Journal of Turkish Science Education, 2023, *20*(3), 468-489.
DOI no: 10.36681/tused.2023.027

Journal of Turkish Science Education
http://www.tused.org
© ISSN: 1304-6020

Effect of instructional design based on cognitive load theory on students' performances and the indicators of element interactivity

Nesli Kala[1] and Alipaşa Ayas[2]

[1]*Dede Korkut Education Faculty, KafkasUniversity, Türkiye, nkala0702@gmail.com, ORCID ID: 0000-0002-4608-708X*
[2]*Education Faculty, Bilkent University, Türkiye, apayas@bilkent.edu.tr ORCID ID: 0000-0002-4898-2918*

**This research was derived from the first author's PhD dissertation.*

ABSTRACT
Thermodynamics is one of the most complex topics in chemistry. Cognitive Load Theory claims that the complexity of a subject is mainly due to element interactivity - how many elements an individual must organise simultaneously in her/his working memory to master a topic. The simultaneous processing of various chemistry and mathematics concepts to learn thermodynamics puts a strain on the working memory capacity of the learner. Accordingly, what kind of change occurs in a learner's cognitive processes according to the level of element interactivity is an issue that needs to be investigated. The aim of this study is to reveal the basic indicators of element interactivity and investigate the effects of instructional design on understanding subjects with different element interactivity levels. With this objective in mind, educational software comprising eight distinct sessions for instructional design was developed in accordance with the Cognitive Load Theory. The sample consisted of 37 freshmen who were taking classes in the Chemistry Department of a public university in Turkey. The instructional design was implemented with the experimental group while the control group followed the lecturer's instructional design. The results indicate that, in terms of the cognitive load in the learning process, the study time and the learning at the retention and transfer level are among the basic indicators of the element interactivity. This study also determined that the instructional design that is developed according to Cognitive Load Theory can provide effective learning at the retention and transfer levels in subjects with high element interactivity.

RESEARCH ARTICLE

ARTICLE INFORMATION
Received:
11.04.2023
Accepted:
14.08.2023

KEYWORDS:
Cognitive load theory, the indicators of element interactivity, complexity, effective learning, thermodynamics.

To cite this article: Kala, N., & Ayas, A. (2023). Effect of instructional design based on cognitive load theory on students' performances and the indicators of element interactivity. *Journal of Turkish Science Education, 20*(3), 468-489.

Introduction

Complex issues in science are challenging for both students and science educators. The significance of instructional design becomes most apparent in the teaching of difficult science topics. One of the theories that emphasise the significance of instructional design is Cognitive Load Theory (CLT). CLT encompasses the complex cognitive tasks that are manifested through the quantity and interaction of the information that is processed simultaneously, as well as the cognitive processes that take place prior to the initiation of the learning (Paas et al., 2004). CLT is fully contingent upon the understanding that working memory has a limited capacity, meaning that learning environments must be regulated in a way that allows the cognitive load to be carefully distributed over the working memory (Chandler & Sweller, 1991). The current research focuses on instructional design development and indicators of complexity according to CLT.

Cognitive Load Theory

The cognitive load is a poly-dimensional structure that encompasses the specific memory load that affects the learner while a task is being performed (Paas & Merriënboer, 1994). CLT describes the cognitive load with reference to intrinsic load, extraneous/ineffective load, and germane/effective load (Clark et al., 2006; Sweller et al., 1998, 2019). Intrinsic cognitive load is contingent upon both the nature or complexity of the subject that is being learned and the learner's level of experience (Große & Renkll, 2007; van Merriënboer & Ayres, 2005; Sweller et al., 1998, 2019). The extraneous load arises as a result of poorly designed instructional materials and design that put a strain on the working memory. CLT provides a range of design principles that can reduce the extraneous cognitive load. Lastly, the germane cognitive load is manifested in the processes that allow the mental structures to be created and organised (van Merriënboer & Ayres, 2005). Learning can occur when the sum of internal, external and germane cognitive loads does not exceed the capacity of working memory (Sweller et al., 1998, 2019).

Related literature suggests that CLT is effective in learning complex scientific concepts (Carlson et al., 2003; Chang & Yang, 2010; Cierniak et al., 2009; Große & Renkll, 2007; Kala & Okal, 2016; Mousavi et al., 1995; Weng et al., 2018). For instance, thermodynamics concepts are particularly difficult to understand, but they hold great significance in the fields of sciences and engineering (Carson & Watson, 2002; Mulop et al., 2012; O'Connell, 2019; Sreenivasulu & Subramaniam, 2013). The literature shows that students (Cochran, 2005; Kırtak, 2010; Sözbilir, 2002; Sözbilir & Bennett, 2007), as well as teachers and lecturers (Galili & Lehavi, 2006; Kruger et al., 1992; Pinto et al., 2005), have difficulty with basic thermodynamics concepts.

O'Connell (2019) claims that there are five main challenges that explain why individuals tend to fail at learning thermodynamics. These challenges are "scope and level", "mathematical abstraction", "an incomplete discipline based only on equations, not numbers", "laws are always true; models are imperfect but necessary" and "solving problems". According to CLT, the complexity of the subject is related to its "element interactivity" (Clark et al., 2006; Große & Renkll, 2007; Sweller et al., 2019; Sweller, 2020).

Complexity and Element Interactivity

For a learner, complexity may be based on intrinsic cognitive load such as the nature of the subject and learner expertise or may arise from material design or teaching procedures (Chen, Paas & Sweller, 2021; Chen et al., 2018; Sweller et al., 2019). In a subject such as thermodynamics, which the majority accept as difficult, the complexity which can be explained by the element interactivity arising from the nature of the subject plays a leading role.

Element interactivity is the coordination of many information elements working simultaneously to complete a task (Clark et al., 2006; Sweller, 2020). Element interactivity is accepted as a criterion of complexity and is closely related to learner characteristics (Chen et al., 2018; Sweller, 2020). A concept with low element interactivity for an expert student may be high element interactivity since the novice student's schema is limited in this subject (Chen et al., 2017). Grading of element interactivity is a major challenge for studies. For example, Deng et al. (2021) accepted as high element interactivity concepts the concepts of aerobic and anaerobic, which were reached by the integration of the oxygen and density schemes that the students had previously been exposed to, and the complex metabolic system. The degree of element interactivity can be defined by the number of items that are required to understand the subject (Kalyuga et al., 2003; Sweller, 2020). However, it is difficult to determine the number of items in a learning situation, because the number is determined by both the complexity of the information to be learned and the student's prior knowledge (Chen et al., 2017). In this case, after determining students' prior knowledge, it is necessary to decide on low or high element interactivity. Unlike low element interactivity materials, high element interactivity materials consist of elements that heavily interact, meaning they cannot be learned in isolation (Sweller, 2010). Accordingly, since heat, work, system and system types in thermodynamics can be learned in isolation, these concepts are considered low element interactive in current research. However, the learner must understand many different concepts (including system, system types, work, energy, internal energy, the first law of thermodynamics, mole, reaction, reaction types) and have basic mathematical skills in order to learn the concept of enthalpy. Moreover, these items must be processed simultaneously in the working memory. There are two main reasons that this study has chosen to focus on the subject of thermodynamics: 1) As stated in the literature, it is a very complex subject to learn, and 2) As seen in the concepts of system and enthalpy, this topic includes many concepts at various element interactivity levels.

According to CLT, the higher the element interactivity is, the heavier the working memory load will be (Sweller, 2010, 2020). For this reason, element interactivity is the main learning variable considered in many studies. For example, Chen et al. (2021) examined the effectiveness of worked examples and problem-solving tasks in learning chemical formulas. For this purpose, they designed their experiments at two element interactivity levels, low and high. Darejeh et al. (2021) compared the effect of narrative types (no-narrative, familiar and unfamiliar context) on cognitive load while learning a complex productivity software application. Researchers used two types of materials in low and high element interactivity for each narrative type. Buchin (2021) used high and low element interactivity tasks in his research focusing on the effect of prior knowledge on retrieval. In these studies, which are very valuable in terms of literature, element interactivity is a variable of experimental design. Although the aims of these studies are different, they provide some clues about the indicators of element interactivity. Experimental designs consisting of more than one session at different element interactivity levels are needed to investigate indicators of element interaction. Experimental designs in the literature, including the aforementioned studies, consist of 1-2 sessions. Despite the importance of item interaction for CLT, no available comprehensive study explores the kind of indicators teachers encounter in the learning environment if the subject has high element interactivity. Ngu & Phan (2016) analysed the element interactivity levels of these equation types based on the hierarchy and complexity level of linear equations as well as the learners' working memory load. This research, however, is in the form of a theoretical article that based on the researchers' own opinions. Experimental and comprehensive studies are thus needed in order to more fully understand element interactivity. Accordingly, this paper is primarily aimed at developing an instructional design that can provide effective learning according to CLT and to determine the basic indicators of element interaction based on this design. With this in mind, it is thought that this study will contribute to the literature on both determining the indicators of element interaction and how an effective instructional design can be developed based on the element interaction level. The research questions that this study will attempt to respond to are as follows:

1- Regarding instructional design;

a. Is there a significant difference between student scores on the Retention Tests, Transfer Tests, Thermodynamics Achievement Post-Test, and Cognitive Load Scales within and between two groups, provided that the Thermodynamics Achievement Pre-Test and The Digit Span Memory Test scores are controlled?

b. How do students' effective learning scores change at the level of retention and transfer according to both instructional designs?

