Items:	0
Source:	KUM – own development
Notes:	none

Name of the item	*M*	*SD*	r_{it-i}
fb_fit_know_c	1.72	0.83	.74
fb_fit_perf_c	2.06	0.82	.47
fb_fit_self_c	1.77	0.80	.57
fb_fit_skil_c	1.68	0.80	.71

Scale			
N	*M*	*SD*	α
188	1.81	0.65	.81

Sample item	
fb_fit_skil_c	The feedback correctly reflects my mathematical skills.

Name of the variable:	feed_fit_s
Prompt in the tool	see 5.3.1
Scaling:	0 = disagree, 1 = somewhat disagree, 2 = somewhat agree, 3 = agree
Reversed Items:	0
Source:	KUM – own development
Notes:	none

Name of the item	*M*	*SD*	r_{it-i}
fb_fit_know_s	1.55	0.83	.75
fb_fit_perf_s	2.07	0.83	.53
fb_fit_self_s	1.60	0.77	.60
fb_fit_skil_s	1.48	0.84	.75

Scale			
N	*M*	*SD*	α
188	1.68	0.67	.83

Sample item	
fb_fit_skil_s	The feedback correctly reflects my mathematical skills.

5.4 Information identified in the feedback

5.4.1 Information on expected performance

Name of the variable:	feed_exp_c
Prompt in the tool	What do you take from the feedback?
Scaling:	0 = disagree, 1 = somewhat disagree, 2 = somewhat agree, 3 = agree
Reversed Items:	0
Source:	KUM – own development
Notes:	none

Name of the item	*M*	*SD*	r_{it-i}
fb_goal_expkn_c	2.28	0.75	.73
fb_goal_expsk_c	2.25	0.79	.73
fb_goal_impkn_c	2.13	0.80	.73
fb_goal_impsk_c	2.21	0.75	.67
fb_goal_reqkn_c	2.20	0.78	.79
fb_goal_reqsk_c	2.21	0.76	.76

Scale			
N	*M*	*SD*	α
188	2.21	0.63	.90

Sample item	
fb_goal_impsk_c	The feedback shows which skills are important in a mathematics study program for me.

Name of the variable:	feed_exp_s
Prompt in the tool	What do you take from the feedback?
Scaling:	0 = disagree, 1 = somewhat disagree, 2 = somewhat agree, 3 = agree
Reversed Items:	0
Source:	KUM – own development
Notes:	none

Name of the item	*M*	*SD*	r_{it-i}
fb_goal_expkn_s	2.10	0.83	.78
fb_goal_expsk_s	2.06	0.84	.78
fb_goal_impkn_s	1.88	0.84	.84
fb_goal_impsk_s	1.74	0.92	.73
fb_goal_reqkn_s	1.87	0.88	.83
fb_goal_reqsk_s	1.81	0.87	.78

Scale			
N	*M*	*SD*	α
188	1.91	0.74	.93

Sample item	
fb_goal_impsk_s	The feedback shows which skills are important in a mathematics study program for me.

5.4.2 Criterial feedback information

Name of the variable:	feed_crit_c
Prompt in the tool	What do you take from the feedback?
Scaling:	0 = disagree, 1 = somewhat disagree, 2 = somewhat agree, 3 = agree
Reversed Items:	0
Source:	KUM – own development
Notes:	none

Name of the item	*M*	*SD*	r_{it-i}
fb_stat_crit1_c	2.32	0.74	.60
fb_stat_crit2_c	2.11	0.78	.63
fb_stat_crit3_c	2.18	0.72	.68
fb_stat_crit4_c	2.43	0.68	.77
fb_stat_crit5_c	2.30	0.76	.74

Scale			
N	*M*	*SD*	α
188	2.27	0.59	.86

Sample item	
fb_stat_crit4_c	The feedback shows me which skills I should develop concerning the different topics.

Name of the variable:	feed_crit_s
Prompt in the tool	What do you take from the feedback?
Scaling:	0 = disagree, 1 = somewhat disagree, 2 = somewhat agree, 3 = agree
Reversed Items:	0
Source:	KUM – own development
Notes:	none

Name of the item	*M*	*SD*	r_{it-i}
fb_stat_crit1_s	1.77	1.04	.71
fb_stat_crit2_s	1.63	1.02	.78
fb_stat_crit3_s	1.76	0.95	.76
fb_stat_crit4_s	1.86	0.95	.86
fb_stat_crit5_s	1.69	1.04	.73

Scale			
N	*M*	*SD*	α
1881	1.74	0.85	.91

Sample item	
fb_stat_crit4_s	The feedback shows me which skills I should develop concerning the different topics.

5.4.3 Social feedback information

Name of the variable:	feed_sozi_c
Prompt in the tool	see 5.4.2
Scaling:	0 = disagree, 1 = somewhat disagree, 2 = somewhat agree, 3 = agree
Reversed Items:	0
Source:	adapted from different sources, e.g. Gibbs, 2003
Notes:	none

Name of the item	*M*	*SD*	r_{it-i}
fb_stat_sozi1_c	1.00	1.05	.91
fb_stat_sozi2_c	1.06	1.06	.93
fb_stat_sozi3_c	1.05	1.02	.93
fb_stat_sozi4_c	1.07	1.00	.90

Scale			
N	*M*	*SD*	α
188	1.05	0.98	.97

Sample item	
fb_stat_sozi1_c	The feedback shows me how good I am compared to other participants of MOAS.

Name of the variable:	feed_sozi_s
Prompt in the tool	see 5.4.2
Scaling:	0 = disagree, 1 = somewhat disagree, 2 = somewhat agree, 3 = agree
Reversed Items:	0
Source:	adapted from different sources, e.g. Gibbs, 2003
Notes:	none

Name of the item	*M*	*SD*	r_{it-i}
fb_stat_sozi1_s	2.57	0.69	.77
fb_stat_sozi2_s	2.59	0.68	.84
fb_stat_sozi3_s	2.53	0.72	.84
fb_stat_sozi4_s	2.41	0.76	.73

Scale			
N	*M*	*SD*	α
188	2.52	0.63	.91

Sample item	
fb_stat_sozi1_c	The feedback shows me how good I am compared to other participants of MOAS.

5.4.4 Dimensional feedback information

Name of the variable:	feed_dime_c
Prompt in the tool	see 5.4.2
Scaling:	0 = disagree, 1 = somewhat disagree, 2 = somewhat agree, 3 = agree
Reversed Items:	0
Source:	KUM – own development
Notes:	none

Name of the item	*M*	*SD*	r_{it-i}
fb_stat_dime1_c	2.32	0.70	.70
fb_stat_dime2_c	2.32	0.67	.81
fb_stat_dime3_c	2.20	0.74	.78
fb_stat_dime4_c	2.28	0.73	.76

Scale			
N	*M*	*SD*	α
188	2.28	0.62	.89

Sample item	
fb_stat_dime3_c	The feedback shows me concerning which topics I am stronger or weaker.

Name of the variable:	feed_dime_s
Prompt in the tool	see 5.4.2
Scaling:	0 = disagree, 1 = somewhat disagree, 2 = somewhat agree, 3 = agree
Reversed Items:	0
Source:	KUM – own development
Notes:	none

Name of the item	*M*	*SD*	r_{it-i}
fb_stat_dime1_s	2.27	0.67	.69
fb_stat_dime2_s	2.25	0.71	.76
fb_stat_dime3_s	2.23	0.78	.73
fb_stat_dime4_s	2.24	0.69	.80

Scale			
N	*M*	*SD*	α
188	2.25	0.61	.88

Sample item	
fb_stat_dime3_s	The feedback shows me the topics that I am stronger or weaker in.

5.5 Consequences of feedback perception

5.5.1 Information on potential actions and consequences in the feedback

Name of the variable:	feed_con_info_c
Prompt in the tool	What does the feedback mean for your preparation concerning the study program? The statements refer to the current reading of the feedback as well as to the use of the feedback in near future.
Scaling:	0 = disagree, 1 = somewhat disagree, 2 = somewhat agree, 3 = agree
Reversed Items:	0
Source:	adapted from Nieskens et al., 2011
Notes:	none

Name of the item	*M*	*SD*	$r_{it\text{-}i}$
fb_con_inf1_c	2.13	0.81	.68
fb_con_inf2_c	1.98	0.81	.67
fb_con_inf3_c	2.14	0.80	.73
fb_con_inf4_c	2.36	0.68	.65

	Scale		
N	*M*	*SD*	α
188	2.15	0.64	.84

Sample item	
fb_con_inf1_c	Feedback gives me hints to prepare for a mathematics study program.

Name of the variable:	feed_con_info_s
Prompt in the tool	What does the feedback mean for your preparation concerning the study program? The statements refer to the current reading of the feedback as well as to the use of the feedback in the near future.
Scaling:	0 = disagree, 1 = somewhat disagree, 2 = somewhat agree, 3 = agree
Reversed Items:	0
Source:	adapted from Nieskens et al., 2011
Notes:	none

Name of the item	*M*	*SD*	r_{it-i}
fb_con_inf1_s	1.58	1.02	.77
fb_con_inf2_s	1.49	0.98	.79
fb_con_inf3_s	1.60	0.97	.73
fb_con_inf4_s	1.81	0.96	.66

Scale			
N	*M*	*SD*	α
188	1.62	0.84	.88

Sample item	
fb_con_inf1_s	Feedback gives me hints to prepare for a mathematics study program.

