

**The Influence of Planning an Assignment on Students' Cognitive Load and Academic
Achievement**

Student: Anniek de Kort

Student number: 4160134

Supervising lecturer: Dr. F. Kirschner

Second assessor: Dr. J.J.H.M. Janssen

Date: 16-07-2019

Abstract: 198 words

Words: 8398 (max. 8000)

Abstract

This study aims to provide more insight into students' cognitive processes while planning a learning task and students' academic achievement. Research shows that Self-Regulated Learning-skills (SRL-skills) can be valuable for learning. Interventions intended to improve SRL-skills are aimed at all skills related to the theory of SRL. However, planning may be the most influential element of this theory because it influences learning strategies afterward.

Two between-subject design experiments were conducted to answer the question 'What influence does a planning phase intervention have on students' experienced cognitive load (CL) and academic achievement during the learning phase?' A mixed-methods design was used to gain insight into participants' planning phase. Experiment 1 showed no difference in intrinsic CL. However, significant differences in germane CL, extraneous CL and academic achievement were found. Academic achievement was higher for students who did not plan their learning task. Experiment 2 found no differences in all types of CL and academic achievement. Several limitations can explain these findings. Participants in experiment 1 worked together; students in experiment 2 worked individually. Also, participants in the experimental group did not find the planning assignment suitable for the given learning task.

Lastly, the planning intervention could be more comprehensive.

Keywords: Self-Regulated Learning, Cognitive Load Theory, planning, academic achievement, mental effort

The Influence of Planning an Assignment on Students' Cognitive Load and Academic Achievement

Students in higher education need to rely more on their abilities to monitor learning because they must regulate their behavior more than students in primary education (De Bruin & Van Merriënboer, 2017). Students can achieve this by using different strategies, like goal setting, help-seeking, or planning. Zimmerman (2002) described these strategies as Self-Regulated Learning (SRL), where self-generated thoughts, feelings, and actions are systematically oriented towards the attainment of students' goals. Teaching students SRL-skills is a viable method for promoting SRL (Zumbrunn, Tadlockk, & Roberts, 2011). Especially planning is a proven prerequisite of effective studying (Vrugt & Oort, 2008). Planning involves the selection of (sub)goals, scheduling, managing one's time, and assigning resources that affect performance (Vrugt & Oort, 2008; Zimmerman, 2002). Students who have underdeveloped planning skills regularly have lower academic achievements (Zimmerman & Pons, 1986). This because students' planning influences the use of learning strategies during learning tasks and is, therefore, valuable for learning (Wirth, Künsting, & Leutner, 2009; Zimmerman, 2002). Therefore, the planning of learning tasks should be taken into account by instructional designers.

Studies on the effect of SRL on academic achievement are showing mixed results. A recent focus on SRL introduced by Raaijmakers, Baars, Paas, Van Merriënboer, and Van Gog, (2018) and Van Gog, Kester, and Paas (2010) in which they look at SRL from a cognitive perspective using Cognitive Load Theory (CLT), could provide new interesting views on these mixed results. Furthermore, there is little empirical evidence on the role of the planning phase in the effects on academic achievement even though the planning phase is affecting all phases of learning (Vrugt & Oort, 2008; Zimmerman, 2002).

This study will add to the existing knowledge on the influence of planning and aims to provide more insights into the effects of planning a learning task on students' cognitive

processes and their academic achievement using CLT. Outcomes of this study can help instructional designers to make more thoughtful decisions about whether to include planning interventions in education.

Self-Regulated Learning Phases and the Effect on Learning

SRL is a cyclical process, where feedback from prior performances is used to make adjustments during current performances (Zimmerman, 2002). Making adjustments is crucial due to constantly changing factors which students need to observe and monitor. SRL is, therefore, an ongoing proactive process in which students must be engaged in their learning (Zimmerman, 2008).

According to Pintrich (2000), SRL consist of four phases: the planning phase; the monitoring phase; the control phase; and the reaction and reflection phase. The planning phase consists of task analysis and self-motivation, which refers to processes and beliefs that occur before students make an effort to learn. This phase consists of goal setting, strategic planning, beliefs about learning, and outcome expectations (Zimmerman & Pons, 1986). Students are reflecting these skills constantly, which may cause changing or updating their plans during the next phases (Pintrich, 2000). These skills thus influence other phases and can, therefore, be seen as an essential part of SRL. In order to succeed in a task, students first have to set goals and construct efficient plans that lead to goal achievement. Research suggests that the more efficient plans and strategies students can construct, the more likely they are to succeed (Heikkilä & Lonka, 2006).

Results of research on the effect of using SRL-skills on academic achievement are mixed. On the one hand the quality and quantity of students' use of SRL-skills, also planning, have been found to positively correlate with academic achievement (Heikkilä & Lonka, 2006; Kitsantas, Winsler, & Huie, 2008; Rytönen, Parpala, Lindblom-Ylänne, Virtanen, & Postareff, 2012). These studies, however, all focus on the effect of SRL in general, and not the planning

phase in particular. Planning is an essential part of SRL, and therefore, the expectation is that this phase has a significant contribution to the found results.

Bol, Campbell, Perez, and Yen (2015) developed SRL training for math. Here, students had to set goals and were asked to complete a time management schedule for the week. They had to review and apply study habits during the week. At the end of the week, students reflect on their goals and set new goals for the following week. They found that this training significantly improved students' academic achievement for math. Tuckman and Kennedy (2011) studied students who followed a course on different SRL-skills. These students overcame procrastination, built self-confidence, and were able to manage their lives better. They also had a higher academic achievement for their first four school terms. Wibrowski, Matthews, and Kitsantas (2017) found similar results. They followed students who had an intervention aimed to support students' motivational beliefs, use of SRL-skills, and academic achievement for four years. They found that students reported higher levels of motivation and study skills in their first year. Besides, students who were enrolled in this intervention had similar or higher academic achievement than students who were not enrolled. However, these differences diminished by the time students graduated.

On the other hand, not all studies are positive. For example, Pepper (2017) researched students who received four workshops with a total duration of seven hours, where they learned SRL-skills. Results of this study show that there was no significant difference in academic achievement between students who followed the workshops and students who did not. This may be caused by a lack of time. Students may not be able to use SRL-skills and their cognitive abilities simultaneously. Raaijmakers et al. (2018) developed an intervention specifically for self-assessment and task-selection in biology. They found that students' use of these SRL-skills did improve for biology, but not for math. Contradictory to what Schunk and Zimmerman (2011) theorize, they suggest that students cannot transfer SRL-skills. The reason is that, during

the SRL phase in math, students could have insufficient cognitive abilities to think about what they have learned during the training for biology problems and simultaneously think how that would translate to these new tasks. Van Gog et al. (2010) studied students who monitored their performance while performing a task and found that monitoring a task significantly decreased performance. They state that SRL-skills during a learning task can be seen as an additional task, which increases students' use of cognitive abilities.

Studies for SRL-skills are mostly positive; however, some studies have found that these skills are non-transferrable or have found no effect on academic achievement. Furthermore, these studies focus primarily on multiple phases of the SRL process instead of on each phase individually. The lack of effects might have been caused by students' cognitive abilities and may, therefore, be an essential factor to consider in SRL (Pepper, 2017; Raaijmakers et al., 2018; Van Gog et al., 2010).

Cognitive Abilities and Cognitive Load Theory

Human cognitive architecture (HCA) explains cognitive abilities (Sweller, 2008). HCA states there is an unlimited long-term memory which interacts with more limited working memory (Kirschner, Kester, & Corbalan, 2010). Information in long-term memory is stored and organized in schemata which incorporates multiple elements of information into a single element with specific functions which reduces working memory load (Sweller, Van Merriënboer, & Paas, 1998). These characteristics have implications for effective instructional design.

CLT considers this to provide theoretical insight when instruction is effective for learning (Sweller, 2008; Sweller, 2011). According to CLT, learning will occur when information is successfully processed into working memory, and new elements of information are incorporated, combined, or adapted into consisting schemata (Kirschner et al., 2010). Therefore, instruction is effective when it encourages schemata creation while making

optimal use of cognitive resources. CLT specifies three types of cognitive load (CL). These three types are additive and cannot exceed the available capacity of working memory (Sweller, 2008). First, intrinsic CL is embedded in instructional materials and therefore, cannot be altered when learning materials have many elements; students' experience a high intrinsic CL. Second, extraneous CL is caused by features of the materials that are not necessary for learning (Bergman et al., 2015; Sweller, 2008). The third type is germane CL, which can result in a useful alteration to long-term memory, and therefore enhances learning (Paas, Renkl, & Sweller, 2003; Sweller et al., 1998). Germane CL stimulates schemata building due to the instructional design of instructional materials.

In recent years, discussions arose about the empirical evidence of germane CL. Several studies found that germane CL was hard to measure and therefore could better be described as part of intrinsic CL (Haji, Rojas, Childs, De Ribaupierre, & Dubrowski, 2015; Choi, Van Merriënboer, & Paas, 2014). However, Leppink, Paas, Van der Vleuten, Van Gog, & Van Merriënboer, (2013) developed a way to measure experienced intrinsic, extraneous, and germane CL. In their questionnaire, they were aware of the methodological issues. However, their findings suggest that intrinsic, extraneous, and germane CL are significantly correlated. Also, germane CL will give relevant information about the way SRL-skills are valuable for schemata building and is therefore included in this study.

Combining Self-Regulated Learning and Cognitive Load Theory

Taking a cognitive perspective on SRL could add knowledge to the effect of interventions for SRL-skills on cognitive processes. Students' planning (i.e., goal setting, strategic planning, beliefs about learning, and outcome expectations) is a prerequisite for effective studying because their planning influences the use of effective learning strategies during learning tasks (Vrugt & Oort, 2008; Wirth et al., 2009; Zimmerman, 2002). Although not much is known about the influence of planning on students' cognitive processes. These

skills can have some important implications for the cognitive processes that occur while learning. In practice, making a planning does not interfere with instructional material itself. Instructional material will be the same with or without a planning. Therefore, intrinsic CL will not be interfered (Bergman et al., 2015). Students will be better prepared when they plan for a learning task and can, therefore, process new information better. Due to better processing, students will experience more germane CL (Wirth et al., 2009). Students who plan do not need to process SRL-skills during the learning task, because they have already done that before the learning task. This will decrease students' experienced extraneous CL (Van Gog et al., 2010). Also, because the additive nature of germane and extraneous CL (Sweller, 2008), it is logical that students who plan a learning task, experience less extraneous CL. Studies found that students' academic achievement was higher when they participated in an intervention for SRL-skills (Tuckman & Kennedy, 2011; Wibrowski et al., 2017). This may be a consequence of higher germane and lower extraneous CL. Lastly, students who plan a learning task, set goals, can process new information better and are better prepared than students who do not plan a learning task.

The research question answered in this study is: What influence does a planning phase intervention have on students' experienced CL and academic achievement during the learning phase? In this intervention, students who made a planning before making a learning task will be compared with students who did not.

Several hypotheses will be tested in this study. Firstly, it is expected that students who participate in the planning phase intervention will experience the same level of intrinsic CL than students who did not participate (H1). Students will receive the same instructional material; therefore, intrinsic CL will not differ. Secondly, students who participate in the planning phase intervention will experience more germane CL than students who do not (H2). This is expected because students who participate in the planning phase intervention are better

prepared to process new information than students who are not participating. Thirdly, because the planning phase intervention is facilitating learning, the expectation is that students who participate in the planning phase intervention experience lower extraneous CL than students who are not (H3). Students who will not participate in the planning phase intervention need to process their SRL-skills during the learning task, which will increase their experienced extraneous CL. This is also a logical consequence of higher germane CL because of the additive nature of the different types of CL. Finally, it is expected that students who participate in the planning phase intervention will have higher academic achievement than students who are not participating (H4). This expectation comes from several studies that found higher academic achievement when students participate in SRL-skills interventions (Tuckman & Kennedy, 2011; Wibrowski et al., 2017). Also, students who participate in the planning phase are expected to have higher germane CL, can process new information better, and are better prepared than students who do not participate in the planning phase which leads to higher academic achievement.

In this study, two between-subject experiments with a mixed-method design will be conducted to test the formulated hypotheses. Students from one sample had to work in pairs to finish their learning task, while the other group worked individually. Both groups had different learning tasks. Studying different groups provides more information about the influence of a planning phase intervention on individuals as well as groups and different learning tasks.

Experiment 1

Method

Design. This study had a mixed-method design, where qualitative and quantitative data were used (Tashakkori & Teddlie, 2003). Qualitative data in this study provided insight into participants' planning phase by asking them what and how they made decisions regarding their planning.

The research was conducted with a between-subject design. Planning was taken as the independent variable, participants received a planning phase intervention, or they did not. Dependent variables were intrinsic CL, extraneous CL, germane CL, and academic achievement.

Participants. Participants of this study were 25 students (11 male, 14 female) who signed up for a statistics course from Utrecht University with the level Bachelor Advanced. Age was not measured in questionnaires; however, students in this cohort had a mean age of 21. The study was conducted during a practical assignment of the course, where participants worked in pairs. All participants received the same lecture before the practical assignment and therefore had the same level of prior knowledge. Students could not participate in this research if they were retaking the course.

An a priori power analysis was computed using G*Power 3.1 (“G*Power: Statistical Power Analyses for Windows and Mac”, n.d.). For a large effect (Cohen, 1988), 108 participants were needed. A power size of 0.33 was achieved, which is considered small.

Participants were randomly assigned to the experimental group ($n = 14$, 9 male, 5 female) and control group ($n = 11$, 2 male, 9 female). In the control group, there was one threesome that made the practical assignment. Participants could voluntarily participate in this study by signing in via a form on BlackBoard (see Appendix A) or by signing in before they made the practical assignment. Participants all signed up voluntarily and received no compensation for participating. Before starting the experiment, participants signed an informed consent form (see Appendix B).

Instruments. Several instruments were used to gain information about participants’ experienced CL and academic achievement.

Practical assignment. The practical assignment consisted of several questions about the previously attended lecture and was made by the course coordinator. Participants had to answer

these questions using the statistical program *R*. Participants received the practical assignment and a document online where they could fill in their answers (see Appendix C and D).

Experienced CL. After the practical assignment, experienced CL was measured via an online questionnaire containing the ten-point mental effort rating (Leppink et al., 2013). This questionnaire contains three questions about the experienced intrinsic CL (for example, '*subjects in this assignment were very complex*'). These questions reveal a Cronbach's alpha of .81, which means that the reliability of these items is good. Question 4, 5, and 6 measure extraneous CL (for example, '*the instructions and/or explanations during the activity were very unclear*'). These questions show a Cronbach's alpha of .75, which means that the reliability is acceptable. Lastly, there were four questions about the experienced germane CL (for example, '*the assignment has really helped me to understand the concepts that have been discussed*'). Cronbach's alpha of these questions is .82, which is considered good. Overall, this questionnaire is a reliable and valid instrument to measure CL. The ten-point mental effort rating consisted of ten items and was scored on an 11-point scale from (0) *not at all the case* to (10) *completely the case* (see Appendix E). The ten-point mental effort rating has to be filled in as quickly as possible after a learning task is completed to preserve validity. Participants received this questionnaire five minutes before the end of the practical via email. One participant did not fill in the ten-point mental effort rating; other participants filled in the ten-point mental effort rating within ten minutes.

Academic achievement. Academic achievement was measured with grades from the practical assignment. The researcher graded participants' practical assignments using answering models provided by the course coordinator (see Appendix F and G). To determine the grades of the practical assignment, first the formula $100 / \text{number of questions} = x$ was used. Next, grades were determined using $\text{questions correct} \cdot x / 100$. Questions which were not

answered correctly, or not answered at all, were considered wrong. Practical assignments were graded between 0 (*very poor*) and 10 (*excellent*).

Planning phase assignment. All participants received a printed assignment they had to make during the planning phase of the practical assignment. Participants in the control group received an assignment with questions that they had to answer with good substantiations and did not involve any planning activity (see Appendix H). For example, students had to estimate how many people lived on the earth in 1763 and gave an elaborated solution for the housing problem in the Netherlands. These types of questions were used so students could not plan their practical assignment.

The experimental group received an assignment to make a planning. (see Appendix I). This format used questions from the Self-Regulation of Learning Self-Report Scale (SRL-SRS) (Toering, Elferink-Gemser, Jonker, Van Heuvelen, & Visscher, 2012). The original SRL-SRS contains 46 items in six subscales (planning, self-monitoring, evaluation, reflection, effort, and self-efficacy) and is intended to measure SRL in learning domains. In this assignment, only the subscale of planning was used. The format contained questions about time-management (*‘How much time do you have to finish this assignment’*), goal setting (*‘Which learning goals did you formulate for this assignment’*), which knowledge and competencies students should use to achieve these goals, and questions about students’ problem solving skills (*‘If you run into a problem, how are you going to solve this’*).

A random sample (n = 4, 4 male) was asked to participate in a short interview (see Appendix J), to gain more information about the planning phase assignment. This interview was structured with 11 questions. The first question *‘can you state your name and student number’* was asked to determine which practical assignment belonged to the pair of participants. Besides, several questions were asked to determine how seriously participants

were taking the planning phase assignment, and why (*'On a scale to 0 to 10, where 0 is not the case at all and 10 is totally the case, how seriously did you fill in the assignment'*).

Questions were based on the SRL-skills described by Hong and O'Neil (2001) and Toering et al., (2012): participants strategies (*'How did you fill in the assignment'* and *'Did you watch the practical assignment before you fill in the assignment'*), goal setting (*'Which goals did you set, and why'*), adjusting the plan (*'Did you adjust your plan while making the practical assignment, if yes, how?'*) and *'How many times did you look back at your assignment while making the practical assignment'*), reflecting on planning (*'Did you reflect on your assignment while making the practical assignment'* and *'What did you like and did not like about the assignment?'*), and problem solving skills (*'Did you have trouble making the practical assignment, if yes, how did you solve these problems?'*).

Procedure. First, students were asked to participate in this study on Blackboard. Because the response rate was low, the researcher decided to actively ask students to participate during a previous practical assignment.

Participants were situated in a room where other students were making their practical assignments. When students entered the room, they were asked to fill in their informed consent forms. After that, the researcher gave the participants their planning phase assignment and briefed them about the study. In this briefing, the researcher told participants they had 15 minutes to make their planning or questions in pairs, and then they had to continue making the practical assignment. After the practical assignment, all participants received an email with the ten-point mental effort scale. Participants had 90 minutes to finish the practical assignment. After participants finished the practical assignment, or when time ran out, they sent their practical assignment to the researcher, even if they did not finish. Participants also submit their planning phase assignment. Five minutes before the end of the practical, the researcher sent an email with a link to the ten-point mental effort rating and asked participants to send their

practical assignment via email and hand in their planning phase assignment. The researcher also asked participants from the experimental group if they had time to answer a few questions which were recorded. When participants were willing to be interviewed, the researcher walked to a quiet part of the hallway and conducted the interview. After three and seven days, the researcher emailed students a reminder to send their practical assignment.

When data were collected, the researcher graded the practical assignments and inserted the data into SPSS and transcribed the interviews.

Data analyses. First, one participant did not fill in the ten-point mental effort rating. Therefore, one pair was deleted from the study. Only eight pairs of students emailed their practical assignment to the researcher; the remaining practical assignments were reported as missing data.

To answer the research question, four independent Samples *t*-test were performed with academic achievement, intrinsic, extraneous, and germane CL as dependent variables and planning as an independent variable. Based on the hypotheses, it was decided to use a one-tailed test ($\alpha = .05$).

Four criteria were tested before conducting the Independent Samples *t*-test. The first criterion was methodological: data had to be interval or ratio. This criterion was not violated. During the research, the second criterion was also not violated because participants did not participate twice, nor did participants influence others. Then, normality was tested with an analysis in SPSS. Lastly, the homogeneity of variance was tested. When all assumptions were not violated, the independent *t*-test was conducted. However, academic achievement was not normally distributed; therefore, a Mann-Whitney *U* test was conducted. Intrinsic CL had substantially more variance in the control group. Therefore, Welch's *t*-test was conducted.

Furthermore, qualitative data were transcribed; subsequently, data were analyzed using NVivo. The method of analytical coding (Richards & Morse, 2012) was used to analyze the

data. The topics that were searched for were the SRL-skills for the planning phase described by Hong and O'neil (2001) and Toering et al. (2012): learners' strategies, goal setting, adjusting the plan, reflecting on planning, and problem-solving skills. Codes were created to analyze whether participants filled in their planning and if they had a negative or positive attitude towards the planning phase assignment.

Results

Four independent samples *t*-tests were used to compare experienced CL and academic achievement by participants in the experimental and the control group. Also, quantitative data were analyzed to determine which SRL-skills participants were using and if they made their planning phase assignment seriously.

Intrinsic Cognitive Load. Preliminary assumption testing indicated that both the experimental ($M = 7.89, SD = .77$) and control group ($M = 8.30, SD = 1.39$) intrinsic CL were all normally distributed. However, in the scores for intrinsic CL, there was substantially more variance in the control group. Consequently, Welch's *t*-test was used to compare the intrinsic CL average from the experimental group to the control group. The *t*-test was non-significant, $t(15.34) = -.88, p = .198$, one-tailed, $d = -.37$, 95% CI [-1.42, .59].

Germane Cognitive Load. Shapiro-Wilk statistic was non-significant, indicating that the assumption of normality was not violated, thus equal variances can be assumed. The *t*-test was statistically significant, with the experimental group ($M = 4.41, SD = 1.87$) reporting germane CL 2.19 higher, 95% CI [.58, 3.79] than the control group ($M = 2.22, SD = 1.83$), $t(21) = 2.83, p < .05$, one-tailed, $d = 1.18$.

