

Open University of Cyprus
Faculty of Pure and Applied Sciences

Master's Programme of Study
Cognitive Systems

Master's Dissertation



**The Impact of Socratic Cognitive Assistants
on Students' Intrinsic Motivation Towards Critical Thinking
and Rational Decision Making**

Elitza Voutcheva

Supervisor

Dr. Konstantinos Tsagkaridis

May 2023

Abstract

This study examines the impact of Socratic Cognitive Assistants (SCAs) on students' intrinsic motivation towards applying critical thinking and rational decision-making. SCAs are software programs that use Socratic questioning to stimulate critical thinking and help learners develop better decision-making skills. The study hypothesizes that the use of SCAs will lead to increased intrinsic motivation towards critical thinking and rational decision-making among students. An additional variable, individuals' