Using Kintsch's Text Comprehension Model to Identify CS Students' Conceptions and Misconceptions

Christina Kyriakou[a], Agoritsa Gogoulou[b] and Maria Grigoriadou
Department of Informatics and Telecommunications, National and Kapodistrian University of Athens, Athens, Greece

Keywords: Introductory Computer Science Courses, Data Storing and Manipulation, Main Memory, Program Execution, Kintsch's Construction-integration Text-comprehension Model, Conceptions, Misconceptions.

Abstract: Recent research attempts to figure out the hard points of Computer Science (CS) discipline curricula that seem to trouble students. The presented work focuses on assessing first-year CS students' knowledge on issues related to fundamental computer architecture concerning main memory organization, operation, and role on program execution. Formulating questions based on Kintsch's text comprehension model theory is a promising diagnostic tool to identify students' conceptions and misconceptions. The paper discusses the Kintsch model as the basis for formulating meaningful questions, analyzes students' answers, and attempts to categorize and explain the revealed misconceptions. The emerging misconceptions may be helpful for effective learning design and appropriate educational material.

1 INTRODUCTION

The in-depth learning of the internal computer operations is of great significance not only for their value but additionally because when students internalize successfully such lower-level information, they gain coherent knowledge of computer operation (Clements, 2000). The non-observable nature of internal computer operations renders learning of computer architecture and organization a problematic issue.

Efficient instruction should consider students' difficulties in comprehending the concepts being taught. This notion is supported by the results of extensive research of cognitivists and educators. For more than four decades, researchers have explored how students learn and what affects their understanding (Bransford, Brown, & Cocking, 2000; Vosniadou, 2001). All findings converge on the great significance of prior constructed students' mental models (Johnson-Laird 1983; Ben-Ari, 2001) or prior knowledge frameworks (Davis, Maher, and Noddings, 1990) for the acquisition and assimilation of new knowledge (Ausubel, 1968; Posner et al. 1982; Vosniadou, & Brewer, 1987). When students are exposed to new knowledge or situations, they use their pre-existing mental models to understand and explain the new concepts. The scientific validity of the previous mental models or knowledge frameworks is a prerequisite in incorporating the new knowledge without misunderstandings. Often students' prior empirical representations of the "real world" are not compatible with scientific representations (Von Glaserfeld, 1995). Under these conditions, misconceptions may be created in students' minds. Such misconceptions or, at best synthetic models are created when students try to incorporate new incompatible knowledge into an existing knowledge structure (Vosniadou 1994; 2003; 2007) or when familiar terms are used in unfamiliar contexts (Clancy, 2004). Misconceptions may be identified by the student's explanations on answering questions or engaging in activities of understanding relevant to the specific concepts. Mistakes, faulty or incomplete answers may reflect what exists in a student's mind and should be considered precious feedback for the instruction and learning process.

Following the theories mentioned above and the cognitive approaches to learning, during the last decades, research has focused on the hard points of Computer Science discipline curricula that seem to trouble students either by their inherent complexity or

[a] https://orcid.org/0000-0002-6835-4783
[b] https://orcid.org/0000-0001-9025-0254

Kyriakou, C., Gogoulou, A. and Grigoriadou, M.
Using Kintsch's Text Comprehension Model to Identify CS Students' Conceptions and Misconceptions.
DOI: 10.5220/0011084000003182
In *Proceedings of the 14th International Conference on Computer Supported Education (CSEDU 2022) - Volume 2*, pages 384-394
ISBN: 978-989-758-562-3; ISSN: 2184-5026

by their interrelatedness with other complicated concepts. The focus has been on introductory programming topics and less on computer architecture and organization topics (Herman, Loui, & Zilles, 2009; 2010; 2011; Herman, Kaczmarczyk, Loui, & Zilles, 2011).

Porter et al. (2013) created a preliminary test of conceptual understanding for an upper-division computer architecture course. They designed nine high-level concept questions on topics including the performance implications of deeper pipelines, the roles of various cache components, and performance analysis of single-cycle, multi-cycle, and pipelined processors. The questions were administered to four separate computer architecture courses at two different institutions and were intended to be correctly answered by any student passing the class. Disappointingly, the results showed a large discrepancy between what instructors thought students were learning and what they had learned. The scholars suggest a further inquiry into student understanding of core course concepts in Computer Architecture.

Grigoriadou and Kanidis (2001) investigated CS students' conceptions on computer cache-memory-related topics with the aid of questionnaires and interviews. The scholars identified the difficulties several students had in understanding the organization and operation of cache-memory. This fact raised concerns to scholars about the depth of students' memory-related knowledge. Another work of the same researchers (2002) on the conceptions of secondary education students indicated a lack of knowledge on computer memory hierarchy, organization, and operation.

The findings and suggestions of the studies mentioned above made the need for efficient instruction of memory-related topics apparent to the authors. The notion that the CS students should be aware of how a computer operates and that encompasses a deep knowledge about memory, motivated authors to conduct a sequence of studies. The ultimate aim of these studies was to support students in their learning of memory-related topics. The first step towards this aim was to investigate students' understanding of memory.

