RESEARCH

Open Access

Refining questionnaire-based assessment of STEM students' learning strategies

Birgit Griese[1*], Malte Lehmann[2] and Bettina Roesken-Winter[2]

Abstract

Background: Learning strategies are considered a key aspect for academic performance not only for science, technology, engineering and mathematics (STEM) students. Refining their assessment thus constitutes a worthwhile research purpose. The aim of this study is to examine the 69-item LIST questionnaire (ZDDP 15:185-200, 1994) with the aim of shortening it while keeping its factor structure and thus its potential for describing learning behaviour and for identifying significant changes therein. This includes exploring if reduced scales remain internally reliable, both in terms of reliability measures and content, and to examine if they stay sensitive enough to capture developments in a pre-post design.

Results: Our cohorts consist of STEM students ($N = 2374$) from different engineering courses at Ruhr-Universität Bochum in Germany, typically predominantly males, some with insufficient background in mathematics or non-native speakers of German. The data was analysed using various statistic methods, e.g. reliability measurement and confirmatory factor analysis. Our findings show that about half of the original items (36 out of 69) are sufficient, reliability holds (Cronbach's $a > 0.7$) and more variance is explained (56.17 % as compared to 45.65 %). Most content-related changes occurring when eliminating so many items survive critical scrutiny.

Conclusions: The study shows that it is possible to refine and considerably shorten the LIST questionnaire to the effect that all but one factor are kept intact. This will simplify future research enormously. Furthermore, the refined factor structure suggests reconsidering the postulate of metacognition as an easily accessed facet of learning behaviour—thus implying promising research perspectives.

Keywords: Engineering education; Mathematics courses; Learning strategies; LIST questionnaire; Metacognition

Background

In science, technology, engineering and mathematics (STEM) education, the passage from secondary to tertiary education is considered problematic, which is especially true regarding the challenges students encounter when being confronted with university mathematics (cf. Gueudet 2008; Henn et al. 2010). Students' failure rates in STEM fields and particularly in engineering subjects are alarmingly high in many countries. In Germany, for instance, almost 48 % of engineering students fail in their first year of university studies (Heublein et al. 2012). In relation to other subjects, the gap between school and university mathematics seems to be extremely high and causes difficulties for students taking mathematics courses. The

obstacles can be categorized in three categories, thus stressing the wide range of aspects: epistemological and cognitive difficulties, sociological and cultural difficulties and didactical difficulties (de Guzmán et al. 1998). Their dramatic character is depicted in the word "abstraction shock", referring to the fact that university mathematics is adding a formal world to the mathematics encountered at school (e.g. Artigue et al. 2007; Tall 2004). The challenge of understanding (and influencing) how the learning of mathematics at university works is often addressed by the use of cognitive development theories, as mathematics is commonly regarded as a rational and cognitively demanding subject (cf. Dreyfus 1995). Studies therefore often elaborate on cognitive difficulties and conceptual obstacles experienced by students in how mathematics is communicated to them (Artigue et al. 2007), particularly referring to the formal level of university mathematics and the prevalent role of proofs (Selden and Selden 2005).

* Correspondence: birgit.griese@rub.de
[1]Department of Mathematics, Ruhr-Universität Bochum, 44801 Bochum, Germany
Full list of author information is available at the end of the article

Griese *et al. International Journal of STEM Education* (2015) 2:12

Page 2 of 12

But mastering a university course containing mathematics requires more than a talent for abstract thought and formalities; it needs a combination of general skills and attitudes such as self-organisation, perseverance and frustration tolerance, as well as subject-specific abilities and meta-level capacities (cf. Pintrich et al. 1993; Weinstein and Palmer 2002; Wild and Schiefele 1994). It is interesting to detect in how far school leavers possess these different features, how these improve during the first year at the university and, importantly, what can be done to efficiently encourage their development. Students' meta-level learning behaviour is crucial, taking account of the words of de Guzmán et al. (1998) who state:

> Students' success is linked to a great extent to their capacity of developing "meta-level" skills allowing them, for instance, to self-diagnose their difficulties and to overcome them, to ask proper questions to their tutors, to optimize their personal resources, to organize their knowledge, to learn to use it in a better way in various modes and not only at a technical level. (p. 760)

Thus, a promising perspective is provided by exploring general and meta-level skills in terms of learning strategies (e.g. Wild 2005; Rach and Heinze 2011), whose investigation allows for revealing both the cognitive dispositions as well as affective barriers and pathways—and finally the interrelations between them.

Numerous research and instruction projects that attend to these issues can be found in many countries (cf. Dunn et al. 2012). The causality between adequate learning strategies and successful learning seems well established (for example, cf. Erdem Keklik and Keklik 2013). Depending on the specific research interest, previous studies have different emphases, for example, affective aspects or problem-solving skills. Most of these studies include data collection via questionnaires. To meet the challenge of assessing learning strategies appropriately, in the past, questionnaires were developed aiming at capturing different facets of these skills. In this paper, we report on refining and adapting one commonly used German questionnaire for assessing students' learning strategies. Our research is dedicated to gaining a reliable and valid instrument that simultaneously allows for making good economic sense. Often, research is not restricted to merely capturing students' learning strategies but also includes interventions, leading to a pre-post design with repeated surveys, so one essential goal lies in keeping the load for students within reasonable limits.

All data was collected within the scope of the design-based research study MP^2-Math/Plus/Practice, which focuses on supporting engineering students in mathematics (cf. Griese et al. 2013; Dehling et al. 2014). Among other issues, the project aims at supporting those with insufficient learning strategies and motivation by attempting to remedy these obstacles. To reveal the influence of the project interventions, the study draws on capturing students' learning strategies and their development in the first semester at university.

Learning strategies
Learning strategies as indications for learning behaviour
Recently, Blömeke et al. (2015) contributed to modelling competence "as a process, a continuum with many steps in between" (p. 7). In particular, they emphasized the following perspective:

> Thus, we suggest that trait approaches recognize the necessity to measure behaviorally, and that behavioral approaches to competence recognize the role of cognitive, affective and conative resources. At this time, we encourage research on competence in higher education emanating from either perspective and paying attention particularly to the steps in between. (Blömeke et al. 2015, p. 7)

Accordingly, in the paper at hand, learning strategies are understood as all kinds of planned and conscious learning behaviour and the attitudes behind it, involving observable actions (e.g. solving tasks, asking questions, taking notes) as well as thought processes (e.g. planning, reflecting) on the basis of both cognitive and affective-motivational dispositions. This extends to the lack of planning and conscious actions as it also presents a characteristic of an individual's learning strategy.