2- Regarding the indicators of element interactivity;

How do students' achievement at the level of retention and transfer, mental effort, and study time change depending on the level of element interactivity of the topic?

Methods

Research Design

An experimental research design was chosen for this study. The subjects were randomly assigned to experimental and control groups (Cohen et al., 2007; Kline, 2009). The same lecturer (Prof. Dr.) taught both the experimental and control groups, although his teaching roles were different. He taught the whole class with the material he developed in the control group. However, in the experimental group, he summarised the topic using the software only to the students who asked for his help individually.

An eight-session educational software for the experimental group's instructional design was developed in accordance with CLT. The developed instructional design was implemented in the experimental group, while the instructional design developed by the lecturer was implemented in the control group. The lecturer was left free to decide how to develop an instructional design. For example, the lecturer decided which learning theory and model would be chosen and how the lecturer-student and student-student interaction would be. The instructor was given an outcome table on which he would focus according to the courses, and he was asked to make designs based on instructional technologies. As seen in Table 1, two different instructional designs used in the study are summarised in seven items, taking into account the points stated by Smith and Ragan (1999).

Table 1

Features of the Instructional Designs Used in the Experimental and Control Groups.

	Instructional design for the experimental group	Instructional design for the control group
Course outcomes	Course outcomes for eight courses were determined by the researchers at the beginning of the research.	The same course outcomes were shared with the lecturer for the control group to improve the instructional design.
Teaching Style	Individual teaching was carried out in which students could progress at their own individual learning pace. Students were responsible for their own learning.	An expository teaching strategy was carried out by the lecturer.
The Lecturer's Role	The students learned the subject from the educational software, but the lecturer attended the classes to support the students. He provided guided learning support in situations where the students needed it.	The lecturer was utterly in the role of an instructor. He solved many subject-related sample questions by giving a presentation and discussing the subject with his/her students.
Physical Conditions	The computer lab and individual headphones were used. The students learned individually from the developed software.	Lessons were taught in a lecture hall with a computer, sound system, projector, and sliding board.
Content and material	An eight-session software was developed according to the principles of CLT.	An eight-session PowerPoint presentation developed by the lecturer.
In-Class Interactions	The students established a dialogue with the lecturer when needed and received the necessary feedback. The students were also encouraged to quietly share their work with peers.	The students established a dialogue with the lecturer when needed and received the necessary feedback. The students were also encouraged to quietly share their work with peers.
Assessment and evaluation	Thermodynamics Achievement Test, Retention Tests, Transfer Tests, and Cognitive Load Scales were used.	Thermodynamics Achievement Test, Retention Tests, Transfer Tests, and Cognitive Load Scales were used.

The Sample

The sample consists of 37 freshmen chemistry students taking a General Chemistry II course in the chemistry department of a public university in Turkey. The sample in the study was chosen on a voluntary basis. Although the sample was 44 students at the beginning, the research was carried out with a smaller group because some students could not volunteer to participate in the sessions regularly. The students were randomly assigned to either the experimental group or the control group. The experimental group had 18 students (12 female and 6 male) and the control group had 19 students (11 female and 8 male).

Data Collection Instruments

The data was collected using (1) Thermodynamics Achievement Test, (2) Retention Tests, (3) Transfer Tests, (4) Cognitive Load Scales and (5) Digit Span Memory Test. An explanation for each of the tests is provided in the following section.

Thermodynamics Achievement Test

As seen in Table 2, the Thermodynamics Achievement Test (TAT) incorporates the basic concepts and four laws of thermodynamics. The TAT includes 27 items that are based on these basic concepts and laws; the items have been validated by a panel of experts, including chemists and chemical educators. The pilot test was implemented for 67 freshmen after test corrections were made based on those expert recommendations (approximately two weeks after the thermodynamics unit of the General Chemistry II class). An item analysis process was conducted for the test items, which took into account item difficulty and item discrimination (Ebel, 1967). As a result, four of the items were removed from the test, meaning that the final version of the TAT consisted of 23 items. Based on the analyses that were conducted, the Pearson correlation was calculated at 0.93 and the Spearman Brown split half test correlation was calculated at 0.96. The mean difficulty of the TAT was calculated as 0.48, while its mean discrimination was calculated as 0.33. The TAT was used to both determine the students' prior knowledge of the subject before the implementation of the test and to observe their development following the test.

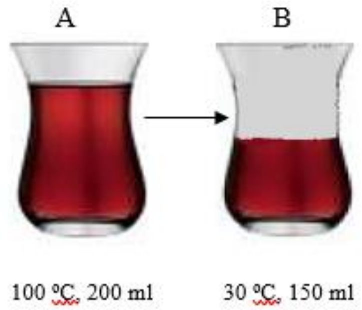
Retention Test

The Retention Test (RT) was developed to assess the students' knowledge after the instruction. Unlike TAT, RTs contain questions only in the knowledge level of Bloom Taxonomy's (1956). The purpose of RTs is not to measure the transfer of knowledge, but to measure what has been learned in that lesson at a basic level. Initially, each of the RTs contained three open-ended questions about the topics that were covered in each session. Two experts in the field of chemistry education examined the clarity of each RT form and whether the questions were relevant to the topics covered in the lesson. In the pilot study, the trial forms of the RTs were also tested. The questions that were answered by almost all of the students were identified, as well as the questions that the students had difficulty answering. In this way, an item with low discrimination was removed from each RT form. As a result, eight RTs were developed in the research, each containing two open-ended questions. Examples of the questions in the RTs are shown below.

- Explain the first law of thermodynamics.
- In what ways can the internal energy of a non-volume-controlled system (e.g. frictionless piston) be changed?

Transfer Test

Since the purpose of CLT is to facilitate the transfer of knowledge from the external environment to the long-term memory and from the long-term memory to the external environment (Sweller, 2020), transfer-level learning was also measured with a separate test. A total of eight Transfer Tests (TT) that corresponded to eight sessions were developed to assess whether the students in the sample were able to apply their learned knowledge to various situations. Each TT contains both near and far transfer questions. If a question could be solved using knowledge of any topic in chemistry, it was considered a *near transfer* question. However, if a question could be solved using knowledge in daily life or in a discipline other than chemistry, this question was considered a *far transfer* question. There are numerous CLT research using near and far transfer questions or tasks in the literature (Buchin, 2021; Chen et al., 2021; Große & Renkll, 2007; Paas et al., 1994). The process that was followed in the development of the RTs was also employed for the TTs. Even though the TTs initially contained three open-ended questions about the subjects of the session, one question from each test was excluded based on the implementation of the pilot test as well as expert opinions. Therefore, each TT contains two open-ended questions. Examples of the questions that appeared in the TTs are shown below.



- Identify whether there is a change in the internal energy of the tea glass between B position and A position.

- $3\ Fe_2O_{3(s)} + H_{2(g)} \rightleftharpoons 2\ Fe_3O_{4(s)} + H_2O_{(g)}$

When the standard gibbs free energy ((ΔG_r^o) of the reaction with an equilibrium constant 2,5. 10^5 at 330 K is - 29 kJ, what can be said for the direction of the reaction at this temperature (R= 8.3145 J / K. mol)?

Cognitive Load Scale

The cognitive load scale (CLS) was developed by Paas and van Merriënboer (1993) in order to assess the difficulty levels of given tasks. Consisting of just a single item, the CLS is a rating scale with 9 categories. The reliability of the scale (otherwise known as the Cronbach Alfa internal consistency coefficient) was calculated at 0.90. When Kılıç and Karadeniz (2004) implemented a Turkish adaptation of the CLS, their Cronbach Alfa internal consistency coefficient was 0.78; also, their Spearman Brown split half test correlation was 0.79.