Extraneous Cognitive Load. Shapiro-Wilk statistic was significant, indicating that the assumption of normality was not violated. Levene's test was also non-significant, thus equal variances can be assumed. The *t*-test was significant, with the experimental group ($M = 5.94$,

$SD = 1.57$) reporting extraneous CL to be -1.44 lower, 95% CI $[-2.93, .03]$ than the control group ($M = 7.39$, $SD = 1.84$), $t(21) = -2.04$, $p < .05$, one-tailed, $d = -.85$.

Academic Achievement. Preliminary assumption testing indicated that the achievement of both experimental group ($M = 4.35$, $SD = 2.47$) and control group ($M = 6.82$, $SD = .78$) were not normally distributed. Therefore, a Mann-Whitney U test was conducted. This test indicated the experimental group ($Mean Rank = 6.75$, $n = 8$) did not have significantly higher academic achievement than the control group ($Mean Rank = 11$, $n = 8$), $U = 18$, $z = -1.75$ (corrected for ties), $p < .05$, one-tailed

Planning. To analyze quantitative data, several codes were created.

Strategies. Two participants did not look into the practical assignment before making the planning '*We just made the planning collaboratively*'. Two other participants did look into the practical assignment before making their planning.

Goal setting. Two participants did not set a clear goal '*we just want to start and see how far we would come*', however, they did write down the actual goal of the practical assignment '*we had the goal to have better insight in the confirmatory factor analyses on multi groups, because that was the goal of the practical assignment*'. The other two participants set a time-related goal for themselves '*we wanted to be done before three-thirty*'.

Adjusting the plan. All participants did not adjust their planning while making their practical assignment.

Reflecting on the plan. All participants did not look back into their planning after they made it. One student mentioned that '*I completely forgot the planning*'.

Problem-solving skills. Two participants did not have any problems with the practical assignment. However, they had a plan when they would come across problems '*we would ask the practical teacher*'. Two participants did come across problems; they resolved this in the following manner: '*we looked back into the lecture slides*'.

Planning phase assignment attitude. Participants answered that they filled in the planning somewhat seriously ‘*I think we wrote down some useful things, but we did follow the manual*’. Two participants were relatively positive about the planning phase assignment ‘*I liked that the assignment gave me some tools*’ and ‘*It was easy that you could just fill it in*’. One student thought it would make it easier to achieve your goals when you write them down, ‘*I believe that, because we wrote it down, we were ready before three-thirty. I think you are more active*’. Two participants had a negative attitude towards the planning phase assignment: ‘*It had no added value*’.

Discussion

Experiment 1 was conducted to answer the research question ‘What influence does a planning phase intervention have on students’ experienced CL and academic achievement during the learning phase?’ The experiment provided more insight into the influence of a planning phase intervention on students’ experienced CL and academic achievement when working in pairs to finish a learning task. To answer this research question, several hypotheses were tested.

First, there was no effect of the planning intervention on intrinsic CL. This finding was in line with the hypotheses that intrinsic CL would not differ (H1). Therefore, this hypothesis can be accepted. This could mean that this planning intervention does not interfere with the instructional material. Second, there was a significant effect of the planning intervention on germane CL. The finding is in line with the hypothesis that students who were participating in the planning phase intervention would experience more germane CL than students who do not (H2). Therefore, this hypothesis can be accepted. There is a chance that students who planned their learning tasks are better prepared to process new information, and experience more germane CL than students who do not. Third, there was a significant effect of the planning intervention on extraneous CL. This finding was in line with the hypothesis that students who

were participating in the planning phase intervention would experience less extraneous CL than students who did not (H3). Therefore, this hypothesis can be accepted. There is a chance that students who did not plan their learning task needed to process their SRL-skills during the learning task, which can increase their experienced extraneous CL. Higher extraneous CL is also a logical consequence of lower germane CL

These results are interesting for instructional designers, teachers, and future researchers. It seems that the planning intervention used in this experiment has a positive effect on students' experienced CL when working in pairs. Therefore, instructional designers can use this intervention when they want pairs of students to plan their learning task. Future research can focus more in-depth on students' used skills, to provide more insights into the effects of a planning intervention on different types of CL.

Finally, there was a significant effect of the planning intervention on academic achievement. Students who did not participate in the planning intervention received higher grades than students who did participate in the planning phase intervention. This contradicted the hypothesis and, therefore, this hypothesis can be rejected. This finding is contradictory to several studies that found higher academic achievement when students participate in SRL-skills interventions (e.g., Tuckman & Kennedy, 2011; Wibrowski et al., 2017). Also, it is remarkable that students experienced higher germane CL but scored lower on academic achievement, because germane CL should be beneficial for learning (Paas et al., 2003; Sweller et al., 1998). Higher germane CL can indicate that students are better prepared and better able to process new information. These findings are contradictory to previous research and can be explained using the conducted interviews.

Students did take time to fill in their planning phase assignment. However, not all SRL-skills were addressed in the planning phase of these students. For example, students did not adjust or reflect on their planning. One student mentioned that he forgot about the planning

phase assignment. This could mean that students did not fill in the planning altogether, and therefore did not address all SRL-skills for the planning phase described by Hong and O'neil (2001) and Toering et al. (2012). It is also possible that students who participated in the planning phase intervention did not have enough time to process their cognitive abilities and SRL-skills at the same time, similarly as students in the study conducted by Pepper (2017). This could also have led to lower academic achievement. Because academic achievement was lower in the experimental group, the intervention given in this experiment should be carefully approached when students receive a grade for academic achievement in future research.

Summarizing, this study indicates that a planning phase intervention may have a positive effect on overall experienced CL. However, this effect does not translate into higher academic achievement. Even though students who participated in the planning phase intervention experienced more germane CL and less extraneous CL, they had an overall lower academic achievement than students who did not participate in the planning phase intervention.

The results of this study could also have been influenced by several limitations. First, students were responsible for emailing their assignments after the practical assignment. The response rate to this appeal was lower than expected, because students forgot, thought their partner emailed the practical assignment already, or did not respond anymore. Second, students received an email with the ten-point mental effort rating; the researcher had no view on how students filled in this questionnaire. Some students filled in the questionnaire later than others, which could have influenced the experienced CL. The ten-point mental effort rating was an online questionnaire. It is more likely that people fill in an online questionnaire with socially desirable answers (Fang, Prybutok, & Wen, 2016). Only four participants were willing to participate in the interview. Other participants did not want to because of time pressure.

Another experiment was set up for a different group of students. The overall setup for the experiment was the same. However, some changes were made regarding the limitations of

experiment 1. Unfortunately, due to practical reasons, not all limitations could be taken into account while setting up experiment 2.

All questionnaires and assignments were handed out in person for experiment 2. Participants were asked to fill these in after the practical. This could mean that the response rate is higher in experiment 2. Participants did not have to hand in their practical assignment because they could not be graded. Academic achievement was measured with self-measurement in experiment 2. Because only a few participants were willing to participate in the interview in experiment 1, open questions were added to the questionnaire. Therefore, the response rate could be higher. In experiment 2, students work individually to finish their learning task. This will provide more information about the influence of a planning phase intervention on individuals.

Experiment 2

Method

Design. The research design did not differ from experiment 1.

Participants. Participants in this study were 29 students (25 female, 4 male, $M_{age} = 20.1$ years, $SD = 1.44$) who signed up for a statistics course from Utrecht University with the level Bachelor Elaborating. The study was conducted during a practical assignment of the course, where participants worked individually. Participants just had a lecture about the information they need for the practical assignment and had, therefore, the same level of prior knowledge. Students could not participate in this research if they were retaking the course.

An a priori power analysis was computed using G*Power 3.1 (“G*Power: Statistical Power Analyses for Windows and Mac”, n.d.). For a large effect (Cohen, 1988), 108 participants were needed. A power size of 0.35 was achieved, which is considered small.

Participants were randomly assigned to the experimental group ($n = 18$, 4 male, 14 female) and the control group ($n = 11$, 11 female). Participants could voluntarily participate by

signing in before the practical assignment and received an informed consent form before they started this study (Appendix K).

Instruments. Similar as in experiment 1, several instruments were used to gain information about participants' experienced CL and academic achievement. All instruments were handed out in a printed form.

Practical assignment. The practical assignment consisted of several questions about the previously attended lecture made by the course coordinator. Participants could open the practical assignment via BlackBoard and had to answer questions using the statistical program SPSS. In contrast with the practical assignment in experiment 1, the students were not allowed to collaborate.

Experienced CL. After the practical assignment, the experienced CL was measured with the ten-point mental effort rating (Leppink et al., 2013). The ten-point mental effort rating also contained questions about the demographic information, participants' planning, and self-measurement (see Appendix L).

Academic achievement. This practical assignment was not graded; therefore, academic achievement was measured using self-measurement. Participants were asked to fill in a self-measurement question. Participants could grade themselves between 0 (*very poor*) and 10 (*excellent*). Self-measurement can be seen as a formative assessment where students can reflect on the quality of their work, judge the degree to which it reflects explicitly stated goals or criteria, and revise accordingly (Andrade & Cizek, 2010). The internal consistency of self-measurement is typically high (Ross, Rolheiser, & Hogaboam-Gray, 2002). The validity of self-measurement is mixed (Ross, 2006). However, Arthur (1995) stated that self-measurement is more useful for learners as a formative tool rather than a summative tool. Because participants were not graded for their practical assignment, using their self-measurement was considered as a reliable method to measure academic achievement.

Planning phase assignment. The planning phase assignments for the experimental and control group did not differ from experiment 1. However, the planning format was adapted from plural to singular because students worked individually (see Appendix L).

Because students had to attend a lecture directly after the practical, participants could not participate in an interview. Therefore, participants from the experimental group were asked to answer a question regarding their planning (*'How seriously did you fill in the planning'*). This question was on the same paper as the ten-point mental effort rating.

Procedure. The research procedure did not differ from experiment 1.

Data analyses. Data analyses were similar as in experiment 1. Four criteria were tested before conducting the Independent Samples *t*-test. When all assumptions were not violated, the independent *t*-test was conducted. Germane CL and academic achievement were not normally distributed. Therefore, for these variables, a Mann-Whitney *U* test was conducted.

In this experiment, one participant did not fill in their questionnaire completely and was therefore deleted from the study. Quantitative data has been analyzed by transcribing all answers using three codes in NVivo: positive, neutral, and negative.

Results

Four independent samples *t*-tests were used to compare experienced CL and academic achievement from participants in the experimental group to experienced CL and academic achievement from participants in the control group. Quantitative data were analyzed to determine if participants were making the planning phase assignment seriously.

Intrinsic Cognitive Load. Preliminary assumption testing indicated that both the experimental group ($M = 7.79, SD = .85$) and control group ($M = 7.96, SD = 1.14$) intrinsic CL were all normally distributed. Neither Shapiro-Wilk statistic or Levene's test were significant, indicating that the assumption of normality was violated, and equal variances can be assumed. The *t* test was non-significant, $t(27) = -.44, p = .332$, one-tailed, $d = .17$, 95% CI [-.99, .64].

Germane Cognitive Load. Preliminary assumption testing indicated that the control group ($M = 6.93$, $SD = 1.58$) was normally distributed. However, the experimental group ($M = 6.43$, $SD = 1.45$) was not normally distributed. Therefore, a Mann-Whitney U test was conducted. This test indicated the experimental group ($Mean Rank = 14.08$, $n = 18$) did not experience less germane CL than the control group ($Mean Rank = 16.5$, $n = 11$), $U = 82.5$, $z = -.75$ (corrected for ties), $p = .228$, one-tailed.

Extraneous Cognitive Load. For extraneous CL, neither Shapiro-Wilk statistic nor Levene's test were significant. The assumption of normality was not violated, and equal variances can be assumed. The t test was non-significant, with the experimental group ($M = 4.48$, $SD = 1.63$) reporting extraneous CL to be 0.9 higher, 95% CI [-2.33, .52] than the control group ($M = 3.58$, $SD = 2.10$), $t(27) = -1.30$, $p = .102$, one-tailed, $d = 0.48$.

Academic Achievement. Preliminary assumption testing indicated that both experimental group ($M = 5.83$, $SD = 1.34$) and control group ($M = 6.36$, $SD = 1.29$) academic achievement were not normally distributed. Therefore, a Mann-Whitney U test was conducted. This test indicated the experimental group ($Mean Rank = 13.83$, $n = 18$) did not have significantly higher academic achievement than the control group ($Mean Rank = 16.91$, $n = 11$), $U = 78$, $z = -.99$ (corrected for ties), $p = .159$, one-tailed.

Planning. To analyze if participants took the planning phase assignment seriously, the question at the end of the questionnaire was analyzed. Two participants did not take the planning phase assignment seriously '*not very seriously, planning was not necessary for this assignment*' and '*for this assignment, planning was not necessary*'. Two participants did take the planning phase assignment seriously but did not make it very detailed '*I did what I wrote down, in that respect, seriously. It was not very elaborate*' and '*It was a bit vague, you just had to make the assignment, you start with A, and you will see where you end up. Maybe short, but seriously*'. Thirteen participants took the planning phase assignment seriously.

Discussion

Experiment 2 was conducted to answer the research question ‘What influence does a planning phase intervention have on students’ experienced CL and academic achievement during the learning phase?’ This experiment provided more insight into the influence of a planning phase intervention on students’ experienced CL and academic achievement when working individually to finish a learning task. To answer this research question, several hypotheses were tested.

First, there was no effect of the planning intervention on intrinsic CL. This finding is in line with the hypotheses that intrinsic CL would not differ (H1). Therefore, this hypothesis can be accepted. This could mean that this planning intervention before a learning task does not interfere with the instructional material. Second, there was no effect of the planning intervention on germane CL. This finding is not in line with the hypothesis that students who were participating in the planning phase intervention would experience more germane CL than students who do not (H2). Therefore, this hypothesis can be rejected. There is a chance students’ in this group did not have the cognitive resources to simultaneously think about their planning phase intervention and the learning task they had to finish. These findings were not in line with the findings of Wirth et al. (2009). They suggest that students who plan their learning task are better prepared to process new information and will, therefore, experience more germane CL than students who do not plan. Also, germane CL and extraneous CL are additive (Sweller, 2008). Therefore, another explanation can be found in the results of extraneous CL. Third, there was no effect of the planning intervention on extraneous CL. This finding is not in line with the hypothesis that students who were participating in the planning phase intervention would experience less extraneous CL than students who do not (H3). Therefore, this hypothesis can be rejected. This indicates that the planning phase intervention was not necessary for learning (Bergman et al., 2015; Sweller, 2008). This can also explain the results for lower germane CL

because this type of CL is enhancing learning (Paas et al., 2003). The findings of the germane and extraneous CL may indicate that this planning phase intervention was hampering students' learning because they may have not enough cognitive resources to simultaneously think about their planning phase intervention and the learning task they had to finish. This is supported by the question about how seriously students in the experimental group filled in their planning phase intervention. The majority of the students took the assignment seriously. However, two students did think the planning phase assignment was not suitable for the practical assignment. The planning phase intervention may not have been suitable for this learning task, which could increase extraneous CL. These findings suggest that this planning phase intervention may not be suitable for learning tasks which students have to finish individually.

Finally, there was no effect of the planning intervention on academic achievement. However, students who did not participate in the planning intervention received higher grades than students who did participate in the planning phase intervention. This was not in line with the hypotheses that students who participate in the planning phase intervention will have a higher academic achievement (H4). Therefore, this hypothesis can be rejected. Even though this was not the expectation, it is a logical consequence of the results of germane and extraneous CL. Because germane CL was lower in the experimental group, and extraneous CL was higher, this could mean that the planning phase intervention was hampering learning and therefore, academic achievement was lower.

Students in experiment 2 experienced the same level of intrinsic CL. However, students in the experimental group had lower academic achievement, which may be caused by lower germane and higher extraneous CL. Therefore, it is recommended that this planning phase intervention should not be used when students must finish a learning task individually.

The results of this study could have been influenced by several limitations. Firstly, students did not receive a grade for their practical assignment but were asked to use self-

measurement so academic achievement could be measured for this study. Research shows that students who are using self-measurement need to be trained to learn how to grade accurately and lower performing students tend to inflate their grades (Sadler & Good, 2006). Also, due to practical reasons, students were not interviewed but had to answer a question about how seriously they filled in their planning phase intervention. Answers to this question were short and not as fulfilling as interviews. Therefore, conclusions drawn about the used SRL-skills in this experiment should be interpreted cautiously.

Summary and Concluding Discussion

Two experiments were conducted to answer the research question ‘What influence does a planning phase intervention have on students’ experienced CL and academic achievement during the learning phase?’ These experiments will provide more insight into the influence of a planning phase intervention on students’ experienced CL and academic achievement when working individually or in pairs to finish a learning task. To answer this research question, several hypotheses were tested.

Both experiments showed no difference in intrinsic CL. This was in line with the expectations (H1), and therefore, the first hypothesis can be accepted. This indicates that this planning phase intervention does not interfere with the learning task.

Second, germane CL was significantly higher when students planned their learning task in a group but were lower when students worked alone. This finding was not entirely in line with the hypothesis that students who were participating in the planning phase intervention would experience more germane CL than students who do not (H2). This hypothesis can be, therefore, partly accepted, partly rejected.

Third, extraneous CL was significantly lower when students planned their learning task in a group but were higher when students worked alone. This finding was not entirely in line with the hypothesis that students who were participating in the planning phase intervention

would experience less extraneous CL than students who do not (H3). This hypothesis can be, therefore, partly accepted, partly rejected.

Germane and extraneous CL has an additive nature. The differences between these two types of CL will be explained together. The differences between these types of load between the two groups may be explained by research conducted by Kirschner, Paas, and Kirschner (2011). In this study, they found that group members learned in a more efficient way for high-complexity tasks than individual learning. This could be explained by the theory that information processes could be divided amongst all working memories of group members. This division of information processes could also have been happening in experiment 1. In experiment 2 students worked individually, and therefore, students in experiment 2 could not have been divided information processes and experienced lower germane CL and higher extraneous CL.

Lastly, academic achievement was lower in the experimental group in both experiments. This finding is not in line with the hypothesis that students who participate in the planning phase intervention will have a higher academic achievement (H4). Therefore, this hypothesis can be rejected. An explanation of these findings may be that the intervention used in this study was not impactful enough. Studies who did find an effect on training SRL-skills all had more extensive interventions, like a training or a whole course (e.g., Hofer & Yu, 2003; Tuckman & Kennedy, 2011).

In sum, this planning phase intervention could harm students' academic achievement. When students are working in pairs, it may have a positive effect on students' experienced CL, but a negative effect when students are working individually. This intervention can be useful for learning tasks where students will not be graded.

These findings can be explained by several limitations. First, the sample size of this study was unsatisfactory. With a larger sample size, more conclusions can be drawn, because

power will be higher, and more conclusions can be drawn. Also, this sample consisted of mainly female students, which could have influenced the results. It could be that this intervention may have more influence on male students because research suggests that female students show more goal-setting and planning strategies than male students (Pajares, 2002). Second, students who were participating in the planning phase intervention did not address all SRL-skills for the planning phase described by Hong and O'neil (2001) and Toering et al. (2012). Students also thought the planning phase intervention was not suitable for the practical assignment, and not all students were taking it seriously. Third, academic achievement measurement could be improved. Some students in experiment 1 did not have sufficient time to finish their practical assignment, which could have been influenced their grades. In experiment 2, students were asked to use self-measurement for their practical assignment. This could have made academic achievement less valid and reliable (Sadler & Good, 2006).

Future research, therefore, should focus on a larger, more mixed sample size. Similar to research conducted by Tuckman and Kennedy (2011) and Wibrowski et al. (2017), first-year university students would be a satisfactory group because they can benefit from the learned skills mostly. Besides, the assignment for the planning phase could be more comprehensive. A short course to teach students how to plan their assignments would be even more satisfactory. In this course, all SRL-skills from the planning phase could be addressed more deeply. In this study, it would be interesting to investigate the differences between students who work in groups and students who work individually more deeply. This can be done by using think-aloud protocols for students who are planning an assignment; this way, more insights in students' planning phases are obtained (Zimmerman, 2008). When researchers design new studies in this field, they should always choose interviews above questionnaires, because data will be richer and therefore, more conclusions can be drawn. Lastly, the academic achievement should be carefully measured. It could be helpful to use Grade Point Averages for students who did follow

a course or intervention at the beginning of the year to research if they have higher academic achievement than students who did not.

Even though this research did not fully support all hypotheses, the planning phase and CL will always be an essential part of learning. Planning an assignment is an essential part of learning and can help students help regulate their learning (Zimmerman, 2002). Regulating learning is a prerequisite for higher academic achievement. This study tried to find an explanation regarding the mixed results in why some interventions for SRL-skills work and others do not. More research is necessary to find explanations which aspects of SRL-skills encourages students' cognitive abilities and therefore academic achievement. This knowledge is necessary to provide students with the best education possible.

