

### **Universiteit Leiden**

### **Opleiding Informatica**

Learning software design: Is abstraction ability key?

Name: Date:

Claire E. Stevenson 20/03/2015

1<sup>st</sup> supervisor: Prof. dr. Michel R.V. Chaudron 2<sup>nd</sup> supervisor: Dave Stikkelorum, MSc.

**BACHELOR THESIS** Leiden Institute of Advanced Computer Science (LIACS) Leiden University Niels Bohrweg 1 2333 CA Leiden The Netherlands

#### Abstract

This study examined the role of domain knowledge and abstract reasoning ability on undergraduate computer science students' software design learning and ability using the LIACS software design test. The study comprised a pretestposttest design. Participants (N=151) were administered pretests that measured (1) software design skills on topics separation of concerns, cohesion & coupling, maintenance, reuse and dependency; (2) domain knowledge (UML); and (3) abstract reasoning ability respectively. During posttest the LIACS software design test was administerd again; this was conducted after the students followed an undergraduate software engineering course. We found that inital software design ability was related to both abstract reasoning ability and domain knowledge; however, improvement in software design ability was somewhat related to domain knowledge, but not to abstract reasoning skills. More specifically with regard to the topics measured we found that *maintenance* and reuse were the most difficult pretest topics, but also showed the greatest improvement from pretest to posttest. In the future administering the LIACS software design test may be helpful for educators so that content can be tailored to the knowledge and skills the students already possess as well as evaluate how effective teaching of particular topics has been. Future research should investigate whether increasing students' knowledge of UML improves their chances of learning from a software design course.

#### 1. Introduction

Software plays an essential role in our daily lives; from cell phones to kitchen appliances, and from furniture production to writing a paper. In order to create excellent software it must be well designed and good software designers are needed to avoid the high costs of software maintenance and error fixing. Software design "aims at the description of the basic features of the future computer system and prescribes the functions the system should perform" (p. 373)[1]. The question this paper explores is which factors influence ability and learning of software design in novices.

The two cognitive factors we investigated were fluid intelligence (i.e., abstract reasoning) and domain knowledge, both of which are related to problem solving skills and the development of expertise in a plethora of domains such as chess, computer programming or knowledge of baseball [2]–[5]. As Siau & Tan (2005) indicate in their review of human cognition factors that play a role in conceptual modeling for software design, domain knowledge and abstract reasoning ability are considered important aspects that play a role at level of the individual software designer[6]. However, this has yet to be investigated in the domain of software design ability. Evidence is provided by Leung & Bollejou (2006) based on their empirical study of design errors of novice analysts; they defined a set of frequent (syntactic and semantic) errors that were thought to be due deficits in one or more cognitive factors[7]. For example, syntactic errors could be caused by insufficient domain knowledge of UML notation whereas semantic errors may be due to inaccuracies in abstract reasoning and/or working memory constraints.

Kramer (2007) poses the question why some software engineers are able to design clear and eloquent systems while others, with similar education and conceptual modeling tools, are not; he argues that abstraction ability is key[8]. Abstraction is the ability to solve novel problems often measured by intelligence tests such as the Raven Progressive Matrices (RPM) that require one to complete visual patterns by inducing abstract rules. In research on the development of expertise abstract reasoning ability is said to plays an important role with regard to a person's initial skill level and is also required for the first cognitive phase of skill acquisition[9]. Working memory, defined as the ability to process, store and retrieve information, also appears to be a factor predicting the acquisition of software development skills[6], [10]–[12]. However, working memory is strongly linked to abstract reasoning ability and some researchers argue that these two psychological constructs may be one and the same [13]–[15]. Given the similarities between working memory and abstract reasoning as well as the advice of Bergersen and colleagues (2011) to include a more pure measure of g(i.e., general intelligence often indictaed a abstract reasoning ability) such as the Raven's Advanced Progressive Matrices (RAPM) to examine person factors influencing software development skills, we chose to use only RAPM in this study.

Domain knowledge is considered a main driving force behind skill acquisition - often referred to as the 'knowledge is power' hypothesis[2]. A few studies have found a relationship between software programming skills and prior experience[4], [16]. Bergersen et al. (2011) determined that domain knowledge was the underlying (i.e. mediating) source in this relationship - where people with more experience inherently have greater domain knowledge and thus perform better on software programming measures[12]. Indeed, domain knowledge showed a strong relationship with programming skills.

In the present study we examined how domain knowledge and abstract reasoning ability relate to students software design ability before and after a software engineering course. UML (Unified Modeling Language) is the current standard for visual abstract software design. For this reason we specifically focus on UML design ability and use the LIACS UML design skills test developed by Stikkelorum and colleagues (2013)[17]. By analyzing students performance and change on five item categories using statistical modeling techniques from item response theory (IRT)[18] we aimed to answer questions about the role of domain knowledge and abstract reasoning ability on: (1) initial performance on a software design test and (2) learning from a course on software design as measured by improvement on the LIACS software design test. In short, is abstract reasoning ability really the key predictor of which students can learn to design software or is "knowledge is power" and thus domain knowledge the driving force? Based on prior research in expertise development we expect abstract reasoning ability to play the largest role in novice software designer's initial ability, whereas domain knowledge is expected to predict both initial ability and performance change.

#### 2. Method

#### 2.1. Participants

This study included 274 first year software engineering students (93% male) recruited from two university courses; one in Gothenburg, Sweden and the other in Utrecht, The Netherlands. Participants agreed to participation through an online informed consent form prior to completing the pretest. 248 students completed all pretest assessments and 151 students completed the posttest assessment.

#### 2.2. Design & Procedure

The study comprised a pretest-posttest design. Each participant was administered three pretest assessments that measured (1) design skills, (2) domain knowledge and (3) abstract reasoning ability respectively. The posttest was conducted after the students followed a freshman undergraduate software engineering course at his/her university. The posttest consisted of the same software design skills task (assessment 1) administered at pretest.

#### 2.3. Materials

#### 2.3.1. LIACS Software Design Skills test

The LIACS Software Design test utilized in this study comprised 20 multiplechoice questions assessing the understanding of software design concepts[19]. The main concepts measured were design principles (separation of concerns, cohesion & coupling), quality (maintainability, reuse) and architecture (dependency); see Table 1 for an overview of which concepts were assessed per item. The items generally consisted of an example design and a question about consequences of changes in the design and/or a comparison of the quality two or more designs; see Figure 1 for an example. A complete list of the items including solutions and main measured concepts can be found in Appendix A. The designs were presented in Unified Modeling Language (UML<sup>1</sup>); this popular modeling language was chosen as it is used in the students' courses on software design.

Item	Separation of	<b>Cohesion &amp;</b>	Maintenance	Reuse	Dependency
	concerns	coupling			
1				Х	Х
2					Х
3					Х
4					Х
5					Х
6			Х		
7					Х
8	Х				
9		Х			
10	Х				
11		Х	Х		
12		Х			Х
13			Х		
14				Х	
15	Х	Х			
16			Х	Х	
17	Х		Х		
18				Х	
19			Х		
20	Х			Х	

Table 1.Concepts measured per item from LIACS Design Skills test.

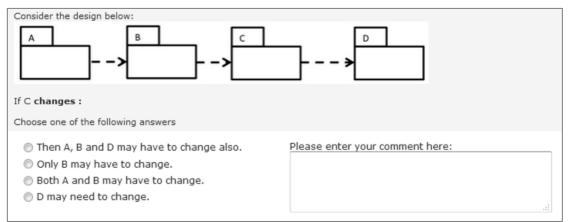


Figure 1. Example item from LIACS Design Skills test.

The initial version of this test comprised 10 items that were evaluated by experts according to criteria of clarity, complexity and difficulty level and piloted in a

<sup>&</sup>lt;sup>1</sup> http://www.uml.org

population of 60 first year university computer science students; a quantitative validation of the test items can be found in Stikkelorum et al. (2013)[17]. These items formed the basis for the items used in the present study. Additional items were created and rated by experts. A quantitative evaluation of the psychometric quality of the version of the items used in the present study can be found in section 3.1 of this paper.

#### 2.3.2. UML knowledge quiz

A UML knowledge quiz was created specifically for this experiment in order to assess domain knowledge that formed a prerequisite for understanding and correctly answering the diagrams used in the LIACS Software Design Skills test. The quiz consisted of 22 items focusing on the syntax and semantics of the Unified Modeling Language (UML). See Figure 2 for an example; the actual items can be found in Appendix B.