Research on the significance of learning strategies in mathematics education has its roots in contributions highlighting the role of affect, motivation and beliefs (cf. McLeod 1992, as a starting point), as all cognitive processes involve affective stances that moderate the tension between modes of intuitive and analytical thinking (e.g. Fischbein 1987; Stavy and Tirosh 2000). In particular, the theory of dual processes in cognitive psychology has been adapted to mathematics education, and the role of affective variables has been pointed out in this context (e.g. Evans 2007). These perspectives provide fresh views on learning processes and have done much to reach a deeper understanding of the obstacles involved. Findings reveal that students' cognitive reflection, as a metacognitive variable, their beliefs about mathematics, and their self-efficacy, are all correlated positively and significantly with mathematical achievement (Gómez-Chacón et al. 2014). There is also evidence that metacognition impacts positively on learning strategies which in turn influences achievement (Griese et al. 2011).

In summary, it can be said, therefore, that this has led to a fortified interest in certain kinds of learning strategies.

Griese et al. International Journal of STEM Education (2015) 2:12

Page 3 of 12

In the context of mathematics, overcoming motivational and affective barriers with the help of meta-skills, e.g. self-regulation, has become an important issue. What is more, mathematics demands the use of effective planning as well as organized and consistent work (cf. Rach and Heinze 2011). More than many other subjects, mathematics is cognitively challenging and needs motivational perseverance, thus representing an ideal research area for the influence of interventions addressing learning strategies, both on a general and a meta-level. However, the goals when assessing students' learning behaviour in science are various: taking an inventory, describing the development, comparing or improving learning behaviour (cf. Lovelace and Brickman 2013), and so are the research interests, in mathematics, popularly performance prediction or the identification of at-risk students. Apart from more time-consuming methods, this has led to a great variety of questionnaires in many languages.

Questionnaires for assessing learning strategies
Questionnaires with different focal points (according to the background of the authors and their research interests) originate from this variety (cf. Pintrich et al. 1993; Weinstein and Palmer 2002). Schellings (2011) gives a comprehensive overview from an international and a Dutch perspective. Though her work is based on the text-heavy learning of history, the general categories of learning behaviour can be applied to other subjects as well. Differentiating between motivational and cognitive aspects when dealing with learning strategies is a widely accepted concept (cf. Nenniger 1999) and is in keeping with the understanding of affective aspects as a key issue.

In the following, approaches to capture learning strategies which have influenced subsequent research fundamentally will be outlined. The selection includes only those which reflect the importance of affective and motivational issues. Pintrich et al. (1993) developed a questionnaire "to measure college undergraduates' motivation and self-regulated learning" (Artino 2005, p. 3), the *Motivated Strategies for Learning Questionnaire* (MSLQ). The MSLQ measures motivation and self-regulated learning in general *and* for a particular course by means of six motivation and nine learning strategies subscales. Initially, Pintrich and De Groot (1990) started by postulating a five latent factor structure comprising expectancy, value, affect, learning strategies and self-regulation. The items that were developed for operationalizing these constructs later formed the basis for the 15 subscales mentioned above. MSLQ has been applied in many research studies (Duncan and McKeachie 2005), partly aiming at developing a new conceptualization with respect to the significance of the single sub-scales (Dunn et al. 2012; Hilpert et al. 2013). MSLQ's reliability has proved "robust", and its predictive validity to actual course performance is considered "reasonable" (Pintrich et al. 1993, p. 801).

The *Approaches to Studying Inventory* (ASI) by Entwistle and Ramsden (1983) and its refinements (ASSIST by Tait et al. 1998, ALSI by Entwistle and McCune 2004) feature the main distinction of categorizing learning behaviour as being of either *strategic* (deep) or of *apathetic* (surface) approach. The dichotomy forms the inventory's two main factors which in turn contain up to 16 subscales, depending on the version of the questionnaire. Although the authors do not group their items into motivational and (meta-)cognitive scales, the object of research is nearly identical to that of MSLQ users. A specific feature of ASI and its variations is the idea to measure not only the desired learning behaviour (*strategic* approach) but also what is hypothesized as less success-oriented (*apathetic* approach). This produces a multifarious picture of learners' behaviour.

Another well-known instrument to capture students learning strategies is the *Learning and Study Strategies Inventory* (LASSI) by Weinstein and Palmer (2002). LASSI covers thoughts, behaviours, attitudes and beliefs in relation to successful learning that can also be fostered by interventions. Its ten scales are classed into affective strategies, goal strategies and comprehension monitoring strategies, thus covering cognitive, metacognitive (particularly, self-regulative), affective and motivational aspects. LASSI is not only used for research purposes but is also recommended to students to use for themselves in order to get feedback on their strengths and weaknesses. LASSI's reliability coefficients (Cronbach's α) for its different scales are reported to score between 0.86 and 0.68, the lowest often being considered insufficient (Weinstein et al. 1987). Its validity to academic performance depends on the specific scale, e.g. Cano (2006) found, using multiple regression, that two scales (namely *Affective Strategies* and *Goal Strategies*) contributed to academic performance, whereas one (*Comprehension Monitoring Strategies*) did not.

All questionnaires described so far resort to self-assessment of student behaviour. It must be conceded that this entails the weakness that "the learner's perceptions of his or her strategies are measured" (Schellings 2011, p. 94), which need not coincide precisely with the strategies themselves. In this context, it is interesting to compare self-reported learning behaviour (especially concerning metacognition) to the results gained with other methods, e.g. *Thinking Aloud* (David 2013). When it comes to affective aspects, however, the learner's subjective perspective is what counts. Other problems, like assessing the sufficiency or efficiency of study time or effort, might be harder to overcome. In our context, beginner students might initially judge their efforts as

Griese *et al. International Journal of STEM Education* (2015) 2:12

Page 4 of 12

sufficient (in relation to what they have been used to at school), but later rate them as inadequate (when their frame of reference has shifted after some months at university). This can result in a seeming decrease, although objectively, they have actually increased their efforts. Particularly in pre-post designs, this must be taken into account when interpreting results. Some authors thus favour using a retrospective pretest data design to measure programme effectiveness (Nimon et al. 2011; Lam and Bengo 2003) which is easily applicable to inventories measuring learning strategies.

