Creating the Digital Logic Concept Inventory

Geoffrey L. Herman, Michael C. Loui, and Craig Zilles

glherman,loui,zilles@illinois.edu, University of Illinois at Urbana-Champaign

ABSTRACT

A concept inventory (CI) is a standardized assessment tool that evaluates how well a student's conceptual framework matches the accepted conceptual framework of a discipline. In this paper, we present our process in creating and evaluating the alpha version of a CI to assess student understanding of digital logic. We have checked the validity and reliability of the CI through an alpha administration, follow-up interviews with students, analysis of administration results, and expert feedback. So far the feedback on the digital logic concept inventory is positive and promising.

Categories and Subject Descriptors

K.3.2 [**Computers & Education**]: Computer & Information Science Education—*Computer Science Education*

General Terms

Human Factors

Keywords

Curriculum, Concept Inventory, Assessment, Logic Design

1. INTRODUCTION AND BACKGROUND

A significant difficulty in judging the success of pedagogical interventions is the ability to directly compare performance between cohorts and institutions. Because of this difficulty, computing educators have issued a general call for the adoption of assessment tools to critically evaluate computing education research [2].

Concept inventories are standardized assessment tools that evaluate how well a student's conceptual framework matches the accepted conceptual framework of a discipline. It is critical to accurately assess conceptual understanding, so that instructors can match instruction to their students' needs. Increasing conceptual learning is important, because students who can organize facts and ideas within a consistent conceptual framework are able to learn new information quickly and are able to more easily apply what they know to new applications [3].

The potential for good assessment tools is clear from the impact of the Force Concept Inventory (FCI), a multiple-choice test in which students choose between the Newtonian conception of force and common misconceptions. The FCI demonstrated that even students who had excelled on conventional examinations failed to answer the simple, conceptual questions on the FCI correctly, and it exposed fundamental flaws in instruction [10]. The results of administrations of the FCI to thousands of students led physics instructors to develop and adopt "interactive engagement" pedagogies [6]. Due to the impact of the FCI, "concept inventory" (CI) tests are being actively developed for a number of science and engineering fields [4].

In this paper, we report the development and validation of a digital logic concept inventory (DLCI). We do not provide the whole DLCI, however. For security reasons, the DLCI is available only on paper by request to the authors. Because CIs are not yet commonly used and there are misconceptions about their use, we first define what a CI is and is not.

A CI is a short multiple-choice test that can classify the examinee as someone who thinks in accordance with accepted conceptions in a discipline or in accordance with common misconceptions [11].

A CI is a standardized test. It must meet the demands of statistical analysis and be broadly applicable to many programs. A CI covers each concept multiple times to strengthen the validity and reliability of measurement. This requirement contrasts with a typical classroom exam, which may cover each concept only once during the exam. To be considered a successful instrument, a CI must also be approved by content experts and be widely adopted.

A CI is not a comprehensive test of everything a student should know about a topic after instruction. CIs selectively test only critical concepts of a topic [11]. If students demonstrate understanding of these critical concepts, then it is reasonable to believe they satisfactorily understand all other concepts of the topic. For example, the FCI tests only a student's knowledge of force after a course in mechanics, which also covers topics such as inertia, momentum, and energy [9].

A CI may complement but not replace a final examination, because a CI is not comprehensive.

A CI is not a teacher evaluation. CIs are intended to measure the effectiveness of teaching methods independent of teacher qualifications [1, 9]. As such, a CI can stimulate the adoption of new pedagogies as it provides an objective measure to compare pedagogies.

A CI reveals students' misconceptions. It does not evaluate their problem solving skills.

2. CONSTRUCTION OF THE DLCI

The DLCI was created by incorporating previously found digital logic misconceptions [7, 8], results from administrations of an alpha version DLCI to 203 students, and a survey of digital logic instructors [5]. The concepts tested by the DLCI were selected using instructor feedback and guidance; the distractors (wrong answers) were created from misconceptions and administration results from the DLCI.

2.1 Concept selection

We used a Delphi consensus rating of what instructors believed to be the important and difficult concepts that students should know after completing a first course in digital logic [5] to choose topics for inclusion. We chose topics that were considered to be highly important by the instructors, but we chose topics independent of the Delphi difficulty ratings in order to check whether the general perceptions of instructors matched reality for the students.

The DLCI has been administered and revised. The current version has 19 items covering topics from four main categories: number representations, combinational logic, functionality of medium-scale integration (MSI), and state and sequential logic. Some topics from the Delphi ratings were not included in the DLCI because they are design based rather than conceptual. The list below shows what concepts are tested and how many items cover each of those concepts (note: the number of items in each sub-category do not add up to the number of items for the larger categories because some items cover multiple sub-categories):

- Number representations (4)
    - Understanding the relationship between representation (pattern) and meaning (value) (3)
    - Conversions between number systems (2)
    - Overflow (1)
- Combinational logic (4)
    - Converting verbal specifications to Boolean expressions (3)
    - Incompletely specified functions (1)
- Functionality of MSI components (2)
- State and sequential logic (9)
    - Converting verbal specifications to state diagrams (2)
    - State transitions (3)
    - Timing diagrams relation state machines (3)
    - Knowing how a sequential circuit corresponds to a state diagram (4)
    - Memory organization (2)

The number of items per concept was motivated by two factors: instrument reliability and student interviews. In order to bolster the reliability of a standardized test, it is necessary for multiple items to test a single concept or ability. So while one concept may be more important than another, the importance of that topic might not correlate with the number of items addressing that concept. A second factor also determined the number of items per concept: the misconceptions found during student interviews. In the

The next two questions refer to a sequential circuit T that has 0 inputs, 3 flip-flops, and 2 outputs.

18) What is the maximum number of distinct states T can potentially be in over time?

a) 6 states   b) 8 states   c) 10 states

e) 16 states   e) 32 states   f) _____

19) At an instant of time, how many states is T in?

a) 1 state   b) 2 states   c) 3 states

d) 4 states   e) 5 states   f) _____

Figure 1: DLCI items with write-in responses

DLCI, all distractors should reflect student misconceptions discovered through interviews and early administrations of the DLCI. Items were written only if compelling misconceptions were found to create effective distractors.

2.2 Item creation

We created the distractors for the DLCI from misconceptions found during problem solving interviews with students [7, 8]. Additional misconceptions were gathered during the alpha administration. On the alpha DLCI, students were given the opportunity to write-in their own answers if they did not like any of the pre-written answers. This option allowed us to uncover misconceptions that did not surface during the interviews. If several students provided the same write-in answer, we deemed the write-in answer to be a new misconception that needed to be investigated with later follow-up interviews. Revised versions of the DLCI include common write-in answers as standard distractors.

2.3 Example item constructions

We explain how we constructed two items of the DLCI.

Students hold multiple misconceptions about the number of possible states in a sequential circuit. Students struggle to understand what allows a circuit to have state, and which circuit components compose state. Some students believe that combinational logic has state, and other students assert that the number of inputs or outputs would affect how many states a sequential circuit might have. Finally, students know that each flip-flop in a system has a state, but they have difficulty with abstracting flip-flop states to system state: the system has only one composite state encoded by the flip-flops [8].

These misconceptions led to the creation of two items on the DLCI (see Figure 1). Item 18 tests whether students correctly understand which components compose state and whether they understand the exponential relationship between flip-flops and circuit state. Item 19 tests whether the students can abstract from individual flip-flop states to the composite circuit state as well as whether inputs and outputs are part of the state. We found that 79% of 108 students answered item 18 correctly and 62% answered item 19 correctly. The write-in response for these items revealed a misconception that had not been found during interviews as 10% of students decided to write-in the answer '0' for both items 18 and 19. Because so many students chose '0' for the answer to 18 and 19, the newest version of the item now includes "0 states" as a standard distractor. Follow-up interviews are needed to know *why* students chose '0'.

For test construction purposes, we decided that each item should have at least four choices to minimize the effect of

7) Alice and Bob have the following requirements for their sandwiches.

Alice must have a sandwich with bacon by itself.

Bob must have a sandwich that does not have both lettuce and bacon.

Which set of Boolean expressions correctly specifies all sandwiches that satisfy their individual requirements?

Use the following variables.

b = bacon; l = lettuce; t = tomato

a) $\boldsymbol{Alice} = b$ $\boldsymbol{Bob} = l \oplus b$

b) $\boldsymbol{Alice} = b$ $\boldsymbol{Bob} = \overline{l\,b}$

c) $\boldsymbol{Alice} = b\,\bar{l}\,\bar{t}$ $\boldsymbol{Bob} = l \oplus b$

d) $\boldsymbol{Alice} = b\,\bar{l}\,\bar{t}$ $\boldsymbol{Bob} = \overline{l\,b}$

Figure 2: Compound concept DLCI item

guessing, which can artificially inflate scores. This requirement was challenging because many concepts had only one misconception and hence only one possible distractor. Using the same process described previously, we created the item in figure 2. Interviews have revealed that when students were asked open-ended questions, they conflated the NAND concept and XOR concept. Students also typically omitted negated variables when translating English specifications into boolean expressions [7]. In both of these cases and many other Boolean problems, we identified only one misconception for each concept. Because we wanted most distractors to be based in real student misconceptions and because we wanted to fully assess students' understanding of Boolean algebra, we decided to combine multiple concepts into one item. While it is ideal to examine only one concept per item, we decided it was better to test multiple concepts per item to fully assess what misconceptions students hold rather than practically eliminate items on Boolean logic.