References

- Andrade, H., & Cizek, G. J. (Eds.). (2010). *Handbook of formative assessment*. Routledge.
- Arthur, H. (1995). Student self-evaluations: How useful? How valid? *International Journal of Nursing Studies*, 32(3), 271-276. [https://doi.org/10.1016/0020-7489\(94\)00043-J](https://doi.org/10.1016/0020-7489(94)00043-J)
- Bergman, E. M., de Bruin, A. B., Vorstenbosch, M. A., Kooloos, J. G., Puts, G. C., Leppink, J., & Van der Vleuten, C. P. (2015). Effects of learning content in context on knowledge acquisition and recall: a pretest-posttest control group design. *BMC medical education*, 15(1). doi:10.1186/s12909-015-0416-0
- Bol, L., Campbell, K. D. Y., Perez, T., & Yen, C.-J. (2015). The effects of self-regulated learning training on community college students' metacognition and achievement in developmental math courses. *Community College Journal of Research and Practice*, 40(6), 480–495. <https://doi.org/10.1080/10668926.2015.1068718>
- Choi, H. H., Van Merriënboer, J. J., & Paas, F. (2014). Effects of the physical environment on cognitive load and learning: towards a new model of cognitive load. *Educational Psychology Review*, 26(2), 225-244. doi:10.1007/s10648-014-9262-6
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*, 2nd edition
Lawrence Erlbaum Associates: Hillsdale, NJ, USA.
- De Bruin, A. B., & Merriënboer, van J. J. (2017). Bridging cognitive load and self-regulated learning research: A complementary approach to contemporary issues in educational research. *Learning and Instruction*, 51, 1-9.
<https://doi.org/10.1016/j.learninstruc.2017.06.001>
- Fang, J., Prybutok, V., & Wen, C. (2016). Shirking behavior and socially desirable responding in online surveys: A cross-cultural study comparing Chinese and American samples. *Computers in human behavior*, 54, 310-317.
<https://doi.org/10.1016/j.chb.2015.08.019>

G*Power: Statistical Power Analyses for Windows and Mac (n.d.). Retrieved

<http://www.gpower.hhu.de/>

Haji, F. A., Rojas, D., Childs, R., De Ribaupierre, S., & Dubrowski, A. (2015). Measuring cognitive load: performance, mental effort and simulation task complexity. *Medical education, 49*(8), 815-827. doi:10.1111/medu.12773

Heikkilä, A., & Lonka, K. (2006). Studying in higher education: students' approaches to learning, self-regulation, and cognitive strategies. *Studies in higher education, 31*(1), 99-117. <https://doi.org/10.1080/03075070500392433>

Hofer, B. K., & Yu, S. L. (2003). Teaching self-regulated learning through a "Learning to Learn" course. *Teaching of Psychology, 30*(1), 30-33. doi:10.1207/S15328023TOP3001_05

Hong, E., & O'Neil Jr, H. F. (2001). Construct validation of a trait self-regulation model. *International Journal of Psychology, 36*(3), 186-194. doi:10.1080/00207590042000146

Kirschner, F., Kester, L., & Corbalan, G. (2010). Cognitive load theory and multimedia learning, task characteristics, and learning engagement: The current state of the art. *Computers in Human Behavior, 27*(1), 1-4 doi:10.1016/j.chb.2010.05.003

Kirschner, F., Paas, F., & Kirschner, P. A. (2011). Task complexity as a driver for collaborative learning efficiency: The collective working-memory effect. *Applied Cognitive Psychology, 25*(4), 615-624. doi:10.1002/acp.1730

Kitsantas, A., Winsler, A., & Huie, F. (2008). Self-Regulation and Ability Predictors of Academic Success During College: A Predictive Validity Study. *Journal of Advanced Academics, 20*(1), 42-68. <https://doi.org/10.4219/jaa-2008-867>

Leppink, J., Paas, F., Van der Vleuten, C. P., Van Gog, T., & Van Merriënboer, J. J. (2013). Development of an instrument for measuring different types of cognitive load.

Behavior research methods, 45(4), 1058-1072.

<https://doi.org/10.3758/s13428-013-0334-1>

Paas, F., Renkl, A., & Sweller, J. (2003). Cognitive load theory and instructional design:

Recent developments. *Educational psychologist*, 38(1), 1-4.

doi: 10.1207/S15326985EP3801_1

Pajares, F. (2002). Gender and Perceived Self-Efficacy in Self-Regulated Learning. *Theory*

Into Practice, 41(2), 116–125. https://doi.org/10.1207/s15430421tip4102_8

Pepper, C.A. (2017). *Fostering Self-Regulated learning at an Independent School* (Doctoral dissertation). Retrieved from John Hopkins University.

<http://jhir.library.jhu.edu/handle/1774.2/44726>

Pintrich, P. R. (2000). The role of goal orientation in self-regulated learning. In *Handbook of self-regulation* (pp. 451-502). Academic Press.

Raaijmakers, S. F., Baars, M., Paas, F., Van Merriënboer, J. J., & Van Gog, T. (2018).

Training self-assessment and task-selection skills to foster self-regulated learning: Do trained skills transfer across domains?. *Applied cognitive psychology*, 32(2),

270-277. <https://doi.org/10.1002/acp.3392>

Richards, L., & Morse, J. M. (2012). *Readme first for a user's guide to qualitative methods*.

Sage.

Ross, J. A. (2006). The reliability, validity, and utility of self-assessment. *Practical*

Assessment, Research and Evaluation, 11(10), 1-13

Ross, J. A., Hogaboam-Gray, A., & Rolheiser, C. (2002). Student self-evaluation in grade 5-6

mathematics effects on problem-solving achievement. *Educational Assessment*, 8(1),

43-58. https://doi.org/10.1207/S15326977EA0801_03

Rytkönen, H., Parpala, A., Lindblom-Ylänne, S., Virtanen, V., & Postareff, L. (2012). Factors

- affecting bioscience students' academic achievement. *Instructional Science*, 40(2), 241-256. doi:10.1007/s11251-011-9176-3
- Sadler, P. M., & Good, E. (2006). The impact of self-and peer-grading on student learning. *Educational assessment*, 11(1), 1-31. doi:10.1207/s15326977ea1101_1
- Schunk, D. H., & Zimmerman, B. (2011). *Handbook of Self-Regulation of Learning and Performance*. Taylor & Francis.
- Sweller, J. (2008). Human cognitive architecture. In M. Spector et al. (Eds.), *Handbook of research for educational communications and technology* (3rd ed., pp. 369–383). Lawrence Erlbaum Associates.
- Sweller, J. (2011). Cognitive load theory. In *Psychology of learning and motivation* (Vol. 55, pp. 37-76). Academic Press. doi:10.1007/978-1-4419-8126-4
- Sweller, J., Van Merriënboer, J. J., & Paas, F. G. (1998). Cognitive architecture and instructional design. *Educational psychology review*, 10(3), 251-296. <https://doi.org/10.1023/A:1022193728205>
- Tashakkori, A., & Teddlie, C. (2010). *SAGE Handbook of Mixed Methods in Social & Behavioral Research*. SAGE Publications.
- Toering, T., Elferink-Gemser, M. T., Jonker, L., van Heuvelen, M. J., & Visscher, C. (2012). Measuring self-regulation in a learning context: Reliability and validity of the Self-Regulation of Learning Self-Report Scale (SRL-SRS). *International Journal of Sport and Exercise Psychology*, 10(1), 24-38. doi:10.1002/acp.1726
- Tuckman, B. W., & Kennedy, G. J. (2011). Teaching Learning Strategies to Increase Success of First-Term College Students. *The Journal of Experimental Education*, 79(4), 478–504. <https://doi.org/10.1080/00220973.2010.512318>
- Van Gog, T., Kester, L., & Paas, F. (2010). Effects of concurrent monitoring on cognitive

- load and performance as a function of task complexity. *Applied Cognitive Psychology*, 25(4), 584–587. <https://doi.org/10.1002/acp.1726>
- Vrugt, A., & Oort, F. J. (2008). Metacognition, achievement goals, study strategies and academic achievement: pathways to achievement. *Metacognition and Learning*, 3(2), 123-146. <https://doi.org/10.1007/s11409-008-9022-4>
- Wibrowski, C. R., Matthews, W. K., & Kitsantas, A. (2017). The role of a skills learning support program on first-generation college students' self-regulation, motivation, and academic achievement: A longitudinal study. *Journal of College Student Retention: Research, Theory & Practice*, 19(3), 317-332. doi:10.1177/1521025116629152
- Wirth, J., Künsting, J., & Leutner, D. (2009). The impact of goal specificity and goal type on learning outcome and cognitive load. *Computers in Human Behavior*, 25(2), 299-305 doi: 10.1016/j.chb.2008.12.004
- Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. *Theory into practice*, 41(2), 64-70. doi:10.1207/s15430421tip4102_2
- Zimmerman, B. J. (2008). Investigating self-regulation and motivation: Historical background, methodological developments, and future prospects. *American educational research journal*, 45(1), 166-183. doi: 10.3102/0002831207312909
- Zimmerman, B.J., & Pons, M. M. (1986). Development of a Structured Interview for Assessing Student Use of Self-Regulated Learning Strategies. *American Educational Research Journal*, 23(4), 614–628. <https://doi.org/10.3102/00028312023004614>
- Zumbrunn, S., Tadlock, J., & Roberts, E. D. (2011). Encouraging self-regulated learning in the classroom: A review of the literature. *Metropolitan Educational Research Consortium (MERC)*, 1-28.

Appendix A
Form BlackBoard

OPROEP DEELNAME MASTERTHESIS ONDERZOEK

Beste student,

Wie ben ik?

Mijn naam is Anniek de Kort, masterstudent Educational Sciences hier aan de Universiteit Utrecht. In het kader van mijn master thesis doe ik, onder begeleiding van dr. Femke Kirschner, een onderzoek naar *self regulated learning*

Wat is het doel van het onderzoek?

Mijn onderzoek gaat over de invloed van zelfregulatie op de resultaten van een statistiek practicum. Daarnaast ben ik geïnteresseerd in de invloed deze opdracht op de cognitieve belasting die studenten ervaren.

Wat houdt het onderzoek in?

Ik zal je tijdens het komende statistiek practicum (in week 9 of 10) vragen om 15 min. voorafgaand aan het practicum kort de tijd te nemen om een opdracht te maken. Vervolgens maak je het practicum als gebruikelijk. Aan het einde van het practicum vul je een korte vragenlijst in van ongeveer vijf minuten. Daarnaast kan het zijn dat je aan het einde van het practicum gevraagd wordt kort een aantal vragen te beantwoorden welke zullen worden opgenomen. De opnames zullen vertrouwelijk behandeld worden. Gesprekken worden geanonimiseerd en gedoceerd zodat geen enkele data is terug te leiden naar individuen. Enkel gegevens die nodig zijn om het onderzoek uit te voeren zullen worden gebruikt.

Het onderzoek zal niet meer dan 20 minuten van jullie kostbare tijd kosten, maar zal mij enorm helpen! Via onderstaande link kun je meer informatie vinden en je inschrijven voor het onderzoek. Let op! Je kan alleen meedoen als ook je partner van de practica zich in wil schrijven voor het onderzoek.

https://docs.google.com/forms/d/e/1FAIpQLSfBt7pTrWmQ09p9GphUx1-GJq5WBIDgldQrg6vlxKYjhyslmg/viewform?usp=sf_link

Appendix B
Informed Consent Experiment 1

Informed Consent studenten practicum A&E

Betreft: Onderzoek naar de invloed van zelfregulatie op de resultaten van een statistiek practicum en de rol van cognitieve belasting hierbij.

Ik heb de uitnodiging voor de proefpersoon gelezen, waarin informatie stond over het onderzoek, en kon aanvullende vragen stellen aan de onderzoeker (Anniek de Kort, a.i.m.dekort@uu.nl) wanneer ik iets niet begreep. Mijn vragen zijn genoeg beantwoord en ik had genoeg tijd om te beslissen of ik meedoe aan dit onderzoek.

Ik weet dat meedoen geheel vrijwillig is. Ik weet dat ik op ieder moment kan beslissen om toch niet mee te doen, daarvoor hoef ik geen reden te geven.

Ik geef hierbij toestemming om onderstaande gegevens te gebruiken:

- de opdracht die ik voorafgaand aan het practicum zal maken met een medestudent,
- de opdracht gemaakt tijdens het practicum,
- de antwoorden die ik gaf op de vragen van de vragenlijst na afloop van het practicum,
- de antwoorden die ik gaf op de vragen van het interview, die gingen over het maken van de opdracht voorafgaand aan het practicum en het gebruik van deze opdracht tijdens het practicum.

Naam proefpersoon:

Handtekening:

Datum : __ / __ / __

Ik verklaar hierbij dat ik deze proefpersoon volledig heb geïnformeerd over het onderzoek.

Naam onderzoeker (of diens vertegenwoordiger):

Handtekening:

Datum: __ / __ / __

Meetinvariantie

In dit deel van het tweede practicum voer je multi-groep confirmatieve factoranalyses uit in R. Aan de hand van opdrachten pas je de al opgedane kennis over multi-groep confirmatieve factoranalyses toe en verkrijg je de vaardigheid om zelf multi-groep confirmatieve factoranalyses uit te voeren. De opdrachten vergroten je inzicht in wat een multi-groep confirmatieve factoranalyse is en wat je er in de (onderzoeks)praktijk aan hebt.

Lever voor aankomend weekend, dus voor aanstaande zaterdag, samen met degene met wie je het practicum doet, één elektronische versie van de uitwerking van het practicum in via SafeAssignment op Blackboard. De practicumbegeleiders kijken de uitwerkingen van de practica na en geven feedback.

- Gebruik de Word-file **A&E Practicum 2.docx** om je antwoorden in te noteren. Deze Word-file kun je vinden op Blackboard in de map Practicum 2 onder Course Content - Practicum materialen.
- Noteer bij het rapporteren van getallen maximaal twee decimalen.
- Geef, waar nodig, een korte toelichting bij de antwoorden.
- Antwoord niet uitsluitend met een getal. Noteer bijvoorbeeld: De gemiddelde leeftijd is 13.34 jaar.
- Gebruik bij een statistische toets een nominaal significantieniveau van 5%.

We maken in dit deel van het tweede practicum gebruik van de databestanden **Angst.txt** en **moodschl.txt**. Je kunt deze bestanden vinden op Blackboard in de map Databestanden (onder Course Content - Practicum materialen). Kopieer deze bestanden naar de op de U:-schijf aangemaakte map 'A&E Computerpractica'.

1. Invoeren van de data in R

Open R. Voordat je R weer gaat gebruiken voor statistische analyses, is het weer handig om aan R de werkmap 'A&E Computerpractica' op te geven, d.m.v. `Change dir...`, enz. Verder is het verstandig om in een nieuw script-bestand te werken. Open een nieuw script-bestand en bewaar dit bestand meteen in je werkmap 'A&E Computerpractica' onder een toepasselijke naam.

Importeer het databestand **Angst.txt** in R door het geven van het volgende commando

```
Angst<-read.table("Angst.txt",header=TRUE)
```

Typ **Angst** in de R Console en druk op de Return-toets om de data te bekijken. Het databestand bevat de scores van 150 respondenten op vier items die zijn geconstrueerd om angst voor dieren, existentiële angst, sociale angst en fobische angst te meten. Elk van de vier items bevat een stelling en aan iedere respondent is gevraagd voor iedere stelling op een vijfpuntsschaal aan te geven in welke mate de stelling op hem of haar van toepassing is. De inhoud van de stellingen is in dit practicum niet van belang en worden ook niet

gegeven. De stellingen zijn wel allemaal zo geformuleerd dat een hogere score een hogere angst aangeeft. De itemscores hoeven dus niet te worden gehercodeerd. De vier itemscores zijn A1, A2, A3 en A4. In principe zijn de itemscores van ordinaal meetniveau, maar voor het gemak wordt er in dit practicum van uitgegaan dat de itemscores van interval meetniveau zijn. Verder bevat het bestand de groepsvariabele `lftd`. Deze groepsvariabele deelt de steekproef op in twee groepen. De eerste groep (aangeduid met score 1) bestaat uit de respondenten met een leeftijd lager dan de gemiddelde leeftijd in de hele steekproef en de tweede groep (aangeduid met score 2) bestaat uit de respondenten met een hogere leeftijd dan de gemiddelde leeftijd in de hele steekproef.

2. Multi-groep confirmatieve factoranalyses

Er wordt verondersteld dat de vier items één en dezelfde gemeenschappelijke angst-factor meten in beide groepen. Deze veronderstelling kan worden onderzocht door de vier vormen van meetinvariantie te toetsen met behulp van multi-groep confirmatieve factoranalyses. Vervolgens kan een aanvullende multi-groep confirmatieve factoranalyse worden uitgevoerd om te onderzoeken of er een verschil in gemiddelden op de gemeenschappelijke angst-factor bestaat tussen de twee groepen.

De multi-groep confirmatieve factoranalyses kunnen worden uitgevoerd met behulp van het R pakket `lavaan`. Installeer `lavaan` als deze nog niet op de computer staat waar je nu mee werkt. Laad `lavaan` vervolgens.

Het eerste multi-groep model dat op de data wordt gepast, is het multi-groep één-factormodel met configurale invariantie over de twee groepen. Je kunt dit model op de data passen door het geven van de volgende drie commando's.

```
mod0<-"f1=~A1+A2+A3+A4"  
fit0<-cfa(mod0,data=Angst,group="lftd")  
summary(fit0,fit.measures=T)
```

Met het eerste commando wordt de specificatie van het één-factormodel weggeschreven onder de naam `mod0`. In het tweede commando heeft de functie `cfa()` nu drie argumenten. Het derde argument is `group="lftd"` en wordt gebruikt om aan te geven dat een multi-group analyse moet worden uitgevoerd en dat `lftd` de groepsvariabele is. Door het geven van het tweede commando wordt het gespecificeerde model op de data van iedere groep afzonderlijk gepast en worden de resultaten weggeschreven onder de naam `fit0`.

Bekijk de 'output' en beantwoord de volgende vragen.

Vraag 1. Hoeveel mensen zitten er in beide groepen?

Vraag 2. Welke waarden vind je voor de toetsingsgrootheid, het aantal vrijheidsgraden en de p-waarde?

Vraag 3. Moet het model volgens de toets worden verworpen of niet? Waarom?

Vraag 4. Past het model goed op de data volgens de CFI? Waarom?

Vraag 5. Past het model goed op de data volgens de TLI? Waarom?

Vraag 6. Past het model goed op de data volgens de RMSEA? Waarom?

Vraag 7. Op welke manier zijn er schalen aan de latente variable toegekend? Waarom?

Omdat er een multi-groep analyse is uitgevoerd, vind je in de ‘output’ nu ook schattingen van de intercepten naast de geschatte factorladingen en schattingen van varianties. Bovendien wordt onder **Intercepts**: ook het gemiddelde van de gemeenschappelijke factor gegeven. De intercepten en het gemiddelde van de factor zijn in deze analyse nog niet van belang, maar straks als er andere vormen van meetinvariantie worden opgelegd wel.

Vraag 8. Bekijk de schattingen van de parameters nog eens goed. Is er een parameter waarvoor de schattingen in beide groepen hetzelfde zijn (vastgelegde waarden zijn geen schattingen en hebben daarom ook geen **Std.Err**)?

Het tweede multi-groep model dat op de data wordt gepast, is het multi-groep één-factormodel met zwakke factoriële invariantie over de twee groepen. Je kunt dit model op de data passen door het geven van de volgende drie commando's.

```
mod1<-"f1=~A1+A2+A3+A4"  
fit1<-cfa(mod1,data=Angst,group="lftd",group.equal=c("loadings"))  
summary(fit1,fit.measures=T)
```

In het tweede commando heeft de functie `cfa()` nu vier argumenten. Het vierde argument is `group.equal=c("loadings")` en wordt gebruikt om aan te geven dat de geschatte factorladingen voor de groepen hetzelfde moeten zijn. Pas het model op de data en controleer in de output of de geschatte factorladingen daadwerkelijk hetzelfde zijn voor de groepen.

Vraag 9. Wat is zwakke factoriële invariantie?

Vraag 10. Welke waarden vind je voor de toetsingsgrootte, het aantal vrijheidsgraden en de p-waarde?

Vraag 11. Moet zwakke factoriële invariantie volgens de toets worden verworpen of niet? Waarom?

Vraag 12. Past het zwakke factoriële invariantie-model goed op de data volgens de CFI? Waarom?

Vraag 13. Past het zwakke factoriële invariantie-model goed op de data volgens de TLI? Waarom?

Vraag 14. Past het zwakke factoriële invariantie-model goed op de data volgens de RMSEA? Waarom?

Het derde multi-groep model dat op de data wordt gepast, is het multi-groep één-factormodel met sterke factoriële invariantie over de twee groepen. Je kunt dit model op de data passen door het geven van de volgende drie commando's.

```
mod2<-"f1=~A1+A2+A3+A4"
fit2<-cfa(mod2,data=Angst,group="lftd",
           group.equal=c("loadings","intercepts"))
summary(fit2,fit.measures=T)
```

Het vierde argument van de functie `cfa()` is nu `group.equal=c("loadings","intercepts")` en wordt gebuikt om aan te geven dat zowel de geschatte factorladingen als de geschatte intercepten voor de groepen hetzelfde moeten zijn. Pas het model op de data en controleer in de output of de geschatte factorladingen en intercepten daadwerkelijk hetzelfde zijn voor de groepen.

Vraag 15. Wat is sterke factoriële invariantie?

Vraag 16. Welke waarden vind je voor de toetsingsgrootte, het aantal vrijheidsgraden en de p-waarde?

Vraag 17. Moet sterke factoriële invariantie volgens de toets worden verworpen of niet? Waarom?

Vraag 18. Past het sterke factoriële invariantie-model goed op de data volgens de CFI? Waarom?

Vraag 19. Past het sterke factoriële invariantie-model goed op de data volgens de TLI? Waarom?