5.5.2 Planned actions and consequences based on the feedback

Name of the variable:	feed_con_act_c
Prompt in the tool	see 5.5.1
Scaling:	0 = disagree, 1 = somewhat disagree, 2 = somewhat agree, 3 = agree
Reversed Items:	0
Source:	KUM – own development
Notes:	none

Name of the item	*M*	*SD*	r_{it-i}
fb_con_act1_c	2.23	0.83	.75
fb_con_act2_c	2.14	0.87	.86
fb_con_act3_c	2.22	0.89	.74
fb_con_act4_c	2.14	0.87	.86

Scale			
N	*M*	*SD*	α
188	2.18	0.77	.91

Sample item	
fb_con_act1_c	Because of the feedback, I will repeat again some of the content before the study program starts.

Name of the variable:	feed_con_act_s
Prompt in the tool	see 5.5.1
Scaling:	0 = disagree, 1 = somewhat disagree, 2 = somewhat agree, 3 = agree
Reversed Items:	0
Source:	KUM – own development
Notes:	none

Name of the item	*M*	*SD*	r_{it-i}
fb_con_act1_s	2.15	0.95	.73
fb_con_act2_s	2.03	0.93	.79
fb_con_act3_s	2.18	0.88	.75
fb_con_act4_s	2.06	0.90	.86

	Scale		
N	*M*	*SD*	α
188	2.11	0.81	.90

Sample item	
fb_con_act1_s	Because of the feedback, I will repeat again some of the content before the study program starts.

5.5.3 Reflection of study choice based on the feedback

Name of the variable:	feed_con_ref_c
Prompt in the tool	see 5.5.1
Scaling:	0 = disagree, 1 = somewhat disagree, 2 = somewhat agree, 3 = agree
Reversed Items:	0
Source:	adapted from different sources, e.g. Köller et al., 2017 and Nieskens et al., 2011
Notes:	none

Name of the item	*M*	*SD*	r_{it-i}
fb_con_ref1_c	0.75	0.91	.84
fb_con_ref2_c	0.69	0.89	.86
fb_con_ref3_c	0.62	0.83	.86
fb_con_ref4_c	1.09	1.02	.73

	Scale		
N	*M*	*SD*	α
188	0.79	0.82	.92

Sample item	
fb_con_ref1_c	Because of the feedback, I will think about my study choice once more.

Name of the variable:	feed_con_ref_s
Prompt in the tool	see 5.5.1
Scaling:	0 = disagree, 1 = somewhat disagree, 2 = somewhat agree, 3 = agree
Reversed Items:	0
Source:	adapted from different sources, e.g. Köller et al., 2017 and Nieskens et al., 2011
Notes:	none

Name of the item	*M*	*SD*	r_{it-i}
fb_con_ref1_s	0.80	0.97	.84
fb_con_ref2_s	0.70	0.90	.87
fb_con_ref3_s	0.65	0.91	.86
fb_con_ref4_s	1.05	1.00	.77

Scale			
N	*M*	*SD*	α
188	0.80	0.86	.93

Sample item	
fb_con_ref1_s	Because of the feedback, I will think about my study choice once more.

5.5.4 Stability of study choice in view of feedback

Name of the variable:	feed_con_stab_c
Prompt in the tool	see 5.5.1
Scaling:	0 = disagree, 1 = somewhat disagree, 2 = somewhat agree, 3 = agree
Reversed Items:	1
Source:	adapted from different sources, e.g. Köller et al., 2017 and Nieskens et al., 2011
Notes:	none

Name of the item	*M*	*SD*
fb_con_stab1_c	1.54	0.93
fb_con_stab2_c (reversed)	1.84	1.00

Scale			
N	*M*	*SD*	*r*
188	1.69	0.77	.41

Sample item	
fb_con_stab1_c	The feedback confirms my study choice for a program with a focus on mathematics.

Name of the variable:	feed_con_stab_s
Prompt in the tool	see 5.5.1
Scaling:	0 = disagree, 1 = somewhat disagree, 2 = somewhat agree, 3 = agree
Reversed Items:	1
Source:	adapted from different sources, e.g. Köller et al., 2017 and Nieskens et al., 2011
Notes:	none

Name of the item	*M*	*SD*
fb_con_stab1_s	1.54	0.94
fb_con_stab2_s (reversed)	1.83	1.01

Scale			
N	*M*	*SD*	*r*
188	1.68	0.80	.50

Sample item	
fb_con_stab1_s	The feedback confirms my study choice for a program with a focus on mathematics.

6 The KUM scales

The items and scales for mathematical knowledge were generated against the background of two aims of MOAS, that is i) to model knowledge regarding different facets and levels and then ii) to analyze the role of different facets of knowledge for being successful in university.

6.1 Review of prior research and test concept

6.1.1 Mathematics knowledge scales

Prior research on the secondary-tertiary transition has highlighted students' mathematical knowledge as one important factor for the success of this transition. The importance mirrors cognitivist and constructivist perspectives on learning, as learners reconstruct new information encountered at university based on their existing knowledge about concepts that are related to the new information. Thus, to study mathematics at university, learners most likely need appropriate prior knowledge to benefit from academic learning opportunities.

To investigate the impact of student's mathematical knowledge, it thus seems mandatory to identify knowledge that may be required as a basis during these learning processes as well as a basis for coping with the demands of typical situations during the first year of mathematics-related degree programs and corresponding lectures. As analysis and linear algebra are usually the two central content areas in the first year of mathematics programs, knowledge related to these areas appears as essential. As described above, we conceptualized prior knowledge regarding each of these two areas from two perspectives:

i) *Top-Down perspective*
Based on the contents of undergraduate mathematics lectures, mathematical concepts, principles, and procedures that are regarded as essential to comprehend and make sense of the contents are identified. Corresponding measurement instruments of *top-down prior knowledge* embed these contents in items similar to later situations in university, thus approximating later practices.

ii) *Bottom-Up perspective*
Based on the contents of the secondary school curriculum, mathematical concepts, principles, and procedures that are central in the school context and regarded as essential to comprehend and make sense of undergraduate mathematics lectures are identified. Corresponding measurement instruments of *bottom-up prior knowledge* embed these contents in items similar to known situations from the school context.

Based on these perspectives, the first step for item generation were theoretical analyses of i) the mathematical content covered in typical analysis and linear algebra lectures (top-down perspective) and ii) the mathematical content covered in secondary school (bottom-up perspective). In both perspectives, we identified mathematical knowledge that is in the focus of the school context or the university context and has a theory-based potential for high predictivity for success in early undergraduate mathematics, in particular real analysis and linear algebra lectures. These analyses were also based on more general research findings in the context of the transition from school to university, for example, the results of MaLeMINT (Deeken et al., 2020). Based on this background, we decided to design four different scales:

- Knowledge of Calculus (KUM-CA; Bottom-up perspective)
- Knowledge of Analytic Geometry (KUM-AG; Bottom-up perspective)
- Knowledge of Analysis (KUM-A; Top-down perspective)
- Knowledge of Linear Algebra (KUM-LA; Top-down perspective)

For each scale, multiple different mathematical topics were selected for item generation, as an exhaustive assessment of all contents, especially regarding the bottom-up perspective, was impossible due to time constraints and also appeared unreasonable as some secondary mathematics contents appeared more likely to be predictive for being successful in university than other mathematical contents. For example, students' handling of inequalities, in particular also including norms/absolute values, was included as an important prerequisite for university mathematics as corresponding knowledge is highly important to keep up in introductory university mathematics courses, for example regarding the concept of ε-δ-continuity. In contrast, knowledge about elementary geometry, for example regarding the formula for the volume of a cone, was regarded as less relevant for students' success in undergraduate mathematics courses. Summing up, we assume the following topics as relevant for success in university mathematics programs (some concepts belong to both fields): Calculus & Analysis – real number, term, inequality, function, limit, derivation, integral; Analytic Geometry & Linear Algebra – equation, linear equation system, vector, linear independent, linear combination, vector operation, line, straight, number set, group.

Prior research from mathematics education on the secondary-tertiary transition has repeatedly highlighted features of school mathematics that differ from features of mathematics as it occurs in first-year university mathematics programs (e.g., Tall, 2008). School-mathematics, according to these works, is more related to working with specific examples, calculation procedures, and the application of these procedures on more or less authentic real-world problems. Considering this description through models of knowledge qualities, school mathematics might be characterized as putting a focus on automatization of procedures and often on more informal connections between different entities of a knowledge network. In terms of a systematic hierarchical ordering of mathematical concepts, school mathematics might be considered less "deep" than university mathematics but more intensively connected to meaningful representations of the concepts. University mathematics, in contrast, is often considered more "abstract", focusing less on examples and more on the specific mathematical relations between the concepts and generic representations of these concepts. These findings at least partially correspond to the distinction of different qualities of knowledge (e.g., de Jong and Ferguson-Hessler, 1996), in particular the distinction of conceptual knowledge ("knowing that") and procedural knowledge ("knowing how") (see Förtsch et al, 2018).