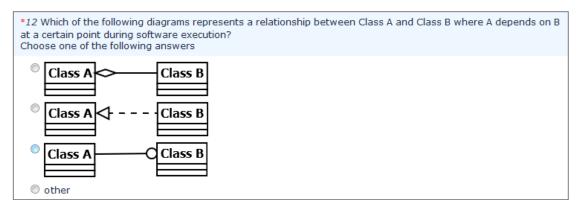


Figure 2. Example item from the UML quiz used to measure domain knowledge.

#### 2.3.3. Abstract Reasoning

A computerized short-form of the original Raven's Advanced Progressive Matrices[20] was administered to measure non-verbal abstract reasoning. This particular short form comprised the even items extracted from the original 36item RAPM. The RAPM is a visuospatial reasoning task often used to measure fluid intelligence and requires participants to induce abstract rules and relations between geometric figures presented in a 3x3 matrix to complete a visual pattern by choosing among nine multiple-choice alternatives. Short forms of the RAPM are used to reduce administration time without losing psychometric properties or predictive validity[21].

#### 3. Results

# 3.1. Psychometric properties of software design and domain knowledge measures

Prior to examine the role of cognitive factors on the students' performance on the LIACS software design test and UML domain knowledge quiz the psychometric properties of these two new instruments were evaluated as measurement scales using both classical test theory (CTT) and item response (IRT) theory[22]. Both methods are used to evaluate test scores and item responses. CTT is more commonly utilized as the metrics are relatively easy to compute; however, IRT has distinct advantages such as being population independent so that item difficulty and discrimination can be evaluated without bias - important in the present situation as the population is rather specific with limited differences in software design ability.

According to CTT item difficulty is determined by the proportion correct responses per item. Item discrimination represents how sensitive an item is to actual differences between test-takers and is thus a means of examining an item's impact on the internal consistency of a test. In CTT item discrimination is represented by the point-biserial correlation between performance on a specific item and the test as a whole. In IRT item difficulty and discrimination can be computed using the two-parameter logistic (2PL) IRT model. In this IRT model the chance that an item is solved correctly (i.e., y = 1) depends on the difference between the latent ability ( $\theta$ ) of the test-taker p and the difficulty ( $\beta$ ) and discrimination ( $\alpha$ ) of item i:

$$P(Y_{pi} = 1 | \theta_p, \beta_i, \alpha_i) = \frac{e^{(\alpha_i(\theta_p - \beta_i))}}{1 + e^{(\alpha_i(\theta_p - \beta_i))}}$$
  
where  $\theta_p \sim N(0, \sigma_{\theta^2})$ ,  $\beta_i \sim N(0, \sigma_{\beta^2})$ , and  $\alpha_i \sim N(0, \sigma_{\alpha^2})$  (1)

In this section we evaluate item difficulty and discrimination with both CTT and IRT; furthermore, reliability and construct validity are addressed from the CTT perspective.

#### 3.2.1. LIACS Software Design Skills

Item difficulty (CTT, proportion correct) ranged from .19 to .80 (M=.58, SD=.17) for the pretest and from .25 to .87 (M=.67, SD=.16) for the posttest; this indicates that there was likely some improvement from pretest to posttest and that the items were generally solved correctly at above chance level (> .25 on items with 4 multiple-choice options). Ideally the 2PL IRT item difficulty parameters represent a range of difficulties so that the estimation of ability is accurate for persons with a range of latent aptitude (see Table 2).