LIST inventory

For our research at Ruhr-Universität Bochum in Germany, the decision fell for the German LIST questionnaire (*Learning Strategies at University*, Wild and Schiefele 1994), which is based on the same classification as MSLQ and takes up aspects from LASSI as well. LIST was invented for measuring learning strategies of medium generality, between learning styles and learning tactics (Wild 2000). The instrument distinguishes between cognitive, metacognitive and resource-related learning strategies and comprises dimensions of learning strategies grouped accordingly. This mirrors the acceptance of this taxonomy in the German-speaking community (Wild 2000). Just like its English predecessors, this approach originates from educational research and thus is not subject-specific. However, in STEM education, the instruments are frequently used to assess students' learning behaviour on a general level while combining the results with subject-related measures (Lovelace and Brickman 2013). The LIST questionnaire for measuring learning strategies in academic studies was first compiled in the 1990s (Wild and Schiefele 1994) and has since been modified and tested several times. LIST has been applied in the context of many subjects, mathematics among them (cf. Liebendörfer et al. 2014, for an overview), with overall satisfying results with regard to reliability (Wild and Schiefele 1994 found Cronbach's α between 0.64 for *Metacognition* and 0.90 for *Attention*) and validity (Wild 2000, 2005). In the following, we explore in detail how the LIST takes up scales and items from MSLQ and LASSI, in order to understand its origins and to illustrate its structure.

Apart from *Motivation*, the scales from LIST are derived directly from MSLQ, although the number of items varies. Some items in LIST are translations of MSLQ items. In addition to LIST, MSLQ seems very differentiated in terms of motivation, it sports six *Motivation* scales (*Intrinsic Goal Orientation, Extrinsic Goal Orientation, Task Value, Control of Learning Beliefs, Self-Efficacy for Learning and Performance, Test Anxiety*) comprising 31 items. LIST does not have items with the label *Motivation*

as such, but LIST's six items on *Attention* (which are all reverse coded) and eight items on *Effort* more or less cover this aspect, for example, "I work late at night or at the weekends if necessary". And other LIST scales, in particular, the resource-related ones, are meant to measure the degree of motivation a student possesses when preparing for an important exam, with items like "I fix the hours I spend daily on learning in a schedule". The main difference between the two questionnaires is that MSLQ puts more emphasis on including different aspects of motivation as *Goal Orientation* or *Control of Learning Beliefs*. For LIST, on the other hand, the aim was to clearly keep apart cognitive and motivational aspects.

LASSI (Weinstein and Palmer 2002) also separates cognitive aspects but has much less communalities with LIST. LASSI scales partly cover the same contents though holding different names, e.g. *Concentration* and *Attitude* (LASSI) compared to *Attention* (LIST). The numbers of items in a scale are different, too: there are 3 to 8 in LIST (if *Metacognitive Strategies* are divided into three scales, 4 to 8 if not), 3 to 12 (3 to 8) in MSLQ, and a constant 8 items in all LASSI scales. This results in considerable differences in analogous scales between LIST and MSLQ: LIST has 31 in *Cognitive Strategies* whereas MSLQ has 19 in the respective scales. According to the inventors of LIST, scales were expanded in order to reach better reliability (Wild 2000). All three questionnaires use Likert scales, ranging from five points (LIST) over six (LASSI) to seven (MSLQ). An overview on how LIST is based on MSLQ and LASSI is provided in Table 1.

As the comparison of sub-scales and items of the three instruments reveals, there are ample reasons for aiming at compactifying and balancing LIST, particularly when planning to use it with large samples. One side of the coin is receiving a reliable and valid instrument to access learning strategies. However, the next thing to consider is that the distribution, filling in and collection of the LIST questionnaire in a lecture hall holding several hundred students lasts almost 30 min, precious time that both lecturers and students would rather spend on mathematics. In addition, many studies aim at assessing learning strategies in a pre-post design at least twice, meaning that the drop-out rate is a serious issue to consider in any study design. That is, if there were a possibility to reduce the time to 15 min, participants' cooperation would be easier to gain.

Study perspectives and research questions

The purpose of the project MP^2-Math/Plus/Practice is to support engineering students in their first year at the university by enhancing their learning strategies and motivation. The length of MP^2-Math/Plus/Practice, which has completed its fifth year, permits to change the project

Griese *et al. International Journal of STEM Education* (2015) 2:12

Page 5 of 12

Table 1 Synoptic table for LIST's roots in MSLQ and LASSI (item numbers)

LIST	MSLQ	LASSI
Critical Checks (8)	Critical Thinking (5)	
Elaborating (8)	Elaboration (6)	Information Processing (8)
Organizing (8)	Organization (4)	Selecting Main Ideas (8)
Repeating (7)	Rehearsal (4)	
Attention (6)	Motivation	Attitude (8)
Effort (8)	Effort Regulation (4)	Concentration (8)
Learning Environment (6)	Time / Study Environment Management (8)	
Time Management (4)		Time Management (8)
Peer Learning (7)	Peer Learning (3)	
Reference (4)	Help Seeking (4)	Study Aids (8)
Planning (4)	Metacognitive Strategies (12)	Self-Testing (8)
Monitoring (4)		
Regulating (3)		
	…	…

interventions according to the principles of educational design research (cf. McKenney and Reeves 2012). The fact that MP^2-Math/Plus/Practice employs the LIST questionnaire in a pre-post design allows for conducting meta-analyses and prompted the study at hand which aims at exploring if the LIST questionnaire can be shortened while keeping the factor structure comparable to the original, and therefore its potential for describing learning strategies and significant changes therein:

(RQ 1): Do reduced LIST scales remain internally reliable both in terms of reliability measures and content?

(RQ 2): Are the reduced scales still sensitive enough to capture developments in a pre-post design?