The preliminary study investigated the level of understanding that first-year undergraduates of the Department of Informatics and Telecommunications of the National and Kapodistrian University of Athens have on memory operations after attending an introductory CS course. The authors collected and analyzed students' answers to open-ended questions administered in the written final exams at the end of the semester in four different academic years. More specifically, the focus was on the concepts related to the communication between main memory and CPU, the special-purpose registers, and the machine language program's execution. The results demonstrated that although most students can recall basic knowledge on these topics, they have great difficulty applying and combining this knowledge to make inferences and give correct answers (Kyriakou & Grigoriadou, 2016).

The present study is the sequel to the preliminary research. Learning about the role of the main memory and its architecture in implementing the von Neumann model is also fundamental to understanding computer operation. This idea motivated the authors to continue the investigation of the students' understanding focusing on these memory-related topics.

This study was also conducted at the same department on first-year students. The investigation is based on Kintsch's Construction-Integration Model of Text Comprehension. The presented work attempts to contribute to the field of Computer Science Education as it proposes an approach of formulating questions in order to assess students' conceptions and provides a categorization of the identified conceptions/ misconceptions.

The rest of the paper is structured as follows. Section Two outlines the underlying text comprehension theory. In Section Three, the empirical study is described. Afterward, the study's limitations and the implications of the research results are discussed, and the paper concludes with further research directions.

2 THEORY BACKGROUND

To investigate students' conceptions on the combination of issues relevant to main memory and program execution, a text comprehension model was sought that relies on questions to assess the degree of comprehension accomplishment. The theory of the Construction-Integration model of text comprehension seemed to fill our requirements. (Kintsch, 1988). Kintsch's theory of Construction-Integration model of text comprehension is Kintsch's extension of his and van Dijk's pre-existed comprehension models (Kintsch and van Dijk, 1978; van Dijk and Kintsch, 1983) and is based on the notion that the comprehension process is moderated by individual differences, such as prior knowledge, abilities, preferences, and strategies, mainly stressing the role of previous knowledge.

According to this cognitive model of discourse understanding, a reader develops two distinct levels of text representation during a text comprehension process which the text base and situation models can describe. The text base model corresponds to the propositional representation of a text, both at the level of the micro- and macrostructure. The situation model corresponds to a representation of the text integrated with other knowledge like events, actions, persons, and generally the situation a text is about. “A situation model may incorporate previous experiences, and hence also previous text bases, regarding the same or similar situations. At the same time, the model may incorporate instantiations from more general knowledge from semantic memory about such situations.” (van Dijk and Kintsch, 1983).

Several types of measures can be used to assess the level of a reader’s text understanding, i.e., to evaluate the extent to which the text base and situation models have been developed (McNamara, Kintsch, Songer, and Kintsch, 1996).

These include text base measures which are free-recall and text-based questions. Free-recall is when the reader is asked to recall the text without explanations. While text-based questions are based entirely on the text content, and the necessary information for their answer is stated in the original text and requires only a single sentence from it.

Additionally, the other being the situation model measures which include problem-solving questions, elaborative-inference questions, bridging-inference questions, and the sorting task. In problem-solving questions, a reader is asked to apply information from the text to a novel situation, which requires a well-formed situation model. Inference questions need inferencing of some kind or analytic reasoning. More specifically, when a reader is asked elaborative-inference questions, he should combine information from the text with outside knowledge, which can be achieved even with a surface situational understanding. On the other hand, in bridging-inference questions, the necessary information is stated in two or more sentences in the text. The reader should combine them and infer their unstated relations to answer the question, which requires a deeper situational understanding and a solid text base. In a sorting task, a reader is asked to relate the concepts presented in the text, which reflects, at least in part, the situation model.

3 THE EMPIRICAL STUDY

The present study aimed to apply Kintsch's theory of the Construction-Integration model of text comprehension and analyze first-year CS students' answers to assess students’ knowledge on main memory organization, operation, and role during program execution and derive, record, and identify possible students' misconceptions. The main research question of the empirical study is:
Are the open-ended bridging-inference questions formulated based on Kintsch's theory of the Construction-Integration model of text comprehension, effective in investigating students’ conceptions?

The study was conducted at the Department of Informatics and Telecommunications of the National and Kapodistrian University of Athens in the undergraduate course “*Introduction to Informatics and Telecommunications*.” The course is delivered through two-hour lectures weekly during the first semester, and students take a written examination at the end. The topics covered are i) Data Storage, ii) Data Manipulation, iii) Operating Systems, iv) Networking and the Internet, and v) Security. The topics involved in this study are Data Storage and Data Manipulation. Data Storage covers issues associated with data representation and storage within a computer like bits storage, main memory organization and capacity, mass storage, representation of information as bit patterns, and binary system. Data Manipulation covers themes related to the way a computer manipulates data and communicates with peripheral devices, the basics of computer architecture, and includes the way computers are programmed in machine language instructions (Brookshear, 2009).