The CLS was used in this study, though some alterations were made in response to the topics that were covered in each session. Since the content varied from session to session, the number of learning tasks that were given to the students also varied. As such, a total of eight CLSs were formed as a result of the tasks that were to the students in each session. While the base of questions for each item of the scale was unchanged, the addenda (concept, principles, and law) were altered in light of the content.

The Digit Span Memory Test

Miller (1956) showed that working memory can process a limited number of items (5-9). Since CLT focuses on working memory and the finitude thereof, the study places particular importance on the overall determination of the students' memory width. The memory width can determine how many items an individual can process. Therefore, digit span tests are used in research focused on working memory such as CLT (Berends, & van Lieshout, 2009; Klingner et al., 2011; Miller, 1956). The Digit Span Memory Test (DSMT) that was used in this study was translated into Turkish by retrieving the source codes that Sezgin (2009) wrote in the Java programming language for the website http://www.dushkin.com/connectext/psy/ch07/digitspan.mhtml. A reliability study was conducted with the permission of the website authorities that determined that the test has a high level of reliability [r = 0,78, p<0,001].

A Brief Description of Material Development Processes for Instructional Designs

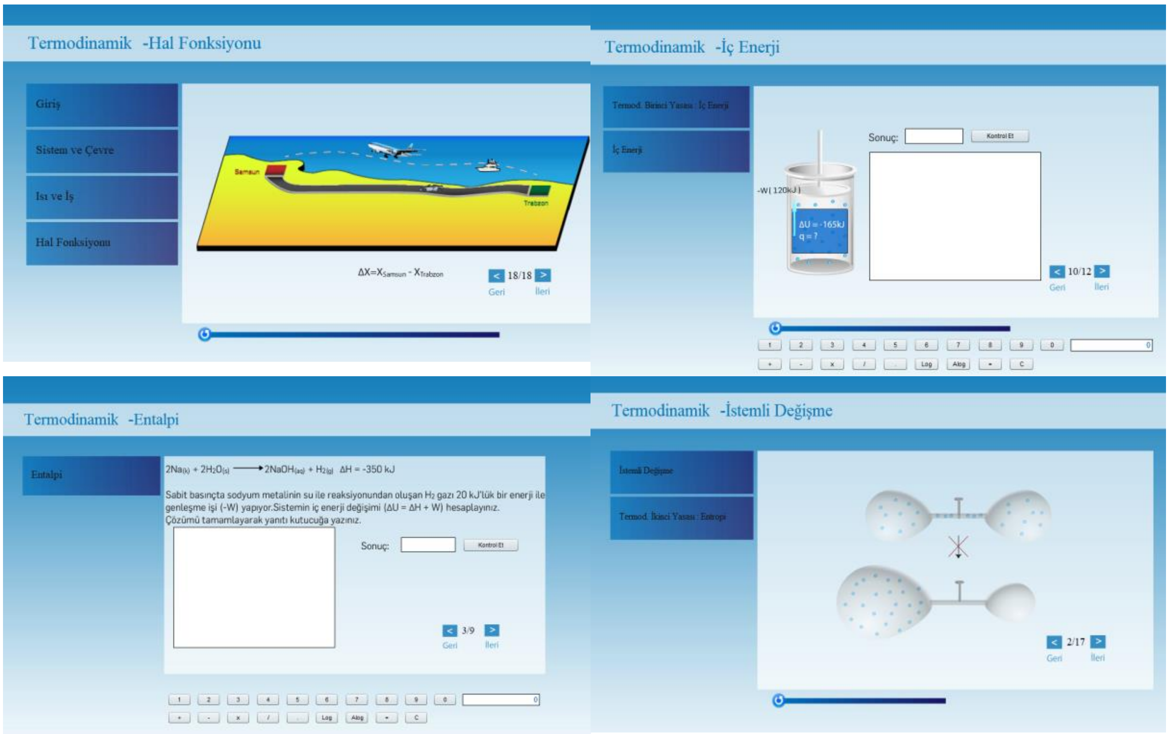
The Development of the Educational Software for the Experimental Group

In this study, educational software was developed to be used in the instructional design, which was itself created with CLT considerations in mind. The scope of the software is based on the observations that are carried out in the classes of four lecturers who taught the General Chemistry II course in the Chemistry Department of the Faculty of Sciences. During this period, interviews were conducted with the lecturers and a group of students. The content was divided into eight parts, as the lecturers who taught General Chemistry II stated during their interviews that they allocated 8-9 hours for the thermodynamics unit. In light of the preliminary study, it was decided to develop the educational software as a guided learning module based on learner interaction. This software is designed in such a way that the student can both learn the subject and solve exercises related to the subject. Moreover, with this software, the student will be able to get feedback according to his/her problem solution. The content of educational software was arranged according to some principles of CLT, including the goal-free effect, worked example effect, completion problem effect, split-attention effect, modality effect, and redundancy effect. For the scaffolding in the software, worked examples of problems and the completion problems were presented. With regard to the completion problems, the software offered feedback based on the student's answers. In addition, the texts in the interfaces with visual presentations, according to the modality effect, were presented as voiced (Figure 1). In the application process, each student was provided with individual headphones to ensure that they were not distracted by the other computers. As the learners were responsible for transitioning between interfaces, the students could progress according to their individual pace. Also, in order to prevent the risk of the application process being affected by internet interruptions, the software was desktop-based (meaning that it was not reliant on internet connectivity or downloads).

After 8 experts, 4 of whom were in the field of chemistry and 4 in the field of educational sciences (2 chemistry education, 1 curriculum development, and 1 instructional technology), analysed the draft software, some changes were made in line with their suggestions. The pilot implementation then took place. All of these steps were intended to improve the material before the main study was rendered. The number of interfaces for each session of the software and the subjects each session contains are shown in Table 2. The interface examples of the developed software are shown in Figure 1.

Figure 1

Some Interface Examples of the Software



The Development of the Instructional Material for the Control Group

A few weeks before the implementation, the content and the course outcomes of the eight sessions were given to the lecturer. This lecturer is one of the experts whose opinion we take into account when dividing the teaching content into eight sessions. In addition, the data collection tools that were used in the research were provided to the lecturer at the onset of the study. The lecturer was told that he could improve upon his own instructional design and use technology at his discretion. The lecturer was informed that he could develop a completely computer-based design if he chose to and that he could be given technical support if needed. However, as the lecturer preferred to use the computer and projector in the presentation phase of the sessions, he chose to prepare PowerPoint presentations for eight sessions.

Procedures

The Pilot Study

After the draft form of the software was developed in the research, a pilot study was conducted. The pilot study was conducted with one of the four lecturers whose course was observed during the software development process. This lecturer was supported to develop his own instructional design in eight sessions by following the steps mentioned earlier. During the pilot study, the issues that were not understood by the students in the software, the RTs, and TTs were determined, and these tools were revised. The pilot study was carried out with approximately 60 freshmen chemistry students. However, the main study was carried out with a smaller group since fewer students were enrolled in this department in the year of the main application. Since the number of students decreased, the lecturer who taught the course in the pilot study could not take the General Chemistry II course the following year. The main study was carried out with one of the four lecturers whose course was observed during the software development process and whose opinion was sought as a field expert during the software development process. Depending on the lecturer, the instructional design also changed, but with the same steps in the pilot study, a second lecturer developed his own instructional design. The steps of the main research are presented in detail below.