Vraag 20. Past het sterke factoriële invariantie-model goed op de data volgens de RMSEA? Waarom?

Onder sterke factoriële invariantie is nu ook het verschil tussen de gemiddelden van de gemeenschappelijke angst-factor in de twee groepen geschat. Dit verschil vind je in de output van de tweede groep onder **Intercepts**: bij **f1**. Omdat het gemiddelde van de gemeenschappelijke factor in de eerste groep op nul is gezet, is het geschatte gemiddelde van de gemeenschappelijke factor in de tweede groep gelijk aan het geschatte gemiddelde verschil op de gemeenschappelijke factor tussen de twee groepen.

Vraag 21. Hoe groot is het geschatte verschil tussen de gemiddelden van de gemeenschappelijke angst-factor in de twee groepen?

Vraag 22. In welke groep is volgens het geschatte gemiddelde verschil het gemiddelde op de gemeenschappelijke angst-factor hoger? Waarom?

Vraag 23. Is het geschatte gemiddelde verschil statistisch significant? Waarom? Welke nulhypothese kan hiermee worden getoetst? Wat is de conclusie?

Dit verschil kan ook worden getoetst met een chi-kwadraatverschiltoets. Dat levert in dit geval hetzelfde resultaat op omdat er maar twee groepen zijn, maar als er meer dan twee groepen zijn dan kan de chi-kwadraatverschiltoets worden gebruikt om in één keer te toetsen of er bewijs is voor gemiddelde verschillen tussen de groepen op de gemeenschappelijke factor (een soort ANOVA, maar dan met een latente afhankelijke variabele). Voor het uitvoeren van de chi-kwadraatverschiltoets wordt eerst het sterke factoriële invariantie-model waarin beide gemiddelden op nul worden gezet op de data gepast. Dit doe je door het geven van de volgende commando's.

```
mod3<-"f1=~A1+A2+A3+A4
f1~c(0,0)*1"
fit3<-cfa(mod3,data=Angst,group="lftd",
          group.equal=c("loadings","intercepts"))
summary(fit3,fit.measures=T)
```

Met de extra regel `f1~c(0,0)*1` in het eerste commando wordt aangegeven dat de gemiddelden van de factor in beide groepen op nul moeten worden gezet (als er drie groepen zijn dan moet `c(0,0)` worden vervangen door `c(0,0,0)`, als er vier zijn door `c(0,0,0,0)`, enz.). Vervolgens kunnen we het verschil tussen de gemiddelden van de gemeenschappelijke angst-factor in de twee groepen toetsen door het geven van het volgende commando.

```
anova(fit3,fit2)
```

Vraag 24. Welke waarden vind je voor de toetsingsgrootte, het aantal vrijheidsgraden en de p-waarde?

Vraag 25. Welke nulhypothese kan er met de chi-kwadraatverschiltoets worden getoetst als er drie groepen waren geweest?

Vraag 26. Wat kan er worden gedaan om te onderzoeken welke groepen van elkaar verschillen als de chi-kwadraatverschiltoets een significant resultaat oplevert in het geval van drie of meer groepen?

Het laatste multi-groep model dat op de data wordt gepast, is het multi-groep één-factor model met strikte factoriële invariantie over de twee groepen. Je kunt dit model op de data passen door het geven van de volgende drie commando's.

```
mod4<-"f1=~A1+A2+A3+A4
A1~~c(u1,u1)*A1
A2~~c(u2,u2)*A2
A3~~c(u3,u3)*A3
A4~~c(u4,u4)*A4"
fit4<-cfa(mod4,data=Angst,group="lftd",
```

```
group.equal=c("loadings","intercepts"))
summary(fit4,fit.measures=T)
```

De extra regels in het eerste commando

```
A1~~c(u1,u1)*A1
A2~~c(u2,u2)*A2
A3~~c(u3,u3)*A3
A4~~c(u4,u4)*A4
```

zorgen ervoor dat ook de varianties van de unieke factoren over groepen aan elkaar gelijk worden gezet. Pas het model op de data en controleer in de output of de geschatte factorladingen, intercepten en unieke varianties daadwerkelijk hetzelfde zijn voor de groepen.

Vraag 27. Wat is strikte factoriële invariantie?

Vraag 28. Welke waarden vind je voor de toetsingsgrootte, het aantal vrijheidsgraden en de p-waarde?

Vraag 29. Moet strikte factoriële invariantie volgens de toets worden verworpen of niet? Waarom?

Vraag 30. Past het strikte factoriële invariantie-model goed op de data volgens de CFI? Waarom?

Vraag 31. Past het strikte factoriële invariantie-model goed op de data volgens de TLI? Waarom?

Vraag 32. Past het strikte factoriële invariantie-model goed op de data volgens de RMSEA? Waarom?

Vraag 33. Welke testscores kunnen er nu worden gebruikt om de waarde van iemand uit de eerste groep op de gemeenschappelijke factor met die van iemand uit de tweede groep te kunnen vergelijken? Waarom?

Het databestand `moodschl.txt` bevat metingen van stemmingen van 200 studenten net voor het volgen van college. De data zijn verkregen door studenten drie stemmingsevaluatie-items voor te leggen (bijvoorbeeld: ‘Op dit moment is mijn humeur slecht . . . goed’) en drie items over de mate van geactiveerdheid (bijvoorbeeld: ‘Op dit moment voel ik me passief . . . actief’). Voor ieder item konden de studenten op een negenpuntsschaal aangeven in hoeverre de stelling op dat moment op hun van toepassing was. Verder werd de studenten gevraagd hun sekse aan te geven.

Er wordt verondersteld dat hetzelfde standaard confirmatieve twee-factormodel geldt voor zowel mannen als vrouwen. Bovendien wordt ervan uitgegaan dat de items in beide groepen dezelfde twee latente variabelen meten: stemmingsevaluatie en geactiveerdheid. Pas

het multi-groep standaard confirmatieve twee-factormodel met configurale invariantie over de twee groepen op de data en beantwoord de volgende vragen.

Vraag 34. Hoeveel mensen zitten er in beide groepen (0 = vrouw, 1 = man)?

Vraag 35. Past het model met configurale invariantie op de data? Waarom?

Pas nu het confirmatieve twee-factormodel met zwakke factoriële invariantie over de twee groepen op de data.

Vraag 36. Past het model met zwakke factoriële invariantie op de data? Waarom?

Pas vervolgens het confirmatieve twee-factormodel met sterke factoriële invariantie over de twee groepen op de data.

Vraag 37. Past het model met sterke factoriële invariantie op de data? Waarom?

Onder sterke factoriële invariantie zijn nu ook verschillen in de gemiddelden van de twee gemeenschappelijke factoren tussen de twee groepen geschat.

Vraag 38. Hoe groot zijn de verschillen in de gemiddelden van de gemeenschappelijke factoren tussen de twee groepen?

Toets deze verschillen met een chi-kwadraatverschiltoets.

Vraag 39. Welke waarden vind je voor de toetsingsgrootte, het aantal vrijheidsgraden en de p-waarde?

Vraag 40. Wat is de nulhypothese en kunnen we die verwerpen?

Vraag 41. Interpreteer de resultaten.

Pas ten slotte het confirmatieve twee-factormodel met strikte factoriële invariantie over de twee groepen op de data.

Vraag 42. Past het model met strikte factoriële invariantie op de data? Waarom?

Factoranalyse van een MTMM-matrix

In dit deel van het tweede practicum voer je confirmatieve factoranalyses uit om de convergente en divergente validiteit van een aantal attitude metingen te onderzoeken. We maken in dit deel van het practicum gebruik van het databestand `Ostrom.txt`. Je kunt dit bestand vinden op Blackboard in de map Databestanden (onder Course Content - Practicum materialen). Kopieer dit bestand naar de op de U:-schijf aangemaakte map 'A&E Computerpractica'.

1. Invoeren van de data in R

Open een nieuw script-bestand en bewaar dit bestand meteen in je werkmap ‘A&E Computerpractica’ onder een toepasselijke naam.

Importeer het databestand `Ostrom.txt` in R door het geven van het volgende commando

```
ostrom<-read.table("Ostrom.txt",header=FALSE)
```

In bovenstaand commando staat nu `FALSE` omdat op de eerste regel in het databestand geen namen voor de variabelen staan, maar meteen data. Als je het databestand bekijkt, dan zie je dat je nu een correlatiematrix in R hebt geïmporteerd. In deze correlatiematrix staan de steekproefcorrelaties tussen twaalf variabelen.

Deze steekproefcorrelaties zijn berekend met de scores die in 1969 in de VS door Ostrom zijn verzameld door aan 189 respondenten stellingen voor te leggen over het kerkelijk instituut. De stellingen verschilden van elkaar met betrekking tot drie componenten: een affectieve component (A), een gedragscomponent (B) en een cognitieve component (C). Verder werden voor elk van de drie componenten meerdere stellingen voorgelegd die met behulp van vier verschillende methoden werden gescoord: Thurstone methode (T), Likert methode (L), Guttman method (G) en zelfbeoordelingen (S). Uiteindelijk konden dus twaalf subtestscores worden geconstrueerd. De correlaties tezamen vormen een multitrekk-multimethode-matrix.

Voordat de juiste namen aan de variabelen worden gegeven, moet de ingelezen data-frame eerst worden omgezet naar een matrix. Dit doe je door het geven van het volgende commando.

```
ostrom<-as.matrix(ostrom)
```

Aan de rijen en kolommen moeten dezelfde namen worden gegeven. Voor de huidige correlatiematrix moet je dit nu doen door de volgende twee commando's te geven.

```
rownames(ostrom)<-c("AT","BT","CT","AL","BL","CL","AG","BG","CG","AS","BS","CS")  
colnames(ostrom)<-c("AT","BT","CT","AL","BL","CL","AG","BG","CG","AS","BS","CS")
```

Check of alles goed is gegaan door het geven van het volgende commando.

```
ostrom
```

2. Analyses

De volgende analyses worden uitgevoerd om een indruk te krijgen van de convergente en divergente validiteit van alle metingen in het onderzoek van Ostrom. Allereerst wordt het algemene alternatieve hypothesemodel op de data gepast. Dit doe je door de volgende drie commando's uit te voeren.

```

modA<-"A=~AT+AL+AG+AS
B=~BT+BL+BG+BS
C=~CT+CL+CG+CS
T=~AT+BT+CT
L=~AL+BL+CL
G=~AG+BG+CG
S=~AS+BS+CS
A~~0*T # onafh. tussen A en T
A~~0*L # onafh. tussen A en L
A~~0*G # onafh. tussen A en G
A~~0*S # onafh. tussen A en S
B~~0*T # onafh. tussen B en T
B~~0*L # onafh. tussen B en L
B~~0*G # onafh. tussen B en G
B~~0*S # onafh. tussen B en S
C~~0*T # onafh. tussen C en T
C~~0*L # onafh. tussen C en L
C~~0*G # onafh. tussen C en G
C~~0*S" # onafh. tussen C en S
fitA<-cfa(modA,sample.cov=ostrom,sample.nobs=189)
summary(fitA,fit.measures=TRUE)

```

Het eerste commando (alles tussen " en ") is de modelspecificatie. A, B en C zijn de trekken (de te meten latente variabelen). De trekken zijn de drie componenten: affectie, gedrag en cognitie. T, L, G en S zijn de vier methoden: Thurstone, Likert, Guttman en zelfbeoordeling. In het tweede commando moet nu worden opgegeven dat een covariantiematrix (in dit geval een correlatiematrix) moet worden geanalyseerd. Dit wordt gedaan door het argument `sample.cov=ostrom`. Daarnaast moet in dit commando ook de steekproefgrootte worden opgegeven. Dit wordt gedaan door het argument `sample.nobs=189`. Het geven van de drie commando's kan een waarschuwingsbericht geven, welke kan worden genegeerd.

Als tweede wordt het model zonder trekken (de componenten) op de data gepast. Dit doe je met de volgende commando's.

```

mod1<-"T=~AT+BT+CT
L=~AL+BL+CL
G=~AG+BG+CG
S=~AS+BS+CS"
fit1<-cfa(mod1,sample.cov=ostrom,sample.nobs=189)
summary(fit1,fit.measures=TRUE)

```

Ook nu kun je het waarschuwingsbericht negeren. Met het volgende commando krijg je de resultaten voor het toetsen van de nulhypothese dat er geen convergente validiteit is.

```
anova(fit1,fitA)
```

Vraag 43. Is er bewijs voor convergente validiteit? Waarom?

Om divergente validiteit te kunnen toetsen moet het volgende model met perfect gecorreleerde trekken op de data worden gepast.

```
mod2<-"A=~NA*AT+AL+AG+AS
B=~NA*BT+BL+BG+BS
C=~NA*CT+CL+CG+CS
T=~AT+BT+CT
L=~AL+BL+CL
G=~AG+BG+CG
S=~AS+BS+CS
A~~1*A # de var van A op 1
B~~1*B # de var van B op 1
C~~1*C # de var van C op 1
A~~0*T
A~~0*L
A~~0*G
A~~0*S
B~~0*T
B~~0*L
B~~0*G
B~~0*S
C~~0*T
C~~0*L
C~~0*G
C~~0*S
A~~1*B # perfecte afh. tussen A en B
A~~1*C # perfecte afh. tussen A en C
B~~1*C" # perfecte afh. tussen B en C
fit2<-cfa(mod2,sample.cov=ostrom,sample.nobs=189)
summary(fit2,fit.measures=TRUE)
```

Het resultaat voor het toetsen van de nulhypothese dat er geen divergente validiteit is, kan nu worden gekregen met het volgende commando.

```
anova(fit2,fitA)
```

Vraag 44. Is er bewijs voor divergente validiteit? Waarom?

Tenslotte willen we weten of er methode-effecten zijn. Om dit te onderzoeken moet het volgende model zonder methode-factoren op de data worden gepast.

```
mod3<-"A=~AT+AL+AG+AS
B=~BT+BL+BG+BS
C=~CT+CL+CG+CS"
```

```
fit3<-cfa(mod3,sample.cov=ostrom,sample.nobs=189)
summary(fit3,fit.measures=TRUE)
```

Het resultaat voor het toetsen van de nulhypothese dat er geen methode-effecten zijn, kan worden gekregen met het volgende commando.

```
anova(fit3,fitA)
```

Vraag 45. Is er bewijs voor methode-effecten? Waarom?

Afsluiting

Bewaar je veranderde script-bestanden op de U:-schijf.

Neem bij vragen of onduidelijkheden contact op met de practicumbegeleiders.

Laat bij vertrek je werkplek opgeruimd achter.

Practicum 2: Assessment en Evaluatie 2018-2019

Naam :

Studentnummer :

Emailadres :

Naam :

Studentnummer :

Emailadres :

Meetinvariantie

2. Multi-groep confirmatieve factoranalyses

Vraag 1. Hoeveel mensen zitten er in beide groepen?

Antwoord:

Vraag 2. Welke waarden vind je voor de toetsingsgrootte, het aantal vrijheidsgraden en de p-waarde?

Antwoord:

Vraag 3. Moet het model volgens de toets worden verworpen of niet? Waarom?

Antwoord:

Vraag 4. Past het model goed op de data volgens de CFI? Waarom?

Antwoord:

Vraag 5. Past het model goed op de data volgens de TLI? Waarom?

Antwoord:

Vraag 6. Past het model goed op de data volgens de RMSEA? Waarom?

Antwoord:

Vraag 7. Op welke manier zijn er schalen aan de latente variabele toegekend? Waarom?

Antwoord:

Vraag 8. Bekijk de schattingen van de parameters nog eens goed. Is er een parameter waarvoor de schattingen in beide groepen hetzelfde zijn (vastgelegde waarden zijn geen schattingen en hebben daarom ook geen Std.Err)?

Antwoord:

Vraag 9. Wat is zwakke factoriële invariantie?

Antwoord:

Vraag 10. Welke waarden vind je voor de toetsingsgrootheid, het aantal vrijheidsgraden en de p-waarde?

Antwoord:

Vraag 11. Moet zwakke factoriële invariantie volgens de toets worden verworpen of niet? Waarom?

Antwoord:

Vraag 12. Past het zwakke factoriële invariantie-model goed op de data volgens de CFI? Waarom?

Antwoord:

Vraag 13. Past het zwakke factoriële invariantie-model goed op de data volgens de TLI? Waarom?

Antwoord:

Vraag 14. Past het zwakke factoriële invariantie-model goed op de data volgens de RMSEA? Waarom?

Antwoord:

Vraag 15. Wat is sterke factoriële invariantie?

Antwoord:

Vraag 16. Welke waarden vind je voor de toetsingsgrootheid, het aantal vrijheidsgraden en de p-waarde?

Antwoord:

Vraag 17. Moet sterke factoriële invariantie volgens de toets worden verworpen of niet? Waarom?

Antwoord:

Vraag 18. Past het sterke factoriële invariantie-model goed op de data volgens de CFI? Waarom?

Antwoord:

Vraag 19. Past het sterke factoriële invariantie-model goed op de data volgens de TLI? Waarom?

Antwoord:

Vraag 20. Past het sterke factoriële invariantie-model goed op de data volgens de RMSEA? Waarom?

Antwoord:

Vraag 21. Hoe groot is het geschatte verschil tussen de gemiddelden van de gemeenschappelijke angst-factor in de twee groepen?

Antwoord:

Vraag 22. In welke groep is volgens het geschatte gemiddelde verschil het gemiddelde op de gemeenschappelijke angst-factor hoger? Waarom?

Antwoord:

Vraag 23. Is het geschatte gemiddelde verschil statistisch significant? Waarom? Welke nulhypothese kan hiermee worden getoetst? Wat is de conclusie?

Antwoord:

Vraag 24. Welke waarden vind je voor de toetsingsgrootte, het aantal vrijheidsgraden en de p-waarde?

Antwoord:

Vraag 25. Welke nulhypothese kan er met de chi-kwadraatverschiltoets worden getoetst als er drie groepen waren geweest?

Antwoord:

Vraag 26. Wat kan er worden gedaan om te onderzoeken welke groepen van elkaar verschillen als de chi-kwadraatverschiltoets een significant resultaat oplevert in het geval van drie of meer groepen?

Antwoord:

Vraag 27. Wat is strikte factoriële invariantie?

Antwoord:

Vraag 28. Welke waarden vind je voor de toetsingsgrootte, het aantal vrijheidsgraden en de p-waarde?

Antwoord:

Vraag 29. Moet strikte factoriële invariantie volgens de toets worden verworpen of niet? Waarom?

Antwoord:

Vraag 30. Past het strikte factoriële invariantie-model goed op de data volgens de CFI? Waarom?

Antwoord:

Vraag 31. Past het strikte factoriële invariantie-model goed op de data volgens de TLI? Waarom?

Antwoord:

Vraag 32. Past het strikte factoriële invariantie-model goed op de data volgens de RMSEA? Waarom?

Antwoord:

Vraag 33. Welke testcores kunnen er nu worden gebruikt om de waarde van iemand uit de eerste groep op de gemeenschappelijke factor met die van iemand uit de tweede groep te kunnen vergelijken? Waarom?

Antwoord:

Vraag 34. Hoeveel mensen zitten er in beide groepen?

Antwoord:

Vraag 35. Past het model met configurale invariantie op de data? Waarom?

Antwoord:

Vraag 36. Past het model met zwakke factoriële invariantie op de data? Waarom?

Antwoord:

Vraag 37. Past het model met sterke factoriële invariantie op de data? Waarom?

Antwoord:

Vraag 38. Hoe groot zijn de verschillen in de gemiddelden van de gemeenschappelijke factoren tussen de twee groepen?

Antwoord:

Vraag 39. Welke waarden vind je voor de toetsingsgrootte, het aantal vrijheidsgraden en de p-waarde?

Antwoord:

Vraag 40. Wat is de nulhypothese en kunnen we die verwerpen?

Antwoord:

Vraag 41. Interpreteer de resultaten.

Antwoord:

Vraag 42. Past het model met strikte factoriële invariantie op de data? Waarom?

Antwoord:

Factoranalyse van een MTMM-matrix

2. Analyses

Vraag 43. Is er bewijs voor convergente validiteit? Waarom?

Antwoord:

Vraag 44. Is er bewijs voor divergente validiteit? Waarom?

Antwoord:

Vraag 45. Is er bewijs voor methode-effecten? Waarom?

Antwoord:

Betrouwbaarheid van testcores

In dit practicum voer je een betrouwbaarheidsanalyse uit. Met een betrouwbaarheidsanalyse krijg je meestal een schatting van de communaliteit van de gebruikte testscore. De schatting van de communaliteit van de testscore geeft dan een indruk van de betrouwbaarheid van de testscore. In de betrouwbaarheidsanalyse kun je besluiten één of meerdere items uit een verzameling items te verwijderen om zo een testscore te krijgen waarvoor de betrouwbaarheid waarschijnlijk hoger is. Als de schatting voor de gebruikte coëfficiënt voor een bepaalde verzameling items optimaal is, bereken dan voor deze verzameling items de testscore (de ongewogen somscore). De verdeling van deze testscore bekijk je door enkele beschrijvende statistieken te berekenen en een bijpassende grafiek te maken. Tot slot zul je met de testscore en informatie over de betrouwbaarheid van de testscore een intervallschatting van een individuele ware score bepalen.

Lever voor de deadline (uiterlijk 16 maart) samen met degene met wie je het practicum doet, één elektronische versie van de uitwerking van het practicum in via SafeAssignment op Blackboard. De practicumbegeleiders kijken de uitwerking van het practicum na en geven feedback.