Differentiating knowledge of facts and knowledge of procedures has a long tradition in psychology (e.g., Anderson, 1983) and resonates in mathematics education, for example in Skemp's (1976) distinction between relational understanding (similar to conceptual knowledge) and instrumental understanding (similar to a superficial form of procedural knowledge). Even though concrete definitions vary, conceptual knowledge usually refers to a network of general facts, concepts, and principles while procedural knowledge covers sequences of mental or concrete actions to achieve a specific goal (cf. Rittle-Johnson et al., 2015). With respect to measurement, it is widely agreed that procedural knowledge will show on tasks in a domain which participants have solved frequently before whereas conceptual knowledge is best assessed with unfamiliar tasks (Rittle-Johnson et al., 2015). While procedural knowledge is usually restricted to solve well-delineated types of problems or sub-problems, conceptual knowledge can be applied more flexibly and broadly across a range of familiar and unfamiliar tasks. Other works use the distinction between "declarative knowledge" and "compiled knowledge" or "encapsulated knowledge" (De Jong and Ferguson-Hessler, 1996; Schmidt and Rikers, 2007). Both conceptual knowledge and procedural knowledge are assumed to occur in declarative as well as compiled or encapsulated forms. Similar to the works of Jukic and Dahl (2012), we are interested in distinguishing different types of knowledge and therefore use the terms conceptual and procedural knowledge. Approaches from mathematics education have proposed an integrated modelling of conceptual and procedural knowledge, for example as procepts (Gray and Tall, 1994), or their mutual relations, for example the process of treating a known procedure as a new mental object (reification, Sfard, 1991). Overall, research has underlined that it is difficult to empirically separate conceptual and procedural knowledge (Schneider, 2006; Rittle-Johnson et al., 2015) and that the distinction may not be meaningful for contents that are well learned.

Taking these considerations and distinctions into account, it did not appear meaningful for the MOAS project to further differentiate the scales and the generated items regarding conceptual and procedural knowledge. Still, the scales focusing on students' mathematical knowledge from a bottom-up perspective may be more closely connected to the notion of procedural knowledge, as it is measured in rather familiar tasks similar to those from the school context. In contrast, students' mathematical knowledge from a top-down perspective may be more closely connected to the notion of conceptual knowledge, as it is measured in rather unfamiliar tasks similar to the situations students may later encounter at university.

Finally, for the scales focusing on students' mathematical knowledge from a bottom-up perspective, task complexity related to arithmetic demands and the number of steps or procedures that must be executed to solve the task has been considered explicitly. Williams and Clarke (1997) refer to this as "numerical complexity", which is driven by the types and combinations of operations required to perform a task and is independent of the conceptual complexity of the task, which is based on the specific concepts handled in the task (see further Stillman & Galbraith, 2003). As an elementary example, the conceptual difficulty of solving the tasks 136:4 and 123456789876544:4 is equivalent as both tasks can be solved via a division algorithm, however the arithmetic demand of the latter task is higher, thus leading to a higher item complexity (see further Pantsar, 2019). Moreover, also the status of the tasks as routine and non-routine tasks, based on the curricula for secondary education, was considered. In particular, task complexity was considered lower if the task rather represents a routine task, that is a demand that has already been encountered before, whereas the complexity of non-routine task, i.e. tasks that can be classified as a problem (in the sense of problem-solving research; see e.g., Dörner, 1979; Schoenfeld, 1985), was considered higher.

6.1.2 Logic scale

Knowledge of logic as addressed in KUM primarily comprises knowledge underlying normatively correct logical reasoning. The focus is primarily on procedural and strategic knowledge and less on conceptual knowledge about logical constructs or corresponding notations. In this vein, verbal logical reasoning tasks are used to measure knowledge.

Logical reasoning has been studied in connection to learning in upper secondary and undergraduate mathematics from two perspectives. From a perspective of the theory of formal discipline (Attridge & Inglis, 2013; Handley et al., 2007; Inglis & Simpson, 2008, 2009; Morsanyi, Kahl, & Rooney, 2017; Morsanyi, McCormack, & O'Mahony, 2018), the assumption has been studied to which extent studying mathematics contributes to building up abstract logical reasoning skills. From a perspective of prerequisites for successful learning in undergraduate studies, logical reasoning has been studied as a predictor of success on specific mathematical tasks, for example proving (Sommerhoff, 2017). Less research is available on the role of logical reasoning for succeeding in undergraduate mathematics courses. One goal of including knowledge about logic in KUM is to address this gap.

Regarding this open question, a range of logical structures can be assumed to be relevant for undergraduate mathematics learning. Introductory mathematics courses and transition courses usually focus on what actually characterizes a valid mathematical statement, for example structural features of such statements such as junctors (and, or, implication, equivalence) and quantifiers (in particular universal and existence quantifiers) (e.g., Reichersdorfer et al., 2014).

Conditional reasoning: In terms of prior research, the gradual development of conditional reasoning from primary school to adolescence has attracted substantial attention (Janveau-Brennan & Markovits, 1999). From this perspective, logical reasoning on tasks covering the four basic logical forms of conditional reasoning (MP: Modus Ponens, MT: Modus Tollens, DA: Denial of the Antecendent, AC: Acceptance of the Consequent, cf. Datsogianni et al., 2020) has been studied. Results indicate that even elementary school students show valid conditional reasoning on some logical forms in specific familiar contexts (e.g., MP tasks in categorical contexts such as "If an animal is a cat, then it has legs"; Markovits, 2000; Markovits & Thompson, 2008). Other studies have shown that only about a third of adult participants systematically answered other forms (DA, AC) in a normatively correct way (Christophorides, Spanoudis, & Demetriou, 2016; Gauffroy & Barrouillet, 2009; Moshman, 1990; Markovits, 2014; Ricco, 2010). Strongest problems are typically observed for less familiar contexts, such as counterfactual ("If you throw a feather at a window, then it will break."), artificial ("On planet varius, if the trees swibble, the weather will frase.") or abstract ("If A, then B") contexts. Even though conditional reasoning can be considered to be at the heart of mathematical proof, Sommerhoff (2017) found no relation between conditional reasoning in an abstract context and undergraduate students' proof skills.

Less research is available on students' skills to identify the equivalence or non-equivalence of two given conditionals. For example, it has been reported repeatedly in the mathematics education literature that students do not differentiate between an implication and its converse (e.g., Küchemann & Hoyles, 2009). In this context, the equivalence of a conditional (if p, then q) and its contrapositive (if not q, then not p) is a central logical relation, while other conditionals (if not p, then not q; if q, then p) are not equivalent to the original statement (if p, then q). However, little systematic research is available on the identification of equivalent and non-equivalent statements. Other junctors than implications in conditional statements (e.g., conjunction, disjunction) have been used less frequently in the past to assess logical reasoning (cf. Leighton, 2004) so that these structures have not been considered for the KUM at the current point of development.

Reasoning with syllogisms and quantified statements: Regarding quantifiers, a long-standing line of psychological research covers syllogistic reasoning. Categorical syllogisms usually consist of two statements which describe a relation between a *subject* and a *predicate* implicitly via a *middle term* (e.g., "All M are P. All S are M."), while the correct conclusion reflects this relation explicitly ("All S are P."; cf. Leighton, 2004). The relations are expressed by one of four *moods* that structurally relate loosely to logical quantifiers: "all A are B" (universal quantifier), "some A are B" (existence quantifier), "no A are B" (universal quantifier and negation) and "some A are not B" (existence quantifier and negation). In syllogistic reasoning tasks, usually a number (e.g., three) alternative conclusions and the option "no inference can be made" is provided for participants to select the correct answer from. There is evidence that participants answer tasks with counterfactual or improbable prerequisites by drawing on their contextual knowledge about the subject, the predicate, the middle term and their relations, rather than using the normatively correct interpretation of the statements (cf. Leighton, 2004 for an overview). Some researchers argue that this reflects a form of pragmatic or adaptive rationality which may be more relevant for decisions in ill-defined, complex real-world tasks (Evans & Feeney, 2004) than making normatively correct inferences. However, for success in mathematics degree programs which focus on strictly defined concepts and their logical relations, reasoning in line with the normative interpretation of syllogistic statements can be assumed to be a relevant skill.

As for conditionals, identifying equivalent and non-equivalent statements involving quantifiers may be a relevant task in undergraduate mathematics learning. In particular statements involving one quantifier and a negation ("For all x not p(x)." is equivalent to "There is no x so that p(x).") or the combination of two quantifiers ("For all x, there is a y, such that p(x,y)." is a weaker statement than "There is a y, such that for all x we have p(x,y).") have a number of equivalent and non-equivalent forms. In advanced undergraduate calculus or "analysis" courses, statements involving a combination of quantifiers frequently occur, for example, when epsilon-definitions of convergence or continuity are introduced. Dealing flexibly with equivalent forms of these kinds of statements and their negations is necessary in many proofs and justifications in this context. However, beyond the study of Barkai et al., (2009) on teacher education students' reasoning with single quantifiers, only little research on undergraduate students' skills in dealing with quantifiers is available.

Based on the described results, the following decisions were made in the design of the KUM logic test:

1. *Logical structures*
   Regarding logical structures, the test items are restricted to conditionals, existence and universal quantifiers, and negations. In syllogistic reasoning tasks, the traditional wording of the existence quantifier (some … are …) was used to keep the connection to prior research. In other tasks, existence quantifiers were explicitly presented in the

form of verbal statements (there is a … such that …) as it is frequent in mathematical practice (and not quantifier symbols such as $\exists, \forall$). Separate task sets addressed conditionals and quantifiers, with negations occurring in both of the task sets.

2. *Task types*

   For both logical structures, two task types were designed:

   (A) Inference tasks which present two statements and a number of possible conclusions together with the option "no conclusion is possible". Participants are asked to select the normatively correct conclusion from the alternatives. For conditional reasoning tasks, a conditional was provided and a statement about the antecedent or the consequent of the conditional. In this way, all four logical forms were covered. The answer alternatives asked whether the other part of the conditional (consequent of antecedent) could be concluded to be true, to be false, or if no conclusion was possible. For inference tasks with quantifiers, typical syllogistic reasoning tasks were used.