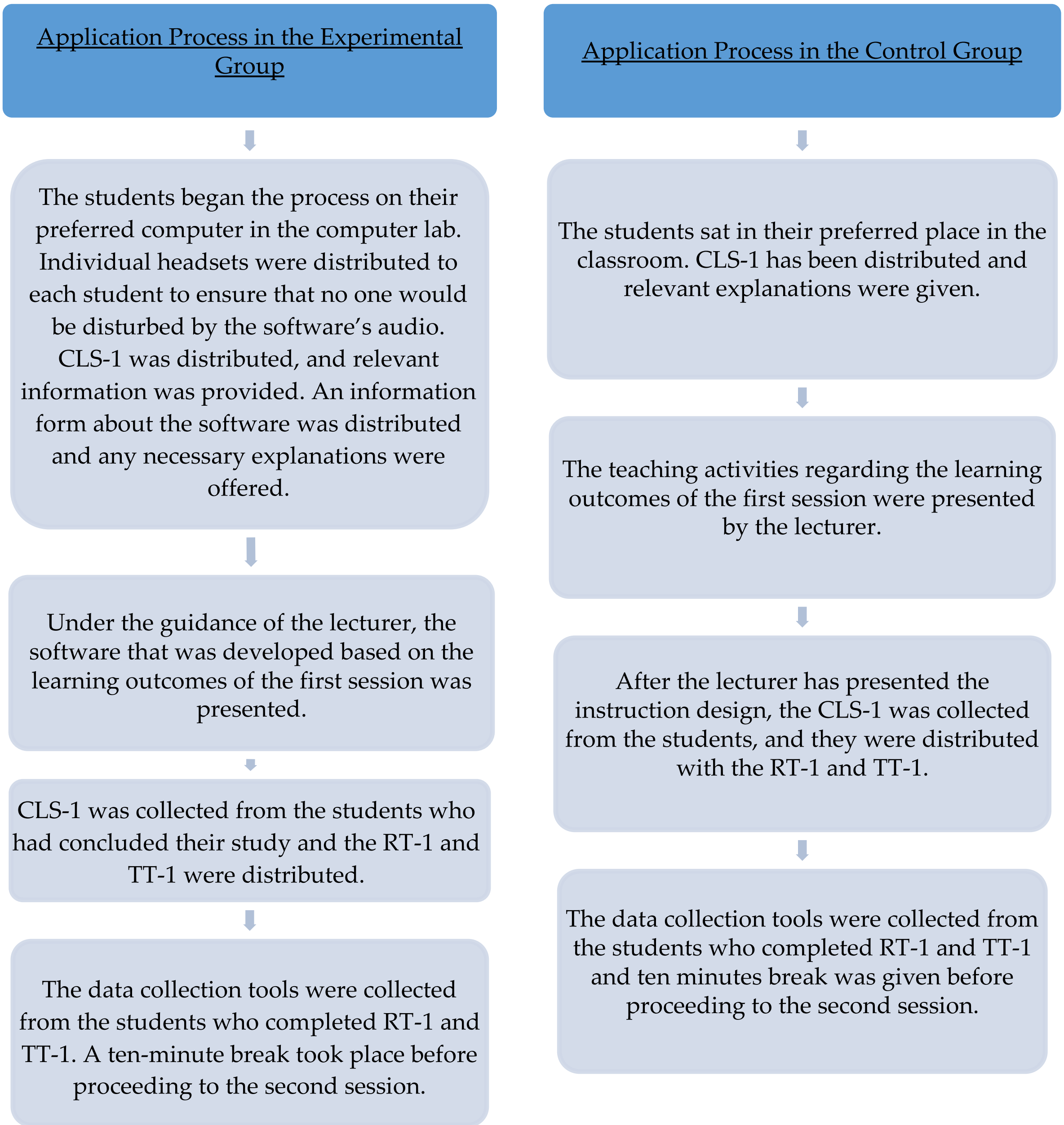
The Main Study

After the students in the sample were randomly assigned into experimental and control groups, the TAT was implemented as a pre-test in order to determine the students' prior knowledge on the thermodynamics unit. Then, the DSMT was taken from the website that was specified in the relevant section above and was given to each student individually in order to determine their memory width. The main study started after the DSMT was implemented.

The software that was developed according to CLT was used in the experimental group. The lecturer presented all of the content himself in the control group and provided guidance in the learning process when needed. In addition, the students in the control group were encouraged to seek out and provide peer support. In the experimental group, the students primarily studied by themselves on their computers, though they also had access to available support from their peers and the lecturer. The students completed two sessions each day. As is highlighted in the flow diagram, there was a ten-minute break between each session. One week after completing the experiment, the TAT was applied to both groups as a post-test. The flow diagram of the first session is provided in Figure 2 for exemplary purposes, with the exact same steps followed in every session of the application process.

Figure 2

The Flow Diagram of the Application Process in the Experimental and Control Groups First Session Sample



The students in the experimental group were free to work on the software as long as they preferred, meaning that certain sessions took more time. Table 2 shows how much time the students in the experimental group dedicated to learning the subject matter for each session. In the control group, how long the session would be allocated was under the control of the lecturer. Although the experimental design was planned as eight-course hours for both groups, an additional four-course hours applied course was given in the control group after the planned training was completed. Since the RTs, TTs and CLSs were applied during the sessions, the extra training of the control group did not affect the results of these three tools. However, since TAT was applied after the training was completed, it may have affected the outcome of TAT in favour of the control group. Since this effect was in favour of the control group, the study was continued. As a result, the experimental group used eight-course hours while the control group used 12-course hours (8 hours for the tutorial and 4 hours for the practice) for the training.

Analysis

In this study, the data obtained from the TAT, RTs, TTs, CLSs and DSMT were analysed using a statistical package software.

Analysis of the Data Obtained from the RT and the TT

As previously mentioned, this study applied eight RTs and eight TTs after each session. Answer keys and rubrics were also developed for each test. The researchers compared the scores they gave after evaluating the TTs and TTs according to these rubrics. In cases where there was a scoring difference, a consensus was reached by discussing. Each test was rated for a total of 100 points. The RT and the TT, which were both used to assess academic achievement, were used in tandem with the CLS to calculate the students' effective learning scores.

Analysis of the Data Obtained from the CLS

A total of eight CLSs were created for each of the students' tasks. The mean scores of the CLS were calculated after each session and a cognitive load score for each student was obtained. The CLS was used to determine the load and, together with the RTs and TTs, to calculate the students' effective learning score.

Analysis of the TAT

In analysing the TAT, a score was determined for each student by designating 1 for correct answers and 0 for blank or wrong answers. The maximum score a student could achieve was 23. Since the TAT was used before and after the study as a pre- and post-test, it was designated as covariate in the analysis for the overall study. In addition, the data that was obtained from the TAT post-test was used in tandem with the cognitive load score in order to calculate the effective learning score.

Statistical Analysis

Firstly, to determine the equivalence of the groups in advance, the study carried out independently sampled t-tests to determine how the groups varied in terms of prior knowledge about the thermodynamics unit, and memory width. Although the research was carried out with a small group, parametric tests were used because preconditions were met. In the literature, there is much CLT research in which parametric tests are applied to small groups (Chen et al., 2021; Darejeh et al., 2021; Leahy et al., 2015). Multi-factor variance analysis (MANOVA) was carried out to determine of the effect of the independent variable, of which effect had been researched, over variables such as the RTs, the TTs, and the CLSs (Cohen et al., 2007). Since CLT attests that both memory width and prior knowledge are among the elements that affect learning, the DSMT and the TAT (pre-test) were designated as covariant.

The data obtained from the sessions were analysed descriptively to determine the indicators of element interaction. In addition, to interpret the effectiveness of instructional design in the experimental group in comparison to the control group, the effective learning scores (E) of the students were calculated based on their cognitive load Z (CLZ) and the performance Z (PZ) scores. According to CLT, effective learning is the emergence of high learning performance without too much cognitive load. The effective learning score ($E = PZ - CLZ / \sqrt{2}$) was developed by Paas and van Merriënboer (1993, 1994) to evaluate the relative effectiveness of various learning methods.

With regard to the second research question of the study, the subjects in the thermodynamics unit were classified according to the element interactivity level. This classification is in line with the opinions of two chemistry experts (lecturers) who have taught General Chemistry II at the university for over ten years each. The lecturers were informed about the concept of element interactivity, at which point a table with sessions and topics was provided to them. The lecturers were asked to divide the sessions into high and low interactive elements based on the session topics. After the lecturers opined that some of the sessions contained topics that had very high interactive elements, we decided to categorise these elements as either low, high or very high. For example, since all of the concepts that were included in the first session could be taught independently from each other, they were classified as having low interactive elements. In the second session, the students were asked to learn the first law of thermodynamics and internal energy, which required converging the eight concepts from the first session that were stored in working memory with basic mathematics knowledge; for this reason, the second session was considered to have high interactive elements. As the third and eighth sessions introduced such concepts as enthalpy and free energy change, these sessions were classified as having very high interactive elements. This is because the students were required to process several elements from working memory, and because these concepts are very abstract, meaning they can only be understood via mathematical equations. It should be noted that this classification system does not consider the subjects in thermodynamics as they compare to other chemistry subjects, but rather how the subjects compare within the unit itself. Table 2 captures the number of interfaces for each session of the software, the subjects each session contains, and the element interaction level of these subjects.