- Gebruik de Word-file **A&E Practicum 3.docx** om je antwoorden in te noteren. Deze Word-file kun je vinden op Blackboard in de map Practicum 2 onder Course Content - Practicum materialen.
- Noteer bij het rapporteren van getallen maximaal twee decimalen.
- Geef, waar nodig, een korte toelichting bij de antwoorden.
- Antwoord niet uitsluitend met een getal. Noteer bijvoorbeeld: De gemiddelde leeftijd is 13.34 jaar.
- Gebruik bij een statistische toets een nominaal significantieniveau van 5%.

We maken in dit practicum weer gebruik van het gegevensbestand **SCHOOLPLEZIER.dat**. Als het goed is, heb je dit bestand samen met **A&E Codeboek Schoolplezier.pdf** in het eerste practicum gekopieerd naar de op de U:-schijf aangemaakte map 'A&E Computerpractica'. Zo niet, dan kun je de bestanden vinden op Blackboard in de map Databestanden (onder Course Content - Practicum materialen).

Ter herinnering, in het bestand **SCHOOLPLEZIER.dat** zijn de gegevens van een onderzoek naar het schoolplezier van leerlingen op de basisschool ingevoerd. Het bestand bevat onder meer de scores van de respondenten op **item1** tot en met **item8**. Elk van de acht items bevat een stelling en de inhoud van de stellingen is te vinden in het bestand **A&E Codeboek Schoolplezier.pdf**. De acht items worden gebruikt voor de operationalisatie van het theoretische begrip 'schoolplezier' en zijn voorgelegd aan een aselechte steekproef leerlingen uit de populatie leerlingen uit groepen 3 - 8 van het reguliere basisonderwijs. Aan de leerlingen is gevraagd voor iedere stelling op een vierpuntsschaal aan te geven in welke mate de stelling op hun van toepassing is. Hierdoor zijn ordinale gegevens verkregen, maar in dit practicum gaan we er omwille van de oefening vanuit dat de afstanden tussen opeenvolgende schaalpunten gelijk zijn en dus dat de itemscores interval meetniveau hebben.

Uit de acht items moet een zo betrouwbaar mogelijke testscore voor het meten van schoolplezier worden geconstrueerd. Voor deze testscore moet gelden dat een hoge waarde correspondeert met veel schoolplezier en een lage waarde met weinig schoolplezier. Het bepalen van de items die opgenomen moeten worden in deze testscore, doen we met de resultaten van de betrouwbaarheidsanalyse.

1. Invoeren van de data in R

Open R. Voordat je R weer gaat gebruiken voor statistische analyses, is het weer handig om aan R de werkmap ‘A&E Computerpractica’ op te geven. Klik op File in de menubalk en selecteer Change dir.... Browse vervolgens naar de map ‘A&E Computerpractica’ op de U:-schijf en klik op OK. Verder is het verstandig om in een nieuw script-bestand te werken. Om een nieuw script-bestand te openen, klik op File in de menubalk en selecteer New script. Bewaar dit nieuwe script-bestand meteen in je werkmap ‘A&E Computerpractica’ door op File in de menubalk te klikken en Save as... te selecteren. Bewaar het script-bestand onder een toepasselijke naam.

Om de gegevens in het databestand SCHOOLPLEZIER.dat met behulp van R te kunnen analyseren, moet het bestand weer in R worden ingelezen. Kopieer en plak het volgende commando in je script-bestand

```
D<-read.table("SCHOOLPLEZIER.dat",header=TRUE)
```

en voer het commando uit.

Zoals uit het eerste practicum bleek, zijn niet alle itemscores op dezelfde manier gecodeerd. Omdat er testcores geconstrueerd gaan worden, zullen we eerst de gegevens van de eerste zes items hercoderen. Hiervoor moeten we het pakket `recoder` laden. Voer dus het commando `library(recoder)` uit. Als je nu achter een nieuwe computer zit, kan het zijn dat het pakket `recoder` niet is geïnstalleerd. Zorg er dan dus voor dat je voor het laden van `recoder`, eerst weer het commando `install.packages("recoder")` hebt gegeven. Voer nu voor het hercoderen van de eerste zes items de volgende commando's uit.

```
D$item1<-recoder(D$item1,"1:4;2:3;3:2;4:1")
D$item2<-recoder(D$item2,"1:4;2:3;3:2;4:1")
D$item3<-recoder(D$item3,"1:4;2:3;3:2;4:1")
D$item4<-recoder(D$item4,"1:4;2:3;3:2;4:1")
D$item5<-recoder(D$item5,"1:4;2:3;3:2;4:1")
D$item6<-recoder(D$item6,"1:4;2:3;3:2;4:1")
```

2. Betrouwbaarheidsanalyse

Om een indicatie van de betrouwbaarheid van de ongewogen somscore van de acht items voor het meten van schoolplezier te krijgen, kunnen de steekproefschattingen van verschillende coëfficiënten worden gebruikt, zoals Cronbachs alfa, Guttman's labda-2 en coëfficiënt

omega. R functies waarmee schattingen van deze coëfficiënten kunnen worden berekend zitten niet standaard in R. R pakketten waar wel zulke functies in zitten, zijn `CTT` en `psych`. Installeer het pakket `CTT` door het geven van het commando

```
install.packages("CTT")
```

Laad het pakket `CTT` vervolgens door het geven van het commando

```
library(CTT)
```

De functie in het pakket `CTT` waarmee een schatting van Cronbachs alfa voor de ongewogen somscore van de itemscores kan worden berekend is `itemAnalysis()`. Geef het commando

```
itemAnalysis(D[,5:12])
```

Vraag 1. Van hoeveel personen zijn de itemscores gebruikt bij het berekenen van de schatting van Cronbachs alfa voor de ongewogen somscore?

Vraag 2. Wat is de geschatte waarde van Cronbachs alfa voor de ongewogen somscore?

Om te bepalen of er items zijn die in aanmerking komen voor verwijdering uit de testscore voor Schoolplezier, maak je gebruik van een combinatie van de volgende vier criteria:

1. Het item discrimineert niet/heeft een te kleine standaardafwijking.
2. De item-restcorrelatie van het item is lager dan .3.
3. Verwijderen van het item heeft een stijging van Cronbachs alfa tot gevolg.
4. Het item past inhoudelijk niet goed bij de overige items.

Gegevens die hiervoor nodig zijn, kunnen worden verkregen door het geven van het commando

```
itemAnalysis(D[,5:12])$itemReport
```

In de nu verschenen tabel staan in de derde kolom onder het kopje `pBis` de geschatte item-restcorrelaties en in de vijfde kolom onder het kopje `alphaIfDeleted` de geschatte Cronbachs alfa's voor de restscores.

Vraag 3. Voor welk item geldt dat de item-restcorrelatie te laag is?

Vraag 4. Is Cronbachs alfa voor de restscore van dit item groter dan Cronbachs alfa voor de totale testscore?

Vraag 5. Is er een inhoudelijk argument om item 7 toch in de testscore op te nemen?

Vraag 6. Is er een statistisch argument om item 7 toch in de testscore op te nemen?

Vraag 7. Hoe kwalificeer je de schatting van Cronbachs alfa voor de testscore volgens de criteria van de COTAN (zie college slides)?

Vraag 8. Welke toepassing van de test (tests voor belangrijke beslissingen op individueel niveau, tests voor minder belangrijke beslissingen op individueel niveau of tests voor onderzoek op groepsniveau) is dus verantwoord?

Met behulp van het R pakket `psych` kunnen naast Cronbachs alfa ook andere coëfficiënten worden geschat. De functies die je hiervoor kunt gebruiken zijn `alpha()`, `guttman()` en `splitHalf()`. Installeer en laad het pakket `psych`.

Vraag 9. Wat is de geschatte waarde van Guttman's λ^2 voor de ongewogen somscore?

Voor het berekenen van een schatting van coëfficiënt omega moet eerst een confirmatief of een exploratief factormodel op de data worden gepast. Voor het passen van een exploratief factormodel kan de functie `fa()` van het R pakket `psych` worden gebruikt. Omdat we ervan uitgaan dat de acht items maar één gemeenschappelijke factor meten, passen we een één-factormodel op de itemscores en schrijven we de resultaten weg onder de naam `out` door het geven van het commando

```
out<-fa(D[,5:12],nfactors=1,covar=TRUE,fm="ml")
```

In bovenstaand commando wordt met het argument `covar=TRUE` aangegeven dat de covariantiematrix van de itemscores moet worden gebruikt in plaats van de correlatiematrix. Anders zal een schatting van coëfficiënt omega voor de ongewogen som van de gestandaardiseerde itemscores worden gegeven en niet voor de ongewogen som van de ongestandaardiseerde itemscores. Met het argument `fm="ml"` wordt aangegeven dat de parameters met de 'maximum likelihood' methode moeten worden geschat. Als je de resultaten van de factoranalyse wilt bekijken dan typ je `out` in de R Console en druk je op de Return-toets. In de kolom `ML1` onder `Unstandardized loadings` staan de geschatte factorladingen. In de kolom `u2` onder `Unstandardized loadings` staan de geschatte varianties van de item-unieke factoren. Deze schattingen zijn nodig voor het berekenen van coëfficiënt omega onder het één-factormodel. Kopieer en plak de volgende commando's direct onder het bovenstaande commando voor de factoranalyse in je script-bestand.

```
L<-out$loadings[,]
t<-diag(out$residual)
e<-rep(1,length(t))
Sigma<-L%*%t(L)+diag(t)
omega<-t(e)%*%L%*%t(L)%*%e/t(e)%*%Sigma%*%e
omega
```


Selecteer alle hierboven staande commando's en druk op de sneltoets Ctrl + r om ze uit te voeren. Het getal dat verschijnt is de schatting van coëfficiënt omega voor de ongewogen somscore $X_1 + X_2 + X_3 + X_4 + X_5 + X_6 + X_7 + X_8$ onder het één-factormodel.

Vraag 10. Wat is de geschatte waarde van coëfficiënt omega voor de ongewogen somscore onder het één-factormodel?

Stel dat we zouden denken dat items 4 en 5 het minst belangrijk zijn, dat items 2 en 6 belangrijker zijn en dat items 1, 3, 7 en 8 het belangrijkste zijn voor het meten van schoolplezier. In dat geval zouden we verschillende gewichten willen toekennen aan verschillende items door in plaats van de ongewogen somscore een gewogen somscore als testscore te gebruiken. We zouden dan bijvoorbeeld kunnen kiezen voor de gewogen somscore $3X_1 + 2X_2 + 3X_3 + X_4 + X_5 + 2X_6 + 3X_7 + 3X_8$. Een schatting van coëfficiënt omega voor deze gewogen somscore onder het één-factormodel kan nu worden gekregen door het geven van de volgende commando's

```
w<-c(3,2,3,1,1,2,3,3)
t(w)%*%L%*%t(L)%*%w/t(w)%*%Sigma%*%w
```

Vraag 11. Wat is de geschatte waarde van coëfficiënt omega voor deze gewogen somscore onder het één-factormodel?

Vraag 12. Als je de antwoorden van vragen 10 en 11 met elkaar vergelijkt, kun je dan concluderen dat je moet spreken van de communaliteit/betrouwbaarheid van de test of juist van de communaliteit/betrouwbaarheid van een testscore? Waarom?

3. Testscore Schoolplezier

Uit de betrouwbaarheidsanalyse zouden we kunnen concluderen dat alle items mogen worden opgenomen in de testscore voor Schoolplezier.

Omdat er geen vooropgestelde ideeën zijn over hoe belangrijk de verschillende items zijn voor het meten van schoolplezier, wordt de ongewogen somscore van de itemscores als testscore gehanteerd. Nu willen we voor elke case (kind) in de steekproef deze testscore berekenen. De ongewogen somscores van alle cases kunnen worden berekend en bewaard onder de naam X door het geven van het commando

```
X<-apply(D[,5:12],1,sum)
```

Als je testscore X in data-frame D wilt opnemen dan geef je het commando `D<-cbind(D,X)`.

Vraag 13. Wat is theoretisch de laagst mogelijke waarde op X?

Vraag 14. Wat is theoretisch de hoogst mogelijke waarde op X?

Bekijk de verdeling van testscore X via een histogram met normaalverdeling. Deze kunnen worden opgevraagd door het geven van de commando's

```
hist(X, freq=FALSE)
x<-seq(min(X, na.rm=TRUE), max(X, na.rm=TRUE), 0.01)
curve(dnorm(x, mean(X, na.rm=TRUE), sd(X, na.rm=TRUE)), add=TRUE)
```

Gebruik de functies `mean()` en `sd()` om het steekproefgemiddelde en de steekproefstandaardafwijking van testscore X te berekenen.

Vraag 15. Wat is het steekproefgemiddelde van testscore X ?

Vraag 16. Wat is de steekproefstandaardafwijking van testscore X ?

Vraag 17. Wat betekent een standaardafwijking?

Vraag 18. Hebben de leerlingen, afgaande op de verdeling van de testscore, plezier in school?

Vraag 19. Wat zou de betekenis van een hoge testscore zijn geweest als we niet items 1 t/m 6 hadden gehercodeerd maar items 7 en 8?

4. Standaardmeetfout

De standaardafwijking van de toevallige meetfout in de populatie wordt ook wel de standaardmeetfout genoemd. Uit de klassieke testtheorie volgt dat de standaardmeetfout gelijk is aan $\sigma_E = \sigma_X \sqrt{1 - \rho_{XT}^2}$. Een steekproefschatting van de standaardmeetfout wordt verkregen door

$$s_E = s_X \sqrt{1 - \hat{\rho}_{XT}^2},$$

waarin s_X de standaardafwijking van testscore X in de steekproef is en $\hat{\rho}_{XT}^2$ staat voor de geschatte betrouwbaarheid/communaliteit van testscore X , zoals bijvoorbeeld de steekproefschatting van Cronbachs alfa, Guttman's lambda-2 of coëfficiënt omega. Gebruikmakend van de geschatte waarde van coëfficiënt omega voor testscore X , kan de standaardmeetfout van testscore X makkelijk in R worden berekend door het geven van het commando

```
se<-sd(X, na.rm=T)*sqrt(1-omega)
```

Door de geschatte standaardmeetfout te noteren in een rapport krijgt de lezer een indicatie van de nauwkeurigheid van de meting.

Vraag 20. Hoe hoog is de geschatte standaardmeetfout van testscore X ?

Met de geschatte standaardmeetfout en de geschatte betrouwbaarheid/communaliteit van testscore X kun je voor een individu een 95%-betrouwbaarheidsintervalschatting bereke-

nen. Het geschatte 95%-betrouwbaarheidsinterval kan worden verkregen door

$$x - 1.96s_E < t < x + 1.96s_E,$$

waarin x staat voor de testscore van het individu, 1.96 de standaardnormale score is die hoort bij 95% onder de standaardnormale verdeling en t de onbekende ware score is van het individu. Voor bijvoorbeeld een kind met testscore 15 op schoolplezier kan in R gemakkelijk een schatting van het 95%-betrouwbaarheidsinterval worden berekend door het geven van de commando

```
c(15-qnorm(.975)*se,15+qnorm(.975)*se)
```

Vraag 21. Welke schatting van het 95%-betrouwbaarheidsinterval wordt er berekend voor een kind met testscore 15?

Stel dat het mogelijk is om bij een kind honderd keer de test af te nemen en dat het kind telkens niet weet wat het de vorige keer heeft geantwoord (onafhankelijke waarnemingen). Vervolgens bepaal je bij al die honderd testcores de 95%-betrouwbaarheidsintervalschatting rondom de geobserveerde testscore.

Vraag 22. Hoeveel (ongeveer) van die intervallen zal dan de ware score van dit kind bevatten?

5. De betrouwbaarheid van subtestcores

Als je naar de inhoud van de acht items kijkt, dan is het niet onredelijk te vermoeden dat de acht items twee latente variabelen meten. Het lijkt erop dat items 1, 2, 4 en 7 vooral de tevredenheid met de school ten opzichte van een voorstelbare andere school meten en dat items 3, 5, 6 en 8 vooral de tevredenheid met de dingen die er op school gebeuren, meten. Deze vermoedens of veronderstellingen kunnen worden onderzocht door het passen van een standaard confirmatief factormodel op de data.

Laad `lavaan` door het geven van het commando `library(lavaan)`. Knip en plak vervolgens de volgende code in je script-bestand

```
model1<-"f1=~item3+item5+item6+item8
        f2=~item1+item2+item4+item7"
fit1<-cfa(model1,data=D)
```

en voer ze uit. Om te onderzoeken hoe betrouwbaar verschillende (sub)testcores zijn, zullen nu schattingen van de communaliteiten van die (sub)testcores worden berekend. De communaliteit van een testscore is de proportie variantie van die testscore die door de gemeenschappelijke factoren wordt verklaard. De communaliteit van een testscore geeft dus aan hoe goed die testscore de gemeenschappelijke factoren meet.

De communaliteiten van de itemscores kunnen worden berekend door het geven van de volgende commando's

```
Theta<-inspect(fit1,"est")$theta
Sigma<-fitted(fit1)$cov
diag(Sigma-Theta)/diag(Sigma)
```

Met het eerste van de hier bovenstaande commando's wordt uit de passingsresultaten in `fit1` de matrix met de geschatte varianties van de item-unieke factoren gehaald en opgeslagen onder de naam `Theta`. Met het tweede commando wordt uit de passingsresultaten de geschatte geïmpliceerde covariantiematrix van de itemscores gehaald en opgeslagen onder de naam `Sigma`. Het laatste commando gebruikt deze twee matrices om schattingen van de communaliteiten van de itemscores te berekenen.

Vraag 23. Voor welk item is de schatting van de communaliteit het hoogst?

Vraag 24. Voor welk item is de schatting van de communaliteit het laagst?

Voor het berekenen van de communaliteiten van de twee subtestscores $S_1 = X_3 + X_5 + X_6 + X_8$ en $S_2 = X_1 + X_2 + X_4 + X_7$ moeten eerst de overeenkomstige vectoren met gewichten worden aangemaakt (een vector is een matrix met maar één rij of één kolom). Dat doe je door het geven van de volgende commando's

```
w1<-c(1,1,1,1,0,0,0,0)
w2<-c(0,0,0,0,1,1,1,1)
```

De hier bovenstaande vectoren met gewichten moeten worden gebruikt omdat X_3 , X_5 , X_6 en X_8 de eerste vier itemscores zijn die aan `lavaan` zijn opgegeven en X_1 , X_2 , X_4 en X_7 de laatste vier. Vervolgens kan voor beide subtestscores coëfficiënt omega worden berekend door het geven van de commando's

```
t(w1)%*(Sigma-Theta)%*w1/t(w1)%*Sigma%*w1
t(w2)%*(Sigma-Theta)%*w2/t(w2)%*Sigma%*w2
```

Vraag 25. Hoe hoog is de proportie variantie van subtestscore S_1 die wordt verklaard door de gemeenschappelijke factoren?

Vraag 26. Hoe hoog is de proportie variantie van subtestscore S_2 die wordt verklaard door de gemeenschappelijke factoren?

Hetzelfde kunnen we doen voor de ongewogen som van alle itemscores. De vector met gewichten voor deze testscore wordt verkregen door het geven van het commando

```
w<-rep(1,ncol(inspect(fit1,"data")))
```

Een schatting van coëfficiënt omega voor de ongewogen somscore onder het standaard confirmatieve twee-factormodel kun je dan berekenen door het geven van het commando

$$t(w) \frac{(\Sigma - \Theta)}{w} / t(w) \Sigma w$$

Vraag 27. Hoe hoog is de proportie variantie van de totale ongewogen som van de itemscores die wordt verklaard door de gemeenschappelijke factoren?

Vraag 28. Wat valt je op als je naar de geschatte communaliteiten van de itemscores, de subtestscores en de ongewogen somscore kijkt?

Ook voor Thurstone factorscores kunnen de communaliteiten worden berekend. Dat doe je door het geven van de volgende commando's.

```
L<-inspect(fit1,"est")$lambda
P<-inspect(fit1,"est")$psi
O<-Sigma-Theta
TS1<-diag(P*t(L)*solve(Sigma)*0*solve(Sigma)*L*P)
TS2<-diag(P*t(L)*solve(Sigma)*L*P)
TS1/TS2
```

Vraag 29. Als je de berekende communaliteiten van de subtestscores S_1 en S_2 met die van de Thurstone factorscores vergelijkt, welke testcores lijken dan betrouwbaarder te zijn?

Afsluiting

Bewaar je veranderde script-bestand op de U:-schijf.

Neem bij vragen of onduidelijkheden contact op met de practicumbegeleiders.

Laat bij vertrek je werkplek opgeruimd achter.

Practicum 3: Assessment en Evaluatie 2018-2019

Naam :

Studentnummer :

Emailadres :

Naam :

Studentnummer :

Emailadres :

Betrouwbaarheid van testcores

2. Betrouwbaarheidsanalyse

Vraag 1. Van hoeveel personen zijn de itemscores gebruikt bij het berekenen van de schatting van Cronbachs alfa voor de ongegewogen somscore?

Antwoord:

Vraag 2. Wat is de geschatte waarde van Cronbachs alfa voor de ongewogen somscore?

Antwoord:

Vraag 3. Voor welk item geldt dat de item-restcorrelatie te laag is?

Antwoord:

Vraag 4. Is Cronbachs alfa voor de restscore van dit item groter dan Cronbachs alfa voor de totale testscore?

Antwoord:

Vraag 5. Is er een inhoudelijk argument om item 7 toch in de testscore op te nemen?

Antwoord:

Vraag 6. Is er een statistisch argument om item 7 toch in de testscore op te nemen?

Antwoord:

Vraag 7. Hoe kwalificeer je de schatting van Cronbachs alfa voor de testscore volgens de criteria van de COTAN (zie college slides)?

Antwoord:

Vraag 8. Welke toepassing van de test (tests voor belangrijke beslissingen op individueel niveau, tests voor minder belangrijke beslissingen op individueel niveau of tests voor onderzoek op groepsniveau) is dus verantwoord?

Antwoord:

Vraag 9. Wat is de geschatte waarde van Guttman's λ^2 voor de ongewogen somscore?

Antwoord:

Vraag 10. Wat is de geschatte waarde van coëfficiënt omega voor de ongewogen somscore onder het één-factormodel?

Antwoord:

Vraag 11. Wat is de geschatte waarde van coëfficiënt omega voor deze gewogen somscore onder het één-factormodel?

Antwoord:

Vraag 12. Als je de antwoorden van vragen 37 en 38 met elkaar vergelijkt, kun je dan concluderen dat je moet spreken van de communaliteit/betrouwbaarheid van de test of juist van de communaliteit/betrouwbaarheid van een testscore? Waarom?

Antwoord:

3. Testscore Schoolplezier

Vraag 13. Wat is theoretisch de laagst mogelijke waarde op X?

Antwoord:

Vraag 14. Wat is theoretisch de hoogst mogelijke waarde op X?

Antwoord:

Vraag 15. Wat is het steekproefgemiddelde van testscore X?

Antwoord:

Vraag 16. Wat is de steekproefstandaardafwijking van testscore X?

Antwoord:

Vraag 17. Wat betekent een standaardafwijking?

Antwoord:

Vraag 18. Hebben de leerlingen, afgaande op de verdeling van de testscore, plezier in school?

Antwoord:

Vraag 19. Wat zou de betekenis van een hoge testscore zijn geweest als we niet items 1 t/m 6 hadden gehercodeerd maar items 7 en 8?

Antwoord:

4. Standaardmeetfout

Vraag 20. Hoe hoog is de geschatte standaardmeetfout van testscore X?

Antwoord:

Vraag 21. Welke schatting van het 95%-betrouwbaarheidsinterval wordt er berekend voor een kind met testscore 15?

Antwoord:

Vraag 22. Hoeveel (ongeveer) van die intervallen zal dan de ware score van dit kind bevatten?

Antwoord:

5. De betrouwbaarheid van subtestscores

Vraag 23. Voor welk item is de schatting van de communaliteit het hoogst?

Antwoord:

Vraag 24. Voor welk item is de schatting van de communaliteit het laagst?

Antwoord:

Vraag 25. Hoe hoog is de proportie variantie van subtestscore S_1 die wordt verklaard door de gemeenschappelijke factoren?

Antwoord:

Vraag 26. Hoe hoog is de proportie variantie van subtestscore S_2 die wordt verklaard door de gemeenschappelijke factoren?

Antwoord:

Vraag 27. Hoe hoog is de proportie variantie van de totale ongewogen som van de itemscores die wordt verklaard door de gemeenschappelijke factoren?

Antwoord:

Vraag 28. Wat valt je op als je naar de geschatte communaliteiten van de itemscores, de subtestscores en de ongewogen somscore kijkt?

Antwoord:

Vraag 29. Als je de berekende communaliteiten van de subtestscores S_1 en S_2 met die van de Thurstone factorscores vergelijkt, welke testcores lijken dan betrouwbaarder te zijn?

Antwoord:

Appendix E
Ten-point Mental Effort Scale Experiment 1

Name _____

Gender: M / W

Student number:

Ik werk in de practica samen met...

Werkgroep:

De onderwerpen in het practicum waren erg complex.

0 1 2 3 4 5 6 7 8 9 10

helemaal niet het geval

helemaal het geval

Het practicum bevatte opdrachten die ik heb ervaren als complex.

0 1 2 3 4 5 6 7 8 9 10

helemaal niet het geval

helemaal het geval

Het practicum bevatte concepten en definities die ik heb ervaren als complex.

0 1 2 3 4 5 6 7 8 9 10

helemaal niet het geval

helemaal het geval

De instructies en/of uitleg gedurende de activiteit waren erg onduidelijk.

0 1 2 3 4 5 6 7 8 9 10

helemaal niet het geval

helemaal het geval

De instructies en/of uitleg waren, in termen van leren, erg inefficiënt.

0 1 2 3 4 5 6 7 8 9 10

helemaal niet het geval

helemaal het geval

De instructies en/of uitleg stonden vol met onduidelijke taal.

0 1 2 3 4 5 6 7 8 9 10

helemaal niet het geval

helemaal het geval

Het practicum heeft me echt geholpen de behandelde begrippen te begrijpen.

0 1 2 3 4 5 6 7 8 9 10

helemaal niet het geval

helemaal het geval

Het practicum heeft mijn kennis en begrip van statistiek echt vergroot.

0 1 2 3 4 5 6 7 8 9 10

helemaal niet het geval

helemaal het geval

Het practicum heeft me geholpen de behandelde formules te begrijpen.

0 1 2 3 4 5 6 7 8 9 10

helemaal niet het geval

helemaal het geval

Het practicum heeft me geholpen de concepten en definities te begrijpen.

0 1 2 3 4 5 6 7 8 9 10

helemaal niet het geval

helemaal het geval

Meetinvariantie

2. Multi-groep confirmatieve factoranalyses

Vraag 1. Hoeveel mensen zitten er in beide groepen?

Antwoord. In de eerste groep zitten 85 personen en in de tweede 65.

Vraag 2. Welke waarden vind je voor de toetsingsgrootheid, het aantal vrijheidsgraden en de p-waarde?

Antwoord. De toetsingsgrootheid heeft de waarde 1.879. Het aantal vrijheidsgraden is 4. De p-waarde is .758.

Vraag 3. Moet het model volgens de toets worden verworpen of niet? Waarom?

Antwoord. Nee, het model kan niet worden verworpen, want de p-waarde is groter dan het gebruikelijke nominale significantieniveau van .05.

Vraag 4. Past het model goed op de data volgens de CFI? Waarom?

Antwoord. De CFI is 1.000. De fit is dus acceptabel, want de waarde is groter dan .95.

Vraag 5. Past het model goed op de data volgens de TLI? Waarom?

Antwoord. De TLI is 1.094. De fit is acceptabel, want de waarde is groter dan .95.

Vraag 6. Past het model goed op de data volgens de RMSEA? Waarom?

Antwoord. De RMSEA is 0.000. De fit is acceptabel, want de waarde is onder 0.05.

Vraag 7. Op welke manier zijn er schalen aan de latente variable toegekend? Waarom?

Antwoord. 'Unit loading identification', de eerste factorlading is op 1 gezet.

Vraag 8. Bekijk de schattingen van de parameters nog eens goed. Is er een parameter waarvoor de schattingen in beide groepen hetzelfde zijn (vastgelegde waarden zijn geen schattingen en hebben daarom ook geen **Std.Err**)?

Antwoord. Nee, er is er niet een.

Vraag 9. Wat is zwakke factoriële invariantie?

Antwoord. Zwakke factoriële invariantie betekent dat de factorladingenmatrix in alle groepen precies hetzelfde is.

Vraag 10. Welke waarden vind je voor de toetsingsgrootheid, het aantal vrijheidsgraden en de p-waarde?

Antwoord. De waarde van de toetsingsgrootheid is 8.399. Het aantal vrijheidsgraden is 7. De p-waarde is .299.

Vraag 11. Moet zwakke factoriële invariantie volgens de toets worden verworpen of niet? Waarom?

Antwoord. Nee, zwakke factoriële invariantie kan niet worden verworpen, want de p-waarde is groter dan het gebruikelijke significantieniveau van .05.

Vraag 12. Past het zwakke factoriële invariantie-model goed op de data volgens de CFI? Waarom?

Antwoord. De CFI is .979. De fit is acceptabel, want deze waarde is groter dan .95.

Vraag 13. Past het zwakke factoriële invariantie-model goed op de data volgens de TLI? Waarom?

Antwoord. De TLI is .965. De fit is acceptabel, want deze waarde is groter dan .95.

Vraag 14. Past het zwakke factoriële invariantie-model goed op de data volgens de RMSEA? Waarom?

Antwoord. De RMSEA is 0.052. De fit is niet acceptabel, maar ook niet onacceptabel, want de waarde is niet kleiner dan .05 en ook niet groter dan .10.

Vraag 15. Wat is sterke factoriële invariantie?

Antwoord. Sterke factoriële invariantie betekent dat zowel de factorladingenmatrix als de vector van intercepten hetzelfde is voor alle groepen.

Vraag 16. Welke waarden vind je voor de toetsingsgrootte, het aantal vrijheidsgraden en de p-waarde?

Antwoord. De waarde van de toetsingsgrootte is 14.885. Het aantal vrijheidsgraden is 10. De p-waarde is .136.

Vraag 17. Moet sterke factoriële invariantie volgens de toets worden verworpen of niet? Waarom?

Antwoord. Sterke factoriële invariantie kan niet worden verworpen, want de p-waarde is groter dan het gebruikelijke nominale significantieniveau van .05.

Vraag 18. Past het sterke factoriële invariantie-model goed op de data volgens de CFI? Waarom?

Antwoord. De CFI is .928. De passing is niet acceptabel, maar ook niet onacceptabel. De waarde is niet lager dan .90.

Vraag 19. Past het sterke factoriële invariantie-model goed op de data volgens de TLI? Waarom?

Antwoord. De TLI is .913. De passing is niet acceptabel, maar ook niet onacceptabel. De waarde is niet lager dan .90.

Vraag 20. Past het sterke factoriële invariantie-model goed op de data volgens de RMSEA? Waarom?

Antwoord. De RMSEA is .081. De passing is niet acceptabel, maar ook niet onacceptabel. De waarde is niet lager dan .05 en ook niet hoger dan .10.

Vraag 21. Hoe groot is het geschatte verschil tussen de gemiddelden van de gemeenschappelijke angst-factor in de twee groepen?

Antwoord. In de output van groep twee onder **Intercepts:** vind je voor **f1** de waarde 0.093. Dit is het verschil tussen de gemiddelden van de gemeenschappelijke factor in de twee groepen.

Vraag 22. In welke groep is volgens het geschatte gemiddelde verschil het gemiddelde op de gemeenschappelijke angst-factor hoger? Waarom?

Antwoord. Volgens het geschatte gemiddelde verschil is het gemiddelde op de gemeenschappelijke angst-factor in de tweede groep hoger dan in de eerste groep omdat het gemiddelde in de eerste groep op nul is gezet en de schatting in de tweede groep positief is.

Vraag 23. Is het geschatte gemiddelde verschil statistisch significant? Waarom? Welke nulhypothese kan hiermee worden getoetst? Wat is de conclusie?

Antwoord. Ja, het geschatte gemiddelde verschil is statistisch significant. De tweezijdige p-waarde is .042. Deze p-waarde kan worden gebruikt voor het toetsen van de nulhypothese dat het gemiddelde verschil gelijk is aan nul tegen de alternatieve hypothese dat het verschil niet gelijk is aan nul.

Vraag 24. Welke waarden vind je voor de toetsingsgrootte, het aantal vrijheidsgraden en de p-waarde?

Antwoord. De waarde van de toetsingsgrootte is 4.9656. Het aantal vrijheidsgraden is 1. De p-waarde is .02586.

Vraag 25. Welke nulhypothese kan er met de chi-kwadraatverschiltoets worden getoetst als er drie groepen waren geweest?

Antwoord. Als er drie groepen waren geweest, dan had de nulhypothese dat de angst-factor gemiddelden van de drie groepen aan elkaar gelijk zijn, getoetst kunnen worden.

Vraag 26. Wat kan er worden gedaan om te onderzoeken welke groepen van elkaar verschillen als de chi-kwadraatverschiltoets een significant resultaat oplevert in het geval van drie of meer groepen?

Antwoord. In het geval van drie of meer groepen kan na een significant toetstingsresultaat worden onderzocht welke groepsgemiddelden op de angst-factor van elkaar verschillen door middel van paarsgewijze vergelijkingen en een correctie voor type I fout inflatie.

Vraag 27. Wat is strikte factoriële invariantie?

Antwoord. In het geval van strikte factoriële invariantie is zowel de factorladingenmatrix, de vector van intercepten als de covariantiematrix van de unieke factoren in alle groepen precies hetzelfde.

Vraag 28. Welke waarden vind je voor de toetsingsgrootte, het aantal vrijheidsgraden en de p-waarde?

Antwoord. De waarde van de toetsingsgrootte is 17.221. Het aantal vrijheidsgraden is 14. De p-waarde is .245.

Vraag 29. Moet strikte factoriële invariantie volgens de toets worden verworpen of niet? Waarom?

Antwoord. Nee, strikte factoriële invariantie kan niet worden verworpen, want de p-waarde is groter dan het gebruikelijke nominale significantieniveau van .05.

Vraag 30. Past het strikte factoriële invariantie-model goed op de data volgens de CFI? Waarom?

Antwoord. De CFI is .952. De passing is acceptabel, want de waarde is groter dan .95.

Vraag 31. Past het strikte factoriële invariantie-model goed op de data volgens de TLI? Waarom?

Antwoord. De TLI is .959. De passing is acceptabel, want de waarde is groter dan .95.

Vraag 32. Past het strikte factoriële invariantie-model goed op de data volgens de RMSEA? Waarom?

Antwoord. De RMSEA is 0.055. De passing is niet acceptabel, maar ook niet onacceptabel, want de waarde is niet kleiner dan 0.05 en ook niet groter dan 0.10.

Vraag 33. Welke test scores kunnen er nu worden gebruikt om de waarde van iemand uit de eerste groep op de gemeenschappelijke factor met die van iemand uit de tweede groep te kunnen vergelijken? Waarom?

Antwoord. Om de waarde van iemand uit de eerste groep op de gemeenschappelijke factor met die van iemand uit de tweede groep te vergelijken, kan nu de ongewogen somscore worden gebruikt omdat er geen reden is om aan te nemen dat er onzuivere itemscores zijn die voor de groepen verschillend meten.

Vraag 34. Hoeveel mensen zitten er in beide groepen?

Antwoord. De eerste groep bestaat uit 90 personen (code 0 voor vrouwen) en de tweede groep uit 110 (code 1 voor mannen).

Vraag 35. Past het model met configurale invariantie op de data? Waarom?

Antwoord. Ja, het model past op de data. De waarde van de toetsingsgrootheid is 10.803. Het aantal vrijheidsgraden is 16. De p-waarde is .821. Het model kan niet worden verworpen want de p-waarde is groter dan .05. De CFI is 1.000. Deze waarde is hoger dan .95. De TLI is ook hoger dan .95. De waarde van de RMSEA is lager dan .05.

Vraag 36. Past het model met zwakke factoriële invariantie op de data? Waarom?

Antwoord. Het model met zwakke factoriële invariantie past op de data. Het model kan niet worden verworpen omdat de p-waarde gelijk is aan .306, de TLI gelijk is aan .967, de CFI gelijk is aan .978 en de RMSEA gelijk is aan .036.

Vraag 37. Past het model met sterke factoriële invariantie op de data? Waarom?

Antwoord. Het model met sterke factoriële invariantie past op de data. Het model kan niet worden verworpen omdat de p-waarde gelijk is aan .368, de TLI gelijk is aan .982, de CFI gelijk is aan .986 en de RMSEA gelijk is aan .027.

Vraag 38. Hoe groot zijn de verschillen in de gemiddelden van de gemeenschappelijke factoren tussen de twee groepen?

Antwoord. Voor de eerste factor is dat 0.07 en voor de tweede factor is dat -0.489.

Vraag 39. Welke waarden vind je voor de toetsingsgrootheid, het aantal vrijheidsgraden en de p-waarde?

Antwoord. De waarde van de toetsingsgrootheid is 2.629. Het aantal vrijheidsgraden is 2. De p-waarde is .269.

Vraag 40. Wat is de nulhypothese en kunnen we die verwerpen?

Antwoord. De nulhypothese is dat de gemiddelden van de twee latente variabelen in beide groepen hetzelfde zijn. Deze nulhypothese kan niet worden verworpen.

Vraag 41. Interpreteer de resultaten.

Antwoord. Er zijn geen verschillen tussen mannen en vrouwen in gemiddelde stemmings-evaluatie en gemiddelde mate van geactiveerdheid.

Vraag 42. Past het model met strikte factoriële invariantie op de data? Waarom?

Antwoord. Het model met strikte factoriële invariantie past op de data. Het model kan niet worden verworpen omdat de p-waarde gelijk is aan .190, de TLI en de CFI gelijk zijn aan .945 en de RMSEA gelijk is aan .047.

Factoranalyse van een MTMM-matrix

2. Analyses

Vraag 43. Is er bewijs voor convergente validiteit? Waarom?

Antwoord. Ja, de nulhypothese van geen convergente validiteit kan worden verworpen, want de p-waarde is kleiner dan .05. De p-waarde is $3.185e-06$, dit betekent 3.185 keer 10 tot de macht -6 oftewel 3.185 gedeeld door 10 tot de macht 6. Een getal delen door 10 tot de macht 6 is hetzelfde als dat getal delen door 1 miljoen. De p-waarde is dus afgerond op drie decimalen gelijk aan .000.

Vraag 44. Is er bewijs voor divergente validiteit? Waarom?

Antwoord. Ja, de nulhypothese van geen divergente validiteit kan worden verworpen. De p-waarde is kleiner dan .05. De p-waarde is $6.216e-05$.

Vraag 45. Is er bewijs voor methode-effecten? Waarom?

Antwoord. Ja, de nulhypothese van geen methode-effecten moet worden verworpen. De p-waarde is kleiner dan .05. De p-waarde is $4.672e-16$.

2. Betrouwbaarheidsanalyse

Vraag 1. Van hoeveel personen zijn de itemscores gebruikt bij het berekenen van de schatting van Cronbachs alfa voor de ongewogen somscore?

Antwoord. 205

Vraag 2. Wat is de geschatte waarde van Cronbachs alfa voor de ongewogen somscore?

Antwoord. .713

Vraag 3. Voor welk item geldt dat de item-restcorrelatie te laag is?

Antwoord. Item 7, de waarde is .295.

Vraag 4. Is Cronbachs alfa voor de restscore van dit item groter dan Cronbachs alfa voor de totale testscore?

Antwoord. Ja, de waarde is .717.

Vraag 5. Is er een inhoudelijk argument om item 7 toch in de testscore op te nemen?

Antwoord. Ja, het item heeft veel met schoolplezier te maken.

Vraag 6. Is er een statistisch argument om item 7 toch in de testscore op te nemen?

Antwoord. Ja, Cronbachs alfa stijgt nauwelijks na verwijdering van het item.

Vraag 7. Hoe kwalificeer je de schatting van Cronbachs alfa voor de testscore volgens de criteria van de COTAN (zie college slides)?

Antwoord. Een testscore met een Cronbachs alfa van .71 is volgens de criteria van de COTAN te kwalificeren als voldoende betrouwbaar voor algemeen beschrijvend gebruik.

Vraag 8. Welke toepassing van de test (tests voor belangrijke beslissingen op individueel niveau, tests voor minder belangrijke beslissingen op individueel niveau of tests voor onderzoek op groepsniveau) is dus verantwoord?

Antwoord. De testscore kan gebruikt worden voor onderzoek op groepsniveau of voor het nemen van minder belangrijke beslissingen op individueel niveau.

Vraag 9. Wat is de geschatte waarde van Guttman's λ^2 voor de ongewogen somscore?

Antwoord. .74

Vraag 10. Wat is de geschatte waarde van coëfficiënt omega voor de ongewogen somscore onder het één-factor model?

Antwoord. .735

Vraag 11. Wat is de geschatte waarde van coëfficiënt omega voor deze gewogen somscore onder het één-factor model?

Antwoord. .639

Vraag 12. Als je de antwoorden van vragen 10 en 11 met elkaar vergelijkt, kun je dan concluderen dat je moet spreken van de communaliteit/betrouwbaarheid van de test of juist van de communaliteit/betrouwbaarheid van een testscore? Waarom?

Antwoord. De communaliteit/betrouwbaarheid hangt af van hoe de itemscores zijn gewogen in de berekening van de testscore. De communaliteit/betrouwbaarheid is dus een eigenschap van de testscore.

3. Testscore Schoolplezier

Vraag 13. Wat is theoretisch de laagst mogelijke waarde op X?

Antwoord. De theoretisch laagst mogelijke testscore is 8.

Vraag 14. Wat is theoretisch de hoogst mogelijke waarde op X?

Antwoord. De theoretisch hoogst mogelijke testscore is 32.

Vraag 15. Wat is het steekproefgemiddelde van testscore X?

Antwoord. 24.90

Vraag 16. Wat is de steekproefstandaardafwijking van testscore X?

Antwoord. 4.82

Vraag 17. Wat betekent een standaardafwijking?

Antwoord. Het is een indicatie voor de spreiding van de scores in de steekproef. Ongeveer 32% van de scores ligt meer dan 1 standaardafwijking van het gemiddelde.

Vraag 18. Hebben de leerlingen, afgaande op de verdeling van de testscore, plezier in school?

Antwoord. Een gemiddelde van 24.90 op een schaal van 8 (geen plezier) tot en met 32 (plezier) duidt erop dat de leerlingen over het algemeen plezier beleven aan school.

Vraag 19. Wat zou de betekenis van een hoge testscore zijn geweest als we niet items 1 t/m 6 hadden gehercodeerd maar items 7 en 8?

Antwoord. Dan zou een hoge score betekenen dat men geen plezier aan school beleeft.

4. Standaardmeetfout

Vraag 20. Hoe hoog is de geschatte standaardmeetfout van testscore X?

Antwoord. 2.483

Vraag 21. Welke schatting van het 95%-betrouwbaarheidsinterval wordt er berekend voor een kind met testscore 15?

Antwoord. [10.13, 19.87]

Vraag 22. Hoeveel (ongeveer) van die intervallen zal dan de ware score van dit kind bevatten?

Antwoord. Ongeveer 95 van de 100 intervallen zal de ware score van dit kind bevatten.

5. De betrouwbaarheid van subtestcores

Vraag 23. Voor welk item is de schatting van de communaliteit het hoogst?

Antwoord. Voor item 4. De schatting van de communaliteit van item 4 is .605.

Vraag 24. Voor welk item is de schatting van de communaliteit het laagst?

Antwoord. Voor item 1. De schatting van de communaliteit van item 1 is .129.

Vraag 25. Hoe hoog is de proportie variantie van subtestscore S_1 die wordt verklaard door de gemeenschappelijke factoren?

Antwoord. De geschatte proportie verklaarde variantie van subtestscore S_1 is .602

Vraag 26. Hoe hoog is de proportie variantie van subtestscore S_2 die wordt verklaard door de gemeenschappelijke factoren?

Antwoord. De geschatte proportie verklaarde variantie van subtestscore S_2 is .551

Vraag 27. Hoe hoog is de proportie variantie van de totale ongewogen som van de itemscores die wordt verklaard door de gemeenschappelijke factoren?

Antwoord. De geschatte proportie variantie van de totale ongewogen som van de itemscores is .721

Vraag 28. Wat valt je op als je naar de geschatte communaliteiten van de itemscores, de subtestcores en de ongewogen somscore kijkt?

Antwoord. Er valt op dat hoe meer itemscores in een testscore worden meegewogen, hoe hoger de communaliteit van de testscore is.

Vraag 29. Als je de berekende communaliteiten van de subtestcores S_1 en S_2 met die van de Thurstone factorscores vergelijkt, welke testcores lijken dan betrouwbaarder te zijn?

Antwoord. De berekende communaliteiten van de twee Thurstone factorscores zijn .789 en .791. Deze waarden zijn aanzienlijk hoger dan de communaliteiten .602 en .551 van S_1 en S_2 . Het lijkt er dus op dat de Thurstone factorscores betrouwbaarder zijn dan de twee subtestcores S_1 en S_2 . De voorkeur gaat dan in eerste instantie uit naar het gebruiken van de Thurstone factorscores.

Appendix H
Planning Phase Assignment Control Group

Namen:

Studentnummers:

Werkgroep:

Beste studenten,

Op de volgende pagina's kunnen jullie een aantal opdrachten vinden. Jullie krijgen 15 minuten de tijd om deze op te lossen. Het is belangrijk dat jullie **geen** mobiele telefoon gebruiken tijdens het maken van deze opdrachten! Na 15 minuten moeten de opdrachten worden ingeleverd bij de onderzoeker. Het op een juiste manier oplossen van opdrachten is belangrijker dan ze allemaal in de tijd af te hebben.

Als er vragen zijn, steek je hand op, of loop naar een van de onderzoekers.

Je zult tien minuten voor het einde van het practicum een vragenlijst ontvangen via e-mail. Vergeet deze niet in te vullen!

Alvast bedankt!

Met vriendelijke groet,

Anniek de Kort

<p>Maak een schatting: hoeveel mensen waren er op de wereld in 1763?</p>	
<p>Istanbul heeft in de geschiedenis veel namen gehad. Schrijf deze op.</p>	
<p>Minister-President Rutte heeft naast deze belangrijke baan ook een andere, weten jullie welke?</p>	

Wat vinden jullie van
klimaatprotesten onder
middelbare scholieren?

Hoe zou Nederland de
huizencrisis op moeten
lossen?

Wat is jullie mening over de volgende stelling:

‘Het is goed dat de Nederlandse staat geen studiefinanciering meer geeft’

Onderbouw jullie mening.

Appendix I
Planning Phase Assignment Experimental Group

Namen:

Studentnummers:

Werkgroep:

Beste studenten,

Op de volgende pagina's vinden jullie een achttal vragen die jullie helpen bij het plannen van het practicum dat jullie zo dadelijk gaan uitvoeren. Omdat jullie de planning samen maken is het belangrijk even goed met elkaar te overleggen over iedere vraag. Jullie krijgen hier 15 minuten voor. Jullie mogen de planning bij het practicum houden. Na afloop van het practicum moeten jullie de planning inleveren bij de onderzoekers. Als er vragen zijn, steek je hand op, of loop naar een van de onderzoekers.

Elk van jullie zal tien minuten voor het einde van het practicum een vragenlijst ontvangen via e-mail. Vergeet deze niet in te vullen!

Alvast bedankt!

Met vriendelijke groet,

Anniek de Kort

<p>Hoeveel tijd hebben jullie om deze opdracht af te ronden?</p>	
<p>Welke leerdoelen hebben jullie gesteld voor deze opdracht? Denk bijvoorbeeld aan de leerdoelen: 'na afloop van het practicum kan ik zelfstandig een bepaalde analyse uitvoeren in R'.</p>	
<p>Wat gaan jullie doen om deze doelen halen?</p>	

<p>Welke vaardigheden of kennis hebben jullie nodig om de doelen te behalen?</p>	
<p>Welke vaardigheden en kennis bevatten jullie, en welke niet?</p>	
<p>Wat gaan jullie doen wanneer jullie vastlopen bij opdrachten?</p>	
<p>Als jullie tegen een probleem aanlopen, hoe gaan jullie dit dan oplossen?</p>	

Schrijf hier jullie plan om het practicum op een goede manier af te ronden.

Daarvoor kunnen jullie de belangrijkste bevindingen uit bovenstaande vragen voor gebruiken.

Appendix J
Interview After Experiment 1

1. *Zouden jullie je naam en studentnummer willen geven.*
2. *Op een schaal van 0 tot 10, waarbij 0 helemaal niet, en 10 heel erg is, hoe serieus hebben jullie de planning ingevuld? Waarom wel of niet?*
3. *Hoe hebben jullie de planning gemaakt?*
4. *Hebben jullie naar de opdracht gekeken voordat jullie de planning gingen maken ?*

5. *Welke doelen hebben jullie gesteld, en waarom?*
6. *Hoe vaak hebben jullie, ongeveer teruggekeken in jullie planning om de opdracht af te ronden? Waarom?*
7. *Hebben jullie je planning aangepast tijdens de opdracht? Zo ja, hoe?*
8. *Hebben jullie vaak teruggedacht aan de planning die jullie gemaakt hebben?*
9. *Liepen jullie tegen problemen aan tijdens de opdracht? Zo ja, hoe hebben jullie deze problemen opgelost?*
10. *Wat vonden jullie fijn of minder fijn aan de planning?*

11. *Zijn er verder vragen, opmerkingen of andere dingen die jullie kwijt willen?*

Goal-setting (Hong & O'Neil, 2001)

- Welke doelen hebben jullie gesteld, en waarom?

Planning actions (Toering et al., 2012)

Following the plan

Adjusting the plan

- Hoe vaak hebben jullie, ongeveer teruggekeken in jullie planning om de opdracht af te ronden? Waarom?
- Hebben jullie je planning aangepast tijdens de opdracht? Zo ja, hoe?

Reflecting on the planning (Hong & O'Neil, 2001)

- Hebben jullie vaak teruggedacht aan de planning die jullie gemaakt hebben?
- Hebben jullie je planning aangepast tijdens de opdracht? Zo ja, hoe?
- Hebben jullie vaak teruggedacht aan de planning die jullie gemaakt hebben?
- Wat vonden jullie fijn en minder fijn aan de planning?

Problem solving skills (Toering et al., 2012)

- Liepen jullie tegen problemen aan tijdens de opdracht? Zo ja, hoe hebben jullie deze problemen opgelost?

Appendix K
Informed Consent Experiment 2

Informed Consent studenten practicum TOE

***Betreft:** Onderzoek naar de invloed van zelfregulatie op de resultaten van een statistiek practicum en de rol van cognitieve belasting hierbij.*

Ik heb de uitnodiging voor de proefpersoon gelezen, waarin informatie stond over het onderzoek, en kon aanvullende vragen stellen aan de onderzoeker (Anniek de Kort, a.i.m.dekort@uu.nl) wanneer ik iets niet begreep. Mijn vragen zijn genoeg beantwoord en ik had genoeg tijd om te beslissen of ik meedoe aan dit onderzoek.

Ik weet dat meedoen geheel vrijwillig is. Ik weet dat ik op ieder moment kan beslissen om toch niet mee te doen, daarvoor hoef ik geen reden te geven.

Ik geef hierbij toestemming om onderstaande gegevens te gebruiken:

- de opdracht die ik voorafgaand aan het practicum zal maken met een medestudent,
- de opdracht gemaakt tijdens het practicum,
- de antwoorden die ik gaf op de vragen van de vragenlijst na afloop van het practicum,
- de antwoorden die ik gaf op de vragen van het interview, die gingen over het maken van de opdracht voorafgaand aan het practicum en het gebruik van deze opdracht tijdens het practicum.

Naam proefpersoon:

Handtekening:

Datum : __ / __ / __

Ik verklaar hierbij dat ik deze proefpersoon volledig heb geïnformeerd over het onderzoek.

Naam onderzoeker (of diens vertegenwoordiger):

Handtekening:

Datum: __ / __ / __

Appendix L
Questionnaire Experiment 2

Naam: _____

Geslacht: M / W (zet een cirkel om wat voor jou van toepassing is)

Leeftijd: _____ (DD/MM/JJJJ)

Studentnummer:

Werkgroep:

Alle vragen gaan over het practicum van deze week (week 17) die zojuist is afgelopen.
Beantwoord de volgende vragen op de volgende schaal:

0 - helemaal niet het geval

10 - helemaal het geval

De onderwerpen in het practicum waren erg complex.

0 1 2 3 4 5 6 7 8 9 10

helemaal niet het geval

helemaal het geval

Het practicum bevatte opdrachten die ik heb ervaren als complex.

0 1 2 3 4 5 6 7 8 9 10

helemaal niet het geval

helemaal het geval

Het practicum bevatte concepten en definities die ik heb ervaren als complex.

0 1 2 3 4 5 6 7 8 9 10

helemaal niet het geval

helemaal het geval

De instructies en/of uitleg gedurende de activiteit waren erg onduidelijk.

0 1 2 3 4 5 6 7 8 9 10

helemaal niet het geval

helemaal het geval

De instructies en/of uitleg waren, in termen van leren, erg inefficiënt.

0 1 2 3 4 5 6 7 8 9 10

helemaal niet het geval

helemaal het geval

De instructies en/of uitleg stonden vol met onduidelijke taal.

0 1 2 3 4 5 6 7 8 9 10
helemaal niet het geval helemaal het geval

Het practicum heeft me echt geholpen de behandelde begrippen te begrijpen.

0 1 2 3 4 5 6 7 8 9 10
helemaal niet het geval helemaal het geval

Het practicum heeft mijn kennis en begrip van statistiek echt vergroot.

0 1 2 3 4 5 6 7 8 9 10
helemaal niet het geval helemaal het geval

Het practicum heeft me geholpen de behandelde formules te begrijpen.

0 1 2 3 4 5 6 7 8 9 10
helemaal niet het geval helemaal het geval

Het practicum heeft me geholpen de concepten en definities te begrijpen.

0 1 2 3 4 5 6 7 8 9 10
helemaal niet het geval helemaal het geval

Welk cijfer zou jij je zojuist gemaakte practicum geven?

0 1 2 3 4 5 6 7 8 9 10

Zeer slecht

Uitstekend

Waarom zou je dit cijfer geven?

Hoe serieus heb je je planning ingevuld?

Heel erg bedankt! Heb je interesse in het ontvangen van de resultaten van mijn onderzoek? Vul dan hier je e-mailadres in. Lever na afloop dit formulier in bij de onderzoeker.

Appendix M FETC Form

APPLICATION FORM FOR THE ASSESSMENT OF A RESEARCH PROTOCOL BY THE FACULTY ETHICS REVIEW BOARD (FERB) OF THE FACULTY OF SOCIAL AND BEHAVIOURAL SCIENCES

General guidelines for the use of this form

1. This form can be used for a single research project or a series of related studies (hereinafter referred to as: "research programme"). Researchers are encouraged to apply for the assessment of a research programme if their proposal covers multiple studies with related content, identical procedures (methods and instruments) and contains informed consent forms and participant information, with a similar population. For studies by students, the FERB recommends submitting, in advance, a research programme under which protocol multiple student projects can be conducted so that their execution will not be delayed by the review procedure. The application of such a research programme must include a proper description by the researcher(s) of the programme as a whole in terms of the maximum burden on the participants (e.g. maximum duration, strain/efforts, types of stimuli, strength and frequency, etc.). If it is impossible to describe all the studies within the research programme, it should, in any case, include a description of the most invasive study known so far.
2. Solely the first responsible senior researcher(s) (from post-doctoral level onwards) may submit a protocol.
3. Any approval by the FERB is valid for 5 years or until the information to be provided in the application form below is modified to such an extent that the study becomes more invasive. For a research programme, the term of validity is 2 years and any extension is subject to approval. The researcher(s) and staff below commit themselves to treating the participants in accordance with the principles of the Declaration of Helsinki and the Dutch Code of Conduct for Scientific Practices as determined by the VSNU Association of Universities in the Netherlands (which can both be downloaded from the FERB site on the Intranet¹) and guarantee that the participants (whether decisionally competent or incompetent and/or in a dependent relationship vis-a-vis the researcher or not) may at all times terminate their participation without any further consequences.
4. The researcher(s) commit themselves to maximising the quality of the study, the statistical analysis and the reports, and to respect the specific regulations and legislation pertaining to the specific methods.
5. The procedure will run more smoothly if the FERB receives all the relevant documents, such as questionnaires and other measurement instruments as well as literature and other sources on studies using similar methods which were found to be ethically acceptable and that testify to the fact that this procedure has no harmful consequences. Examples of studies where the latter will always be an issue are studies into bullying behaviour, sexuality, and parent-child relationships. The FERB asks the researcher(s) to be as specific as possible when they answer the relevant

¹ See: <https://intranet.uu.nl/facultaire-ethische-toetsingscommissie-fetc>

questions while limiting their answers to 500 words maximum per question. It is helpful to the FERB if the answers are brief and to the point.

6. Our FAQ document that can be accessed through the Intranet provides background information with regards to any questions.
7. The researcher(s) declare to have described the study truthfully and with a particular focus on its ethical aspects.

Signed for approval²:

Date:

² The senior researcher (holding at least a doctoral degree) should sign here.

A. GENERAL INFORMATION/PERSONAL DETAILS

1.

a. Name(s), position(s) and department(s) of the responsible researcher(s):

Femke Kirschner, Education Development Assistant, Social and Behavioral Sciences, Educational Development and Training.

b. Name(s), position(s) and department(s) of the executive researcher(s):

Anniek de Kort, Master Student Educational Sciences, Social and Behavioral Sciences.

2. Title of the study or research programme - Does it concern a single study or a research programme? Does it concern a study for the final thesis in a bachelor's or master's degree course?

Master Thesis Educational Sciences. It concerns a single study as final thesis in a master's degree course.

3. Type of study (with a brief rationale):

- experimental

Two groups of students will be researched. The experimental group will get time to plan an assignment for a course ahead, the control group gets time to plan their course ahead. While making the assignment, some groups of students will be recorded. After the assignment, the groups will receive a questionnaire about their perceived cognitive load.

Also, the grades of the assignment are taken into account.

4. Grant provider:

N/A

5. Intended start and end date for the study:

February 4th – April 12.

6. Research area/discipline:

Social and Behavioural Sciences

7. For some (larger) projects it is advisable to appoint an independent contact or expert whom participants can contact in case of questions and/or complaints. Has an independent expert been appointed for this study?³

No, however, participants can contact Femke Kirschner.

8. Does the study concern a multi-centre project, e.g. in collaboration with other universities, a GGZ mental health care institution, a university medical centre? Where exactly will the study be conducted? By which institute(s) are the executive researcher(s)

³ This contact may, in principle, also be a researcher (within the same department, or not) who is able to respond to the question or complaint in detail. Independent is to say: not involved in the study themselves. The FERB upholds that an independent contact is not obligatory, but will be necessary when the study is more invasive.

employed?

No

9. Is the study related to a prior research project that has been assessed by a recognised Medical Ethics Review Board (MERB) or FERB?

No

If so, which? Please state the file number:

N/A

B. SUMMARY OF THE BACKGROUND AND METHODS

Background

1. What is the study's theoretical and practical relevance? (500 words max.):

This study will research the influence of planning on the cognitive load and the academic achievement of students. Self-Regulated Learning (SRL) has been an influential theory in the Educational Sciences. SRL is taking the affective human architecture in mind, and claims that, when students have more SRL skills, they score higher on test, are less anxious about tests, and have better metacognitive skills. Because of this, more interventions to train SRL skills are developed. Studies show that these interventions on SRL can be effective, however, other studies show that interventions can be seen as a secondary task. This can be explained by using a cognitive approach to human architecture. A theory that explains the cognitive human architecture is the Cognitive Load Theory (CLT). This study wants to show if there is more Cognitive Load (CL) when students are using SRL during an assignment and want to know if these theories influence the academic achievement of students. This study is providing theoretical knowledge about the cognitive influence on SRL, this is needed because those two theories are mainly studied separately. This study also has a practical relevance, because instructional designers can take both affective and cognitive aspects of human architecture in mind when designing courses or other educational material.

2. What is the study's objective/central question?

The central question in this study is 'what influence does planning has on students' cognitive load and how does this affect their achievement'.

3. What are the hypothesis/hypotheses and expectation(s)?

Hypothesis 1: Students who plan ahead experience more cognitive load than students who don't plan ahead.

Hypothesis 2: Learning outcomes from students who plan ahead are higher than outcomes from students who do not plan ahead.

The expectations are that students who plan ahead experience more cognitive load during the course but have higher outcomes than students who do not plan ahead.

Design/procedure/invasiveness

4. What is the study's design and procedure? (500 words max.):

When students were waiting for the assignment, the researcher explained the study shortly, whereafter they got time to fill in their informed consent forms. The researcher collected all the forms, whereafter the groups received their assignment. The assignment was made in groups of two.

Students from the experimental group got approximately 15 minutes to plan the assignment ahead using the format, while students from the controlled group got approximately 15 minutes to plan their whole course ahead. After students planned the assignment or the course, the researcher put microphones near 20 groups of students to record their actions during the assignment. The division between recorded groups were 10 in the experimental group and 10 in the control group. After this was done, students made the assignment.

After students were done with the assignment, or when time run out, they filled in the ten-point mental effort scale. A random sample of students from the experimental group were asked to answer a short interview when they were done with the assignment to determine if they actually planned the assignment ahead. Both groups also received a short explanation of this research, with the hypotheses written on it. This debriefing also explained the importance of SRL-skills (see Appendix E).

The assignments were all reviewed by the researcher to determine academic achievement.

5.

a. Which measurement instruments, stimuli and/or manipulations will be used?⁴

The experimental group will receive time to plan the course ahead. They will receive a format for their planning.

- The ten-point mental effort scale. This questionnaire consists of 10 items on a 11-point Likert scale.

The control group will only receive time to plan their course ahead. They will receive a format for their planning.

- The ten-point mental effort scale.

Both groups will receive a demographic questionnaire at the beginning of the study

⁴ Examples: invasive questionnaires; interviews; physical/psychological examination, inducing stress, pressure to overstep important standards and values; inducing false memories; exposure to aversive materials like a unpleasant film, video clip, photos or electrical stimulus; long-term of very frequent questioning; ambulatory measurements, participation in an intervention, evoking unpleasant psychological or physical symptoms in an experiment, denial, diet, blood sampling, fMRI, TMS, ECG, administering stimuli, showing pictures, etc. In case of the use of a device (apparatus) or administration of a substance, please enclose the CE marking brochure for the relevant apparatus or substance, if possible.

b. What does the study's burden on the participants comprise in terms of time, frequency and strain/efforts?

Both groups will receive approximately 15 minutes to plan the assignment or course ahead. After the groups finished the assignment, they will receive 5 minutes to finish the ten-point mental effort scale. Some groups in the experimental group will be asked to answer short questions while being recorded.

c. Will the participants be subjected to interventions or a certain manner of conduct that cannot be considered as part of a normal lifestyle?

No

d. Will unobtrusive methods be used (e.g. data collection of uninformed subjects by means of observations or video recordings)?

No

e. Will the study involve any deception? If so, will there be an adequate debriefing, and will the deception hold any potential risks?

Yes, students from the control group make a planning for their course, this will not be taken into account in this study. All students receive a short explanation of the goal of this study afterwards. In this debriefing, they can read the potential of planning and why it can be important for studying.

6. Will the participants be tested beforehand as to their health condition or according to certain disorders? Are there any inclusion and/or exclusion criteria or specific conditions to be met in order for a participant to take part in this study?

No

7. Risks for the participants -

a. Which risks does the study hold for its participants?

There will be no risks for participants.

b. To what extent are the risks and objections limited? Are the risks run by the participants similar to those in daily life?

Yes, this intervention will be so small, there is no risks in participants' daily life.

8. How does the burden on the participants compare to the study's potential scientific contribution (theory formation, practical usability)?

The burden will be small for participants, while the practical usability is great.

9. Will a method be used that may, by coincidence, lead to a finding of which the participant should be informed?⁵ If so, what actions will be taken in the case of a coincidental finding?

When results show that students who do plan ahead, are perceiving less cognitive load and have higher academic achievement, participants will get informed by an announcement on blackboard.

Analysis/power

10. How will the researchers analyse the data? Which statistical analyses will be used?

Data was analyzed using the statistical program SPSS. First, missing answers were labeled and deleted. Also, students who were recidivists were deleted from the sample, because expertise level needed to be the same. After deleting participants, a total of --- participants were included in this study. Next, items were reversed and total scores for independent variables were calculated. Also, recorded data was transcribed and coded. In order to answer the research question, an independent Samples t Test was performed with CL as dependent variable and SRL the independent variable. To measure if there was a significant increase in academic achievement, a factorial between groups analysis of variance (ANOVA) was conducted with CL as dependent variable and SRL and academic achievement as independent variables.

11. What is the number of participants? Provide a power analysis and/or motivation for the number of participants. The current convention is a power of 0.80. If the study deviates from this power, the FERB would like you to justify why this is necessary:

⁵ For instance: dementia, dyslexia, giftedness, depression, extremely low heartbeat in an ECG, etc. If coincidental findings may be found, this should be included in the informed consent, including a description of the actions that will be taken in such an event.

With a power of 0.80, there will be 102 participants in the sample size (51 controlled group, 51 experimental group).

C. PARTICIPANTS, RECRUITMENT AND INFORMED CONSENT PROCEDURE

1. The nature of the research population (please tick):

1. General population without complaints/symptoms
2. General population with complaints/symptoms
3. Patients or population with a diagnosis (please state the diagnosis)

2. Age category of the participants (please tick):

- 18 years or older
- 16-17 years
- 13-15 years
- 12 years or younger

3. Does the study require a specific target group? If so, justify why the study cannot be conducted without the participation of this group (e.g. minors):

No.

4. Recruitment of participants -

a. How will the participants be recruited?

Students who are enrolled in a university course will be asked to participate in this study.

b. How much time will the prospective participants have to decide as to whether they will indeed participate in the study?

Approximately 20 minutes.

5. Does the study involve informed consent or mutual consent? Clarify the design of the consent procedure (who gives permission, when and how). Does the study involve active consent or passive consent? If no informed consent will be sought, please clarify the reason:

Both researcher and the participant have to agree on participation. The participants all get an informed consent form where they can tick boxes if they want to participate or not.

When participants want to participate, the researcher will check if they are recidivist. If they are recidivists, they will be excluded from the research.

6. Are the participants fully free to participate and terminate their participation whenever they want and without stating their grounds for doing so?

Yes

7. Will the participants be in a dependent relationship with the researcher?

No

8. Compensation

- a. **Will the participants be compensated for their efforts? If so, what is included in this recompense (financial reimbursement, travelling expenses, otherwise). What is the amount?**

There will be no compensation for participants.

- b. **Will this compensation depend on certain conditions, such as the completion of the study?**

N/A

D. **PRIVACY AND INFORMATION**

1.

- a. Will the study adhere to the requirements for anonymity and privacy, as referred to in the Faculty Protocol for Data Storage⁶?

- anonymous processing and confidential storage of data (i.e. storage of raw data separate from identifiable data): **yes/no**
- the participants' rights to inspect their own data: **yes/no**
- access to the data for all the researchers involved in the project: **yes/no**

If not, please clarify.

- b. **Has a Data Management Plan been designed?**

Yes

2.

- a. **Will the participant be offered the opportunity to receive the results (whether or not at the group level)?**

Yes, they can fill in their email address on the informed consent form.

- b. **Will the results of the study be fed back to persons other than the participants (e.g. teachers, parents)?**

No

If so, will this feedback be provided at the group or at the individual level?

No

3.

- a. Will the data be stored on the faculty's data server? **yes/no**

- b. Will the data that can be traced back to the individual be stored separately on the other faculty server available for this specific purpose?

Yes

If not, please clarify where will the data be stored instead?

⁶ This can be found on the Intranet: <https://intranet.uu.nl/wetenschappelijke-integriteit-facultair-protocol-dataopslag>

E. ADDITIONAL INFORMATION

Optional.

F. FORMS TO BE ENCLOSED (CHECKLIST)

- Text (advert) for the recruitment of participants
- Information letter for participant
- Informed consent form for participants
- Written or oral feedback information (debriefing text)
- (Descriptions of) questionnaires
- (Descriptions of) measurement instruments/stimuli/manipulations
- Literature/references

Signature(s):⁷

Date and place:

Name, position:

⁷ The senior researcher (holding at least a doctoral degree) should sign here.

Appendix A - Informed Consent Experimental Group

Beste student,

Wie ben ik?

Mijn naam is Anniek de Kort, masterstudent Educational Sciences aan de Universiteit Utrecht. Ik doe onderzoek in het kader van mijn master thesis, onder begeleiding van Femke Kirschner.

Wat is het doel van het onderzoek?

Mijn onderzoek gaat over de Cognitive Load die studenten ervaren tijdens het volgen van een cursus. Daarnaast wil ik graag meten of het maken van een planning invloed heeft op deze cognitive load en op behaalde resultaten.

Wat houdt het onderzoek in?

Je zal wekelijks een (korte) vragenlijst van ongeveer vijf minuten invullen. Met deze vragenlijst zal ik vragen naar de ervaren cognitive load en naar de gemaakte planning. De vragenlijst zal aan het einde van iedere week aan je gegeven worden.

Privacy en vertrouwelijkheid

Alle gegevens worden vertrouwelijk behandeld en anoniem verwerkt. Dat wil zeggen dat straks in de uitkomsten van het onderzoek niet te zien is welke antwoorden jij gegeven hebt. De gegevens worden alleen voor onderzoeks- en opleidingsdoeleinden gebruikt. Jouw docent krijgt jouw vragenlijsten niet te zien.

Mogelijkheid tot vragen, informatie en toestemming

Als je nog vragen hebt over het onderzoek, stel die dan nu aan één van de onderzoekers of stuur een mail aan: Anniek de Kort, A.I.M.dekort@uu.nl. Voor verdere vragen over de cursus en opdracht die wij maken kun je contact opnemen met: Femke Kirschner, F.C.Kirschner@uu.nl

Als je mee wilt doen aan het onderzoek, vul dan het formulier op de achterzijde in. Je kunt het formulier inleveren bij een van de onderzoekers.

Met vriendelijke groet,

Dr. Femke Kirschner, Anniek de Kort

TOESTEMMINGSVERKLARING voor deelname aan wetenschappelijk onderzoek

Ik heb uitleg gekregen over het onderzoek. De informatie over het onderzoek heb ik goed gelezen. Ik heb mijn eventuele vragen over het onderzoek gesteld. Ik heb goed nagedacht over of ik aan het onderzoek wil deelnemen. Ik mag op ieder moment stoppen met het onderzoek als ik dat wil. Ik hoef niet uit te leggen waarom ik wil stoppen.

Zet hiernaast een kruisje in het vakje dat voor jou van toepassing is

- Ik doe wel mee aan het onderzoek
- Ik doe niet mee aan het onderzoek

Naam :

Werkgroep :

Geboortedatum :

Datum :

Handtekening :

Als je informatie wilt ontvangen over de uitkomsten van het onderzoek, vul dan hier je e-mail adres in:

Appendix B - Informed Consent Control Group

Beste student,

Wie ben ik?

Mijn naam is Anniek de Kort, masterstudent Educational Sciences aan de Universiteit Utrecht. Ik doe onderzoek in het kader van mijn master thesis, onder begeleiding van Femke Kirschner.

Wat is het doel van het onderzoek?

Mijn onderzoek gaat over de cognitive load die studenten ervaren tijdens het volgen van een cursus.

Wat houdt het onderzoek in?

Je zal wekelijks een (korte) vragenlijst van ongeveer vijf minuten invullen. Met deze vragenlijst zal ik meten wat de ervaren cognitive load is van afgelopen week. De vragenlijst zal aan het einde van iedere week aan je gegeven worden.

Privacy en vertrouwelijkheid

Alle gegevens worden vertrouwelijk behandeld en anoniem verwerkt. Dat wil zeggen dat straks in de uitkomsten van het onderzoek niet te zien is welke antwoorden jij gegeven hebt. De gegevens worden alleen voor onderzoeks- en opleidingsdoeleinden gebruikt. Jouw docent krijgt jouw vragenlijsten niet te zien.

Mogelijkheid tot vragen, informatie en toestemming

Als je nog vragen hebt over het onderzoek, stel die dan nu aan één van de onderzoekers of stuur een mail aan: Anniek de Kort, A.I.M.dekort@uu.nl. Voor verdere vragen over de cursus en opdracht die wij maken kun je contact opnemen met: Femke Kirschner, F.C.Kirschner@uu.nl

Als je mee wilt doen aan het onderzoek, **vul dan het formulier op de achterzijde in.** Je kunt het formulier inleveren bij een van de onderzoekers.

Met vriendelijke groet,

Dr. Femke Kirschner, Anniek de Kort

TOESTEMMINGSVERKLARING voor deelname aan wetenschappelijk onderzoek

Ik heb uitleg gekregen over het onderzoek. De informatie over het onderzoek heb ik goed gelezen. Ik heb mijn eventuele vragen over het onderzoek gesteld. Ik heb goed nagedacht over of ik aan het onderzoek wil deelnemen. Ik mag op ieder moment stoppen met het onderzoek als ik dat wil. Ik hoef niet uit te leggen waarom ik wil stoppen.

Zet hiernaast een kruisje in het vakje dat voor jou van toepassing is

- Ik doe wel mee aan het onderzoek

- Ik doe niet mee aan het onderzoek

Naam :

Werkgroep :

Geboortedatum :

Datum :

Handtekening :

Als je informatie wilt ontvangen over de uitkomsten van het onderzoek, vul dan hier je e-mail adres in:



Appendix C - Questionnaire Experimental Group

Name _____

Date of birth: ___/___/____ (DD/MM/YYYY)

Gender: M / W

Please circle your answer

[1] I determine how to solve a problem before I begin.

1 2 3 4

almost never

almost always

[2] I think through in my mind the steps of a plan I have to follow.

1 2 3 4

almost never

almost always

[3] I ask myself questions about what a problem requires me to do to solve it, before I do it.

1 2 3 4

almost never

almost always

[4] I imagine the parts of a problem I still have to complete.

1 2 3 4

almost never

almost always

[5] I carefully plan my course of action to solve a problem.

1 2 3 4

almost never

almost always

[6] I figure out my goals and what I need to do to accomplish them.

1 2 3 4

almost never

almost always

[7] I clearly plan my course of action to solve a problem.

1 2 3 4

almost never

almost always

[8] I develop a plan for the solution of a problem

[16] The activity really enhanced my knowledge and understanding of statistics.

0 1 2 3 4 5 6 7 8 9 10

Not at all the case

completely the case

[17] The activity really enhanced my understanding of the formulas covered.

0 1 2 3 4 5 6 7 8 9 10

Not at all the case

completely the case

[18] The activity really enhanced my understanding of concepts and definitions.

0 1 2 3 4 5 6 7 8 9 10

Not at all the case

completely the case

Appendix D - Questionnaire Control Group

Name _____

Date of birth: ___/___/____ (DD/MM/YYYY)

Gender: M / W

Please circle your answer

All of the following questions refer to the activities from last week (week --) that just finished. Please respond to each of the questions on the following scale (0 meaning not at all the case and 10 meaning completely the case).

[1] The topic/topics covered in the activity was/were very complex.

0 1 2 3 4 5 6 7 8 9 10

Not at all the case

completely the case

[2] The activity covered formulas that I perceived as very complex.

0 1 2 3 4 5 6 7 8 9 10

Not at all the case

completely the case

[3] The activity covered concepts and definitions that I perceived as very complex.

0 1 2 3 4 5 6 7 8 9 10

Not at all the case

completely the case

[4] The instructions and/or explanations during the activity were very unclear.

0 1 2 3 4 5 6 7 8 9 10

Not at all the case

completely the case

[5] The instructions and/or explanations were, in terms of learning, very ineffective.

0 1 2 3 4 5 6 7 8 9 10

Not at all the case

completely the case

[6] The instructions and/or explanations were full of unclear language.

0 1 2 3 4 5 6 7 8 9 10

Not at all the case

completely the case

[7] The activity really enhanced my understanding of the topic(s) covered.

0 1 2 3 4 5 6 7 8 9 10

Not at all the case completely the case

[8] The activity really enhanced my knowledge and understanding of statistics.

0 1 2 3 4 5 6 7 8 9 10

Not at all the case completely the case

[9] The activity really enhanced my understanding of the formulas covered.

0 1 2 3 4 5 6 7 8 9 10

Not at all the case completely the case

[10] The activity really enhanced my understanding of concepts and definitions.

0 1 2 3 4 5 6 7 8 9 10

Not at all the case completely the case

Appendix E - Planning Format

	Goal setting
What is/are your goal(s) in this course?	
How do you plan to achieve these goal(s)?	
	Strategic Planning
How much time do you have for this course?	
How are you planning on dividing your time for this course?	
	Self-efficacy
Which skills do you need to have to achieve your goals?	
Which skills do you already master?	
Which skills do you not yet master?	
What will you do when you have questions regarding these skills?	
	Selection of strategies and methods
Which strategies do you need to have to achieve your goals?	
Which methods do you need to have to achieve your goals?	
What will you do when you have questions regarding these strategies or methods?	

Week	What has to be done?	How do I achieve that?
1		
2		
3		
4		
5		
6		
7		
8		

Appendix F – References

- Bergman, E. M., de Bruin, A. B., Vorstenbosch, M. A., Kooloos, J. G., Puts, G. C., Leppink, J., & van der Vleuten, C. P. (2015). Effects of learning content in context on knowledge acquisition and recall: a pretest-posttest control group design. *BMC medical education*, 15(1), 133. doi:10.1186/s12909-015-0416-0
- Bol, L., Campbell, K. D., Perez, T., & Yen, C. J. (2016). The effects of self-regulated learning training on community college students' metacognition and achievement in developmental math courses. *Community College Journal of Research and Practice*, 40(6), 480-495. doi:10.1080/10668926.2015.1068718
- Choi, H. H., Van Merriënboer, J. J., & Paas, F. (2014). Effects of the physical environment on cognitive load and learning: towards a new model of cognitive load. *Educational Psychology Review*, 26(2), 225-244. doi:10.1007/s10648-014-9262-6
- De Bruin, A. B., & Merriënboer, van J. J. (2017). Bridging cognitive load and self-regulated learning research: A complementary approach to contemporary issues in educational research. *Learning and Instruction*, 51, 1-9.
doi:<https://doi.org/10.1016/j.learninstruc.2017.06.001>
- Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, 41, 1149-1160. doi:10.3758/BRM.41.4.1149
- Flavell, J.H. (1978). Metacognitive development. In J.M. Scandura & C.J. Brainerd, eds., *Structural/Process theories of Complex Human Behaviour* (pp. 213–245). The Netherlands: Sijthoff and Noordhoff.
- Glogger-Frey, I., Gaus, K., & Renkl, A. (2017). Learning from direct instruction: Best prepared by several self-regulated or guided invention activities?. *Learning and Instruction*, 51, 26-35. doi:<https://doi.org/10.1016/j.learninstruc.2016.11.002>
- Goštautaitė, B. (2014). Age, work engagement and individual work performance: the influence of work design (Doctoral dissertation, ISM University of Management and Economics).
- Haji, F. A., Rojas, D., Childs, R., de Ribaupierre, S., & Dubrowski, A. (2015). Measuring cognitive load: performance, mental effort and simulation task complexity. *Medical education*, 49(8), 815-827. doi:10.1111/medu.12773
- Hattie, J., Biggs, J., & Purdie, N. (1996). Effects of learning skills interventions on student

- learning: A meta-analysis. *Review of educational research*, 66(2), 99-136.
doi:<https://doi.org/10.3102/00346543066002099>
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of educational research*, 77(1), 81-112. doi:<https://doi.org/10.3102/003465430298487>
- Heikkilä, A., & Lonka, K. (2006). Studying in higher education: students' approaches to learning, self-regulation, and cognitive strategies. *Studies in higher education*, 31(1), 99-117. doi:<https://doi.org/10.1080/03075070500392433>
- Hofer, B. K., & Yu, S. L. (2003). Teaching self-regulated learning through a "Learning to Learn" course. *Teaching of Psychology*, 30(1), 30-33.
doi:10.1207/S15328023TOP3001_05
- Jelsma, O., & van Merriënboer, J. J. (1990). The ADAPT design model: Towards instructional control of transfer. *Instructional Science*, 19(2), 89-120.
doi:<https://doi.org/10.1007/BF00120698>
- Kalyuga, S., Ayres, P., Chandler, P., & Sweller, J. (2003). Expertise reversal effect. *Educational Psychologist*, 38, 23–31. doi:10.4018/978-1-60566-048-6.ch003
- Kirschner, F., Kester, L., & Corbalan, G. (2010). Cognitive load theory and multimedia learning, task characteristics, and learning engagement: The current state of the art. doi:10.1016/j.chb.2010.05.003
- Kizilcec, R. F., Pérez-Sanagustín, M., & Maldonado, J. J. (2016, April). Recommending self-regulated learning strategies does not improve performance in a MOOC. In *Proceedings of the Third (2016) ACM Conference on Learning@ Scale* (pp. 101-104). ACM. doi:10.1145/2876034.2893378
- Leppink, J., Paas, F., Van der Vleuten, C. P., Van Gog, T., & Van Merriënboer, J. J. (2013). Development of an instrument for measuring different types of cognitive load. *Behavior research methods*, 45(4), 1058-1072.
doi:<https://doi.org/10.3758/s13428-013-0334-1>
- Leppink, J., Paas, F., Van Gog, T., van Der Vleuten, C. P., & Van Merriënboer, J. J. (2014). Effects of pairs of problems and examples on task performance and different types of cognitive load. *Learning and Instruction*, 30, 32-42.
doi:<https://doi.org/10.1016/j.learninstruc.2013.12.001>
- Paas, F., Tuovinen, J. E., Tabbers, H., & Van Gerven, P. W. (2003). Cognitive load measurement as a means to advance cognitive load theory. *Educational psychologist*,

- 38(1), 63-71. doi:10.1207/S15326985EP3801_8
- Panadero, E., Jonsson, A., & Botella, J. (2017). Effects of self-assessment on self-regulated learning and self-efficacy: Four meta-analyses. *Educational Research Review*, 22, 74-98. doi:<https://doi.org/10.1016/j.edurev.2017.08.004>
- Pintrich, P. R., & De Groot, E. V. (1990). Motivational and self-regulated learning components of classroom academic performance. *Journal of educational psychology*, 82(1), 33-40.
- Raaijmakers, S. F., Baars, M., Paas, F., van Merriënboer, J. J., & Van Gog, T. (2018). Training self-assessment and task-selection skills to foster self-regulated learning: Do trained skills transfer across domains?. *Applied cognitive psychology*, 32(2), 270-277. doi:<https://doi.org/10.1002/acp.3392>
- Raaijmakers, S. F., Baars, M., Schaap, L., Paas, F., & Van Gog, T. (2017). Effects of performance feedback valence on perceptions of invested mental effort. *Learning and Instruction*, 51, 36-46. doi:<https://doi.org/10.1016/j.learninstruc.2016.12.002>
- van Merriënboer, J.J.G. & Sweller, J. (2005). Cognitive load theory and complex learning: Recent developments and future directions. *Educational Psychology Review*, 17, 147–177. doi:10.1007/s10648-005-3951-0
- Schmitz, B., & Wiese, B. S. (2006). New perspectives for the evaluation of training sessions in self-regulated learning: Time-series analyses of diary data. *Contemporary educational psychology*, 31(1), 64-96. doi:10.1016/j.cedpsych.2005.02.002
- Sweller, J. (2011). Cognitive load theory. In *Psychology of learning and motivation* (Vol. 55, pp. 37-76). Academic Press. doi:10.1007/978-1-4419-8126-4
- Sweller, J., Van Merriënboer, J. J., & Paas, F. G. (1998). Cognitive architecture and instructional design. *Educational psychology review*, 10(3), 251-296.
- Toering, T., Elferink-Gemser, M. T., Jonker, L., van Heuvelen, M. J., & Visscher, C. (2012). Measuring self-regulation in a learning context: Reliability and validity of the Self-Regulation of Learning Self-Report Scale (SRL-SRS). *International Journal of Sport and Exercise Psychology*, 10(1), 24-38. doi:10.1002/acp.1726
- Tuckman, B. W., & Kennedy, G. J. (2011). Teaching learning strategies to increase the success of first-term college students. *The Journal of Experimental Education*, 79, 478–504. doi:10.1080/00220973.2010.512318
- Wibrowski, C. R., Matthews, W. K., & Kitsantas, A. (2017). The role of a skills learning

support program on first-generation college students' self-regulation, motivation, and academic achievement: A longitudinal study. *Journal of College Student Retention: Research, Theory & Practice*, 19(3), 317-332. doi:10.1177/1521025116629152

Zimmerman, B. J., & Pons, M. M. (1986). Development of a structured interview for assessing student use of self-regulated learning strategies. *American educational research journal*, 23(4), 614-628.

Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. *Theory into practice*, 41(2), 64-70. doi: 10.1207/s15430421tip4102_2