3.1 Educational Material

The educational material used in this introductory course is in written (text) and electronic form. Two course books support students’ learning (Brookshear, 2009; Forouzan, 2003). Electronic material is based on the course books. It includes lecture notes, delivered through an LMS (called “eclass”) and additional educational electronic material, in activity form, offered through SCALE, a web-based activity-oriented learning environment. SCALE supports the knowledge construction process by engaging students in activities that address specific learning goals in the context of fundamental concepts and by providing tutoring and informative feedback (Gogoulou et al., 2007).

In the course framework, students' engagement with the activities in the SCALE is optional and motivated by a pre-defined offered reward of up to one grade out of ten towards the students' final course performance.

3.2 Method

To investigate students' conceptions on the role and organization of the main memory in program execution (von Neumann model implementation), written data were collected from two different groups of students using two diagnostic tools (open-ended questions and small-group discussions). A set of open-ended questions were formulated based on Kintsch's text-comprehension theory. In particular, the bringing-inference type of question was used to assess students' text understanding. Then a categorization scheme was defined using a methodology based on Thematic Analysis, as Braun and Clarke (2006) outlined. Inter-coder reliability of the categorization was assessed, and the categorization was then analyzed and interpreted.

3.3 Participants and Data Collection

In the study, a total of 360 students participated. Group A of 270 students participated in the compulsory final course exams in the academic year 2011-2012. 84% were first-year, 16% were second-year, and seniors. Their responses to a set of questions related to computer memory are the first dataset (Dataset1) that was collected and analyzed. Group B of 90 students took part in an optional mid-term course project in 2016-2017. 92% were first-year, 8% were second-year, and seniors. Their participation was motivated by a one-grade bonus reward (out of ten). The project reviewed the topics of Data Storage and Manipulation after the relevant lectures. The course project provided two datasets to the study (Dataset 2 & 3). Dataset2 consisted of students' written responses to two open-ended questions. As a follow-up method, the small-group discussions provided detailed students' opinions on one of the questions. Students' ideas constitute the Dataset3 of the study.

3.4 Bridging-Inference Questions

Both experimental groups answered bridging-inference questions based on Kintsch's text comprehension model. To answer the questions requires combining knowledge from Data Storage and Data Manipulation and then concluding with inferences. To answer the questions, students should have a good text-based understanding of the topics: 1) Data Storage (the organization and operation of the main memory and hard disk); 2) Data Manipulation (CPU's architecture and operations, machine-language program's execution), and 3) the von Neumann architecture.
Indicative questions are:

- *Many things have changed since the first PCs were made, such as size, electronic circuitry, software, applications. What do you think has remained the same?*
- *A student with a Simple Computer wrote the program, executed it, and then stored it on the hard disk before shutting down the computer.*
  - *a) Describe the format of the program stored on the hard disk.*
  - *b) Another day, when the student wanted to open and edit the program, what actions would the Simple Computer take?*
- *Please comment on the correctness of the following suggestion and justify your answer: "A computer may run using the hard disk when the main memory is missing."*

Here follows the analysis of the latter question, which was articulated to encompass the following research questions (RQ):

RQ1 What do the students believe about the role of main memory on a program execution?

RQ2 How do the students relate the main memory organization (ordered directly-accessed addressable cells/bytes) with the implementation of the sequential execution of a program's instructions (as defined in the von Neumann model)?

RQ3 What is the students' knowledge of main and secondary memory data storage?

Moreover, to answer this question, students should be able to infer the text-unstated relations from the mentioned above topics and to describe: 1) the reasons why the main memory and the hard disk have the particular structure and their role on a program's execution; 2) the communication between the main memory and CPU; 3) the direct access of the main memory contents on cells/byte-level through their addresses and the significance of this capability for the sequential instructions' execution as is specified in the von Neumann's model, 4) the inability to access data/instructions on byte-level when stored in a hard disc. These inferences depend on situational understanding.

By linking all this information, students should conclude that, as the von Neumann architecture defines, a computer cannot run when the main memory is missing. The correct justification is that, contrary to the hard disk structure, only the main memory's organization in ordered addressable cells/bytes enables the distinct storage and fetching of a program's instructions which is essential for the machine cycle (fetching, decoding, and execution).

3.5 Data Analysis

3.5.1 Dataset1 Analysis

Group A students' answers (Dataset1) were filed, manually content-analyzed, and classified according to the correctness and justification.

Each response was initially classified as correct, incorrect, and partially correct by one of the authors. 75% of the students failed to make the proper inferences. These students answered the question either: 1) incorrectly, arguing that a computer may run using only the hard disk when the main memory is missing, or 2) partially correct/incompletely, arguing that a computer cannot run without the main memory, but giving incomplete or wrong reasoning.