$\begin{array}{ c c c c c c c c c c c c c c c c c c c$	Psychometric item properties of LIACS software design test.								
$\begin{array}{c c c c c c c c c c c c c c c c c c c $	Pretest					Postte	st		
$\begin{array}{c c c c c c c c c c c c c c c c c c c $		CTT		IRT		СТТ		IRT	
$\begin{array}{c ccccccccccccccccccccccccccccccccccc$	item	-		β	α	-		β	α
3.63.23-0.66 $0.85$ .70.07-4.39 $0.19$ 4.69.42-0.27 $4.16$ .57.35 $0.55$ $0.62$ 5.71.19-0.56 $10.75$ .74.31-1.31 $0.90$ 6.57.31-0.67 $0.41$ .86.36-1.77 $1.32$ 7.63.18-2.81 $0.28$ .42.25-0.31 $1.07$ 8.80.20-6.48 $0.18$ .62.16-1.41 $0.89$ 9.61.25 $0.53$ $0.56$ .85.20-3.85 $0.51$ 10.74.32 $3.00$ $0.15$ .54.16-1.00 $1.41$ 11.68.24-2.61 $0.29$ .70.25 $5.30$ $0.21$ 12.65.23 $1.18$ $0.30$ .78.37-1.19 $1.36$ 13.44.17-6.23-0.23.87.17-1.17 $1.39$ 14.77.15-2.97 $0.48$ .75.32-1.40 $0.37$ 15.68.07-0.47 $0.30$ .25.10-0.721.6116.39.18-1.74 $0.64$ .74.36-0.33 $0.46$ 17.42.10-1.03 $0.43$ .77.37-3.07 $0.62$	1		.21	0.81	0.73		.26	0.08	0.63
4.69.42 $-0.27$ 4.16.57.350.550.625.71.19 $-0.56$ 10.75.74.31 $-1.31$ 0.906.57.31 $-0.67$ 0.41.86.36 $-1.77$ 1.327.63.18 $-2.81$ 0.28.42.25 $-0.31$ 1.078.80.20 $-6.48$ 0.18.62.16 $-1.41$ 0.899.61.250.530.56.85.20 $-3.85$ 0.5110.74.323.000.15.54.16 $-1.00$ 1.4111.68.24 $-2.61$ 0.29.70.255.300.2112.65.231.180.30.78.37 $-1.19$ 1.3613.44.17 $-6.23$ $-0.23$ .87.17 $-1.17$ 1.3914.77.15 $-2.97$ 0.48.75.32 $-1.40$ 0.3715.68.07 $-0.47$ 0.30.25.10 $-0.72$ 1.6116.39.18 $-1.74$ 0.64.74.36 $-0.33$ 0.4617.42 $10$ $-1.03$ 0.43.77.37 $-3.07$ 0.62	2	.72	.27	-1.25	0.83	.74	.39	-1.02	1.29
5.71.19 $-0.56$ $10.75$ .74.31 $-1.31$ $0.90$ 6.57.31 $-0.67$ $0.41$ .86.36 $-1.77$ $1.32$ 7.63.18 $-2.81$ $0.28$ .42.25 $-0.31$ $1.07$ 8.80.20 $-6.48$ $0.18$ .62.16 $-1.41$ $0.89$ 9.61.25 $0.53$ $0.56$ .85.20 $-3.85$ $0.51$ 10.74.32 $3.00$ $0.15$ .54.16 $-1.00$ $1.41$ 11.68.24 $-2.61$ $0.29$ .70.25 $5.30$ $0.21$ 12.65.23 $1.18$ $0.30$ .78.37 $-1.19$ $1.36$ 13.44.17 $-6.23$ $-0.23$ .87.17 $-1.17$ $1.39$ 14.77.15 $-2.97$ $0.48$ .75.32 $-1.40$ $0.37$ 15.68.07 $-0.47$ $0.30$ .25.10 $-0.72$ $1.61$ 16.39.18 $-1.74$ $0.64$ .74.36 $-0.33$ $0.46$ 17.42 $10$ $-1.03$ $0.43$ .77.37 $-3.07$ $0.62$	3	.63	.23	-0.66	0.85	.70	.07	-4.39	0.19
5.71.19 $-0.56$ $10.75$ .74.31 $-1.31$ $0.90$ 6.57.31 $-0.67$ $0.41$ .86.36 $-1.77$ $1.32$ 7.63.18 $-2.81$ $0.28$ .42.25 $-0.31$ $1.07$ 8.80.20 $-6.48$ $0.18$ .62.16 $-1.41$ $0.89$ 9.61.25 $0.53$ $0.56$ .85.20 $-3.85$ $0.51$ 10.74.32 $3.00$ $0.15$ .54.16 $-1.00$ $1.41$ 11.68.24 $-2.61$ $0.29$ .70.25 $5.30$ $0.21$ 12.65.23 $1.18$ $0.30$ .78.37 $-1.19$ $1.36$ 13.44.17 $-6.23$ $-0.23$ .87.17 $-1.17$ $1.39$ 14.77.15 $-2.97$ $0.48$ .75.32 $-1.40$ $0.37$ 15.68.07 $-0.47$ $0.30$ .25.10 $-0.72$ $1.61$ 16.39.18 $-1.74$ $0.64$ .74.36 $-0.33$ $0.46$ 17.42 $10$ $-1.03$ $0.43$ .77.37 $-3.07$ $0.62$	4	.69	.42	-0.27	4.16	.57	.35	0.55	0.62
7.63.18-2.81 $0.28$ .42.25-0.31 $1.07$ 8.80.20-6.48 $0.18$ .62.16-1.41 $0.89$ 9.61.25 $0.53$ $0.56$ .85.20-3.85 $0.51$ 10.74.32 $3.00$ $0.15$ .54.16-1.00 $1.41$ 11.68.24-2.61 $0.29$ .70.25 $5.30$ $0.21$ 12.65.23 $1.18$ $0.30$ .78.37-1.19 $1.36$ 13.44.17-6.23-0.23.87.17-1.17 $1.39$ 14.77.15-2.97 $0.48$ .75.32-1.40 $0.37$ 15.68.07-0.47 $0.30$ .25.10-0.721.6116.39.18-1.74 $0.64$ .74.36-0.33 $0.46$ 17.4210-1.03 $0.43$ .77.37-3.07 $0.62$	5	.71	.19	-0.56	10.75	.74	.31	-1.31	0.90
8.80.20 $-6.48$ $0.18$ .62.16 $-1.41$ $0.89$ 9.61.25 $0.53$ $0.56$ .85.20 $-3.85$ $0.51$ 10.74.32 $3.00$ $0.15$ .54.16 $-1.00$ $1.41$ 11.68.24 $-2.61$ $0.29$ .70.25 $5.30$ $0.21$ 12.65.23 $1.18$ $0.30$ .78.37 $-1.19$ $1.36$ 13.44.17 $-6.23$ $-0.23$ .87.17 $-1.17$ $1.39$ 14.77.15 $-2.97$ $0.48$ .75.32 $-1.40$ $0.37$ 15.68.07 $-0.47$ $0.30$ .25.10 $-0.72$ $1.61$ 16.39.18 $-1.74$ $0.64$ .74.36 $-0.33$ $0.46$ 17.42 $10$ $-1.03$ $0.43$ .77.37 $-3.07$ $0.62$	6	.57	.31	-0.67	0.41	.86	.36	-1.77	1.32
9       .61       .25       0.53       0.56       .85       .20       -3.85       0.51         10       .74       .32       3.00 <b>0.15</b> .54       .16       -1.00       1.41         11       .68       .24       -2.61       0.29       .70       .25       5.30       0.21         12       .65       .23       1.18       0.30       .78       .37       -1.19       1.36         13       .44       .17       -6.23       -0.23       .87       .17       -1.17       1.39         14       .77       .15       -2.97       0.48       .75       .32       -1.40       0.37         15       .68       .07       -0.47       0.30       .25       .10       -0.72       1.61         16       .39       .18       -1.74       0.64       .74       .36       -0.33       0.46         17       .42      10       -1.03       0.43       .77       .37       -3.07       0.62	7	.63	.18	-2.81	0.28	.42	.25	-0.31	1.07
10.74.32 $3.00$ $0.15$ .54.16-1.00 $1.41$ $11$ .68.24-2.61 $0.29$ .70.25 $5.30$ $0.21$ $12$ .65.23 $1.18$ $0.30$ .78.37-1.19 $1.36$ $13$ .44.17-6.23-0.23.87.17-1.17 $1.39$ $14$ .77.15-2.97 $0.48$ .75.32-1.40 $0.37$ $15$ .68.07-0.47 $0.30$ .25.10-0.721.61 $16$ .39.18-1.74 $0.64$ .74.36-0.33 $0.46$ $17$ .4210-1.03 $0.43$ .77.37-3.07 $0.62$	8	.80	.20	-6.48	0.18	.62	.16	-1.41	0.89
11.68.24-2.61 $0.29$ .70.25 $5.30$ $0.21$ $12$ .65.23 $1.18$ $0.30$ .78.37 $-1.19$ $1.36$ $13$ .44.17 $-6.23$ $-0.23$ .87.17 $-1.17$ $1.39$ $14$ .77.15 $-2.97$ $0.48$ .75.32 $-1.40$ $0.37$ $15$ .68.07 $-0.47$ $0.30$ .25.10 $-0.72$ $1.61$ $16$ .39.18 $-1.74$ $0.64$ .74.36 $-0.33$ $0.46$ $17$ .42 $10$ $-1.03$ $0.43$ .77.37 $-3.07$ $0.62$	9	.61	.25	0.53	0.56	.85	.20	-3.85	0.51
$\begin{array}{cccccccccccccccccccccccccccccccccccc$	10	.74	.32	3.00	0.15	.54	.16	-1.00	1.41
13.44.17-6.23-0.23.87.17-1.171.3914.77.15-2.970.48.75.32-1.400.3715.68.07-0.470.30.25.10-0.721.6116.39.18-1.740.64.74.36-0.330.4617.4210-1.030.43.77.37-3.070.62	11	.68	.24	-2.61	0.29	.70	.25	5.30	0.21
14.77.15-2.970.48.75.32-1.400.3715.68.07-0.470.30.25.10-0.721.6116.39.18-1.740.64.74.36-0.330.4617.4210-1.030.43.77.37-3.070.62	12	.65	.23	1.18	0.30	.78	.37	-1.19	1.36
15.68.07-0.470.30.25.10-0.721.6116.39.18-1.740.64.74.36-0.330.4617.4210-1.030.43.77.37-3.070.62	13	.44	.17	-6.23	-0.23	.87	.17	-1.17	1.39
16.39.18-1.740.64.74.36-0.330.4617.4210-1.030.43.77.37-3.070.62	14	.77	.15	-2.97	0.48	.75	.32	-1.40	0.37
17 .42 <b>10</b> -1.03 0.43 .77 .37 -3.07 0.62	15	.68	.07	-0.47	0.30	.25	.10	-0.72	1.61
	16	.39	.18	-1.74	0.64	.74	.36	-0.33	0.46
18 .19 <b>.16</b> -1.29 0.62 .77 .34 -1.53 0.60	17	.42	10	-1.03	0.43	.77	.37	-3.07	0.62
		.19		-1.29	0.62	.77	.34	-1.53	0.60
19 .54 <b>.05</b> -1.18 0.52 .70 .41 -1.36 1.16				-1.18	0.52				1.16
<u>20</u> .30 .21 -8.60 -0.10 .50 .14 -0.03 0.37	-		.21	-8.60	-0.10	.50	.14	-0.03	0.37

Table 2.

Psychometric item properties of LIACS software design test.

\* p<.05, \*\* p<.01

Item discrimination for CTT is represented by the item-total correlation. A rule-of-thumb in psychometrics is to discard items with an item-total correlation of less than .20. As can be seen in Table 2 nine items in the pretest and six items in the posttest meet this criterion. In the 2PL IRT model the item discrimination index is related to the range of test scores for which the item discriminates best between test-takers; higher values indicate greater discrimination and those

near zero ( $\alpha$ <|.20|) yield little to no information on the test-taker's ability and should be discarded (see Table 2). Based on these values it is advisable to revise items 8, 10 and 20 from the pretest and item 3 from the posttest. It is important to note for analyses presented in section 3.2.2 that items 8, 10 and 20 each measure understanding of the concept *separation of concerns*.

Reliability, i.e. "...the extent to which differences in respondents' observed scores are consistent with differences in their true scores" (Furr & Bacharach, 2014, p. 103), was determined using Cronbach's alpha coefficient of internal consistency; for the 20 item scale at pretest this was  $\alpha$ =.60 and at posttest  $\alpha$ =.69. These reliabilities are considered acceptable[23].

Construct validity, i.e. "...the degree to which test scores can be interpreted as reflecting a particular psychological construct" (Furr & Bacharach, 2014, p. 201), was determined by computing the correlation between posttest IRT scale scores and the students' scores on the software design course final exams. The resulting Pearson's (r=.26, p=.06) for the students from Utrecht who completed both the posttest and the final exam (N=55) is in the expected positive direction and considered weak; this provides some evidence that similar constructs were measured by the LIACS software design test and the respective course exams.

#### 3.2.2. UML Domain Knowledge Quiz

Item difficulty ranged from .00 to .77 (M=.37, SD=.21). Reliability based on Cronbach's alpha coefficient of internal consistency was  $\alpha$ =.69 and is considered good[23]. Table 3 provides an overview of the CTT and IRT item properties. Based on these results it is advisable to remove or revise item 13 as no one solved this correctly and also to revise items 1, 5-7, 11 and 20 given the low item-total correlations.

	СТТ		IRT	
itom	<i>p</i> -value	Item-total	Q	α*
item	<i>p</i> -value	correlation	β	u
1	.08	.02	2.34	1.25
2	.64	.56	-0.54	1.25
3	.68	.58	-0.73	1.25
4	.47	.40	0.18	1.25
5	.18	.15	1.52	1.25
6	.18	.08	1.54	1.25
7	.23	.15	1.23	1.25
8	.17	.22	1.60	1.25
9	.38	.40	0.58	1.25
10	.29	.30	0.93	1.25
11	.15	.19	1.76	1.25
12	.17	.20	1.62	1.25
13	.00	.00	10.07	1.25
14	.77	.57	-1.19	1.25
15	.40	.41	0.48	1.25
16	.52	.46	-0.03	1.25
17	.39	.40	0.51	1.25
18	.59	.59	-0.32	1.25
19	.50	.49	0.07	1.25
20	.13	.05	1.88	1.25
21	.60	.58	-0.38	1.25
22	.29	.22	0.95	1.25

Table 3. Psychometric item properties of UML domain knowledge quiz.

\* The best IRT model for the UML quiz is a special case of the 2PL model, i.e. the Rasch (1PL) model with the same  $\alpha$  value for all items.

#### 3.2. Performance on the LIACS software design test

This was investigated using explanatory IRT[18] analyses where item responses  $(Y_{pi})$  for person p on item i on the LIACS Software design ability test was the dependent variable and j variables were evaluated as predictors (X) using the following regression model:

$$P(Y_{pi} = 1 | X_{pij}, \beta_i) = \frac{e^{(\sum_{j=1}^{J} B_j X_{pij} + \theta_p + \beta_i)}}{1 + e^{(\sum_{j=1}^{J} B_j X_{pij} + \theta_p + \beta_i)}}$$

where  $\theta_p \sim N(0, \sigma_{\theta^2})$  and  $\beta_i \sim N(0, \sigma_{\beta^2})$  (2)

The resulting regression weights (B) for the person predictors (abstract reasoning, domain knowledge) and item predictors (measured concepts: separation of concerns, cohesion & coupling, maintenance, reuse, dependency)

are shown in Table 4. The results of this model are discussed per preditor type (persons, items) in the following two sections.

#### Table 4.

Results of the explanatory item response model evaluating item predictors (measured concepts: separation of concerns, cohesion & coupling, maintenance, reuse, dependency) and person predictors (abstract reasoning, domain knowledge, pretest-to-posttest change). Main item effects were evaluated with the pretest as reference time point.

Predictor	В	SE	р	lower bound	upper bound		
Intercept (overall mean)	-0.02	0.41	.97	-0.42	0.39		
Main effects item predictors							
separation of concerns	-0.23	0.30	.44	-0.53	0.07		
cohesion & coupling	-0.04	0.28	.88	-0.32	0.23		
maintenance	-0.71	0.30	.02	-1.02	-0.41		
reuse	-1.23	0.28	<.001	-1.50	-0.95		
dependency	-0.17	0.33	.60	-0.50	0.16		
Main effects person predictors							
post-pre	0.23	0.36	.53	-0.13	0.59		
abstract reasoning	0.08	0.02	<.001	0.06	0.10		
domain knowledge	0.23	0.06	<.001	0.17	0.29		
Interaction effects item predictors v	vith prea	test-to-	posttest	change			
separation of concerns X post-pre	-1.02	0.19	<.001	-1.20	-0.83		
cohesion & coupling X post-pre	-0.29	0.18	.11	-0.46	-0.11		
maintenance X post-pre	0.95	0.19	<.001	0.76	1.15		
reuse X post-pre	0.88	0.17	<.001	0.71	1.06		
dependency X post-pre	-0.54	0.21	.01	-0.75	-0.33		
Interaction effects person predictors with pretest-to-posttest change							
abstract reasoning X post-pre	0.02	0.02	.29	0.00	0.05		
domain knowledge X post-pre	-0.15	0.08	.06	-0.23	-0.07		

## 3.2.1. Role of abstract reasoning and domain knowledge in learning software design

The main research question was whether abstract reasoning and domain knowledge were able to predict initial ability and learning of software design. The results of the statistical model (see Table 4) showed that the chance of solving an item correctly generally improved from pretest to posttest for a student with average abstract reasoning skills and domain knowledge. Domain knowledge was a significant predictor of pretest performance (B=0.23, SE=0.06, p<.001) and a marginal predictor of improvement from pretest to posttest (B=-0.15, SE=0.08, p=.06); see Figure 3. Abstract reasoning was a significant predictor of pretest performance was a significant predictor of pretest performance (B=0.02, p<.001), but did not predict performance change over time (B=0.02, SE=0.02, p=.06); see Figure 4.

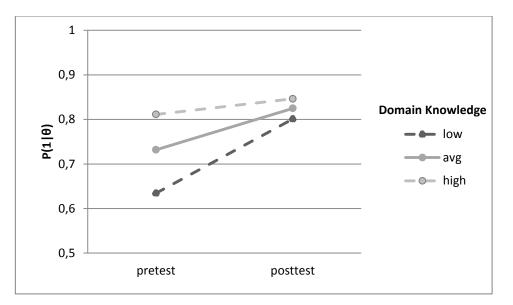


Figure 3. Domain knowledge (low = -2 SD, avg = Mean and high = +2 SD) significantly influenced the chance of solving LIACS software design test items correctly (y-axis) on the pretest and showed a marginal effect in predicting improvement from pretest to posttest (x-axis). Students with low domain knowledge generally performed less well on the software design test, whereas those with higher domain knowledge performed better. The gap between students with lower and higher domain knowledge decreased from pretest to posttest after following a software engineering course.

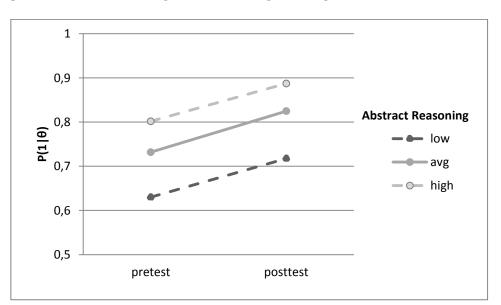


Figure 4. Abstract reasoning ability (low = -2 SD, avg = Mean and high = +2 SD) significantly influenced the chance of solving LIACS software design test items correctly (y-axis) on the pretest, but did not affect improvement from pretest to posttest (x-axis). Students with low abstract reasoning ability generally performed less well on the software design test, whereas those with higher abstract reasoning ability performed better. The differences between students of differing abstract reasoning ability did not change after following a software engineering course, i.e. abstract reasoning ability did not predict learning from before and after the course.

#### *3.2.2. Performance per measured concept*

The LIACS software design test contained items measuring five concepts: separation of concerns, cohesion & coupling, maintainability, reuse and dependency. The results are depicted in Figure 5. Items measuring understanding of the concepts *maintainability* and *reuse* were significantly more difficult than the average item (maintenance: B=-0.71, SE=0.30, p=.02; reuse: B=-1.23, SE=0.28, p<.001), whereas *separation of concerns, cohesion & coupling* and *dependency* were of average difficulty. Performance improvement from pretest to posttest was significant (p<.05) for items measuring understanding of the concepts *maintainability* and *reuse* (maintenance: B=0.95, SE=0.19, p=<.001; reuse: B=0.88, SE=0.17, p<.001); however, for *separation of concerns* and *dependency* there was a significant decline in performance from pretest to posttest (separation of concerns: B=-1.02, SE=0.19, p<.001; dependency: B=-0.54, SE=0.21, p<.01). There were no significant changes in performance on the items measuring understanding of *cohesion & coupling* (B=-0.29, SE=0.18, p=.11).

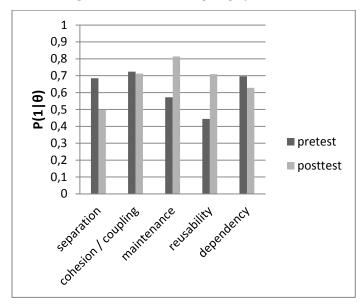


Figure 5. The chance of solving an item correctly (y-axis) that measured a particular concept (x-axis) on the LIACS software design test administered at pretest (light gray columns) and posttest (dark gray columns). Results show that items concerning the topics *maintenance* and *reuse* were more difficult on the pretest than items measuring understanding of the concepts *separation of concerns, cohesion & coupling* and *dependency*. Performance on items measuring understanding of *maintainability* and *reuse* improved, but performance on items measuring understanding of *separation of concerns* and *dependency* declined. There was no change in performance on items measuring *cohesion & coupling*. The presented model was computed with domain knowledge and abstract reasoning fixed at average skill levels.

#### 4. Discussion

The aim of the current study was to examine the role of domain knowledge and abstract reasoning ability on undergraduate computer science students' software design learning and ability. To this end, participants completed assessments of (1) software design ability in which comparisons between and/or judgements of changes to designs were assessed five concepts: separation of concerns, cohesion & coupling, maintainability, reuse and dependency; (2) domain knowledge of the syntax and semantics of the Unified Modeling Language (UML) commonly utilized in software design; and (3) abstract reasoning ability. The software design task was administered to first year computer science students before and after a software engineering course. There were four main findings: (1) inital software design ability was related to both abstract reasoning ability and domain knowledge; (2) improvement in software design ability was somewhat related to domain knowledge, but not to abstract reasoning skills; (3) the topics maintenance and reuse were more difficult on the pretest than separation of concerns, cohesion & coupling and dependency; and (4) understanding of the concepts maintainability and reuse improved from pretest to posttest, but performance on items measuring understanding of separation of concerns and dependency declined. The discussion is organized along these findings.

# 4.1. Abstract reasoning and domain knowledge as predictors software design ability and learning

The students' knowledge of the correct syntax and semantics of the Unified Modeling Language (UML) was used as a proxy to measure their domain knowledge in the field of software design. Based upon prior research on the development of expertise[24] and the "knowledge is power" hypothesis[2] we expected domain knowledge to play a role in both initial software design ability as well as how much students learned about software design after following a course in software engineering. Indeed UML knowledge was a predictor of both initial software design skills as well as improvement from pretest to posttest. This finding is in accordance with the findings of Bergensen et al. (2011) in the domain of software programming where greater knowledge indicated better programming skills in expert software developers.

Abstract reasoning ability, a main component of fluid intelligence, has been postulated to be involved in all learning, especially novel tasks[25], [26]. Therefore, we expected the initial ability and learning of software design of the undergraduate computer science students in our study also to be related to their performance on the Raven's Advanced Progressive matrices - a measure of fluid intelligence. Indeed our results showed that initial understanding of software design principles were significantly predicted by the students' abstract reasoning ability. However, the students' learning from a course in software design as measured by their change in performance on the LIACS software design test administered before and after the course, was not related to their abstract reasoning ability. Perhaps the effect of abstract reasoning, which is a construct similar to working memory, is also mediated by domain knowledge as was the case in Bergensen et al.'s (2011) found that the predictive effect of working memory on software programming skills in experts was mediated by domain knowledge. Abstract reasoning is very similar if not the same construct as working memory[13]–[15]; therefore perhaps domain knowledge also functions as a mediator in the prediction of software design skills learning - i.e., those with better abstract reasoning likely have greater domain knowledge and show the most growth in software design performance. In our statistical model we included abstract reasoning and domain knowledge as predictors at the same level, if domain knowledge is a mediator this could remove evidence the influence of abstract reasoning; future investigations should focus on whether domain knowledge mediates the relationship between abstract reasoning and the learning of software design principles.

#### 4.2. Understanding of measured software design concepts

On the whole the students performed quite well on the LIACS software design test, with approximately 60% chance of solving pretest items correctly and a 70% chance on the posttest - thus on the whole showing significant improvement from pretest to posttest. Students appeared to know more about the design concepts *separation of concerns, cohesion & coupling* and *dependency* during the pretest than the topics *maintenance* and *reuse*. However, the most improvement from pretest to posttest was on these two more difficult topics; this is most likely because of the relatively high chance of solving items on the other three topics correctly - there may have been less room to improve on these particular topics. Strangely, performance on items measuring understanding of *separation of concerns* and *dependency* declined somewhat; this may be due to problems with the items measuring these topics; the psychometric analyses of the test indicate that nearly all of the items measuring separation of concerns should be revised. On the whole our results provide insight into which topics students mastered both before and after the course. In the future administering the LIACS software design test may be helpful for educators so that content can be tailored to the knowledge and skills the students already possess as well as evaluate how effective teaching of particular topics has been.

#### 4.3. Limitations

Some limitations of this study deserve mention and can be informative for future research. First, domain knowledge in the form of a UML quiz was only assessed at pretest. In future research this task should also be administered at posttest in order to (1) assess gains in domain knowledge as a consquence of the course and (2) examine whether the mediating role of domain knowledge on the students' design ability changes during the learning process - a specific prediction of Ackerman's (1998) theory of how skills develop[9].

Second, the instruments we used to assess software design ability and domain knowledge (UML) were created specifically for this study and require further development and evaluation before assuming their reliability and validity. The reliability of these measures were addressed using psychometric classical test theory and item response theory; however, certain revisions to the items are warrented given their low correlation with other related items and/or lack of fit in the item response measurement model. The construct validity was examined by computing correlations with course grades; the moderate correlation indicated that the software design skills test indeed measures a similar construct to the software engineering course exams at both universities.

A third limitation was that we did not account for the role of experience in the students' ability to learn software design. Experience is a good predictor of expertise in novices[1] and likely strongly related to both domain knowledge and initial software design ability as appears to be the case with software programming [12]. A fourth and final limitation to overcome in future research is the inclusion of a control group. By including a control group we can control for the effect of retesting when measuring learning in a pretest-intervention-posttest design.

#### 4.4. Conclusions & Future Directions

Based on this study we can conclude that both domain knowledge and abstract reasoning are key predictors of students' software design ability. Domain knowledge in particular predicts how much students learn from a software design course. Perhaps increasing students' knowledge of UML may improve their chances of learning from a software design course. In the future the effectiveness of improving domain knowledge on course or training outcomes should be investigated.

The present study focused on undergraduate students performance before and after a course on software design; it woud be of particular interest to use the LIACS software design test to examine the performance of complete novices (e.g., psychology students) and undergraduate computer science students before and after following training in software design[27].

#### 5. Acknowledgements

This project was conducted under the supervision of prof. dr. Michel Chaudron and PhD candidate Dave Stikkelorum. Dave conducted nearly all of the data collection for this project and has published a paper on this topic[19]. I am very grateful for the opportunity to collaborate with these researchers on such an interesting project that in addition to completing my computer science study also spans my interests and expertise developed as a researcher in developmental and educational psychology.