   (B) Equivalence judgement tasks which present two statements made by two persons (here Hans and Petra). The participants were asked to judge whether the statements were equivalent ("If Hans is right, then Petra is right, and vice versa."), or Hans' statement was stronger than Petra's ("If Hans is right, then Petra is right, but not vice versa."), or Petra's statement was stronger than Hans' ("If Petra is right, then Hans is right, but not vice versa."), or if the two statements were logically independent. Equivalence judgement tasks for conditionals involved a conditional (e.g., with the structure "If not A, then B.") and either its contrapositive ("If not B, then A."), its converse ("If B, then not A."), or the converse of its contrapositive ("If A, then not B."). Equivalence judgement tasks for quantifiers either contained two statements with one quantifier and a negation each or two statements with two quantifiers (with and without negations). Equivalent and non-equivalent statement pairs were included.

3. *Contexts*

   In terms of contexts, four different kinds of contexts were used: (A) everyday contexts (example item: "If it rains, the street will get wet."), (B) artificial contexts ("On planet varius, if the trees swibble, the weather will frase."), (C) valid mathematical contexts ("If a function f has an extreme value at the position x, then f'(x) = 0.") and (D) pseudo-mathematical contexts ("For every nice function f, there is a position x, such that f(x) = 2."; participants were instructed that "nice functions" should be assumed to be some definable mathematical term and that at least one "nice function" exists).

   Context types (A)-(C) were implemented for conditional reasoning and conditional equivalence judgement tasks. Syllogistic reasoning tasks involved artificial contexts with everyday objects (with statements such as "All Chinese are Teachers."). Equivalence judgement tasks for quantifiers only covered pseudo-mathematical (D) contexts.

The current items in the KUM logic tests are a selection of a larger item universe. The selection was made to cover all major facets of the item model but was still restricted to a number of tasks that could administered in a scaling study.

6.2 Generating level models

All knowledge tests were scaled based on the sample from the scaling study which took place in an introductory course for future university mathematics students (bachelor programs in mathematics and in financial mathematics, upper secondary mathematics teacher education program) at the LMU Munich. As the scales were distributed over three consecutive days, the sample for each scaling varied from 125 to 142 students. Two booklets, containing the same items in reverse orders, were used to reduce sequency effects and effects of missing data. The raw data was scored dichotomously and scaled with the one-dimensional Rasch model using the R package TAM (Robitzsch et al., 2020).

To distinguish different qualities of knowledge within each of the five knowledge facets, we aimed to identify and characterize levels of knowledge. A prominent approach to develop such levels is the bookmark procedure (Mitzel et al., 2001): The items are sorted by their empirical difficulties and are analyzed regarding contrasting demands of the items against the background of the theoretical frameworks (described in each section below). This method leads to a verbal description of the knowledge levels and a list of corresponding items which can be solved using knowledge on the respective level (cf. Rach & Ufer, 2020 for the analysis knowledge test).

6.3 Knowledge of analysis

6.3.1 Scaling results

Scale	*Knowledge of analysis*
WLE mean	*0.00
WLE sd	1.09
WLE reliability	0.77
EAP reliability	0.79
MNSQ infit – mean	1.00
MNSQ infit – max	1.18
MNSQ infit – min	0.83
MNSQ outfit – mean	1.08
MNSQ outfit – max	2.95
MNSQ outfit – min	0.73
item parameter mean	-0.36
item parameter sd	1.40
item parameter max	2.91
item parameter min	-3.50
item parameter SE mean	0.24

* EAP parameters were restrained to zero for scaling.

6.3.2 Items, answer formats, and item parameters

Name of the variable	*Response format*	*Threshold value*	*Level*	*Concept*
score_moas_analysis_1	Single choice	-3.50	1	Derivations
score_moas_analysis_2	Single choice	-1.50	1	Real numbers
score_moas_analysis_3	Complex choice	0.24	3	Derivations
score_moas_analysis_4	Single choice	-0.02	2	Real numbers
score_moas_analysis_5	Single choice	-1.14	2	Real numbers
score_moas_analysis_6	Single choice	-1.99	1	Functions
score_moas_analysis_7	Single choice	-1.11	2	Derivations
score_moas_analysis_8	Single choice	-1.56	1	Functions
score_moas_analysis_9	Open	-2.38	1	Functions
score_moas_analysis_10	Complex choice	1.15	4	Functions
score_moas_analysis_11	Complex choice	-0.52	2	Functions
score_moas_analysis_12	Open	0.04	3	Derivations
score_moas_analysis_13	Open	1.69	4	Equations
score_moas_analysis_14	Complex choice	0.37	3	Functions
score_moas_analysis_15	Single choice	0.02	3	Series
score_moas_analysis_16	Complex choice	0.46	3	Functions
score_moas_analysis_17	Complex choice	0.26	3	Functions
score_moas_analysis_18	Complex choice	-1.19	2	Functions
score_moas_analysis_19	Open	-0.98	2	Functions
score_moas_analysis_20	Complex choice	1.13	4	Functions
score_moas_analysis_21	Complex choice	0.51	3	Derivations
score_moas_analysis_22	Complex choice	0.64	3	Functions
score_moas_analysis_23	Single choice	2.91	4	Derivations
score_moas_analysis_24	Single choice	0.36	3	Sequences
score_moas_analysis_25	Single choice	-1.67	1	Real numbers
score_moas_analysis_26	Single choice	-1.56	1	Rational numbers

6.3.3 Description of knowledge levels

Level	*Level Description*	*# Items*	*Threshold value*	*Item example*
1	Procedural knowledge and knowledge about facts	7	below -1.35	Calculate the first derivation of the function $f: \mathbb{R}^+ \rightarrow \mathbb{R}, f(x) = 3x^4 + x^{\frac{1}{3}} - 6$.
2	Conceptual knowledge incorporating few or disconnected well-known representations	6	-1.35 to -0.00	Let f be $f: \mathbb{R} \rightarrow \mathbb{R}, f(x) = 3^x + 1$. State the y-axis intercept of the graph of f.
3	Connected conceptual knowledge incorporating multiple, connected, but not necessarily formal representations of mathematical concepts	9	0.00 to 0.90	In which interval is the angle α with these conditions: tan(α) > 0 and sin(α) < 0? ☐ $]0°; 90°[= (0°; 90°)$ ☐ $]90°; 180°[= (90°; 180°)$ ☐ $]180°; 270°[= (180°; 270°)$ ☐ $]270°; 360°[= (270°; 360°)$
4	Connected conceptual knowledge, including formal notations and central mathematical practices like proving and defining formally	4	Above 0.90	The value of $\lim_{h \to 0} \frac{\sqrt{2+h}-\sqrt{2}}{h}$ is ☐ 0. ☐ $\frac{1}{2\sqrt{2}}$. ☐ $\frac{1}{\sqrt{2}}$. ☐ ∞.

6.4 Knowledge of linear algebra

6.4.1 Scaling results

Scale	*Knowledge of linear algebra*
WLE mean	*-0.02
WLE sd	1.35
WLE reliability	0.81
EAP reliability	0.82
MNSQ infit – mean	1.00
MNSQ infit – max	1.27
MNSQ infit – min	0.82
MNSQ outfit – mean	1.07
MNSQ outfit – max	3.64
MNSQ outfit – min	0.57
item parameter mean	-0.04
item parameter sd	1.55
item parameter max	3.69
item parameter min	-3.71
item parameter SE mean	0.25

* EAP parameters were restrained to zero for scaling.

6.4.2 Items, answer formats, and item parameters

Name of the variable	*Response format*	*Threshold value*	*Level*	*Concept*
score_moas_linalg_1	Complex choice	-0.82	1	Vector operations
score_moas_linalg_2	Complex choice	-1.46	1	Vector operations
score_moas_linalg_3	Complex choice	-0.24	2	Orthogonal vectors
score_moas_linalg_4	Complex choice	0.99	3	Scalar products
score_moas_linalg_5	Complex choice	1.36	3	Orthogonal vectors
score_moas_linalg_6	Open	0.70	3	Linearly dependent vectors
score_moas_linalg_7	Complex choice	-0.23	2	Linear combinations
score_moas_linalg_8	Complex choice	3.69	4	Linearly dependent vectors
score_moas_linalg_9	Complex choice	2.43	4	Linearly dependent vectors
score_moas_linalg_10	Complex choice	0.86	3	Linearly dependent vectors
score_moas_linalg_11	Complex choice	-0.59	2	Linearly dependent vectors
score_moas_linalg_12	Single choice	-1.23	1	Straights
score_moas_linalg_13	Single choice	-0.30	2	Straights

score_moas_linalg_14	Open	0.01	2	Straights
score_moas_linalg_15	Complex choice	-0.02	2	Linear equation systems
score_moas_linalg_16	Open	-0.62	1	Linear equation systems
score_moas_linalg_17	Single choice	-1.57	1	Linear equation systems
score_moas_linalg_18	Open	-0.55	2	Distances
score_moas_linalg_19	Single choice	-3.71	1	Distances
score_moas_linalg_20	Complex choice	1.13	3	Groups
score_moas_linalg_21	Complex choice	0.07	2	Groups
score_moas_linalg_22	Open	-1.27	1	Groups
score_moas_linalg_23	Open	-2.41	1	Linear functions
score_moas_linalg_24	Complex choice	1.48	4	Linear functions
score_moas_linalg_25	Complex choice	1.22	3	Linear functions