Then the correct answers were removed from Dataset1. Two of the authors collaborated and processed the Dataset1 answers to derive students' conceptions. They went through the answers and coded them using Thematic Analysis methodology. The method is data-driven. Many students' responses included several justification cases. Each different justification case that was observed defined a new category. Categories were modified and refined after repeated reading of the data. Each case was recorded and coded with a small phrase, i.e., "ROM," "hard disk synchronization issue," "RAM volatility," etc.

Then, for category validation purposes, the third author and another experienced instructor (Inst) from the Department specialized in relevant topics like computer architecture, digital logic design, and digital systems, worked independently. They coded 100 randomly chosen answers from Dataset1, and the inter-coder reliability was calculated (see Results section).

Lastly, only when a justification case occurred in more than five student answers was it kept as significant. The relative frequency of each case was calculated, i.e., the percentage of students' responses that mentioned each justification case.

3.5.2 Dataset2 Analysis

In order to investigate the possible re-occurrence and persistence of the students' conceptions about main memory over time and under different conditions, the study was extended to Group B students. Moreover, the motivation was to ensure that students' incomplete answers were due to their lack of knowledge or misunderstandings and not due to a lack of test time or just because they didn't consider it necessary to add more information to their justification.

In particular, Group B students participated in a mid-term optional review project on Data Storage and Manipulation topics after the relevant lectures had been delivered. The first part of the project consisted of several activities in the WbLE supporting the course. The activities included questions relevant to the topics of the lectures. The students worked individually on the activities. Two open-ended bridging-inference questions were posed to Group B students in this context. The previously described question is one of the two.

Again Group B students' answers (Dataset2) were filed, manually content-analyzed, and classified according to the correctness and justification. Classification according to their correctness showed that 82% of the students' answers were either wrong or partially correct. After removing the right answers from Dataset2, two authors followed the same category-development procedure described in 3.5.1.

Afterward, the third author and the Inst working in collaboration used a sample of 40 randomly-chosen answers of Dataset2 and confirmed the categories.

Once again, the justification cases occurring in less than five student answers were considered infrequent and omitted.

3.5.3 Data Validation

For data validation, another input source of students' conceptions was used. In the second part of the optional course project, Group B students were asked to discuss the open-ended bridging-inference question described above in the "eclass" LMS. The students were separated into small groups of 5-6 members (each member holding a different opinion). The guidelines asked students: a) to express their opinions about the open-ended question and agree or disagree with their peers justifying their views, and b) to elicit a team response to the question under consideration.

This process provided insight into students' thinking and conceptions because it required students to explicitly articulate content in their own words. Consequently, it served as an indicator of their understanding or misunderstanding.

Table 1: Outline of the most frequent justification cases of the incorrect and partially correct students answers.

Cases	Type of answer / Description of justification case	Dataset1 (%)	Dataset2 (%)
1	Incorrect Answer: When the main memory is missing, a computer may operate slower using only the hard disk because both are similar storage systems that work at different speeds.	57%	23%
2	Partially Correct Answer: According to von Neumann's architecture, the main memory is one of the fundamental computer subsystems. The hard disk communicates only with the main memory, not the CPU.	17%	12%
3	Partially Correct Answer: A program should be in the main memory for its execution, so the main memory is necessary for a computer's operation.	14%	23%
4	Partially Correct Answer: ROM is part of the main memory, so the computer cannot start up if it is missing.	11%	41%
5	Partially Correct Answer: The main memory is volatile and used for data storage during program execution. A hard disk is a non-volatile storage system, so it would soon be out of space if it replaced the main memory regardless of its capacity.	19%	22%
6	Partially Correct Answer: The electronic circuitry of the main memory renders its speed fast enough for synchronization with CPU circuits. A hard disk has mechanical parts and requires a physical motion for its operation, making its speed extremely slow.	15%	7%

In the end, the screenshots of the online discussions were collected and filed (Dataset3). The authors studied and compared the opinions of each student with their previous answers in the first part of the project (Dataset2). The Dataset2 answers were confirmed and enriched by the extra information provided through the discussion session.

The discussion session served as a follow-up method that provided clear information about students' knowledge since many of them elaborated on their justifications to persuade their peers.

3.6 Results

The most frequent justification cases of the incorrect and partially correct student answers are outlined in Table 1. In both datasets, the same justification cases prevail, regardless of the variation in the percentages of students.

The interest of the study is focused on these most frequent justification cases since it is likely that future instructional interventions will concentrate on the most common students' conceptions. For this reason, we calculated the inter-coder agreement for these cases, and it was 93%.

As shown in Table 1, in both datasets, the distribution of answers with cases 2, 3, 5, and 6 are similar. On the contrary, there is a significant difference in the percentages of cases 1 and 4. Great interest arises from the fact that the percentage of students who answered wrongly in 2011-2012 is double than that of 2016-2017 (57% vs. 23%). A possible explanation is that, before answering the question, the Group B students had been engaged in the relevant activities in SCALE. This probably boosted the refinement of their knowledge. This conjecture aligns with the study results about the positive effects on students' learning after engaging in SCALE (Verginis, et al, 2009).