The evaluations of the MP^2-Math/Plus/Practice interventions, in which the results gained from LIST play an important role, are not the focus of this paper. They will address more content-related and less technical issues, e.g. in what respect students modify their learning strategies and to what extent this can be influenced by the project interventions (for more details, see Griese et al. 2011). We restrict ourselves to refer to MP^2-Math/Plus/Practice only so far as information is needed to pursue the final goal of gaining a more workable version of LIST, making future research easier.

Methods
Participants of the study
LIST has been used for different groups of students from various backgrounds reading all sorts of subjects. Our research focuses on engineering students in their first semester at the university. Out of the students questioned, 77.70 % are males, 22.30 % females. The average age is $M_{age} = 20.34$ years ($SD_{age} = 2.22$ years). Almost one quarter (24.85 %) of the students have a first language different

from German. Only 60.86 % of them attended an advanced course in mathematics at school, and no more than 57.55 % went to the preparation course in mathematics at the university prior to their engineering course. Moreover, 20.33 % of the students neither took part in an advanced mathematics course at school nor did they visit the preparation course at university. These last characteristics classify the academic qualifications of a substantial number of students for the chosen subject of engineering as insufficient.

Study design and instruments
In order to study the development of learning strategies, a pre-post design was decided upon, the post questionnaire consisting of identical items in retrospect, presented in the past tense. The questionnaires were distributed during the break or at the end of the mathematics lecture and collected on-site after completion. This lecture covers what is considered basic mathematics: linear algebra, differential and integral calculus. The approach is axiomatic, with an emphasis on calculation. On the whole, proofs are presented in the lectures but not tested in the examination. The pre-test took place around the second week of the regular lectures; the post-test around the second last week.

The modification of the original LIST used in the research at hand is minimal: the scale *Critical Checks* was eliminated as it did not seem appropriate for mathematics at the beginning of the university. The items referring to the use of reference works were reworded in order to relate to digital sources as well. A new starting item (I study for my courses) was introduced but not used for evaluation—and consequently not included in the following countings of items, which starts at 69 for the

Griese *et al. International Journal of STEM Education* (2015) 2:12

Page 6 of 12

consequently named LIST69. Whereas MSLQ and LIST usually prefer five-point Likert scales, we opted for a four-point Likert scale (with poles *very rarely* and *very often*) in order to get more concise results.

With respect to data collection, the number of filled-in questionnaires varied immensely, see Table 2. In addition to that, some questionnaires were incomplete. Despite all this, well over 2000 data sets were collected. Before starting on shortening LIST69, it underwent exploratory and confirmatory factor analysis as well as tests for internal reliability. This was expected to set the limits for the refining process, as less input cannot carry more information, i.e. unreliable or inconclusive scales in LIST69 were not manipulated to become and well-cut and clear in shorter versions.

Process of data analysis

The objective of shortening LIST69 while keeping as much of the original scales as possible was conducted as follows, balancing consideration of content with statistical calculations: in a first step, the complete sample of data from 2011/2012 until 2013/2014 ($N = 2374$) was randomly split in half. The first half ($N = 1187$) was used to reduce the questionnaire by calculating Cronbach's α values for each original scale, including how they would develop when single items were left out. Thus, a scale was shortened by eliminating the item whose deletion yielded the best alpha. This process was repeated as long as the alphas stayed good (>0.7), preferably until there were only half as many items as before, in order to produce a perceptible pruning. The shortened scales were tested on the second half of the sample ($N = 1187$), meaning their Cronbach's α values were calculated as well. This complete process was repeated three times for three different random splits of the sample, thus enabling an assessment for the stability of the process, in order to minimize the risk of coincidental results. Three rounds also mean that in case of incongruent results, a tendency might be detected.

As a second step, the shortened version of LIST69 (named LIST45 as it contained 45 items) then underwent tests on a second half of the sample, i.e. a principal component analysis (PCA) to explore its structure, followed by some minor changes in item deletion and

Table 2 Number of LIST69 questionnaires (60 or more items answered)

	Pre	Post
2011 / 2012	345 (231)	103 (99)
2012 / 2013	421 (376)	254 (233)
2013 / 2014	955 (780)	296 (253)
Sum	1721 (1387)	653 (585)
Total	2374 (1972)	

further PCA as well as confirmatory factor analysis (CFA). The results were compared to the results of PCA and CFA of LIST69 from before, including model fit. Slight modifications of factor descriptions and changes in the loadings of single items onto factors were expected. As the pre and post scores would be explored separately, Cronbach's α values for single questionnaires were calculated from the complete sample as well as a further test for scale reliability.

The aim of researching the development of learning strategies demands further investigation into whether significant developments can be identified with the help of the shortened questionnaire. For this purpose, the item scores of each scale were combined using the formula $100/3 \cdot [1/n \cdot (x_1 + \ldots + x_n) - 1] = 100/3 \cdot (\bar{x} - 1)$, which renders values between 0 and 100 for a four-point Likert scale (n = number of items in the scale; scale scores under 25 describing rare use, between 25 and under 50 infrequent use, between 50 and under 75 regular use, 75 or more continual use of the learning strategies). Using LIST69 and the reduced data, the differences of means of these combined scores were calculated in dependent t-tests (listwise exclusion of cases, level of significance 0.05).

Results

The exploration of LIST69 identified the subscales of *Metacognition* (*Planning*, *Monitoring* and *Regulating*) as problematic, as all three had Cronbach's $\alpha < 0.7$ from the beginning. Thus, it was only possible to treat it as a complete scale where the eleven items rendered $\alpha = 0.728$. PCA of LIST69 was conducted with orthogonal rotation (varix) and pairwise exclusion of cases on the complete sample. The chosen rotation takes into account that from the construction of LIST, the scales are expected to be unrelated. The Kaiser-Meyer-Olin (KMO) measures proved good, $KMO_{LIST69} = 0.913$, attesting the sample adequacy. All KMO values for individual items were greater than 0.76, which meets the accepted limit of 0.5. Bartlett's test of sphericity χ^2_{LIST69} (2346) = 37,916.098, $p_{LIST69} = 0.000$ indicated that correlations between items were sufficiently large for PCA. For LIST69, 13 components had eigenvalues over Kaiser's criterion of 1 and in combination explained 50.59 % of variance. Traditionally, the LIST questionnaire is expected to render 10 scales (12 if *Metacognition* is regarded as having subscales *Planning*, *Monitoring* and *Regulating*), whereas in our case, Kaiser's criterion of extracting as many components as have eigenvalues >1 renders more. Extraction of ten factors in LIST69 resulted in explaining 45.65 % of variance. All this hints that the intended factor structure should be reconsidered.