6.4.3 *Description of knowledge levels*

Level	*Level Description*	*# Items*	*Threshold value*	*Item example*
1	Procedural knowledge and knowledge about facts	8	below -0.60	Which operations with two vectors $v_1, v_2 \in \mathbb{R}^2$ have a vector as a result, which operations a number? Mark the operations with a vector as result. ☐ $v_1 + v_2$ ☐ $v_1 \circ v_2$ (scalar product) ☐ $a \cdot v_1$ $(a \in \mathbb{R})$ ☐ $v_1 - v_2$
2	Conceptual knowledge incorporating few or disconnected well-known representations	8	-0.60 to 0.50	Which of the following equation systems have exactly one solution in $\mathbb{R}$? Mark them. ☐ $\begin{matrix} x - y = 1 \\ 3x - 3y = 0 \end{matrix}$ ☐ $\begin{matrix} x - y = 1 \\ 2x + 2y = 6 \end{matrix}$ ☐ $\begin{matrix} x - y = 1 \\ -x + y = -1 \end{matrix}$ ☐ $\begin{matrix} x - y = 1 \\ 2x - 2y = 2 \end{matrix}$

3	Connected conceptual knowledge incorporating multiple, connected, but not necessarily formal representations of mathematical concepts	6	0.50 to 1.40	Two vectors are orthogonal to each other in $\mathbb{R}^2$. Which of the following statements are true? Mark them. ☐ Each vector in $\mathbb{R}^2$ can be presented as a linear combination of the two vectors. ☐ The zero vector is only represented by $0 \cdot v_1 + 0 \cdot v_2$ as a linear combination of the two vectors. ☐ The scalar product of the two vectors is 0. ☐ The intersection angle of v_1 with the x-axis is as large as the intersection angle of v_2 with the x-axis.
4	Connected conceptual knowledge, including formal notations and central mathematical practices like proving and defining formally	3	Above 1.40	Which of the following $v \in \mathbb{R}^2$ are for any positive numbers $a, b > 0$ always linear independent from $\binom{a}{b}$? Mark them. ☐ $\binom{1}{2}$ ☐ $\binom{a-b}{b-a}$ ☐ $\binom{a+b}{b}$ ☐ $\binom{3}{-4}$

Table 1 Levels of knowledge in the KUM-LA test with item examples and difficulty parameters

6.5 Knowledge of calculus

6.5.1 Scaling results

Scale	*Knowledge of calculus*
WLE mean	-0.04*
WLE sd	1.12
WLE reliability	0.72
EAP reliability	0.75
MNSQ infit – mean	0.99
MNSQ infit – max	1.23
MNSQ infit – min	0.81
MNSQ outfit – mean	1.00
MNSQ outfit – max	1.28
MNSQ outfit – min	0.61
item parameter mean	-0.72
item parameter sd	0.93
item parameter max	1.07
item parameter min	-2.76
item parameter SE mean	0.25

*EAP parameters were restrained to zero for scaling.

6.5.2 Items, answer formats, and item parameters

Name of the variable	*Response format*	*Threshold value*	*Level*	*Concept*
score_moas_rvinf_01	Open	0.07	3	Equations
score_moas_rvinf_02	Single Choice	-1.50	2	Terms
score_moas_rvinf_03	Single Choice	-2.60	1	Equations
score_moas_rvinf_04	Single Choice	-0.93	2	Equations
score_moas_rvinf_05	Single Choice	0.22	4	Terms
score_moas_rvinf_06	Single Choice	0.14	4	Functions
score_moas_rvinf_07	Single Choice	0.27	4	Terms
score_moas_rvinf_08	Single Choice	-0.59	3	Inequalities
score_moas_rvinf_09	Single Choice	-0.88	3	Inequalities
score_moas_rvinf_10	Single Choice	-2.76	1	Functions
score_moas_rvinf_11	Single Choice	-1.36	2	Equations
score_moas_rvinf_12	Open	-0.37	3	Calculation Rules
score_moas_rvinf_13	Open	-0.75	3	Functions
score_moas_rvinf_14	Open	-1.83	2	Derivations
score_moas_rvinf_15	Single Choice	0.61	4	Derivations
score_moas_rvinf_16	Single Choice	0.03	3	Derivations
score_moas_rvinf_17	Complex Choice	0.07	4	Calculation Rules
score_moas_rvinf_18	Complex Choice	-0.71	3	Calculation Rules

score_moas_rvinf_19	Complex Choice	0.25	4	Calculation Rules
score_moas_rvinf_20	Single Choice	-1.00	2	Calculation Rules
score_moas_rvinf_21	Single Choice	1.07	4	Calculation Rules
score_moas_rvinf_22	Open	-0.68	3	Integrals
score_moas_rvinf_23	Complex Choice	-0.55	3	Calculation Rules
score_moas_rvinf_24	Single Choice	-0.83	3	Limits
score_moas_rvinf_25	Single Choice	-0.55	3	Functions
score_moas_rvinf_26	Single Choice	-1.23	2	Derivations
score_moas_rvinf_27	Single Choice	-0.48	3	Polynomial Divisions
score_moas_rvinf_28	Single Choice	-1.82	2	Linear Equation Systems
score_moas_rvinf_29	Single Choice	-2.14	2	Linear Equation Systems

6.5.3 Description of knowledge levels

Level	*Level Description*	*# Items*	*Threshold value*	*Item example*
1	Elementary processing of mathematical objects and symbols in routine settings. No difficult calculations required.	2	below -2.5	Given a linear function f whose graph has slope 2 and goes through the point P(1|5). What is the functional equation of f? ☐ $f(x) = 2x + 5$ ☐ $f(x) = 2x + 1$ ☐ $f(x) = 2x + 3$ ☐ $f(x) = 2x - 3$
2	Processing of mathematical objects and symbols in routine settings that require a substantial number of calculations or algebraic operations	8	-2.7 to -0.90	Simplify the term $15x^2 - 5x[(x + y)(3 - y)] + 15xy$
3	Processing of mathematical objects and tasks in known but non-routine situations. Knowledge about specific calculation rules without the need to apply them.	12	-0.90 to 0.07	Let $\mathrm{f(x)} = -\frac{1}{2}x + 7$ be a given function with $\mathrm{x} \in \mathbb{R}$. Determine the inverse function $\mathrm{f}^{-1}(x)$. ☐ $\mathrm{f}^{-1}(x) = -2x + 14$ ☐ $\mathrm{f}^{-1}(x) = 2x - 14$ ☐ $\mathrm{f}^{-1}(x) = -2x - 14$ ☐ $\mathrm{f}^{-1}(x) = 2x + 14$
4	Processing of mathematical objects and tasks which require the application of specific, non-routine conceptual knowledge or strategies to be solved.	7	Above 0.07	Simplify the term $\ln(a) + \ln(b) + 3 \cdot \ln(2) - \ln(8b)$ ☐ $\ln\left(\frac{a}{8}\right) + \ln(6)$ ☐ $\ln\left(\frac{a+8+b}{8b}\right)$ ☐ $\ln(a)$ ☐ $\ln(a - 8 - 7b)$

Table 1 Levels of the KUM-CA test with item examples and difficulty parameters

6.6 Knowledge of analytical geometry

6.6.1 Scaling results

Scale	*Knowledge of analytical geometry*
WLE mean	-0.11*
WLE sd	1.25
WLE reliability	0.38
EAP reliability	0.59
MNSQ infit – mean	1.01
MNSQ infit – max	1.15
MNSQ infit – min	0.87
MNSQ outfit – mean	1.02
MNSQ outfit – max	1.43
MNSQ outfit – min	0.74
item parameter mean	-0.17
item parameter sd	1.36
item parameter max	2.82
item parameter min	-2.27
item parameter SE mean	0.28

* EAP parameters were restrained to zero for scaling.

6.6.2 Items, answer formats, and item parameters

Name of the variable	*Response format*	*Threshold value*	*Level*	*Concept*
score_moas_rvag_01	Open	-1.28	1	Vectors
score_moas_rvag_02	Single Choice	-2.27	1	Vectors
score_moas_rvag_03	Single Choice	-1.53	1	Lines
score_moas_rvag_04	Single Choice	0.07	2	Vector Products
score_moas_rvag_05	Open	1.44	3	Vector Products
score_moas_rvag_06	Open	-0.48	2	Dot Products
score_moas_rvag_07	Single Choice	0.58	3	Planes
score_moas_rvag_08	Single Choice	-0.15	2	Lines
score_moas_rvag_09	Single Choice	-0.97	2	Relative Positions
score_moas_rvag_10	Single Choice	1.17	3	Distances
score_moas_rvag_11	Open	2.82	4	Triple Products
score_moas_rvag_12	Single Choice	-0.18	2	Circles
score_moas_rvag_13	Single Choice	-0.11	2	Spheres
score_moas_rvag_14	Single Choice	-1.53	1	Linear Combinations