Table 2

The Content of each Session of the Educational Software and the Approximate Time the Students in the Experimental Group Required to Learn the Subjects

Sessions	The subjects of the session	The level of element interactivity	Number of interfaces in the software	Mean study period of the subjects in the session	Mean time for an interface
First session	System, environment, open system, closed system and isolated system, heat, work and state function	Low element interactivity	18	25 minutes	1.39
Second session	The first law of thermodynamics and internal energy	High element interactivity	12	32 minutes	2.67
Third session	Enthalpy	Very high element interactivity	9	35 minutes	3.89
Fourth session	Spontaneous change, the second law of thermodynamics and entropy	Low element interactivity	17	12 minutes	0.71
Fifth session	Standard entropy, absolute entropy and the third law of thermodynamics	High element interactivity	11	20 minutes	1.82
Sixth session	The total entropy change	High element interactivity	12	25 minutes	2.08
Seventh session	The Gibbs free energy, standard reaction energy and phase shifts	High element interactivity	15	37 minutes	2.47
Eighth session	The free energy change and equilibrium, and zeroth law of thermodynamics	Very high element interactivity	14	43 minutes	3.07

Findings

Findings Related to the First Research Question

Regarding the first research question, firstly, the equivalence of the groups in terms of various variables was tested. The TAT pre-test score of the students in the control group ($\overline{X}$ =4.53, Sd=2.04) is higher than that of the experimental group ($\overline{X}$ =3.88, Sd=2.17), though this difference is not significant [t(35)= .35, p= .36, p> .05]. In addition, there is no significant difference [t(35)= .54, p= .59, p> .05] in terms of their DSMT scores between the students of the experimental ($\overline{X}$ =10.38, Sd=2.12) and the control groups($\overline{X}$ =9.95, Sd=2.76). It is evident, therefore, that both groups are statistically equivalent.

Since RT, TT and CLS were applied in all eight sessions, the arithmetic mean of these data collection tools was taken to perform statistical operations. The mean score and standard deviation for both groups in terms of the RTs, TTs, TAT and CLSs are shown in Table 3.

Table 3

The Descriptive Analysis of the Mean Scores of the RT, TT, and CLS for Both Groups

Data Collection Instruments*	Group	N	$\overline{X}$	Sd
RT	Experimental	18	77.07	11.15
	Control	19	38.48	13.55
TT	Experimental	18	62.96	12.39
	Control	19	29.67	9.44
TAT Post-test	Experimental	18	10.39	3.20
	Control	19	8.16	3.96
CLS	Experimental	18	3.34	1.07
	Control	19	2.70	.94

Note. Max score of the RT and TT are 100. Max score of the CLS is 9. Max score of the TAT is 23.

As captured in Table 3, there was a difference of approximately 40 points between the mean scores of the students in the experimental group for the RT ($\overline{X}$ = 77.07) and those of the control group ($\overline{X}$ = 38.48). Similarly, the mean scores of the students in the experimental group in the TT and TAT post-tests and the CLS are higher than those in the control group.

Since covariance equality was provided at the beginning of this study (according to Box's M test, F(10–5813.61)=0.92, p=0.51, p>0.05) for MANOVA, the Wilks' Lambda test was used to interpret the effect of the instructional design variable over the RT, TT, CLS and TAT post-test application. According to this test result, the memory control variables had a significant relationship with the dependent variables at the moderate level [(λ)= .87, F(4-30)=1.08, η_p2= .13, p= .39, p> .05] and the TAT pre-test control variable had a significant relationship with the dependent variables at the high level [(λ)= .94, F(4-30)= .46, η_p2= .06, p= .77, p> .05]. In addition, the independent variable had a significant effect on the dependent variables [(λ)= .18, F(4-30)= 33.43, η_p2= .82, p= .00, p< .05]. A calculated multi-factor variance analysis (MANOVA) of these findings suggested that there was a significant difference between the RT [F(1-33)= 83.74, η_p2= .72, p= .00, p< .05] and TT [F(1-33)= 78.91, η_p2 = .71, p= .00, p< .05] scores of the students in both groups, though there was no significant difference between the TAT post-test [F(1-33)= 3.42, η_p2 = .09, p= .07, p> .05] and CLS scores of groups [F(1-33)= 3.14, η_p2 = .09, p= .09, p> .05].

The calculation of the effective learning scores is significant because it has an influence over determining the contribution to the learning of the design prepared according to CLT. The performance Z scores were calculated according to the TAT post-test, RTs, TTs, and CLSs and the effective learning scores were calculated in these three standards of the students who learn the subject of thermodynamics through different designs in the experimental and control groups. The results are shown in Table 4.

Table 4

Performance Z Scores, Cognitive Load Z Scores and Effective Learning Scores of the Students in the Experimental and Control Groups

Data Collection Instruments	Group	N	Performance Z Score	Cognitive Load Z Score	Effective Learning Score (E)
RTs	Experimental	18	.86	.31	.39
	Control	19	- .81	- .30	- .37
TTs	Experimental	18	.85	.31	.38
	Control	19	- .81	- .30	- .36
TAT Post-test	Experimental	18	.31	.31	- .01
	Control	19	- .29	- .30	.01

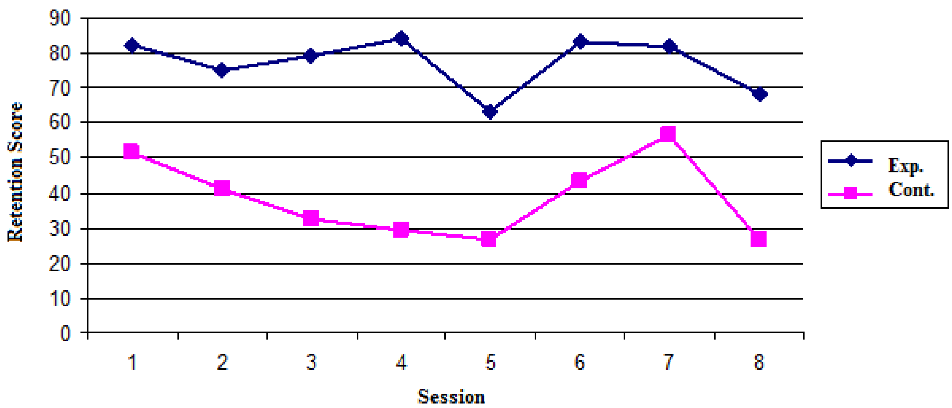
As shown in Table 4, all performance Z scores (RT, TT, and TAT post-test) and cognitive load Z scores of the experimental group are higher than that of the control group. According to the retention and transfer effective learning scores calculated on the basis of these scores, it was detected that the score of the experimental group is higher than that of the control group. Even though the TAT post-test Z score of the experimental group was higher than the control group, the thermodynamic academic achievement effective learning scores (E= - .01) were almost equal to the control group (E= .01) since their mental effort scores are also higher.

Findings Related to the Second Research Question

The distribution of RT, TT, and CLS's according to the sessions was examined in order to determine what changed in the different element interactivity levels. The distribution of the RT scores over each session is shown in Figure 3.

Figure 3

Retention Score Distribution of the Students in the Experimental and Control Groups

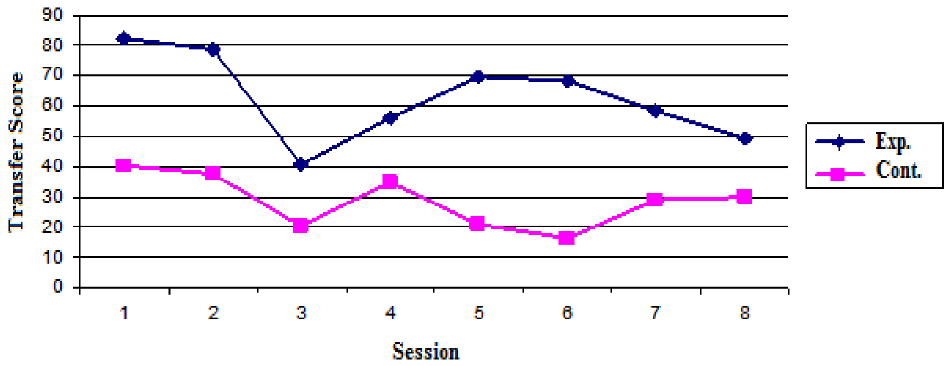


Note. Sessions 1 and 4 include low element interactive subjects; Sessions 2, 5, 6 and 7 include high; and sessions 3 and 8 include very high.