Nevertheless, it is noteworthy that regardless of the distribution variations, both groups of students seem to hold the same conceptions even if they expressed them under different conditions.

3.7 Students' Conceptions and Emerging Misconceptions

The analysis of students' answers revealed their conceptions about main memory, hard disc, program execution, and their interrelations. Here follows the detailed description of these conceptions that answer the research questions of this study and the related misconceptions that appeared.

57% and 23% of the students in the two academic years 2011-2012 and 2016-2017 respectively answered incorrectly that a hard disk could replace the main memory, with the drawback being the slower operation of the computer, considering them both storage systems. Case 1 answers revealed the major misconception of the similarity between the main memory and hard disk. Both are considered storage systems with the same functionality. The fact that each unit has a different role, organization, and method of data storage was ignored or underestimated (RQ1, RQ2, and RQ3).

The majority of the students gave a wide range of reasoning regarding the capability of the computer to run when the main memory is missing. The most common answers were: “hard disk is slower than the main memory so the computer will run slower”; “a hard disk doesn’t have an addressing system like the main memory does, so the storing and reading processes would be slower”; “hard disk doesn’t communicate with the CPU through the bus as the main memory does so the data transfer rate would be lower”; “the implementation of the virtual memory already uses a part of the hard disk and it may use the whole capacity of the hard disk, when the main memory is missing, with the flow of the slower retrieve/store processes in a fully occupied disk”; “cache memory could replace main memory during a program execution, but as it has less capacity the data retrieval from the hard disk would need much more time.”

Even though there are some hints of knowledge about the different structures and organization of the two units (main memory and hard disk) in these justifications, it seems that the students have constructed these fragments of knowledge in the context of their alternative frameworks.

A misconception about the performance of the main memory derives from these justifications. Students attribute the high performance of the main memory either to its cell addressing system, which enables the reading and writing processes to be quick, or to the fast bit transfer rate of the bus connecting the main memory with the CPU. Still, students overlook the fact that this performance results from the electronic circuitries of the main memory.

Either way, all the above justifications ignore the role of the main memory and its organization for program execution (RQ1 and RQ2).

Besides, some students hold a misconception about the virtual memory mechanism, as they consider that it can be implemented with a standalone hard disk when the main memory is missing.

Furthermore, other students include the cache memory in their justifications, stating that it may substitute the main memory. These students consider it feasible to run a program from the cache memory when the main memory is missing, a statement that reveals a misconception about the operation of cache memory.

Students expressing the answers of case 2 superficially recall the von Neumann architecture without providing any further information concerning the research questions.

Case 3 students seem to recognize the role of main memory (RQ1), but no information is provided for RQ2 & RQ3.

Case 4 students claim that the hard disk cannot displace the main memory as the start-up information stored in the ROM is crucial. As evidence to their claims, most of the students mention their practical experience of a non-running computer that causes “beep” sounds in case of RAM failure. The misconception here is that the significance of the main memory seems to be constrained only to a storage place for the start-up information, i.e., the ROM is the only important part.

Moreover, many of the answers mentioning this justification case proposed overcoming the start-up problem by storing the ROM data on the hard disk. It seems that the underlying misconception here is about the similarity of the two units (described above).

Case 5 students state that the main memory is irreplaceable for a computer’s operation because of its volatility. These students seem to acknowledge the significance of the role of the main memory on computer operation (RQ1), but not for the scientifically correct reason (RQ2). Besides, many students mention that hard disks are non-volatile, and they would soon run out of space if they replaced the main memory on program execution. This opinion adds another aspect to the previous misconception, suggesting that only volatility is the crucial difference between the main memory and the hard disk. Students ignore the data storage distinction between the two units (RQ3). A few of them propose to solve the out-of-space disk problem with proper programming, which may erase the unnecessary data from the hard disk whenever necessary. The background of such thoughts is the misconception of similarity between the two units.

Students mentioning the justification case 6 seem to acknowledge the significant role of the main memory (RQ1). They express the scientifically valid opinion that the electronic circuitry of the main memory renders it fast enough for synchronization with the CPU circuits. These students state that the main memory’s high performance is the reason for storing a program’s instructions and data during its execution. They seem to believe that only the main memory's high performance makes it indispensable for computer operation. So it seems that they ignore the relation of the main memory’s unique structure in ordered directly-accessed addressable cells for the execution of the sequential instructions of a program (RQ2). In addition, students think that poor

Table 2: Brief description of students' misconceptions and reasoning. Outcomes concerning the RQs.