The first step of reduction of the questionnaire was expected to lead to a distinct and reliable reduction of

Griese *et al. International Journal of STEM Education* (2015) 2:12

Page 7 of 12

the questionnaire, due to the use of three different random splits. In effect, in seven scales, the reductions were identical in all three random splits. For three scales (*Elaborating*, *Metacognition* and *Time Management*), the reductions suggested that two out of the three random splits were chosen. As expected, when reducing the number of items in a scale, the α values suffer (see Table 3), due to the square of this number being a factor in their calculation. For *Attention*, however, the α profits from the reduced scale, an improvement which will have to pass the test with regard to content—as all scales must. On the whole, the shortened version of LIST69, LIST45, yields acceptable and often good α values, considering the nature of its items (self-reporting rather than testing) and the drastically reduced number of items. They are all exactly in the range suggested by the analyses of the split sample, which once more hints at the stability of the procedure.

However, some items remained problematic in both versions of the questionnaire. The *Metacognition* items did not form a separate scale. This is in keeping with the results of other research (cf. Wild 2000), and the items were therefore candidates for deletion. The items from the *Metacognition* subscale, *Planning*, consistently loaded on the same component as *Time Management*, prompting a fresh inspection of this new scale. This lead to further deletion of items following the procedure described above, which quickly revealed all four *Planning* items as obsolete. *Repeating* items did not satisfy either; the items loaded on different scales. Not surprisingly, the *Monitoring* item that mentions fellow students (monitoring 4, see Additional file 1) loads on *Peer Learning*.

The PCA prompted the reduction to 36 items (keeping the *Repeating* items, but deleting the remaining 9

Metacognition items) in a new questionnaire correspondingly named LIST36, which was expected to yield nine factors that exactly match nine out of the ten scales from the original LIST69 (all but *Metacognition*). A subsequent PCA of LIST36 (varimax rotation, pairwise exclusion of cases, number of factors set to be 9, KMO-$_{\text{LIST36}}$ = 0.821, all KMO values $\geq$ 0.617, χ^2_{LIST36} (630) = 9340.628, p_{LIST36} = 0.000 resulted in Table 4 (for a second half of the randomly split sample) which illustrates how neatly the items load onto the remaining nine factors. Together, these factors explain 56.17 % of variance (whereas, extraction of ten factors in LIST69 only explained 45.65 % of variance, see above). The Cronbach's α values are >0.7 without exception. This raises hopes for an imminent improvement of the instrument. The reduction to LIST36 also solves the problem of multicollinearity which exists in LIST69 (where the determinant of the correlation matrix is 1.151×10^{-9} for the complete sample, cf. Field 2009, p. 660), as it brings the determinant up to 8.765×10^{-5}, meeting the criterion of exceeding 10^{-5}.

All in all, the deletions show slight variations to our first attempt at refining the LIST questionnaire (cf. Griese et al. 2014), when we achieved a version with 32 items and slightly worse alphas, using data from earlier project years. As a side project, oblique rotation (direct oblimin) was also tried: the results stay the same, meaning that the same items load on the same factors as strongly as in orthogonal rotation, the same amount of variance is explained; and the structure matrix does not reveal any disturbing interconnections between the components.

To judge the development of the model fit, a comparative fit index (CFI) and two badness-of-fit indices (root mean square error of approximation (RMSEA) and

Table 3 Cronbach's α for LIST69 and shortened LIST45, N = # used data sets

Scale	LIST69			LIST45		
	#items	α	N	#items	α	N
Organizing	8	.81	1605	3	.74	1969
Elaborating	8	.77	1606	5	.73	1816
Repeating	7	.73	1646	6	.72	1702
Metacognition	11	.73	1513	9	.71	1603
Effort	8	.76	1556	5	.70	1751
Attention	6	.75	1663	3	.82	1899
Time Management	4	.76	1901	3	.71	1956
Learning Environment	6	.70	1721	5	.72	1775
Peer Learning	7	.78	1664	3	.77	1915
Using Reference	4	.77	1920	3	.73	1971
Total	69			45		

Table 4 Summary of PCA for LIST36, item loadings <0.4 suppressed, N = 1187

LIST items	Elaborating	Repeating	Effort	Learning Environment	Attention	Peer Learning	Time Management	Using Reference	Organizing
ela2	0.74								
ela4	0.71								
ela6	0.68								
ela3	0.63								
ela5	0.56								
repeating7		0.70							
repeating6		0.65							
repeating3		0.60							
repeating4		0.56							
repeating2		0.48							
repeating1		0.47							
effort3			0.66						
effort2			0.64						
effort8			0.61						
effort6			0.56						
effort1			0.53						
environ5				0.75					
environ6				0.75					
environ1				0.57					
environ4				0.56					
environ2				0.55					
attention6					0.87				
attention5					0.85				
attention4					0.84				
peer1						0.84			
peer2						0.83			
peer6						0.76			
time3							0.79		
time4							0.77		
time1							0.72		
reference4								0.75	
reference1								0.73	
reference3								0.72	

orga2									0.76
orga5									0.74
orga3									0.69
Percentage of variance	16.52	7.20	6.38	5.75	5.38	4.53	3.86	3.46	3.08
Cronbach's α	0.72	0.72	0.71	0.73	0.82	0.77	0.72	0.73	0.72

Griese et al. International Journal of STEM Education (2015) 2:12

Page 9 of 12

Griese *et al. International Journal of STEM Education* (2015) 2:12

Page 10 of 12

Table 5 Fit indices for CFA for original and shortened questionnaires

	CFI	RMSEA	SRMR	$\chi^2(df)$	p
LIST69	0.8001	0.0420	0.0523	11464.87(2211)	0.00
LIST45	0.8419	0.0448	0.0466	5191.93(900)	0.00
LIST36	0.8364	0.0556	0.0495	4644.44(558)	0.00

standardized root mean residual (SRMR)) were calculated to balance for complexity of the model, sample size and robustness against violations of the underlying distribution (Table 5), revealing acceptable model fit in the shortened questionnaires (cf. Schreiber et al. 2006).