6.6.3 Description of knowledge levels

Level	*Level Description*	*# Items*	*Threshold value*	*Item example*
1	Elementary processing of mathematical objects and symbols in routine settings. No difficult calculations required.	4	below -1.0	Let the points $A = \begin{pmatrix} 1 \\ 5 \\ 4 \end{pmatrix}$ and $B = \begin{pmatrix} 7 \\ 3 \\ 6 \end{pmatrix}$ be given. Determine the centre of the line segment $\overline{AB}$.
2	Processing of mathematical objects and symbols in routine settings that require a substantial number of calculations or algebraic operations	6	-1.0 to 0.50	Determine the intersection of lines $g: \vec{X} = \begin{pmatrix} 2 \\ 2 \\ 0 \end{pmatrix} + \lambda \begin{pmatrix} 1 \\ -1 \\ -3 \end{pmatrix}; \lambda \in \mathbb{R}$ and $g: \vec{X} = \begin{pmatrix} 1 \\ -2 \\ -2 \end{pmatrix} + \mu \begin{pmatrix} -1 \\ 2 \\ 4 \end{pmatrix}; \mu \in \mathbb{R}$.
3	Processing of mathematical objects and tasks in known but non-routine situations.	3	0.50 to 2.80	The plane E is created by the vectors $\vec{a} = \begin{pmatrix} -2 \\ -4 \\ 3 \end{pmatrix}$ and $\vec{b} = \begin{pmatrix} -1 \\ -1 \\ 2 \end{pmatrix}$ and goes through the point $\vec{A} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$. Determine the cartesian equation of the plane.
4	Processing of mathematical objects and tasks which require the application of specific, non-routine conceptual knowledge or strategies to be solved.	1	Above 2.80	Determine the volume of the parallelepiped which is determined by the vectors $\vec{a} = \begin{pmatrix} 2 \\ 0 \\ 3 \end{pmatrix}$, $\vec{b} = \begin{pmatrix} -4 \\ 2 \\ -1 \end{pmatrix}$, and $\vec{c} = \begin{pmatrix} -3 \\ 2 \\ 1 \end{pmatrix}$.

Table 2 Levels of the KUM-AG test with item examples and difficulty parameters

6.7 Knowledge of logic

6.7.1 Scaling results

Scale	*Knowledge of logic*
WLE mean	*0.00
WLE sd	1.07
WLE reliability	0.83
EAP reliability	0.85
MNSQ infit – mean	1.00
MNSQ infit – max	1.23
MNSQ infit – min	0.71
MNSQ outfit – mean	1.02
MNSQ outfit – max	1.59
MNSQ outfit – min	0.66
item parameter mean	-0.57
item parameter sd	0.95
item parameter max	1.76
item parameter min	-2.93
item parameter SE mean	0.22

* EAP parameters were restrained to zero for scaling.

6.7.2 Items, answer formats, and item parameters

Name of the variable	*Response format*	*Threshold value*	*Level*	*Logical structure*	*Task type*	*Context*
score_moas_logic_1	Single choice	-0.44	4	conditional	equivalence judgement	everyday
score_moas_logic_2	Single choice	-0.63	3	conditional	equivalence judgement	everyday
score_moas_logic_3	Single choice	-0.02	4	conditional	equivalence judgement	everyday
score_moas_logic_4	Single choice	-0.20	4	conditional	equivalence judgement	everyday
score_moas_logic_5	Single choice	-1.44	2	conditional	equivalence judgement	mathematical: extrema
score_moas_logic_6	Single choice	0.21	4	conditional	equivalence judgement	mathematical: extrema
score_moas_logic_7	Single choice	0.37	4	conditional	equivalence judgement	mathematical: extrema
score_moas_logic_8	Single choice	-0.61	3	conditional	equivalence judgement	mathematical: extrema
score_moas_logic_9	Single choice	0.18	4	conditional	equivalence judgement	mathematical: lineare independence
score_moas_logic_10	Single choice	-0.79	2	conditional	equivalence judgement	mathematical: lineare independence
score_moas_logic_11	Single choice	-0.64	3	conditional	equivalence judgement	mathematical: lineare independence
score_moas_logic_12	Single choice	0.23	4	conditional	equivalence judgement	mathematical: lineare independence
score_moas_logic_13	Single choice	-0.33	4	conditional	equivalence judgement	fictional
score_moas_logic_14	Single choice	-0.19	4	conditional	equivalence judgement	fictional
score_moas_logic_15	Single choice	-0.35	4	conditional	equivalence judgement	fictional
score_moas_logic_16	Single choice	-0.41	4	conditional	equivalence judgement	fictional
score_moas_logic_17	Single choice	-0.52	4	conditional	inference	everyday
score_moas_logic_18	Single choice	-1.04	2	conditional	inference	everyday
score_moas_logic_19	Single choice	-2.90	1	conditional	inference	mathematical: extrema
score_moas_logic_20	Single choice	-1.14	2	conditional	inference	mathematical: extrema
score_moas_logic_21	Single choice	-0.47	4	conditional	inference	mathematical: lineare independence
score_moas_logic_22	Single choice	-0.95	2	conditional	inference	mathematical: lineare independence
score_moas_logic_23	Single choice	-2.94	1	conditional	inference	fictional
score_moas_logic_24	Single choice	-1.31	2	conditional	inference	fictional
score_moas_logic_25	Single choice	-1.02	2	quantifiers	inference	everyday
score_moas_logic_26	Single choice	-1.14	2	quantifiers	inference	everyday

score_moas_logic_27	Single choice	-0.18	4	quantifiers	inference	everyday
score_moas_logic_28	Single choice	-1.10	2	quantifiers	inference	everyday
score_moas_logic_29	Single choice	-1.45	2	quantifiers	inference	everyday
score_moas_logic_30	Single choice	-0.35	4	quantifiers	equivalence judgement	pseudo-mathematical: nice functions
score_moas_logic_31	Single choice	-0.62	3	quantifiers	equivalence judgement	pseudo-mathematical: nice functions
score_moas_logic_32	Single choice	-0.46	4	quantifiers	equivalence judgement	pseudo-mathematical: nice functions
score_moas_logic_33	Single choice	0.79	5	quantifiers	equivalence judgement	pseudo-mathematical: nice functions
score_moas_logic_34	Single choice	1.76	5	quantifiers	equivalence judgement	pseudo-mathematical: nice functions
score_moas_logic_35	Single choice	-1.20	2	quantifiers	equivalence judgement	pseudo-mathematical: nice functions
score_moas_logic_36	Single choice	-0.63	3	quantifiers	equivalence judgement	pseudo-mathematical: nice functions
score_moas_logic_37	Single choice	-0.34	4	quantifiers	equivalence judgement	pseudo-mathematical: nice functions
score_moas_logic_38	Single choice	-0.34	4	quantifiers	equivalence judgement	pseudo-mathematical: nice functions
score_moas_logic_39	Single choice	1.01	5	quantifiers	equivalence judgement	pseudo-mathematical: nice functions
score_moas_logic_40	Single choice	-1.09	2	quantifiers	equivalence judgement	pseudo-mathematical: nice functions
score_moas_logic_41	Single choice	-0.63	3	quantifiers	equivalence judgement	pseudo-mathematical: nice functions

6.7.3 Description of knowledge levels

Level	Level Description	# Items	Threshold value	Item example
1	Elementary conditional inferences • Conditional inferences in the Modus Ponens form.	2	below -2.17	For every differentiable function f and $x_0 \in \mathbb{R}$, it is true that: • If f has a local extremum at x_0, then $f'(x_0) = 0$. • f has a local extremum at x_0. This information is true, for sure. What can you conclude from this information? ☐ $f'(x_0) = 0$ ☐ $f'(x_0) \neq 0$ ☐ None of the two options above.
2	Conditional and syllogistic inferences and identifying simple equivalent statements • Conditional inferences in the Modus Tollens or Acceptance of the Consequent forms with statements that contain negations. • Syllogistic reasoning with at least one universal quantifier. • Identifying that a conditional without negations is equivalent to its contrapositive. • Identifying and using that statements of the form "for all x not s(x)" are equivalent to statements of the form "there is no x, such that s(x)".	11	-2.17 to -0.71	It is true that: • If it rains, then the street is wet. • The street is not wet. This information is true, for sure. What can you conclude from this information? ☐ It is raining. ☐ It is not raining. ☐ None of the two options above.

3	Advanced conditional reasoning and identifying non-equivalence of statements involving negations • Conditional inferences in the Denial of the Antecedent form. • Identifying that a conditional (possibly with negations) is equivalent to its contraposition. • Identifying that a conditional is logical independent from the converse of its contraposition. • Identifying and using that statements of the form “not for all x s(x)” are equivalent to statements of the form “there is at least one x, such that not s(x)”. • Identifying the correct logical relation between two non-equivalent quantified statements involving a quantifier and a negation.	16	-0.71 to -0.26	Hans and Petra are talking about so-called *nice functions* and their values at different positions of their domain $\mathbb{R}$. • Hans states: Not all nice functions f have the value 2 at position $x = 1$. • Petra states: All nice functions f do not have the value 2 at position $x = 1$. How do the two statements relate to each other logically? ☐ The two statements say the same, logically. If Hans is right, then also Petra is right. If Petra is right, then also Hans is right. ☐ If Hans is right, then also Petra is right (but not necessarily the other way round). ☐ If Petra is right, then also Hans is right (but not necessarily the other way round). ☐ The two statements are logically completely independent. Each of the two could be right, independently of the other one being right or not.
4	Non-equivalence of a conditional and its converse • Identifying the correct logical relation between a conditional (possibly involving negations) and its converse. • [Making syllogistic inferences with a counterintuitive conclusion.]	8	-0.26 to 0.58	Hans and Petra are talking about the weather. • Hans states: If it rains, then the street is not dry. • Petra states: If the street is not dry, then it rains. How do the two statements relate to each other logically? ☐ The two statements say the same, logically. If Hans is right, then also Petra is right. If Petra is right, then also Hans is right. ☐ If Hans is right, then also Petra is right (but not necessarily the other way round). ☐ If Petra is right, then also Hans is right (but not necessarily the other way round).