As seen in Figure 3, students in both groups got lower scores in Sessions 3, 5, and 8 compared to the other sessions. The sessions in which the groups differed are Sessions 2 and 4. The experimental group got low scores in Session 2 and the control group got low scores in Session 4.

Figure 4

Transfer Score Distribution of the Students in the Experimental and Control Groups

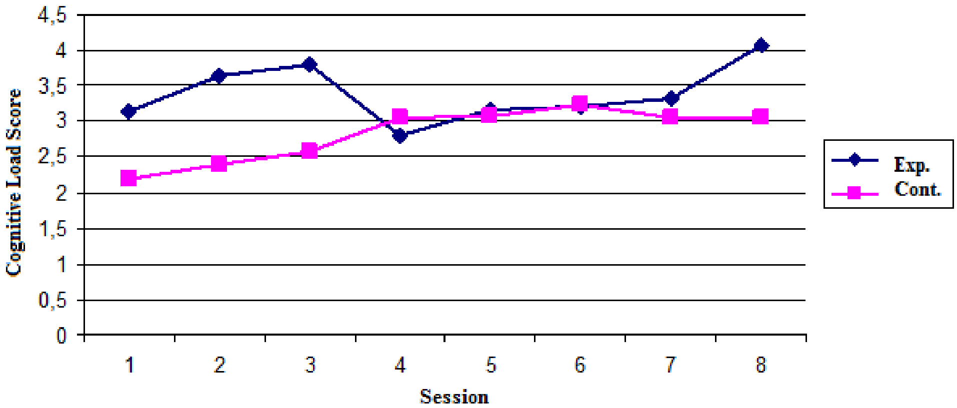


Note. Sessions 1 and 4 include low element interactive subjects; Sessions 2, 5, 6 and 7 include high; and sessions 3 and 8 include very high.

At the transfer level, it is seen in Figure 4 that all students scored lower in Sessions 3, 7, and 8 compared to the other sessions. The students in the experimental group got low scores in Session 4, while the students in the control group got low scores in Sessions 5 and 6. In addition, the transfer scores of the students in both groups were significantly lower in comparison to those of the first two sessions.

Figure 5

Cognitive Load Score Distribution of the Students in the Experimental and the Control Groups



Note. Sessions 1 and 4 include low element interactive subjects; Sessions 2, 5, 6 and 7 include high; and sessions 3 and 8 include very high.

While there were substantial differences in the cognitive load values between the students in experimental and control groups during Sessions 1, 2, 3, and 8, Figure 5 suggests that the cognitive load values in the other four sessions are very close. While the students in the experimental group dedicated a minimum amount of mental effort to learn the subjects that were covered in Session 4, they made the most mental effort in Sessions 3 and 8. Also, the cognitive load of the students in the control group was lower in the first three sessions compared to the final five.

Discussion and Implications

The present research is aimed to develop an instructional design that can provide effective learning in accordance with the indicators of element interaction. When the results of the eight sessions were analysed statistically, we found a significant difference between the mean scores of the RTs and TTs in favour of the experimental group. However, there is no such relevant difference between the mean scores of the TAT and the CLS. The difference between the scores of the students in the experimental group between the pre-and post-test of the TAT is 6.51, while the difference in the control group is 3.63. Although this difference is not statistically significant, it is noteworthy when considering that there are a total of 23 questions in the TAT. Moreover, the lack of difference in TAT can be explained by the fact that the lecturer taught extra 4-course hours to the control group before the TAT. As the application process of the experimental group had been 33% shorter than that of the control group, it could be inferred that CLT is effective in achieving substantial success over a short period of time.

When the memory and TAT pre-test scores of both groups are controlled, a relevant difference between the scores of the RTs, TTs, CLSs and TAT post-test is detected according to the MANOVA test. This observed difference in the students of the experimental group originates from the instructional design, which was developed according to CLT. Indeed, previous literature suggests that instructional designs that are created in accordance with CLT principles succeed in similar difficult-to-learn subjects (Cierniak et al., 2009; Chang & Yang, 2010; Leahy et al., 2003; Moreno & Mayer, 1999; Sweller & Chandler, 1994; Tindall-Ford et al., 1997; Weng et al., 2018). In addition, even though there is a slight difference in the students' cognitive load levels in favour of the control group, the experimental group's students achieved effective learning both at the transfer level and the retention level. In the study by Chen et al. (2021) comparing the Example-Problem and Problem-Example groups, they found that the groups differed statistically in the high element interactivity transfer test, not in the low element interactivity retention test. This finding shows that the real impact of an instructional design emerges in transfer-level learning. While researchers asked only close transfer questions in the transfer test, in the current study, apart from near transfer questions, far transfer questions with much more element interactivity were also asked. This result shows that the instructional design that was created for the experimental group satisfied the study's goal of increasing the germane cognitive loads by reducing the students' extraneous cognitive loads, regardless of the high element interaction of thermodynamics.

In this study, the students in the control group demonstrated low performance and a low cognitive load, while the students in the experimental group demonstrated high performance and a partially high cognitive load. According to the cognitive load classification of Paas and van Merriënboer (1993), the CLS scores of the students in both groups are around 3 over a maximum of 9, meaning that they are in the low category of the scale. In the current study, the students learned concepts in the thermodynamics unit, where there is an excessive amount of element interactivity that is quite difficult to learn. According to CLT, it is inevitable for an individual's intrinsic cognitive load to be high if he or she is learning something that has excessive amounts of element interaction (as is the case here). Also, if the individual has established schemata, performed schema automation, or expanded their existing schemata through adaptation, their germane cognitive load would be high. If the subject of thermodynamics is as difficult as the literature demonstrates (Carson & Watson, 2002; O'Connel, 2019; Sreenivasulu & Subramaniam, 2013), then why did the students in the control group load so little while learning these topics? If the underload in the control group is due to good instructional design (meaning the extraneous cognitive load is low), then why is the success so low?

An analysis of the current Turkish Education System can help inform these queries. Even though the Turkish Ministry of National Education has adopted a constructivist approach for the last two decades, teaching is still mostly teacher-directed. This format gives the students a greater sense of confidence during the lectures and the responsibility of teaching is entirely on the lecturer. With this in mind, even though the students in the control group demonstrated low performance in the retention and transfer levels, it is assumed that the students are loaded less since the learning responsibility has been transferred to the lecturer. In other words, it is worth considering that the students in the control group did not try to learn the subject because they trusted the lecturer and therefore did not concern themselves with the difficulty of the topic. Conversely, we can assume that the students who learned the topic via educational software felt responsible for their own learning and their own self-awareness. Consequently, the low cognitive load may result from good instructional design or when the learning does not take place.

It is understood from the finding that while the students in the experimental group got the highest scores from Sessions 1 and 4 in terms of RTs, they got the lowest scores from Sessions 5 and 8 (Figure 3). At the retention level, the students in the control group got the highest scores from Sessions 1 and 7, and the lowest scores from Sessions 3, 4, 5, and 8 (Figure 3). When these findings were compared with the element interactivity classifications of the experts (Table 2), it was determined that all students in Session 5 and students in the control group in Session 4 scored lower than expected. In addition, the control group students getting the highest score from Session 7 does not match the element interactivity level determined by the experts. According to the TTs, the students in the experimental group got the highest points from Sessions 1 and 2, and the lowest from Sessions 3, 4, and 8 (Figure 4). In the TTs, the students in the control group got the highest scores from Sessions 1 and 2, and the lowest scores from Sessions 3, 5, and 6 (Figure 4). When RTs 4, 5, and 6 and TTs 4, 5, and 6 were examined to interpret these findings, it was seen that these tests included conceptual rather than procedural questions. Contrary to this situation, RTs 2 and 7 and TTs 2 and 7 mostly contain procedural questions. Multiple choice questions are used in university entrance exams in Turkey, which is known to be procedural rather than conceptual (Bekdemir et al., 2010; Birgin & Gürbüz, 2009; Kaya & Keşan, 2012). In addition, during the application process of the control group, it was observed that almost all of the questions used by the lecturer to teach the concepts were procedural. From this point of view, it is thought that especially the students in the control group have difficulty in solving conceptual questions, both because of the influence of the university entrance exam because they are freshmen, and because of the teaching style of the lecturer. As a result, although learning is affected by other variables such as teaching style, it is directly affected by the element interactivity (Figures 3 and 4). In other words, it was determined that as the element interactivity level of the topics learned in the study increased, learning became more difficult and scores decreased.

In the current study, how mental effort changes according to element interactivity was also examined. As can be seen in Figure 5, although the students in the control group had a slightly higher load in the first three sessions, there was not much difference between the sessions in terms of loading. This may be due to the fact that the learning control is in the lecturer and the students cannot focus on their own cognitive processes enough. Unlike the control group, the students in the experimental group spent the most mental effort in Sessions 3 and 8, and the least mental effort in Sessions 1 and 4. In this respect, the loading situation in the experimental group is fully compatible with the element interactivity classifications of the experts. Although there is no study in the literature that empirically examines the relationship between element interactivity and loading, in terms of CLT, cognitive load is expected to be high when an individual is learning complex subjects (Leahy & Sweller, 2016; Blayney et al., 2016; Darejeh et al., 2021). In this study, it is important in terms of the literature to reach the conclusion that more mental effort is spent while learning topics with high element interaction by comprehensive study.

It is possible, as well, that another indicator of element interactivity is the study time of the subject. Since the learning time of each session in the control group was under the control of the lecturer, this variable could only be examined in terms of the experimental group. The studying time per slide of the students in the experimental group is listed from most to least as Session 3, 8, 2, 7, 6, 5, 1, 4 (Table 2). The mental effort of the same group according to the sessions is listed from most to least as Session 8, 3, 2, 7, 6, 5, 1, 4 (Figure 5). From this, it is seen that the order of loading and studying time of the students in the experimental group according to the sessions is almost exactly the same. The first two sessions of these two rankings (Sessions 3, 8) contain very high-item interactive topics, while the last two sessions (Sessions 1, 4) contain low-item interactive topics. In other words, these two rankings completely coincide with the element interactivity level of the experts. Though there are no studies available in the literature that directly examine the relationship between study time, loading, and element interactivity, certain studies have indirectly assessed the relationship between study time and loading (Cierniak et al., 2009; Mousavi et al., 1995). Darejeh et al. (2021), in their research examining learning according to different narrative types, measured test-solving time, not learning time, according to element interactivity level. This research supports the current research because the researchers found that in all narrative types, students loaded less on the low element interactivity subject than the high element interactivity subject and answered the relevant test in a shorter time. Unlike the literature, this research revealed with a broad perspective (eight sessions) that if a topic's element interactivity is high, then the students will dedicate more time and mental effort to learning it. Therefore, the research contains important results in terms of literature.

Conclusion

The present research has two important conclusions. The study showed that the cognitive load in the learning process, the study time, and the learning at the retention and transfer level are the indicators of element interactivity. This research also concludes that instructional design that is developed according to CLT can provide effective learning at both the retention and transfer levels of subjects with high element interactivity. Teachers are responsible for determining a topic's level of element interactivity and developing an appropriate instructional design. The chemistry curriculum for high schools in Turkey contains many topics that have very high element interaction, including thermodynamics. The students there also participate in a university admittance examination that references the entirety of this curriculum at the end of high school (lycee). Based on the fact that the instruction design that was created for the experimental group facilitated effective learning over a shorter period of time when compared to the control group, it is evident that structuring the chemistry curriculum in accordance with CLT will facilitate a higher degree of knowledge over a short period of time.

Limitations of the Study

In quantitative research, sample size is a significant consideration. The main limitation of this study is that it was only able to account for a small sample size (37 volunteer students in total). One of the reasons for this limitation is that the software for thermodynamics was developed according to the General Chemistry-II course in the chemistry department. Since the software was developed for a specific target audience, the sample was limited to first-year students in the chemistry department. Also, some students in the first year did not volunteer to participate in the study because the research would take about a month, including the application of the pre- and post-tests. These students were excluded from the study because volunteering was essential in the study and the study was conducted with 44 students. Those students who did not attend all of the sessions were also excluded from the sample, thereby culminating in a small sample size.

In addition, the indicators of element interactivity are limited to the concepts of achievement at the levels of retention and transfer as they relate to mental effort and study time. Though we were only able to examine general achievement, different levels of learning were also included in order to procure more decisive results. In addition, since the students in the control group learned the topic from a single source (the lecturer), the effect that changes in study time had on element interactivity could not be examined in the control group. Therefore, the indicator was limited by the findings that were obtained from the experimental group. In fact, the research could only determine the indicators of element interactivity based on one group, although a second teaching design was integrated into the study to determine whether different designs could render similar findings.

Funding: This study was not funded by a grant provided by an external source, but the researchers were staff members at the university and researchers' salaries were paid when the research was undertaken.

Conflict of Interest: The authors declare that they have no conflict of interest.

Ethical Approval: All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

References

Bekdemir, M., Okur, M., & Gelen, S. (2010). The effects of 2005 elementary mathematics education curriculum on the elementary seventh grade students' conceptual and procedural knowledge and skills, *Erzincan University Journal of Education Faculty*, *12*(2), 131-147.

Berends, I. E., & Van Lieshout, E. C. (2009). The effect of illustrations in arithmetic problem-solving: Effects of increased cognitive load. *Learning and Instruction*, *19*(4), 345-353. https://doi.org/10.1016/j.learninstruc.2008.06.012

Birgin, O., & Gürbüz, R. (2009). Investigation procedural and conceptual knowledge of Primary second Level Students about the rational numbers (İlköğretim II. Kademe Öğrencilerinin Rasyonel Sayılar Konusundaki İşlemsel ve Kavramsal Bilgi Düzeylerinin İncelenmesi). *Journal of Uludağ University Faculty of Education, 22* (2), 529-550.

Blayney, P. Kalyuga, S., & Sweller, J. (2016). The impact of complexity on the expertise reversal effect: experimental evidence from testing accounting students, *Educational Psychology, 36* (10), 1868–1885. https://doi.org/10.1080/01443410.2015.1051949

Bloom, B. S., Engelhart, M.D., Furst, E.J. Hill, W. H., & Krathwohl, D. R. (1956). *Taxonomy of educatıonal objectıves: The classification of educational goals*. Longman.

Buchin, Z. L. (2021). *Retrieval-based learning and element interactivity: The role of prior knowledge*. [Unpublished Doctoral dissertation]. The University of North Carolina at Chapel Hill.

Carlson, R., Chandler, P., & Sweller, J. (2003). Learning and understanding science instructional material. *Journal of Educational Psychology, 95,* 629-640. https://doi.org/10.1037/0022-0663.95.3.629

Carson, E. M., & Watson, J. R. (2002). Undergraduate students' understandings of entropy and gibbs free energy. *University Chemical Education*, *6*, 4–12.

Chandler, P., & Sweller, J. (1991). Cognitive Load Theory and the format of instruction. *Cognition and Instruction.* 8 (4), 293–332. https://ro.uow.edu.au/edupapers/128

Chang, C.C., & Yang F.Y. (2010). Exploring the cognitive loads of high-school students as they learn concepts in web-based environments. *Computers and Education*, *55*, 673-680. https://doi.org/10.1016/j.compedu.2010.03.001

Chen, O., Castro-Alonso, J. C., Paas, F., & Sweller, J. (2018). Extending cognitive load theory to incorporate working memory resource depletion: evidence from the spacing effect. *Educational Psychology Review, 30*, 483-501, DOI 10.1007/s10648-017-9426-2

Chen, O., Kalyuga, S., & Sweller, J. (2017). The expertise reversal effect is a variant of the more general element interactivity effect. *Educational Psychology Review*, *29*(2), 393-405. https://doi.org/10.1007/s10648-016-9359-1

Chen, O., Paas, F., & Sweller, J. (2021). Spacing and interleaving effects require distinct theoretical bases: A systematic review testing the cognitive load and discriminative-contrast hypotheses. *Educational Psychology Review*, *33*(4), 1499-1522. https://doi.org/10.1007/s10648-021-09613-w

Chen, O., Woolcott, G., & Kalyuga, S. (2021). Comparing alternative sequences of examples and problem-solving tasks: the case of conceptual knowledge. *The Educational and Developmental Psychologist*, *38*(1), 158-170. DOI: 10.1080/20590776.2021.1915098

Cierniak, G., Scheiter, K., & Gerjets, P. (2009). Explaining the split-attention effect: Is the reduction of extraneous cognitive load accompanied by an increase in germane cognitive load?. *Computers in Human Behavior*, *25*, 315–324. https://doi.org/10.1016/j.chb.2008.12.020

Clark, R.C., Nguyen, F., & Sweller, J. (2006). *Efficiency in learning: Evidence-Based guidelines to manage cognitive load*. Published by Pfeiffer.