#### 6. References

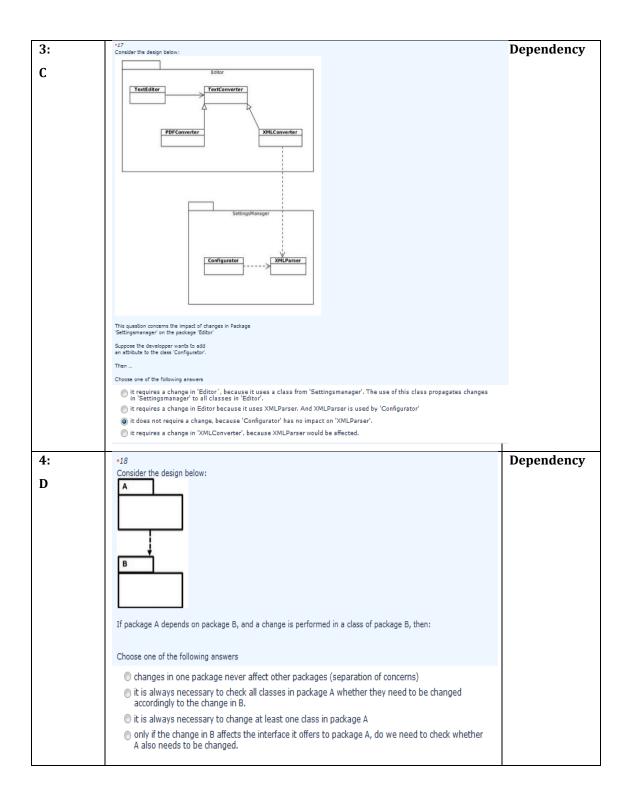
 S. Sonnentag, C. Niessen, and J. Volmer, "Expertise in software design," in *The Cambridge handbook of expertise and expert performance*, no. 1981, 2006, pp. 373–387.

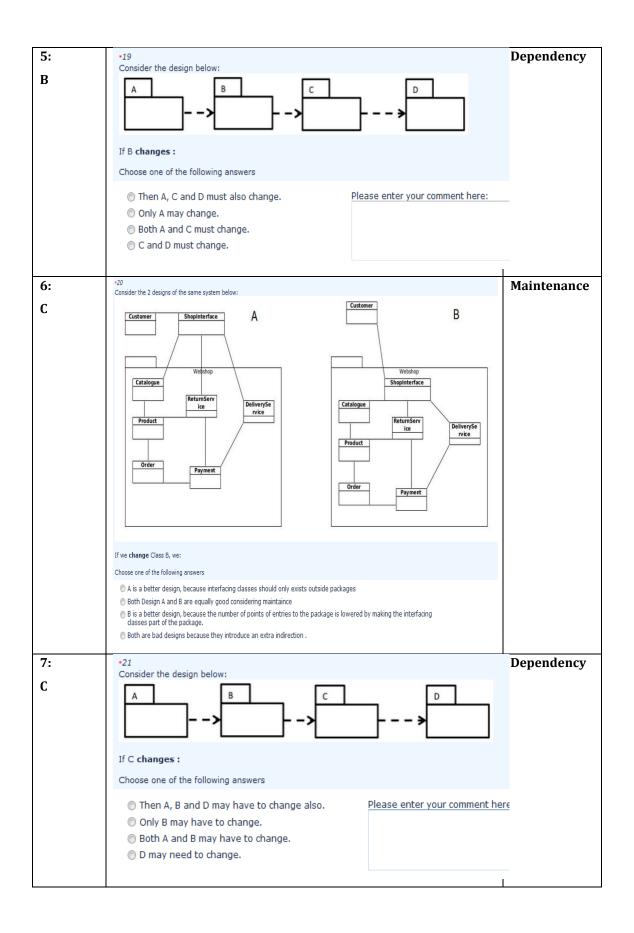
- [2] D. Z. Hambrick and R. W. Engle, "Effects of domain knowledge, working memory capacity, and age on cognitive performance: an investigation of the knowledge-is-power hypothesis.," *Cogn. Psychol.*, vol. 44, no. 4, pp. 339–87, Jun. 2002.
- [3] R. B. Cattell, *Intelligence: its Structure, Growth and Action: its Structure, Growth and Action.* New York, New York, USA: Elsevier, 1987.
- [4] J. R. Anderson, "Skill acquisition: Compilation of weak-method problem situations.," *Psychol. Rev.*, vol. 94, no. 2, pp. 192–210, 1987.
- [5] P. L. Ackerman, "New Developments in Understanding Skilled Performance," *Curr. Dir. Psychol. Sci.*, vol. 16, no. 5, pp. 235–239, Oct. 2007.
- [6] K. Siau and X. Tan, "Improving the quality of conceptual modeling using cognitive mapping techniques," *Data Knowl. Eng.*, vol. 55, no. 3, pp. 343– 365, Dec. 2005.
- [7] N. Bolloju and F. S. K. Leung, "Assisting novice analysts in developing quality conceptual models with UML," *Commun. ACM*, vol. 49, no. 7, pp. 108–112, 2006.
- [8] J. Kramer, "Is abstraction the key to computing?," *Commun. ACM*, vol. 50, no. 4, pp. 36–42, Apr. 2007.
- [9] P. L. Ackerman, "Determinants of individual differences during skill acquisition: Cognitive abilities and information processing.," *J. Exp. Psychol. Gen.*, vol. 117, no. 3, pp. 288–318, 1988.
- [10] V. J. Shute, "Who is Likely to Acquire Programming Skills?," *Journal of Educational Computing Research*, vol. 7. pp. 1–24, 1995.
- [11] W. W. Wittmann and H.-M. Süß, "Investigating the paths between working memory, intelligence, knowledge, and complex problem-solving performances via Brunswik symmetry.," in *Learning and individual differences: Process, trait, and content determinants.*, P. L. Ackerman, P. C. Kyllonen, and R. D. Roberts, Eds. Washington, D.C.: APA, 1999, pp. 77–108.
- [12] G. R. Bergersen and J.-E. Gustafsson, "Programming Skill, Knowledge, and Working Memory Among Professional Software Developers from an Investment Theory Perspective," *J. Individ. Differ.*, vol. 32, no. 1987, pp. 201–209, 2011.
- [13] P. L. Ackerman, M. E. Beier, and M. O. Boyle, "Working Memory and Intelligence : The Same or Different Constructs ?," vol. 131, no. 1, pp. 30– 60, 2005.
- [14] M. Buehner, S. Krumm, and M. Pick, "Reasoning = working memoryp attention," vol. 33, pp. 251–272, 2005.

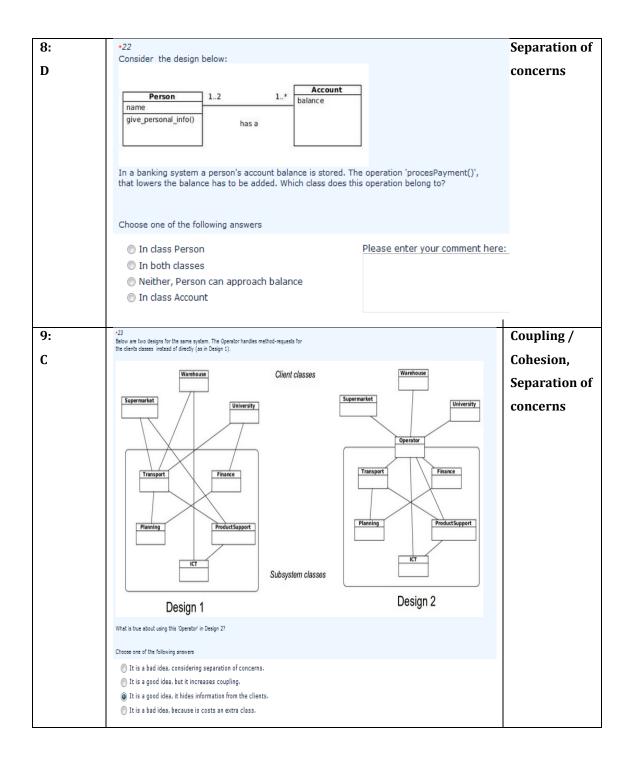
- [15] R. Colom, I. Rebollo, A. Palacios, M. Juan-espinosa, and P. C. Kyllonen, "Working memory is ( almost ) perfectly predicted by g," vol. 32, pp. 277– 296, 2004.
- [16] E. Arisholm and D. I. K. Sjoberg, "Evaluating the effect of a delegated versus centralized control style on the maintainability of object-oriented software," *IEEE Trans. Softw. Eng.*, vol. 30, no. 8, pp. 521–534, Aug. 2004.
- [17] D. R. Stikkolorum, C. E. Stevenson, and M. R. V Chaudron, "Technical report 2013-02," Leiden, The Netherlands, 2013.
- [18] P. A. L. De Boeck and M. Wilson, *Explanatory item response models: A generalized linear and nonlinear approach.* New York, New York, USA: Springer, 2004.
- [19] D. R. Stikkolorum, C. E. Stevenson, and M. R. V Chaudron, "Assessing Software Design Skills and Their Relation With Reasoning Skills," in *EduSymp@ MoDELS*, 2013, pp. 1–8.
- [20] J. Raven, J. C. Raven, and J. H. Court, *Manual for Raven's Progressive Matrices and Vocabulary Scales. Section 4: The Advanced Progressive Matrices.* San Antonion, TX: Harcourt Assessment, 1998.
- [21] D. A. Bors and T. L. Stokes, "Raven's Advanced Progressive Matrices: Norms for First-Year University Students and the Development of a Short Form," *Educ. Psychol. Meas.*, vol. 58, no. 3, pp. 382–398, Jun. 1998.
- [22] S. E. Embretson and S. Reise, *Item response theory for psychologists.* Mahwah, NJ: Erlbaum Publishers., 2000.
- [23] R. M. Furr and V. R. Bacharach, *Psychometrics An Introduction*, 2nd ed. London: SAGE, 2014, p. 442.
- [24] K. A. Ericsson, N. Charness, P. J. Feltovich, and R. R. Hoffman, *Handbook of Expertise*. London: Cambridge University Press, 2006.
- [25] M. C. Voelkle, W. W. Wittmann, and P. L. Ackerman, "Abilities and skill acquisition: A latent growth curve approach," *Learn. Individ. Differ.*, vol. 16, no. 4, pp. 303–319, Jan. 2006.
- [26] R. Primi, M. E. Ferrão, and L. S. Almeida, "Fluid intelligence as a predictor of learning: A longitudinal multilevel approach applied to math," *Learn. Individ. Differ.*, vol. 20, no. 5, pp. 446–451, Oct. 2010.
- [27] D. R. Stikkolorum, M. R. V. Chaudron, and O. de Bruin, "The Art of Software Design, a Video Game for Learning Software Design Principles," in *Models* 2012, 2012.