				☐ The two statements are logically completely independent. Each of the two could be right, independently of the other one being right or not.
5	Relation between universal and existence quantifiers • Identifying the correct logical relation between two non-equivalent statements involving a universal and an existence quantifier.	3	above 0.58	Hans and Petra are talking about so-called *nice functions* and their values at different positions of their domain $\mathbb{R}$. • Hans states: For all nice functions f, there is at least one position x, such that the value of f at the position x is not 2. • Petra states: There is at least one position x, such that not all nice functions have the value 2 at the position x. How do the two statements relate to each other logically? ☐ The two statements say the same, logically. If Hans is right, then also Petra is right. If Petra is right, then also Hans is right. ☐ If Hans is right, then also Petra is right (but not necessarily the other way round). ☐ If Petra is right, then also Hans is right (but not necessarily the other way round). ☐ The two statements are logically completely independent. Each of the two could be right, independently of the other one being right or not.

Table 1 Levels of the KUM-LO test with item examples and difficulty parameters

7 Literature

Anderson, J. R. (1983). *The architecture of cognition*. Harvard University Press.

Attridge, N., & Inglis, M. (2013). Advanced mathematical study and the development of conditional reasoning skills. *PLoS one, 8*, e69399.

Barkai, R., Tabach, M., Tirosh, D., Tsamir, P., & Dreyfus, T. (2009). Modes of argument representation for proving: The case of general proof. In V. Durand-Guerrier, S. Soury-Lavergne, & F. Arzarello (Eds.), *Proceedings of the 6th Congress of the European Society for Research in Mathematics Education* (pp. 271-280), Lyon: France.

Carstensen, B., Klusmann, U., Baum, M., Brouër, B., Burda-Zoyke, A., Degner, I., … Zimmermann, F. (2020). *STePS 2020: Skalenhandbuch zur Dokumentation der Evaluationsinstrumente im Projekt "Lehramt mit Perspektive an der CAU Kiel“ – 4. Messzeitpunkt [STePS 2020: Documentation of evaluation scales in the project “Teacher education with perspectives at the CAU Kiel”*. IPN.

Christoforides, M., Spanoudis, G., & Demetriou, A. (2016). Coping with logical fallacies: a developmental training program for learning to reason. *Child Development, 87*(6), 1856–1876.

Clark, M., & Lovric, M. (2009). Understanding secondary-tertiary transition in mathematics. *International Journal of Mathematical Education in Science and Technology, 40*(6), 755–776.

Datsogianni, A., Sodian, B., Markovits, H., & Ufer, S. (2020). Reasoning with conditionals about everyday and mathematical concepts in primary school. *Frontiers in Psychology, 11*, 531640.

Deeken, C., Neumann, I., & Heinze, A. (2020). Mathematical prerequisites for STEM programs: What do university instructors expect from new STEM undergraduates? *International Journal of Research on Undergraduate Mathematics Education, 6*(1), 23-41..

Dehne, M., & Schupp, J. (2007). Persönlichkeitsmerkmale im Sozio-oekonomischen Panel (SOEP)-Konzept, Umsetzung und empirische Eigenschaften [Personality variables in the socio-economical panel (SOEP) – concept, operationalisation and empirical characteristics]. *Research Notes*, *26*(1), 70.

De Jong, T., & Ferguson-Hessler, M. G. M. (1996). Types and qualities of knowledge. *Educational Psychologist, 31*(2), 105–113.

Dieter, M. (2012). *Studienabbruch und Studienfachwechsel in der Mathematik: Quantitative Bezifferung und empirische Untersuchung von Bedingungsfaktoren* [Drop-out and change of study in mathematics: Quantification and empirical analysis of factors] (Doctoral dissertation). http://duepublico.uniduisburg-essen.de/servlets/DerivateServlet/Derivate-30759/Dieter_Miriam.pdf. Accessed 4 Nov 2019.

Dorier, J.-L., & Sierpinska, A. (2002). Research into the teaching and learning of linear algebra. In D. Holton, M. Artigue, U. Kirchgräber, J. Hillel, M. Niss, & A. Schoenfeld (Eds.), *The Teaching and Learning of Mathematics at University Level* (Vol. 7, pp. 255-273). Springer Netherlands. https://doi.org/10.1007/0-306-47231-7_24.

Durand-Guerrier, V., Boero, P., Douek, N., Epp, S. S., & Tanguay, D. (2012). Examining the role of logic in teaching proof. In G. Hanna & M. de Villiers (Eds.), *Proof and proving in mathematics education* (pp. 369-389). Springer.

Dörner, D. (1979). *Problemlösen als Informationsverarbeitung [Problem solving as information processing]*. Kohlhammer.

Evans, J.S.B., & Feeney, A. (2004). The role of prior belief in reasoning. In J. P. Leighton, & R. J. Sternberg (Eds.), *The Nature of Reasoning* (pp. 78-102). Cambridge University Press.

Förtsch, C., Sommerhoff, D., Fischer, F., Fischer, M. R., Girwidz, R., Obersteiner, A., ... & Neuhaus, B. J. (2018). Systematizing professional knowledge of medical doctors and teachers: Development of an interdisciplinary framework in the context of diagnostic competences. *Education Sciences, 8*(4), 207.

Gauffroy, C., & Barrouillet, P. (2009). Heuristic and analytic processes in mental models for conditionals: an integrative developmental theory. *Developmental Review, 29*(4), 249–282.

Gibbs, G. & Simpson, C. (2003). Measuring the response of students to assessment: the Assessment Experience Questionnaire. *11th Improving Student Learning Symposium.*

Greene, J.A., Sandoval W.A., & Bråten, I (2016). An introduction to epistemic cognition. In J.A. Greene, W.A. Sandoval, & I. Bråten (Eds.), *Handbook of Epistemic Cognition* (pp. 1-15). Routledge.

Gray, E. M., & Tall, D. (1994). Duality, ambiguity, and flexibility: A proceptual view of simple arithmetic. *Journal for Research in Mathematics Education, 26*(2), 115–141.

Gueudet, G. (2008). Investigating the secondary-tertiary transition. *Educational Studies in Mathematics, 67*(3), 237–254.

Hailikari, T., Nevgi, A., & Komulainen, E. (2008). Academic self-beliefs and prior knowledge as predictors of student achievement in mathematics: A structural model. *Educational Psychology, 28*(1), 59–71.

Hailikari, T., Nevgi, A., & Lindblom-Ylänne. (2007). Exploring alternative ways of assessing prior knowledge, its components and their relation to student achievement: A mathematics based case study. *Studies in Educational Evaluation, 33*, 320–337.

Handley, S. J., Capon, A., Beveridge, M., Dennis, I., & Evans, J. S. B. (2007). Working memory, inhibitory control and the development of children's reasoning. *Thinking & Reasoning* 10, 175–195.

Harks, B., Rakoczy, K., Hattie, J., Besser, M. & Klieme, E. (2014). The effects of feedback on achievement, interest, and self-evaluation: the role of feedback's perceived usefulness. *Educational Psychology*, *34*(3), 269-290.

Harks, B., Rakoczy, K., Klieme, E., Hattie, J., & Besser, M. (2014). Indirekte und moderierte Effekte von schriftlicher Rückmeldung auf Leistung und Motivation [Indirect and moderated effects of written feedback on performance and motivation]. In H. Ditton & A. Müller (Eds.), *Feedback und Rückmeldungen: theoretische Grundlagen, empirische Befunde, praktische Anwendungsfelder* (pp. 163-194). Waxmann.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77* (1), 81–112. DOI: 10.3102/003465430298487.

Hattie, J., & Wollenschläger, M. (2014). A conceptualization of feedback. In H. Ditton & A. Müller (Eds.), *Feedback und Rückmeldungen: theoretische Grundlagen, empirische Befunde, praktische Anwendungsfelder* (pp. 135-150). Waxmann.

Heublein, U., & Schmelzer, R. (2018). *Die Entwicklung der Studienabbruchquoten an den deutschen Hochschulen: Statistische Berechnungen auf der Basis des Absolventenjahrgangs 2016* [The development of dropout rates at German universities – Statistical analyses on the basis of the graduates in 2016]. DZHW.

Inglis, M., & Simpson, A. (2008). Conditional inference and advanced mathematical study. *Educational Studies in Mathematics, 67*(3), 187–204.

Inglis, M., & Simpson, A. (2009). Conditional inference and advanced mathematical study: further evidence. *Educational Studies in Mathematics, 72*(2), 185–198.

Janveau-Brennan, G., & Markovits, H. (1999). The development of reasoning with causal conditionals. *Developmental Psychology 35*(4), 904-911.

Jukic, L., & Dahl, B. (2012). University students' retention of derivative concepts 14 months after the course: Influence of 'met-befores' and 'met-afters'. *International Journal of Mathematical Education in Science and Technology, 43*(6), 749–764.

Kauper, T., Retelsdorf, J., Bauer, J., Rösler, L., Möller, J., & Prenzel, M. (2012). *PaLea – Panel zum Lehramtsstudium: Skalendokumentation und Häufigkeitsauszählungen des BMBF-Projektes [PaLea – Penel concerning teacher education: documentation of scales and frequency count of the BMBF-project]*. http://www.palea.uni-kiel.de/wp-content/uploads /2012/04/PaLea%20Skalendokumentation%204_%20Welle.pdf. Accessed 05 March 2014.