	Description of Misconception	Explanation/Reasoning/RQs results
1	The similarity between main memory and hard disc. Both are storage systems with the same functionality.	Regardless of their differences, both systems can store data, so the main memory's replacement by a hard disk will only slow the system's performance. RQ1 & RQ2: The main memory's role and organization on a computer's operation are ignored or underestimated. RQ3: This conception shows ignorance in data storage on both systems.
2	The significant role of the main memory is the storing of start-up information.	The only important part of the main memory is ROM. RQ1: This conception disregards the role of the main memory during program execution.
3	a. Volatility is the reason for using the main memory for data storage during program execution. b. A hard disk is non-volatile, so it cannot replace the main memory during program execution.	Volatility is considered the most significant difference between main memory and hard disk and the reason for storing data in the main memory throughout program execution. RQ1: This conception recognizes the role of main memory RQ2: This conception underestimates the actual reason for this role. The main memory's unique organization in ordered addressable cells allows the direct access of a program's data and instructions for sequential execution. RQ3: This conception shows ignorance in data storage on both systems.
4	a. High-speed performance justifies the use of main memory during program execution. b. Hard disk has such a poor performance that it cannot replace the main memory during program execution.	High-speed performance is the most significant difference between main memory and hard disk and explains the use of the main memory to store data during program execution. RQ1: This conception recognizes the role of main memory RQ2: This conception underestimates the actual reason for this role. The main memory's unique organization in ordered addressable cells allows the direct access of a program's data and instructions for sequential execution. RQ3: This conception shows ignorance in data storage on both systems.
5	Virtual memory mechanism	A standalone hard disk can implement the virtual memory mechanism. RQ1: This conception demonstrates ignorance about the role of main memory during program execution.
6	Cache memory operation	A program can run from the cache memory when the main memory is missing. RQ1: The conception implies ignorance about the role of main memory during program execution.
7	Main memory's performance is dependent on the addressing system or the data transfer rate of the bus.	This conception ignores the positive effects of the main memory's electronic nature/circuitry

performance is the reason for the inability of the hard disk to replace the main memory, which is another misunderstanding about what differentiates the two units. Again the difference in data storage is ignored (RQ3). Furthermore, some of the students propose the SSD type of hard disks as a possible substitute for the main memory, considering this technology advancement a step towards replacing the main memory in program execution. The latter misconception is the source of such thoughts. The reasoning is that since the discrepancy in the performance of the two units is eliminated, the hard disk could serve as the main memory. A synopsis of the results of the study (Table 2) is that students: 1) ignore or underestimate the role of main memory in program execution (RQ1); 2) cannot link the main memory's organization in ordered directly accessed addressable bytes/words with the sequential execution of the program's instructions (RQ2), even when they recognize the role of the main memory in program execution, and 3) in many cases overlook the difference of data storage between the main memory and the secondary memory (RQ3).

3.8 Limitations of the Study

Bouvens stated that identifying students' misconceptions from open-ended tests becomes difficult since language problems make students generally less eager to write their answers in complete sentences (as cited in Kaltakci Gurel, Eryilmaz, and McDermott, 2015). Thus, students' potential language problems may be a limitation since all the input data were based on written responses. Nevertheless, the group-discussion sessions allowed students to re-think their opinions, study peers' views, reflect on different ideas and formulate their responses. Thus, group-discussion sessions served as a diagnostic tool that strengthened the results of the open-ended test.

Group A and Group B students had different course instructors during the two academic years. However, the educational setting of the introductory course remained the same, i.e., the educational material (course-books, supplementary electronic material) and the instructional approach (lectures). Even though the particular course books are broadly used in higher education CS introductory courses worldwide, this may limit the extent to which these results may be replicable in studies using different educational material and possibly different teaching methods.

The fact that the data were collected only from the students at one institution imposes another possible limitation to the study. Nevertheless, the students of this department are considered top-qualified students on a national level, as they achieved very high grades on the national entrance exams. Moreover, the input data of the study were collected with a time interval of five years that strengthens the robustness and time persistence of the results.

4 DISCUSSION AND CONCLUSIONS

The analysis of students' answers to understanding questions reflects what exists in their minds. The bridging-inference type of questions based on Kintsch's text comprehension model provides a suitable context to explore students' ideas on a range of concepts. To answer such questions, the reader needs to combine knowledge from several text parts and infer their unstated relations.

Using a bridging-inference question, the study revealed that a considerable percentage of students have such superficial knowledge on the concepts of main memory organization, operation and role, hard disc structure, data storage, program execution, and their interrelations, that in some cases it is scientifically invalid. These naive or intuitive students' conceptions are possibly based on everyday experience, their social network, and knowledge on Informatics which was acquired during the school years (e.g., "beep" sounds in case of RAM failure, volatility of main memory)