What is more, LIST36's scales are quite stable in the single pre and post surveys conducted in the course of the study; out of the 54 Cronbach's α values (nine scales in six surveys) only a few were <0.7, none were <0.6 (see Table 6).

The comparison of t-tests between the pre and post data of the project years was meant to reveal if the shortened questionnaire kept the significance of the originally used one, and also if there were scales that never produced significant differences. From the data available, it can be deduced that (a) the trends are always identical in the long and shortened questionnaires, (b) no significant developments are lost and (c) there were only two exceptions concerning significance of developments, i.e. *Organizing* in project year 2011/2012 where LIST36 changed the level of significance from 0.118 in LIST69 to under the customary level of 0.05 and similarly for *Peer Learning* in 2013/2014. Additional file 1 gives all LIST items in detail (translation from German by the first author), arranged by scale, with indications of deletion. Table 7 provides an overview of the evaluation of item deletions.

For the metacognitive scales *Monitoring* and *Regulating*, at a closer look, the aspects inquired for are mostly

Table 6 Cronbach's α of LIST36 for the individual surveys

	Pre 2011/12	Post 2011/12	Pre 2012/13	Post 2012/13	Pre 2013/14	Post 2013/14
Organizing	0.72	0.77	0.71	0.77	0.70	0.79
Elaborating	0.67	0.74	0.75	0.72	0.67	0.74
Repeating	0.69	0.78	0.65	0.72	0.71	0.69
Effort	0.68	0.76	0.70	0.70	0.70	0.62
Attention	0.76	0.89	0.80	0.75	0.81	0.85
Time Management	0.63	0.66	0.73	0.70	0.69	0.77
Learning Environment	0.71	0.72	0.74	0.77	0.71	0.72
Peer Learning	0.77	0.78	0.79	0.72	0.74	0.82
Using Reference	0.66	0.73	0.76	0.70	0.71	0.78

mirrored in *Effort*, though more superficially. One item in *Monitoring* (monitoring 4, see Additional file 1) can well be associated with *Peer Learning*, as it contains interacting with fellow students as much as it contains testing oneself. The metacognitive scale *Planning*, whose close connection to *Time Management* was obvious from the data analysis, nevertheless stresses substantial planning rather than simple timing. Its deletion means losing a more effective and more reflected learning strategy. It is doubtful, though, if techniques of this depth can be recorded well in a quick and self-reporting questionnaire. All in all it can be said that deleting the metacognitive scales must be regarded critically.

Discussion

Investigations into an approved and validated questionnaire originally designed to explore learning strategies

Table 7 Overview of content-related aspects of item deletions

Scale	Summary of content-related aspects of item deletion
Organizing	The deletions are acceptable (e.g. the aspect of structuring subject matter to make it easier to remember is contained in orga2 and orga3). However, compiling one's own lists is not included anymore—but writing your own summaries is
Elaborating	The deletions seem particularly appropriate for engineering students, who are either explicitly taught the connections to their other subjects and courses, or they do not stand a chance to discover them on their own
Repeating	The deletion only concerns one item rather referring to texts (reciting paragraphs) which does not fit to mathematics very well
Metacognition	The deletions can be regarded as critical, as here we find aspects that are neither covered by other scales nor is their depth reflected anywhere else, see text for details
Effort	This scale does not lose its focus on perseverance but a more concrete manifestation of this attitude, namely working late and long hours. Comparing one's learning time with others', though, measures not only objective effort but also subjective self-confidence and seems problematic in terms of multidimensionality
Attention	The six original items are paraphrases of each other, so that deleting half of them does not result in any loss of information, as the high Cronbach's α for this scale shows (cf. Table 6)
Time Management, Using Reference	The deleted items are mirrored in those kept
Learning Environment, Peer Learning	The deletions seem unexciting. The deleted item about always sitting at the same place when studying does not match students' reality well, as most have places of study at home *and* meet fellow students on campus. The same applies to comparing notes with fellow students in times of digital uploads

Griese *et al. International Journal of STEM Education* (2015) 2:12

Page 11 of 12

in general have led to a shorter version consisting of only 36 of the original 69 items, which will make it easier to ensure students' cooperation for further research. LIST36 was developed using the data from research into STEM students' learning strategies and mirrors nine out of the ten scales in LIST69 and has proven its reliability and its factors explain more variance. In addition, the shortened questionnaire appears as suitable for use in pre-post designs as its predecessor. If the differences in pre and post scores are a valid measurement for competence, development at all has to be viewed critically, though, as individual reference frames may have shifted. Furthermore, the validity in connection with academic performance will have to be further explored in our future research. Another point is that the generalizability of the new instrument will need to be tested on other populations than engineering students, as some of the content-related issues might be weighted differently for different groups of students.

Leaving aside the numerous calculations, it is nonetheless important to take a close look at which items have suffered deletion, if they contain any specific content that might get lost in the scale or if they involve a shift in content focus. For our purposes, we can keep in mind that we are dealing with first year engineering students and their learning of mathematics, which on the one hand makes some deletions acceptable but which, on the other hand, also represents a limitation. It still needs to be tested if the findings from our sample of first year engineering students can be generalized.

To avoid problems stemming from inter-language or inter-cultural differences, it seemed wise to use a German questionnaire in our research, although a transfer to another country and language can present a test for a questionnaire's validity. International cooperation has been effected with colleagues from Spain (cf. Gómez-Chacón et al. 2015, accepted for CERME9), showing that LIST keeps its qualities when being used in another country: After being translated into English and then into Spanish, the cognitive and metacognitive scales from LIST kept their reliability, an indication for the questionnaire's universal applicability.

In addition to this, it has become apparent that some of our initial assumptions about the impact and measurability of metacognitive behaviour of STEM students might be incorrect, thus guiding further research towards other ways of examining this concept. This aspect represents a point that might interest researchers from other areas, as metacognition and its measurability often are popular research interests.