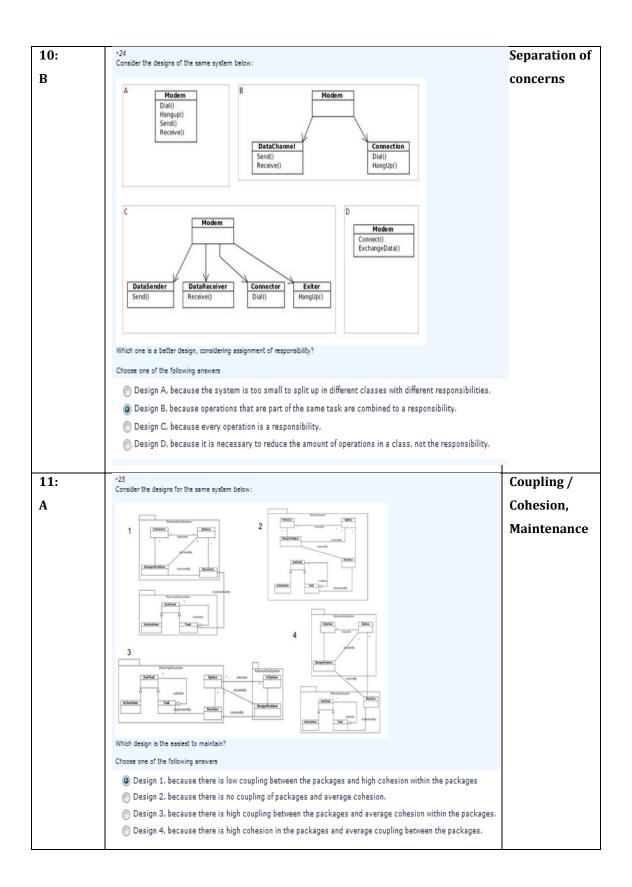
Item:	Content	Concepts
solution		
1:	*15 Consider the design below:	Reuse,
В	Class A Class B If we want to reuse Class B we: Choose one of the following answers	Dependency
	<ul> <li>have to change Class A</li> <li>also need to use Class A</li> <li>do not need to use Class A</li> <li>need to inherit from A</li> </ul>	
2: C	*16 Consider the design below: Class A If we change Class B, we: Choose one of the following answers	Dependency
	<ul> <li>need to change Class A</li> <li>need to inherit from A</li> <li>do not need to change Class A</li> <li>can not do that, because B uses A</li> </ul>	

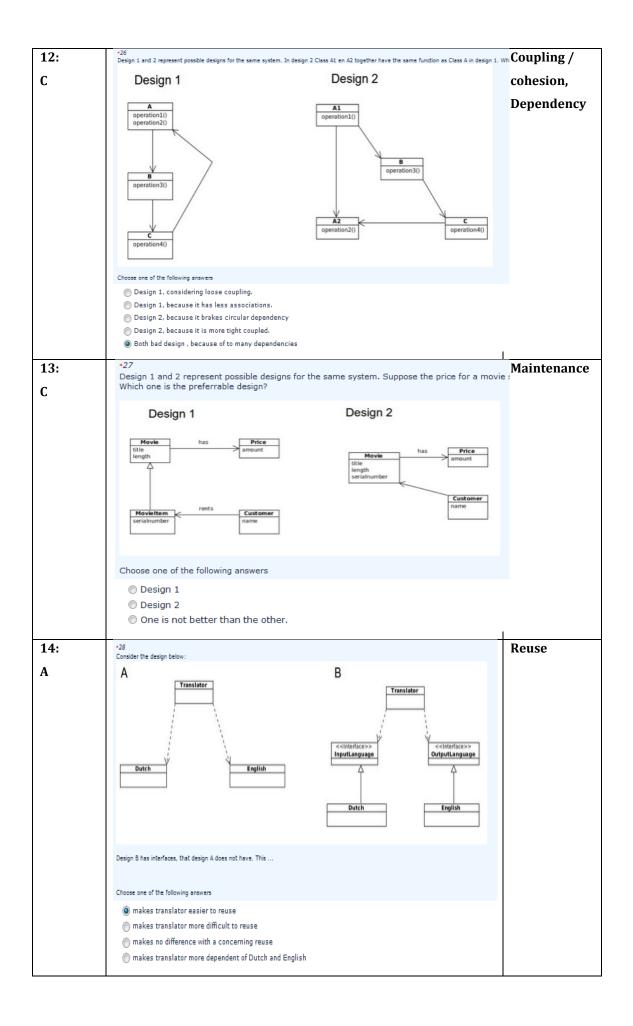
### Appendix A: Software Design Skills Questions