In line with Grigoriadou and Kanidis's (2002) research, a significant percentage of students attribute some of the characteristics of secondary memory to the main memory. What is more, those scholars have shown that although most of the school students acknowledged that data are stored on main memory temporarily (volatility) or that the main memory is a high-speed memory, nearly none recognized the role of main memory addressing or the necessity of storing the program on main memory for its execution. These results resemble the misconceptions of volatility and high-speed performance revealed in the present research. The similarity between the school students' and the CS students' conceptions may be explained by the notion that in discourse comprehension, prior knowledge "provides part of the context within which a discourse is interpreted. The context is thought of as a kind of filter through which people perceive the world" (Kintsch, 1988). What is being argued here is that when the CS students (both newcomers and older ones) were exposed to the new knowledge (in text or oral form, by reading the course books or attending the lectures), they constructed discourse representations constrained by their prior school knowledge, among other factors. An implication of this is the need to diagnose CS students' prior conceptions early on before delivering lectures, plan the sequence of topics instructed, and refine and reorganize preconceptions by using specific instructional interventions

Undoubtedly, another important influencing factor of the learning process is the educational material, as presented in the offered course books. Although both are highly accepted and used worldwide in higher education CS introductory courses, a review of these course-books resulted in the observation that the concepts under investigation are not presented in a satisfactory manner. Some knowledge components are described in detail, others briefly, or even confusingly, and information about their linkage is unclear or missing. For example, the Forouzan course book firstly presents the memory hierarchy – registers, cache, main memory accurately - both in text and visual form (see Figure 5.4) so that the readers can become aware of their underlying similarities. But afterward, while presenting the mass storage systems it includes the main memory in a misleading comparison that possibly confuses readers and enhances the misconception of similarity. It is noteworthy that relevant books (e.g the well-known Patterson-Hennessy book (2005) contain visual representations of the memory hierarchy (see Figures 7.1 and 7.3) that compare computer memory with the mass storage systems, referring to them as memory too. This kind of presentation may mislead readers by creating the impression of the similarity of the systems. One immediate implication of this is the need for teachers to review textbooks or other educational material to detect points that may cause misunderstandings. The teachers should emphasize the misleading references during the lectures to prevent the potential consolidation of misconceptions.

Another aspect regarding the educational material is that most of the questions provided at the end of each section in the two course books are text-based, focusing on memorization or comprehension. Consequently, there is a lack of bridging-inference questions that might help or challenge students thinking about the unstated interrelations of the concepts. The bridging-inference question used in the study proved challenging enough for students to attempt to link the topics. This fact implies that the enrichment of the educational material with this type of question seems a promising practice to enhance students' linking-inference skills and consequently their learning.

Furthermore, the effect of the course-book text coherence on the students' understanding of the concepts under study is not negligible. Studies in the domain of text comprehension (McNamara et al., 1996; McNamara & Kintsch, 1996) and specifically on learning from texts on computer science (Gasparinatou and Grigoriadou, 2013) have concluded that high-knowledge readers benefit from a low cohesion text. Conversely, low-knowledge readers benefit from a high cohesion text. The students of the empirical study were mainly first-year students who are low-knowledge readers and do not seem to benefit from the low cohesion text of the course books. This indication is in agreement with the results of the text comprehension studies. A resulting implication of this issue is the need to revise the educational material under the prism of text coherence. That is, to present content with high cohesion texts to facilitate first-year students' understanding of the concepts under study.

In conclusion, formulating bridging-inference questions based on Kintsch's text comprehension model seems an effective tool to assess and investigate students' understanding of CS curricula. Whether the use of this model and especially of well-designed bridging-inference questions, is a promising tool in the direction of supporting knowledge restructuring and refinement, needs to be further investigated.

REFERENCES

Ausubel, D. (1968). *Educational Psychology*. N.Y, Holt, Rinechart & Winston.

Ben-Ari, M. (2001). Constructivism in computer science education. *J. Comput. Math. Sci. Teach. 20*(1), 45–73.

Bransford, D. J., Brown, L. A., & Cocking, R. R. (editors) (2000). *How People Learn Brain, Mind, Experience, and School*, Committee on Developments in the Science of Learning. National Research Council. National Academy Press.

Braun, V., & Clarke, V. (2006). Using Thematic Analysis in Psychology. *J. Qualitative Research in Psychology*, 3(2), 77-101.

Brookshear, J. G. (2009). Computer Science: An Overview, 10th Edition. Pearson Education, Inc, publ as Addison Wesley Higher Education.

Clancy, M. (2004). Misconceptions and attitudes that interfere with learning to program. In S. Fincher and M. Petre, (eds), *Computer Science Education Research*. Taylor and Francis Group, London.

Clements A., (2000). The Undergraduate Curriculum in Computer Architecture, *Computer Architecture Education.* Micro May/June 2000, pp 10-22

Davis, R., Maher, C. and Noddings, N. (1990). Constructivist views of the teaching and learning of mathematics. *Journal for Research in Mathematics Education*, Monograph No.4.

Forouzan, B. A. (2002). *Foundations of Computer Science – From Data Manipulation to Theory of Computation*. Publ. by Brooks/Cole, Thomson Learning.

Gasparinatou, A., & Grigoriadou, M. (2013) Exploring the effect of background knowledge and text cohesion on learning from texts in computer science. *Educational Psychology*, 33:6, 645-670.