Köller, M. M., Aldrup, K., & Klusmann, U. (2017). Lehrer_in werden?! [To become teacher?!] *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 49*(3), 138–151.

Kopp, B., & Mandl, H. (2014). Lerntheoretische Grundlagen von Rückmeldungen [Theoretical foundations of feedback]. In H. Ditton & A. Müller (Eds.), *Feedback und Rückmeldungen: theoretische Grundlagen, empirische Befunde, praktische Anwendungsfelder* (pp. 29-41). Waxmann.

Kosiol, T., Rach, S., & Ufer, S. (2019). (Which) Mathematics Interest is Important for a Successful Transition to a University Study Program? *International Journal of Science and Mathematics Education, 17*(7), 1359–1380.

Küchemann, D., & Hoyles, C. (2009). From empirical to structural reasoning in mathematics: Tracking changes over time. In D. A. Stylianou, M. L. Blanton, & E. J. Knuth (Eds.), *Teaching and learning proof across the grades* (pp. 171–190). Routledge.

Leighton, J.P. (2004). Defining and describing reasoning. In J. P. Leighton & R. J. Sternberg (Eds.), *The Nature of Reasoning* (pp. 3–11). Cambridge University Press.

Markovits, H. (2000). A mental model analysis of young children's conditional reasoning with meaningful premises. *Thinking & Reasoning 6*(4), 335–347.

Markovits, H. (2014). Conditional reasoning and semantic memory retrieval. In A. Feeney, & V. Thompson (Eds.), *Reasoning as Memory* (pp. 67–84). Psychology Press.

Markovits, H., & Thompson, V. (2008). Different developmental patterns of simple deductive and probabilistic inferential reasoning. *Memory & Cognition, 36*(6), 1066–1078.

Marsh, H. W., Abduljabbar, A. S., Parker, P. D., Morin A. J. S., Abdelfattah, F., Nagengast, B., Möller, J., & Abu-Hilal, M. M. (2015). The internal/external frame of reference model of self-concept and achievement relations: Age-cohort and cross-cultural differences. *American Educational Research Journal*, *52*(1), 168–202.

Mitzel, H. C., Lewis, D. M., Patz, R. J., & Green, D. R. (2001). The bookmark procedure: Psychological perspectives. In G. J. Cizek (Ed.), *Setting performance standards: Concepts, methods and perspectives* (pp. 249–281). Lawrence Erlbaum Assoc.

Morsanyi, K., Kahl, T., & Rooney, R. (2017). The link between math and logic in adolescence: The effect of argument form. In M. E. Toplak, & J. Weller (Eds.), *Individual Differences in Judgment and Decision Making From a Developmental Context* (pp. 166–185). Psychology Press.

Morsanyi, K., McCormack, T., & O'Mahony, E. (2018). The link between deductive reasoning and mathematics. *Thinking & Reasoning, 24*, 234–257.

Moshman, D. (1990). The development of metalogical understanding. In W. Overton (Ed.), *Reasoning, Necessity, and Logic: Developmental Perspectives* (pp. 205–225). Lawrence Erlbaum Associates.

Narciss, S. (2013). Designing and evaluating tutoring feedback strategies for digital learning environments on the basis of the Interactive Tutoring Feedback Model. *Digital Education Review, 23*, 7–26.

Nieskens, B., Mayr, J., & Meyerdierks, I. (2011). CCT - Career Counselling for Teachers. Evaluierung eines Online-Beratungsangebots für Studieninteressierte [CCT – Career Counselling for Teachers: Evaluation of an online-consulting for study-interested persons]. *Lehrerbildung auf dem Prüfstand, 4*(1), 8–32.

Pantsar, M. (2019). Cognitive and computational complexity: Considerations from mathematical problem solving. *Erkenntnis*, 1–37.

Pekrun, R., Goetz, T., Titz, W., & Perry, R. P. (2002). Positive emotions in education. In E. Frydenberg (Ed.), *Beyond coping. Meeting goals, visions, and challenges* (pp. 149–173). Oxford University Press.

Rach, S., & Heinze, A. (2017). The transition from school to university in mathematics: which influence do school-related variables have? *International Journal of Science and Mathematics Education, 15*(7), 1343–1363.

Rach, S., & Ufer, S. (2020). Which prior mathematical knowledge is necessary for study success in the university study entrance phase? Results on a new model of knowledge levels based on a reanalysis of data from existing studies. *International Journal for Research in Undergraduate Mathematics Education*, *6*(3), 375–403.

Rach, S., Ufer, S., & Kosiol, T. (2019). Self-concept in university mathematics courses. In U. T. Jankvist, M. Van den Heuvel-Panhuizen, & M. Veldhuis (Hrsg.), *Proceedings of the Eleventh Congress of the European Society for Research in Mathematics Education* (pp. 1509-1516). Freudenthal Group & Freudenthal Institute, Utrecht University and ERME.

Rakoczy, K., Pinger, P., Hochweber, J., Klieme, E., Schütze, B., & Besser, M. (2019). Formative assessment in mathematics: Mediated by feedback's perceived usefulness and students' self-efficacy. *Learning and Instruction 60*, 154–165.

Rasmussen, C., & Ellis, J. (2013). Who is switching out of calculus and why. In A.M. Lindmeier, & A. Heinze (Eds.), *Proceedings of the 37th Conference of the International Group for the Psychology of Mathematics Education* (Vol. 4, pp. 73–80). PME.

Reichersdorfer, E., Ufer, S., Lindmeier, A., & Reiss, K. (2014). Der Übergang von der Schule zur Universität: Theoretische Fundierung und praktische Umsetzung einer Unterstützungsmaßnahme am Beginn des Mathematikstudiums [The transition from school to university: Theoretical foundations and pratical implementation of study support at the beginning of a mathematics study]. In I. Bausch, et al. (Eds.), *Mathematische Vor-und Brückenkurse* (pp. 37-53). Springer Spektrum.

Ricco, R. B. (2010). The development of deductive reasoning across the lifespan. In W. F. Overton (Ed.), *Biology, Cognition, And Methods Across The Life-Span: Handbook Of Life-Span Development*, Vol. 1 (pp. 391–430). Wiley.

Rittle-Johnson, B., Schneider, M., & Star, J. R. (2015). Not a one-way street: Bidirectional relations between procedural and conceptual knowledge of mathematics. *Educational Psychology Review, 27*, 198–208.

Robitzsch, A., Kiefer, T., & Wu, M. (2020). *TAM: Test Analysis Modules. R package version 3.5-19*. https://CRAN.R-project.org/package=TAM.

Schmidt, H. G., & Rikers, R. M. (2007). How expertise develops in medicine: knowledge encapsulation and illness script formation. *Medical Education, 41*(12), 1133–1139.

Schneider, M. (2006). *Konzeptuelles und prozedurales Wissen als latente Variablen: ihre Interaktion beim Lernen mit Dezimalbrüchen* [Conceptual and procedural knowledge as latent variables: it‘s interaction when learning decimal fractions] (Doctoral Dissertation). https://depositonce.tu-berlin.de/bitstream/11303/1605/2/Dokument_1.zip. Accessed 8 June 2021.

Schoenfeld, A. H. (1985). *Mathematical problem solving*. Elsevier.

Sommerhoff, D. (2017). *The individual cognitive resources underlying students' mathematical argumentation and proof skills* (Doctoral dissertation). LMU Munich.

Sfard, A. (1991). On the dual nature of mathematical conceptions: Reflections on process and objects as different side of the same coin. *Educational Studies in Mathematics, 22*, 1–36.

Skemp, R. R. (1976). Relational understanding and instrumental understanding. *Mathematics Teaching, 77*, 20–26.

Stewart, S. (2017). School algebra to linear algebra: Advancing through the worlds of mathematical thinking. In S. Stewart (Ed.), *And the Rest is Just Algebra* (pp. 219-232). Springer International Publishing.

Stillman, G., & Galbraith, P. (2003). Towards constructing a measure of the complexity of application tasks. S. J. Lamon, W. A. Parker, & K. Houston (Eds.), *Mathematical Modelling: A way of life* (pp. 179-188). Woodhead Publishing.

Strijbos, J.-W., & Müller, A. (2014). Personale Faktoren im Feedbackprozess [Personal characteristics in the feedback process]. In H. Ditton & A. Müller (Eds.), Feedback und Rückmeldungen: theoretische Grundlagen, empirische Befunde, praktische Anwendungsfelder (pp. 83-134). Waxmann.

Tall, D. (2008). The transition to formal thinking in mathematics. *Mathematics Education Research Journal, 20*(2), 5–24.

Ufer, S. (2015). The role of study motives and learning activities for success in first semester mathematics studies. In K. Beswick, T. Muir, & J. Wells (Eds.), *Proceedings of the 39th conference of the International Group for the Psychology of mathematics education* (Vol. 4, pp. 265–272). PME.

Ufer, S., Rach, S., & Kosiol, T. (2017). Interest in mathematics = Interest in mathematics? What general measures of interest reflect when the object of interest changes. *ZDM Mathematics Education, 49*(3), 397-409.

Ulriksen, L., Møller Madsen, L., & Holmegaard, H. T. (2010). What do we know about explanations for drop out/opt out among young people from STM higher education programmes? *Studies in Science Education,46*(2), 209–244.

Williams, G.& Clarke, D (1997). The complexity of mathematics tasks. In N. Scott & H. Hollingsworth (Eds.), *Mathematics: Creating the Future Proceedings of the 16th Biennial Conference of The Australian Association of Mathematics Teachers Adelaide: AAMT* (pp. 451–457).