Conclusions

Depending on the purpose of the individual study, it might still be advisable to keep the *Metacognition* items in the questionnaire for further tests as the decision to consider deleting all of them was a hard one. Our research into the interventions of the project MP^2-Math/Plus/Practice itself had been planned to keep a close eye on metacognitive learning strategies, hypothesizing that they make the difference when it comes to successful learning behaviour. However, the results of our analysis show that the characteristics formerly believed to be separate from more technical learning strategies actually load on already existing factors, and there they do not even contribute relevantly. One logical conclusion might be that metacognitive dispositions are latent factors which work in the background and become apparent only in other learning strategies—where they can then be measured accordingly. Another explanation could lie in the nature of the data acquisition used: metacognitive reflections might be inappropriate for being measured in self-assessing printed questionnaires. Other forms of procuring data, for example, from interviews, videographed peer discussions, classroom observations, flash interviews, reflective essays or learning diaries, might prove more adequate. Further research is required, and our data collected from interviews, learning diaries and other sources will be used for this purpose.

Additional file

Additional file 1: LIST questions (pre). Items (item names) deleted for LIST36.

Authors' contributions
All data was collected in the context of the project MP^2-Math/Plus at Ruhr-Universität Bochum in Germany. MP^2-Math/Plus/Practice was initially (2010 – 2012) supported by Stifterverband für die Deutsche Wissenschaft in cooperation with Heinz-Nixdorf-Stiftung and is now an established part of the academic program. BG designed the study, collected the data, performed statistical analyses, drafted and revised the manuscript. ML performed some and reviewed all statistical analyses and reviewed the manuscript. BR-W participated in designing the study and drafting the manuscript, and supervised all work. All three authors read, re-read, discussed and approved the final manuscript.

Authors' information
Bettina Rösken-Winter is a full professor for Mathematics education at the Humboldt-Universität zu Berlin (Germany). Her research interests include mathematical education in engineering studies, continuous teacher professional development, and design-based research. Birgit Griese is a mathematics teacher and a mathematics education researcher at Ruhr-Universität Bochum (Germany). She is currently working on her PhD thesis investigating engineering students' learning strategies. Malte Lehmann is a mathematics education researcher at the Humboldt-Universität zu Berlin (Germany). He is interested in statistics and is currently working on his PhD project investigating students' problem solving competences.

Author details
[1]Department of Mathematics, Ruhr-Universität Bochum, 44801 Bochum, Germany. [2]Professional School of Education, Humboldt Universität zu Berlin, 10099 Berlin, Germany.

Griese *et al. International Journal of STEM Education* (2015) 2:12

Page 12 of 12

References

Artigue, M, Batanero, C, & Kent, P. (2007). Mathematics thinking and learning at post-secondary level. In F Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 1011–1049). Greenwich, Connecticut: Information Age Publishing.

Artino, AR (2005). A review of the motivated strategies for learning questionnaire. http://eric.ed.gov/?id=ED499083. Accessed 1 December 2014.

Blömeke, S, Gustafsson, JE, & Shavelson, RJ. (2015). Beyond dichotomies. Competence viewed as a continuum. *Zeitschrift für Psychologie, 223*(1), 3–13.

Cano, F. (2006). An in-depth analysis of the learning and study strategies inventory (LASSI). *Educational and Psychological Measurement, 66*(6), 1023–1038.

David, A (2013). *Aufgabenspezifische Messung metakognitiver Aktivitäten im Rahmen von Lernaufgaben* (PhD thesis). Technische Universität, Chemnitz. http://nbn-resolving.de/urn:nbn:de:bsz:ch1-qucosa-133054. Accessed 7 April 2015.

de Guzmán, M, Hodgson, BR, Robert, A, & Villani, V. (1998). Difficulties in the passage from secondary to tertiary education. In G Fischer & U Rehmann (Eds.), *Proceedings of the international congress of mathematicians—documenta mathematica extra volume III* (pp. 747–762). Rosenheim: Geronimo GmbH.

Dehling, H, Glasmachers, E, Griese, B, Härterich, J, & Kallweit, M. (2014). MP²-Mathe/Plus/Praxis, strategien zur vorbeugung gegen studienabbruch. *Zeitschrift für Hochschulentwicklung, 9*(4), 39–56.

Dreyfus, T (Guest Editor) (1995). Advanced mathematical thinking [special issue]. *Educational Studies in Mathematics, 29*(2).

Duncan, TG, & McKeachie, WJ. (2005). The making of the motivated strategies for learning questionnaire. *Educational Psychologist, 40*, 117–128.

Dunn, KE, Lo, WJ, Mulvenon, SW, & Sutcliffe, R. (2012). Revisiting the motivated strategies for learning questionnaire: a theoretical and statistical reevaluation of the metacognitive self-regulation and effort regulation subscales. *Educational and Psychological Measurement, 72*, 312–331.

Entwistle, N, & McCune, V. (2004). The conceptual bases of study strategies inventories. *Educational Psychology Review, 16*(4), 325–346.

Entwistle, N, & Ramsden, P. (1983). *Understanding student learning*. London: Croom Helm.

Erdem Keklik, E, & Keklik, I. (2013). Motivation and learning strategies as predictors of high school students' math achievement. *Cukurova University Faculty of Education Journal, 42*(1), 96–109.

Evans, JSBT. (2007). *Hypothetical thinking. Dual processes in reasoning and judgement*. Hove: Psychology Press.

Field, A. (2009). *Discovering statistics using SPSS* (3rd ed.). Thousand Oaks: Sage Publications.

Fischbein, E. (1987). *Intuition in science and mathematics. An educational approach*. Dordrecht: Kluwer Academic Publishers.

Gómez-Chacón, IM, García-Madruga, JA, Vila, JO, Elosúa, MR, & Rodríguez, R. (2014). The dual processes hypothesis in mathematics performance: beliefs, cognitive reflection, reasoning and working memory. *Learning and Individual Differences, 29*, 67–73.