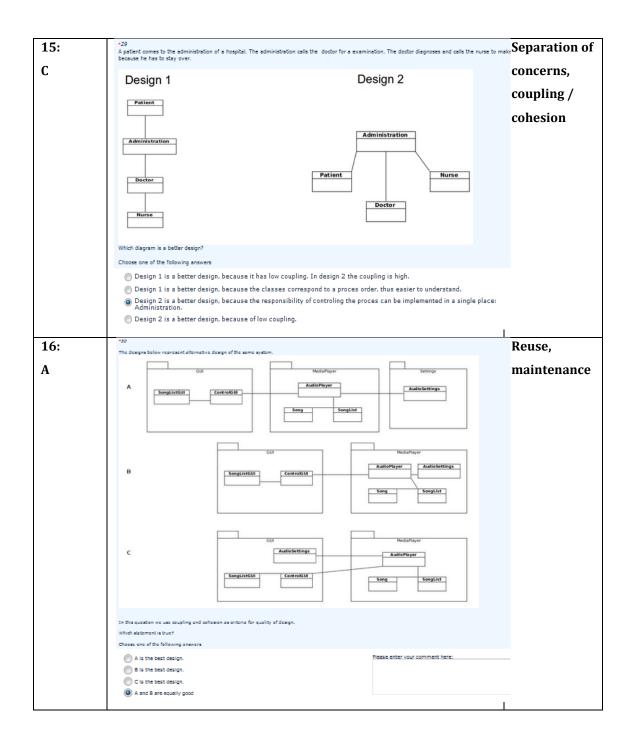


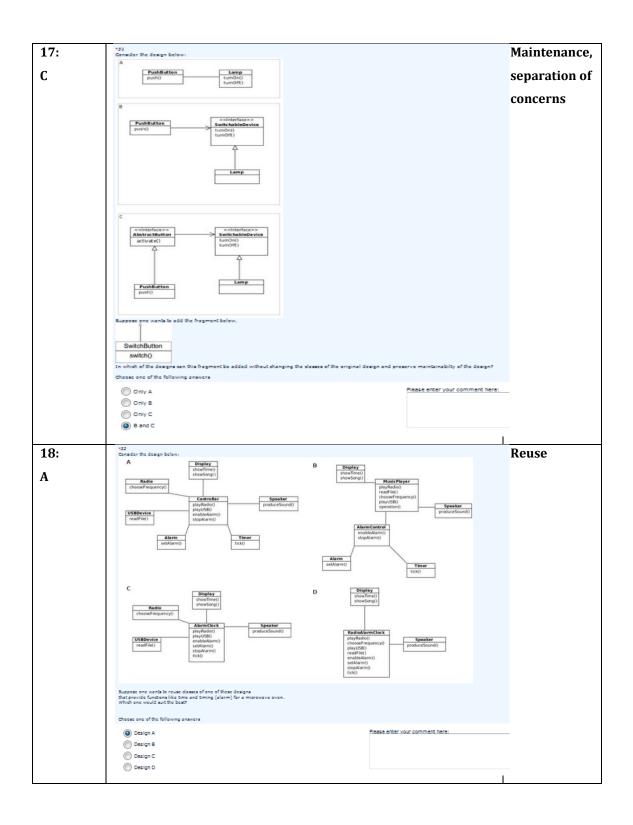


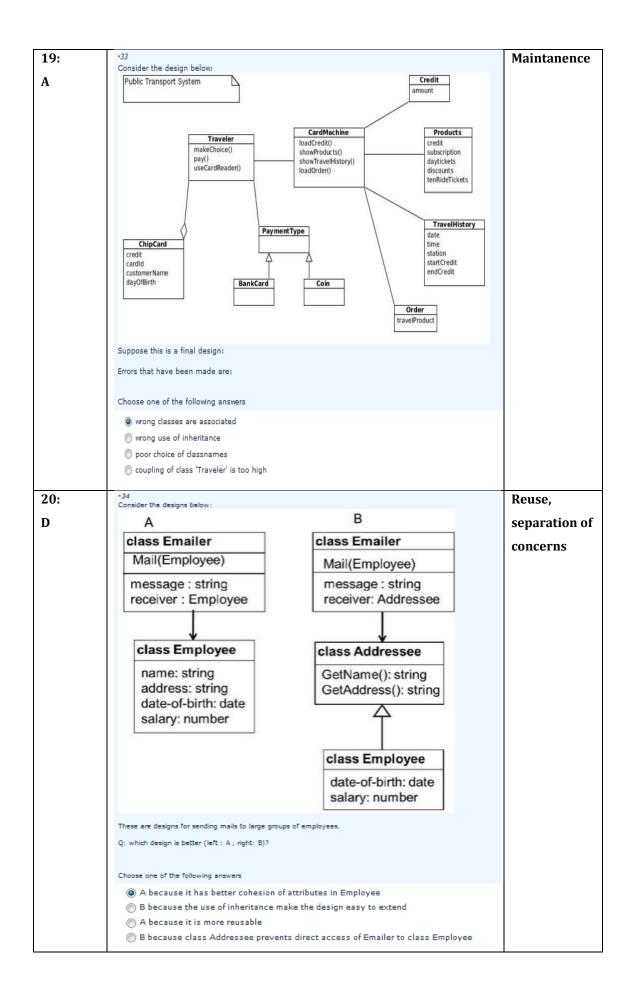












### Appendix B UML knowledge quiz

7       #       0	1							
implementation of entities istates and behavior of entities relationships between entities 2 *9 Which symbol specifies the visiability of a class member? to 7 bublic Private Protected Package Derived Statistics 8 to 10 Which instance level relationship is represented in each of the following diagrams? 8 to 11 Class A Class B I class A Class B <pi a="" b<="" class="" p=""></pi>	1		followi	ng thin	gs except			
8 *10 Which instance level relationship is represented in each of the following diagrams?		interaction between entities						
Public Private Protected Package Derived State 9 Which symbol specifies the visiability of a class member? to 7 8 10 Which instance level relationship is represented in each of the following diagrams? 11 10 Which instance level relationship is represented in each of the following diagrams? 11 Class A Class B 0 0 0 0 0 0 0 0 0 0 0 0 0		implementation of entities						
2       •9 Which symbol specifies the visiability of a class member?         to       •		states and behavior of entities						
to 7 10 10 10 10 10 10 10 11 11 10 10		relationships between entities						
to 7 10 10 10 10 10 10 10 11 11 10 10								
7       # 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	2	*9 Which symbol specifies the visiability of a class me	ember?					
7 - 0 0 0 0 0 0 0 0 ! 0 0 0 0 0 / 0 0 0 0 0 / 0 0 0 0 0 + 0 0 0 0 0 0 + 0 0 0 0 0 + 0 0 0 0 0 + 0 0 0 0 0 0	to		Public	Private	Protected	Package	Derived	Static
8 *10 Which instance level relationship is represented in each of the following diagrams? 8 to 11 Class A Class B O Class B O O O O O O O O O O O O O O O O O O O	7	#	$\odot$	$\bigcirc$	$\odot$	$\bigcirc$	$\odot$	$\bigcirc$
8 *10 Which instance level relationship is represented in each of the following diagrams?  8 *10 Which instance level relationship is represented in each of the following diagrams?  11 Class A Class B O O O O O O O O O O O O O O O O O O O		-					0	
8 *10 Which instance level relationship is represented in each of the following diagrams? 8 to 11 Class A Class B O O O O O O O O O O O O O O O O O O			<u> </u>		-		0	
8 *10 Which instance level relationship is represented in each of the following diagrams?  8 *10 Which instance level relationship is represented in each of the following diagrams?  11 Class A Class B O O O O O O O O O O O O O O O O O O O					-		0	
8 *10 Which instance level relationship is represented in each of the following diagrams? 11 Class A Class B </th <th></th> <th></th> <th>-</th> <th>-</th> <th>-</th> <th><math>\bigcirc</math></th> <th><math>\odot</math></th> <th><math>\odot</math></th>			-	-	-	$\bigcirc$	$\odot$	$\odot$
not shown         8       *10 Which instance level relationship is represented in each of the following diagrams?         to       Image: Class A instance level relationship is represented in each of the following diagrams?         11       Image: Class A instance level relationship is represented in each of the following diagrams?         Image: Class A instance level relationship is represented in each of the following diagrams?         Image: Class A instance level relationship is represented in each of the following diagrams?         Image: Class A instance level relationship is represented in each of the following diagrams?         Image: Class A instance level relationship is represented in each of the following diagrams?         Image: Class A instance level relationship is represented in each of the following diagrams?         Image: Class A instance level relationship is represented in each of the following diagrams?         Image: Class A instance level relationship is represented in each of the following diagrams?         Image: Class A instance level relationship is represented in each of the following diagrams?         Image: Class A instance level relationship is represented in each of the following diagrams?         Image: Class A instance level relationship is represented in each of the following diagrams?         Image: Class A instance level relationship is represented in each of the following diagrams?         Image: Class A instance level relationship is represented in each of the following diagrams?			$\odot$	$\odot$	$\odot$	$\odot$	$\odot$	$\odot$
8 *10 Which instance level relationship is represented in each of the following diagrams? to 11 Class A Class B Class				$\bigcirc$		$\bigcirc$		$\odot$
to 11 Class A Class B O O O O O O O O O O O O O O O O O O		not shown	$\odot$	$\bigcirc$	$\odot$	$\bigcirc$	$\odot$	$\odot$
to 11 Class A Class B O O O O O O O O O O O O O O O O O O	0	*10 Which instance level relationship is represented in	each of t	the follow	vina diaarams	?		-
Image: Class A     Association     Aggregation     Composition     Association       11     Class A     Class B     Image: C	0							
Class A Class B O O O O	to				Aggregation	Composit		
	11	Class A Class B	O		0	O	(	0
Class A Class B		Class A Class B	O		0	$\odot$	(	0
		Class A Class B	0		O	O	(	0
Class A Class B O O O O		Class A Class B	O		O	O	(	0
Class A Class B O O O		Class A Class B	O		O	O	(	0
Class A Class B O O O					-	0		
not shown 💿 💿 💿		Class A Class B	$\odot$		0	0	(	0