3. DLCI ADMINISTRATION

In Spring 2009, 203 students from the University of Illinois at Urbana-Champaign were given 25 minutes to complete the alpha DLCI near the end of the semester, before taking their final exams. These students were taking either CS 231 or ECE 290 (introductory digital logic courses for second year students). The CS 231 and ECE 290 students were given slightly different forms of the DLCI. The two versions differed by which experimental items we included, but 15 items were the same on both versions. Summative statistics on student performance on the DLCI can be seen in Table 1.

Because the inventory needs to be a quick assessment of conceptual knowledge, we measured how long students took to complete the DLCI. All students were able to complete the inventory in the 25 minutes allotted, and several students took less than ten minutes.

4. RELIABILITY AND VALIDITY

We checked the validity and reliability of the DLCI through several methods. All measures and data in this section are preliminary, but demonstrate the development process.

4.1 Reliability

We are estimating reliability in two ways: single administration measures and multiple administration measures.

ECE290	Metric	CS 231
111	N	92
8.3	Mean	8.2
2.4	STD DEV	2.7
8	Median	8
1	Minimum score	2
15	Maximum score	14
Normal	Distribution	Normal

Table 1: Distribution of scores on the alpha DLCI administration

To measure single administration reliability we used the Kuder-Richardson Formula 21 (KR-21). The CS 231 student administration had a KR-21 reliability rating of 0.505 and the ECE 290 student administration had a KR-21 reliability rating of 0.639. These preliminary findings indicate that the inventory has acceptable (above 0.4 is acceptable) single administration reliability after the first administration.

To estimate multiple administration reliability, we conducted follow-up interviews with 11 self-selected students two or three weeks after they took the DLCI the first time. Students retook the DLCI in an exam-like situation. After recompleting the inventory, students retook the DLCI one more time in a modified think-aloud format. We compared the students' scores on the retakes with their original scores. Only one student's scores on the first two administrations of the DLCI varied by more than one point. This student had the lowest score of all students interviewed and performed three points worse on the retake. All other students performed the same or one point better on the second administration. Some improvement on the retest is expected, because students had studied for and/or had taken their final exam between administrations. The time gap between administrations and the small number of students that we have been able to recruit to retake the DLCI prohibits any strong conclusions, but the results thus far are promising.

4.2 Validity

Reliability is a necessary (but not sufficient) condition to show that the instrument is valid. Because the preliminary estimates of reliability are positive, we also checked to see whether the instrument is an unbiased estimator of student conceptual knowledge, accurately reflects student misconceptions, and is perceived by experts to measure digital logic conceptual knowledge.

4.2.1 Bias

We checked for bias on two levels: the inventory as a whole and individual items. A 2-tailed t-test revealed that there was no significant difference (p=.8) between the performance of the CS 231 students and the ECE 290 students, and hence no evidence of bias.

To compare the performance of the different departments on individual items we used the χ^2 test. We found that CS 231 students performed significantly better (p=0.001) on only one item, which focused on student understanding of conditional logic. We hypothesize that CS 231 students performed better on this item, because they, unlike ECE 290 students, are required to take a discrete mathematics course prior to or concurrently with the digital logic

Question	a	b	c	d	e	f
1	73	45	46	20	19	
2	63	13	31	96		
3	126	46	25	5		
4	11	69	91	32		
5	29	58	92	24		
6	70	26	33	74		
7	14	15	83	91		
8	33	51	13	105		
9	19	155	4	2	18	5
10	3	47	3	45	105	
11	29	141	8	24		
12	29	146	12	16		
13	22	36	101	9	31	
14	31	16	121	5	4	22
15	3	189	8	0	0	

Table 2: Item responses from alpha DLCI (highlighted choices represent the correct answer)

Figure 3: Item response curve of DLCI item 7

Quintile	a	b	c	d	# of students
1	0%	0%	33%	67%	43
2	0%	2%	40%	58%	48
3	13%	6%	39%	42%	31
4	9%	14%	49%	28%	43
5	16%	16%	45%	24%	38

Table 3: Student responses for DLCI item 7 by quintile (correct answer highlighted)

course. Because the computer science students are exposed to conditional logic in multiple courses, we believe their prior knowledge improved their performance on this item. Because students performed poorly on this item, and because our follow-up interviews revealed that students chose their answers based on misconceptions, we believe that this item can be used to assess conceptual knowledge, but it may be better placed on the discrete mathematics concept inventory that is also being developed.