Gómez-Chacón, IM, Griese, B, Rösken-Winter, B, & Gonzàlez-Guillén, C. (2015). Engineering students in Spain and Germany—varying and uniform learning strategies. In N Vondrova & K Krainer (Eds.), *Proceedings of the 9th conference of european researchers in mathematics education*. Prague, Czech Republic: Charles University. in press.

Griese, B, Glasmachers, E, Kallweit, M, & Roesken, B. (2011). Engineering students and their learning of mathematics. In B Roesken & M Casper (Eds.), *Current state of research on mathematical beliefs XVII. Proceedings of the MAVI-17 conference (pp. 85–96)*. Bochum: Professional School of Education, RUB.

Griese, B, Lehmann, M, & Roesken-Winter, B. (2014). Exploring university students' learning strategies in mathematics: refining the LIST questionnaire. In S Oesterle, C Nicol, & P Liljedahl (Eds.), *Proceedings of the joint meeting of PME 38 and PME-NA 36, Vol. 6* (p. 131). Vancouver, Canada: PME.

Griese, B, Roesken-Winter, B, Kallweit, M, & Glasmachers, E. (2013). Redesigning interventions for engineering students: learning from practice. In AM Lindmeier & A Heinze (Eds.), *Proceedings of the 37th conference of the international group for the psychology of mathematics education vol. 5* (p. 65). Kiel, Germany: PME.

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

Henn, HW, Bruder, R, Elschenbroich, J, Greefrath, G, Kramer, J, & Pinkernell, G. (2010). Schnittstelle schule - hochschule. In A Lindmeier & S Ufer (Eds.), *Beiträge zum mathematikunterricht 2010* (pp. 75–82). Münster: WTM.

Heublein, U, Richter, J, Schmelzer, R, & Sommer, D. (2012). *Die entwicklung der schwund- und studienabbruchquoten an den deutschen hochschulen: statistische berechnungen auf der basis des absolventenjahrgangs 2010. Forum hochschule: vol. 2012,3*. Hannover: HIS.

Hilpert, J, Stempien, J, van der Hoeven Kraft, KJ, & Husman, J. (2013). Evidence for the latent factor structure of the MSLQ: a new conceptualization of an established questionnaire. *SAGE-Open, 3*, 1–10. doi:10.1177/2158244013510305.

Lam, TC, & Bengo, P. (2003). A comparison of three retrospective self-reporting methods of measuring change in instructional practice. *American Journal of Evaluation, 223*(1), 3–13.

Liebendörfer, M, Hochmuth, R, Schreiber, S, Göller, R, Kolter, J, Biehler, R, Kortemeyer, J, & Ostsieker, L. (2014). Vorstellung eines fragebogens zur erfassung von lernstrategien in mathematikhaltigen studiengängen. In J Roth & J Ames (Eds.), *Beiträge zum mathematikunterricht 2014* (pp. 739–742). Münster: WTM.

Lovelace, M, & Brickman, P. (2013). Best practices for measuring students' attitudes toward learning science. *CBE – Life Sciences Education, 12*, 606–617.

McKenney, SE, & Reeves, TC. (2012). *Conducting educational design research*. New York: Routledge.

McLeod, DB. (1992). Research on affect in mathematics education: a reconceptualization. In DA Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 575–596). New York: Macmillan.

Nenniger, P. (1999). On the role of motivation in self-directed learning. The 'two-shells-model of motivated self-directed learning' as a structural explanatory concept. *European Journal of Psychology of Education, 14*(1), 71–86.

Nimon, K, Zigarni, D, & Allen, J. (2011). Measures of program effectiveness based on retrospective pretest data: are all created equal? *American Journal of Evaluation, 32*(1), 8–28.

Pintrich, P, & de Groot, E. (1990). Motivational and self-regulated learning components of classroom academic performance. *Journal of Educational Psychology, 81*(1), 33–40.

Pintrich, P, Smith, D, Garcia, T, & McKeachie, W. (1993). Reliability and predictive validity of the motivated strategies for learning questionnaire (MSLQ). *Educational and Psychological Measurement, 53*(3), 801–813.

Rach, S, & Heinze, A. (2011). Studying mathematics at the university: the influence of learning strategies. In B Ubuz (Ed.), *Proceedings of the 35th conference of the international group for the psychology of mathematics education* (Vol. 4, pp. 9–16). Ankara, Turkey: PME.

Schellings, C. (2011). Applying learning strategy questionnaires: problems and possibilities. *Metacognition and Learning, 6*(2), 91–109.

Schreiber, JB, Nora, A, Stage, FK, Barlow, EA, & King, J. (2006). Reporting structural equation modeling and confirmatory factor analysis results: a review. *The Journal of Educational Research, 99*(6), 323–337.

Selden, A, & Selden, J. (2005). Perspectives on advanced mathematical thinking. *Mathematical Thinking and Learning, 7*(1), 1–13.

Stavy, R, & Tirosh, D. (2000). *How students (mis-)understand science and mathematics: intuitive rules*. New York: Teachers College Press.

Tait, H, Entwistle, N, & McCune, V. (1998). ASSIST: a reconceptualisation of the approaches to studying inventory. In C Rust (Ed.), *Improving student learning: improving students as learners* (pp. 262–271). Oxford: Oxford Centre for Staff and Learning Development.

Tall, DO. (2004). Building theories: the three worlds of mathematics. *For the Learning of Mathematics, 24*(1), 29–33.

Weinstein, CE, & Palmer, DR. (2002). *Learning and study strategies inventory (LASSI): user's manual* (2nd ed.). Clearwater: H & H Publishing.

Weinstein, CE, Schulte, AC, & Palmer, DR. (1987). *Learning and study strategies inventory (LASSI)*. Clearwater: H & H Publishing.

Wild, KP. (2000). *Lernstrategien im studium: strukturen und bedingungen*. Münster: Waxmann.

Wild, KP. (2005). Individuelle lernstrategien von studierenden. Konsequenzen für die hochschuldidaktik und die hochschullehre. *Beiträge zur Lehrerbildung, 23*(2), 191–206.

Wild, KP, & Schiefele, U. (1994). Lernstrategien im studium. *Ergebnisse zur Faktorenstruktur und Reliabilität eines neuen Fragebogens. Zeitschrift für Differentielle und Diagnostische Psychologie, 15*, 185–200.