Cochran, M. (2005). *Student understanding of the second law of thermodynamics and the underlying concepts of heat, temperature, and thermal equilibrium.* [Unpublished Doctoral Dissertation]. The University of Washington.

Cohen, L., Manion, L., & Morrison, K. (2007). *Research methods in education*, Sixth edition, Taylor and Francis or Routledge.

Darejeh, A., Marcus, N., & Sweller, J. (2021). The effect of narrative-based E-learning systems on novice users' cognitive load while learning software applications. *Educational Technology Research and Development*, *69*(5), 2451-2473. https://doi.org/10.1007/s11423-021-10024-5

Deng, A., Zhang, T., & Chen, A. (2021). Challenges in learning aerobic and anaerobic concepts: an interpretative understanding from the cognitive load theory perspective, *Physical Education and Sport Pedagogy*, *26*(6), 633-648, DOI: 10.1080/17408989.2020.1849595

Ebel, R.L. (1967). The relation of item discrimination to test reliability. *Journal of Educational Measurement*, *4*(3), 125-128.

Galili, L., & Lehavi, Y. (2006). Definitions of physical concepts: A study of physic teachers' knowledge and views. *International Journal of Science Education*, *28*(5), 521-541. https://doi.org/10.1080/09500690500338847

Große, C.S., & Renkll, A. (2007). Finding and fixing errors in worked examples: Can this foster learning outcomes?. *Learning and Instruction*, *17*, 612-634. https://doi.org/10.1016/j.learninstruc.2007.09.008

Kala, N., & Okal, A. S., (2016, September 28-30). *The effect of the material developed according to split attention effect of Cognitive Load Theory on student achievement and loading in the subject of simple machines.* [Paper presentation]. 12th National Science and Mathematics Education Congress, Trabzon.

Kalyuga, S., Ayres, P., Chandler, P., & Sweller, J. (2003). The expertise reversal effect. *Educational Psychologist*, *38* (1), 23–31. https://ro.uow.edu.au/edupapers/136

Kaya, D., & Keşan, C. (2012). Conceptual and operational implementations for numeric section students who are a candidate for university. *Journal of Research in Education and Teaching*, *1*(3), 339-346.

Kılıç, E., & Karadeniz, Ş. (2004). Specifying students' cognitive load and disorientation level in hypermedia. *Educational Administration: Theory and Practice*, *40*, 562–579.

Kırtak, V.N. (2010). *Prospective physics, chemistry and biology teachers' levels of associating thermodynamics laws with daily life and environmental problems*. [Unpublished Master Dissertation). Balıkesir University.

Kline, R.B. (2009). *Becoming a behavioral science researcher: A guide to producing research that matters*, The Guilford Press.

Klingner, J., Tversky, B., & Hanrahan, P. (2011). Effects of visual and verbal presentation on cognitive load in vigilance, memory, and arithmetic tasks. *Psychophysiology*, *48*(3), 323-332. DOI: 10.1111/j.1469-8986.2010.01069.x

Kruger, C., Palacio, D., & Summers, M. (1992). Surveys of English primary teachers' conceptions of force, energy, and materials. *Science Education*, *76* (4), 339–351.

Leahy, W., Chandler, P., & Sweller, J. (2003). When audiotary presentations should and should not be a component of multimedia instruction. *Applied Cognitive Psychology*, *17*, 401-418. DOI: 10.1002/acp.877

Leahy, W., Hanham, J., & Sweller, J. (2015). High element interactivity information during problem solving may lead to failure to obtain the testing effect. *Educational Psychology Review*, *27*(2), 291-304. https://doi.org/10.1007/s10648-015-9296-4

Leahy, W., & Sweller, J. (2016), Cognitive Load Theory and the effects of transient information on the modality effect. *Instructional Science*, *44*, 107–123. DOI 10.1007/s11251-015-9362-9

Miller, G.A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological review*, *63*(2), 81.

Moreno, R., & Mayer, R. (1999). Cognitive principles of multimedia learning: The role of modality and contiguity. *Journal of Educational Psychology, 91*, 358–368.

Mousavi, S.Y., Low, R., & Sweller, J. (1995). Reducing cognitive load by mixing auditory and visual presentation modes. *Journal of Educational Pschology*, *87*, 319-334.

Mulop, N., Yusof, K. M., & Tasir, Z. (2012). A review on enhancing the teaching and learning of thermodynamics. *Procedia-Social and Behavioral Sciences*, *56*, 703–712. https://doi.org/10.1016/j.sbspro.2012.09.706.

Ngu, B.H., & Phan, H.P. (2016). Unpacking the complexity of linear equations from a cognitive load theory perspective. *Educational Psychology Review*, *28*, 95–118. DOI 10.1007/s10648-015-9298-2

O'Connell, J.P. (2019). Challenges to learning and teaching thermodynamics. *Chemical Engineering Education, 53*(1), 2-9.

Paas, F.G.W.C., Renkll, A. & Sweller, J. (2004). Cognitive Load Theory: Instructional implications of the interaction between information structures and cognitive architecture. *Instructional Science*, *32*, 1-8. https://www.jstor.org/stable/41953634

Paas, F.G.W.C., & van Merrienboer, J. J. G. (1993). The efficiency of instructional conditions: An approach to combine mental effort and performance measures. *Human Factors*, *35*(4), 737-743.

Paas, F.G.W.C., & van Merrienboer, J. J. G. (1994). Instructional control of cognitive load in the training of complex cognitive tasks. *Educational Psychology Review*, *6*, 351-372.

Paas, F.G.W.C., van Merriënboer, J. J., & Adam, J. J. (1994). Measurement of cognitive load in instructional research. *Perceptual and motor skills*, *79*(1), 419-430.

Pinto', R., Couso, D., & Gutierrez, R. (2005). Using research on teachers' transformations of innovations to inform teacher education; The case of energy degradation. *Science Education*, *89* (1), 38–55

Sezgin, M.E. (2009). *The effects of multimedia courseware designed based on cognitive theory of multimedia learning on cognitive load, performance levels and retention.* [Unpublished Doctoral Dissertation]. Çukurova University.

Smith, P.L., & Ragan, T.J. (1999). *Instructional design*. 2nd Ed. Toronto: John Wiley and Sons.

Sözbilir, M., & Bennett, J. M. (2007). A Study of Turkish chemistry undergraduates' understandings of entropy. *Journal of Chemical Education*, *84*(7), 1204-1208.

Sözbilir, M. (2002). Turkish chemistry undergraduate students. misunderstandings of Gibbs free energy. *University Chemical Education*, *6*, 73-83.

Sreenivasulu, B., & Subramaniam, R. (2013). University students' understanding of chemical thermodynamics. *International Journal of Science Education*, *35*(4), 601–635. https://doi.org/10.1080/09500693.2012.683460

Sweller, J. (2020). Cognitive load theory and educational technology. *Educational Technology Research and Development*, *68*(1), 1–16. https://doi.org/10.1007/s11423-019-09701-3

Sweller, J. (2010). Element interactivity and intrinsic, extraneous, and germane cognitive load. *Educational Psychological Review*, *22*, 123–138. DOI 10.1007/s10648-010-9128-5

Sweller, J., & Chandler, P. (1994). Why some material is difficult to learn. *Cognition and Instruction*, *12*(3), 185–233. DOI: 10.1207/s1532690xci1203_1

Sweller, J., van Merriënboer, J.J.G., & Paas, F.G.W.C. (1998). Cognitive architecture and instructional design. *Educational Psychology Review, 10*, 251–296.

Sweller, J., van Merriënboer, J.J.G., & Paas, F.G.W.C. (2019). Cognitive architecture and instructional design: 20 years later. *Educational Psychology Review*, *31*, 261–292. https://doi.org/10.1007/s10648-019-09465-5

Tindall-Ford, S., Chandler, P., & Sweller, J. (1997). When two sensory modes are beter than one. *Journal of Experimental Psychology: Applied*, *3*, 257-287. https://doi.org/10.1037/1076-898X.3.4.257

van Merrienboer, J.J.G., & Ayres, P. (2005). Research on Cognitive Load Theory and its design implications for e-learning. *Educational Technology Research and Development*, *53*(3), 5-13.

Weng, C., Otanga, S., Weng, A., & Cox, J. (2018). Effects of interactivity in E-textbooks on 7th graders science learning and cognitive load. *Computers and Education*, *120*, 172–184. https://doi.org/10.1016/j.compedu.2018.02.008