4.2.2 Student content validity

Because we tried to base the distractors in the DLCI on student misconceptions, one measure of validity was to check whether all or most of the distractors were chosen, and whether distractors were chosen because of misconceptions or because of ambiguity in the inventory items. In Table 2 we show that every distractor was chosen by the students except for two distractors on item 15. We also found that items 9 and 15 were answered correctly more than any other items. Both items concerned multiplexers. We suspected that these items revealed that student misconceptions were not widely held for these problems.

After conducting our follow-up interviews, we confirmed that students could accurately and quickly answer item 15 using conceptual arguments, but we learned that many students were answering item 9 correctly using "plug and chug" without understanding the fundamental concept underlying the item. We decided to alter item 9 and drop item 15 from future versions. Using a similar methodology of watching how students solved various items, we determined which other items needed to be altered or dropped.

We also checked to see if better performance on the whole DLCI correlated with better performance on individual items. We evaluated item quality by constructing item response curves (IRCs). Students were grouped into quintiles based on their DLCI total scores. Students in the first quintile had the highest scores, and students in the fifth quintile had the lowest scores. A student's quintile was taken to be the
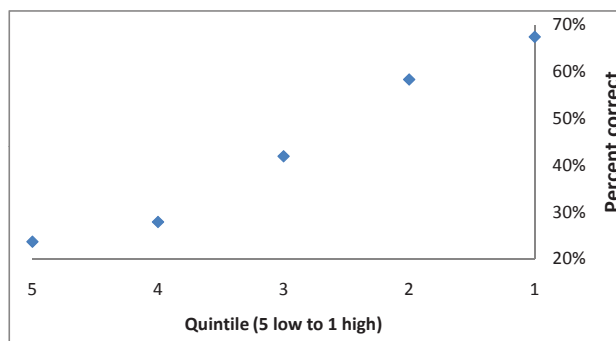
student's level of conceptual knowledge. The IRC for item 7 (Figure 2) is shown in Figure 3 and is typical of all items. This IRC demonstrates the desired correlation between conceptual knowledge and item performance.

In addition to examining the IRCs, we examined which distractors the different quintiles chose. Table 3 shows that students in the first quintile never chose the "bacon-by-itself" misconception distractors (a and b), but students chose these distractors with increasing frequency as the quintile increased. We see similar behavior (although less pronounced) with the "not both" misconception distractors (a and c) as students chose these distractors with increasing frequency as the quintile increased. Students with a stronger conceptual understanding of digital logic don't struggle with omitting visible negated variables, but still struggle with the translation of "not both" into logic, while weaker students struggle with both concepts. The difference in prevalence of the two misconceptions may indicate which concepts need to be stressed more in instruction. In order to accurately assess how common and robust these misconceptions are, however, we need to administer the DLCI to a few thousand students at multiple institutions.

4.2.3 Expert content validity

To check content validity, we collected feedback from 9 experts. We chose experts who had not only taught the material frequently, but who had published textbooks or rigorous pedagogical articles. We strove to find experts with a diversity in race, gender, geography, and type of institution. The experts were asked for feedback on revised items, newly proposed items, and the DLCI as a whole.

On individual items the experts were asked to (1) answer the DLCI item, (2) decide whether the item reflects core

concepts that students should know after completing a first course in digital logic design, and (3) rate the quality of the item. At most one expert chose an answer that we did not expect for all but one item.

For 14 of 15 items in the DLCI, at least eight of nine experts concurred that the item reflected core concepts that students should know after a first course in digital logic. For the 15th item, seven of nine experts agreed it represented core knowledge for a first course. Negative expert comments expressed concerns that students would not know the material in the item after a first course. Based on the Delphi ratings and majority opinion of the experts, we believe that each item on the DLCI accurately reflects what students should know after a first course in digital logic.

The experts also provided feedback on the wording and quality of the items by rating each item on a scale of "do not use," "use with major changes," "use with minor changes," or "use as is." A majority of experts rated every item with "use with minor changes" or "use as is." Only three items received votes of "do not use' and none of these items received more than one vote for "do not use." The "do not use" votes were always accompanied by comments that the item would not be covered in a first course. Because of the general support of the items, we believe that each item was generally well written although we will try to further improve clarity and accuracy.