Gogoulou, A., Gouli, E., Grigoriadou, M., Samarakou, M., and Chinou, D. (2007) "A Web-based educational setting supporting individualized learning, collaborative learning and assessment," *Educ. Tech. Soc. J.,* vol. 10, no. 4, pp. 242–256.

Grigoriadou, M., & Kanidis, E. (2001). Students' approaches to the Computer Cache Memory and their Exploitation in the Development of a Web-based Learning Environment. *In Proc. of 8th PanHellenic Conf. in Informatics, Cyprus,* 472-481.

Grigoriadou M., & Kanidis E. (2002). Secondary education students' conceptions about computer main memory organization and operation. In Proc. of 3rd Pan Hellenic Conference with International Participation:

Information and Communication Technologies in Education. Rhodes, Sep. 2002, pp. 249-258

Herman, G. L., Zilles, C., Loui, M. C. (2009, October). Work in progress: Students' misconceptions about state in digital systems. *Proceeding of 39th ASEE/IEEE Frontiers Education Conference*, pp. T4D1-T4D2.

Herman, G. L., Loui, M. C., Zilles, C. (2010, October). Work in progress: How do engineering students misunderstand number representations? *Proc. 40th ASEE/IEEE Frontiers Educ. Conf.*, pp. T3G1-T3G2.

Herman, G. L., Kaczmarczyk, L., Loui, M. C., Zilles, C. (2011). Discovering students' misconceptions in boolean logic. *Trans. Comput. Educ.*

Herman, G. L., Loui, M. C., Zilles, C. (2011). Students' misconceptions about medium-scale integrated circuits. *IEEE Transactions in Education.*

Johnson-Laird, P. N. (1983). *Mental models: towards a cognitive science of language, inference, and consciousness.* Harvard University Press.

Kaltakci Gurel. D., Eryilmaz, A., & McDermott, L. (2015). A Review and Comparison of Diagnostic Instruments to Identify Students' Misconceptions in Science. *Eurasia Journal of Mathematics, Science and Technology Education,* 11(5), 989-1008.

Kintsch,W. and van Dijk,T.A. (1978). Toward a Model of Text Comprehension and Production. *Psychological Review*, 85(5), 363-394.

Kintsch, W. (1988). The Role of Knowledge in Discourse Comprehension: A Construction-Integration Model. *Psychological Review*, 95(2), 163-182.

Kyriakou C. and Grigoriadou M. (2016, October). Students' cognitive difficulties on basic informatics concepts. *Proc. of 8th CIE2016 – 8th Conference on Informatics in Education*, pp. 314-326.

McNamara, D. S., Kintsch, E., Songer, N. B., and Kintsch, W. (1996). Are good texts always better? Text coherence, background knowledge, and levels of understanding in learning from text. *J. Cognition and Instruction*, 14(1), 1-43.

McNamara, D. S., & Kintsch, W. (1996). Learning from texts: Effects of prior knowledge and text coherence. *Discourse Processes*, 22, 247–288.

Patterson, D. A., & Hennessy, J. L. (2005). *Computer Organization and Design, The Hardware/Software Interface*/3rd Edition. Publ. by Elsevier Inc.

Porter, L., Garcia, S., Tseng, H. W., & Zingaro, D. (2013). Evaluating student understanding of core concepts in computer architecture. *In Proc 18th ItiCSE, Canterbury, UK*, 279–284.

Posner, G. J., Strike, K. A., Hewson, P., W. & Gertzog, W. A. (1982). Accommodation of a Scientific Conception: Towards a Theory of Conceptual Change *J. Science Education*, 66, 211-227.

Van Dijk, T. A., and Kintsch, W. (1983). *Strategies of discourse comprehension.* San Diego, CA:Academic Press.

Verginis I., Gogoulou A., Gouli, E., Boubouka M., and Grigoriadou M. (2009). Enhancing Learning in Introductory Computer Science Courses through SCALE: An empirical study. *IEEE Transactions on Education*, 54(1), pp. 1-13.

Von Glaserfeld, E. (1995). *Radical Constructivism: A way of Knowing and Learning*, London: The Falmer Press.

Vosniadou, S., & Brewer, W. F. (1987). Theories of knowledge restructuring in development. *Review of Educational Research*, 57, 51-67.

Vosniadou, S. (1994). Capturing and modeling the process of conceptual change. *J. Learning and Instruction*, 4 (1), 45-69.

Vosniadou, S. (2001). *How Children Learn* (Educational Practice Series). Brussels, Belgium: International Academy of Education.

Vosniadou, S. (2003). Exploring the relationship between conceptual change and intentional learning. In G.M Sinatra and P.R. Pietrich (Eds.) *Intentional Conceptual Change*. Mahwah, Lawrence Erlbaum Associates, 377-406.

Vosniadou, S. (2007). Conceptual Change and Education, in *Human Development*, vol. 50, 47–54.