Finally, the experts were asked to provide their opinions about the DLCI as a whole. The experts were asked to (1) decide if the DLCI as a whole reflects core conceptual knowledge after a first course in digital logic, (2) comment on the topic coverage, and (3) indicate how confident they would be that a student who performed well on the DLCI would perform well in a digital logic course.

The majority of experts indicated that the DLCI reflects core conceptual knowledge. The majority also indicated that they were moderately confident or greatly confident that the DLCI would be a good predictor of student performance. Most experts noted that the DLCI is not a comprehensive assessment and that it did not test problem solving skills and design. Experts explained lowered confidence ratings because of these comments.

Our experts' comments about the DLCI all reflect concerns that the DLCI is not a typical exam. This commentary emphasizes the necessity of communicating what a CI is and is not. The DLCI does, however, assess whether students possess core conceptual knowledge. The DLCI is well positioned to be widely adopted at most institutions that teach digital logic.

5.   CONCLUSIONS

Based on the expert feedback, student performance, and methods of construction, we believe that the DLCI is on its way to becoming a valid and reliable instrument that can be used at most institutions that teach digital logic. The DLCI is a short (less than 30 minutes to complete), multiple-choice instrument that has passed the initial tests for reliability and validity. Our experts' comments also indicated that the DLCI evaluates core conceptual knowledge that is taught in digital logic at most institutions. The DLCI is meeting not only the structural design qualities we desire for the instrument, but also meets the standards of being broadly applicable and informative. Once fully validated through beta testing at multiple institutions, we believe that the DLCI

will be a standardized instrument that can be used to compare how different pedagogies affect conceptual learning.

In order to motivate the adoption of the DLCI, we will need to accurately communicate what a CI is and is not. Once fully validated through administration to over 3,000 students, we expect that the DLCI will show that even students who perform well on traditional digital logic tasks possess robust, commonly held misconceptions.

6.   ACKNOWLEDGMENTS

7.   REFERENCES

[1] V. L. Almstrum, P. B. Henderson, V. Harvey, C. Heeren, W. Marion, C. Riedesel, L.-K. Soh, and A. E. Tew. Concept Inventories in Computer Science for the Topic Discrete Mathematics. In *ITiCSE '06: Proceedings of the 11th annual SIGCSE Conference on Innovation and Technology in Computer Science Education*, pages 163–167, Bologna, Italy, 2006. ACM press.

[2] M. Ben-Ari. The Concorde doesn't fly anymore. In *the Technical Symposium on Computer Science Education (SIGCSE), keynote address*, Feb 2005.

[3] J. D. Bransford, A. L. Brown, and R. R. Cocking. *How People Learn: Brain, Mind, Experience, and School.* National Academy Press, Washington, DC, 1999.

[4] D. Evans et al. Progress on Concept Inventory Assessment Tools. In *the Thirty-Third ASEE/IEEE Frontiers in Education*, pages T4G–1–T4G–8, Nov 2003.

[5] K. Goldman, P. Gross, C. Heeren, G. Herman, L. Kaczmarczyk, M. C. Loui, and C. Zilles. Identifying Important and Difficult Concepts in Introductory Computing Courses using a Delphi Process. In *Proceedings of the 39th annual ACM technical symposium on Computer Science Education*, pages 256–260. ACM SIGCSE, March 2008.

[6] R. Hake. Interactive-engagement vs traditional methods: A Six-thousand-student Survey of Mechanics Test Data for Introductory Physics Courses. *Am. J. Physics*, 66:64–74, 1998.

[7] G. L. Herman, L. Kaczmarczyk, M. Loui, and C. Zilles. Proof by Incomplete Enumeration and Other Logical Misconceptions. In *Proceedings of the Fourth International Computing Education Research Workshop (ICER 2008)*, Sydney, Australia, September 6-7 2008.

[8] G. L. Herman, C. Zilles, and M. C. Loui. Work in progress: Students' Misconceptions about State in Digital Systems. In *Proceedings of the Frontiers in Education Conferenc*, San Antonio, TX, October 18-21 2009.

[9] D. Hestenes, M. Wells, and G. Swackhamer. Force Concept Inventory. *The Physics Teacher*, 30:141–158, 1992.

[10] J. P. Mestre. Facts and Myths about Pedagogies of Engagement in Science Learning. *Peer Review*, 7(2):24–27, Winter 2005.

[11] G. Rowe and C. Smaill. Development of an Elecromagnetics Course Concept Inventory. In *Proceedings of the 2008 American Society for Engineering Education Annual Conference and Exposition*, pages AC 2008–242–266, 2008.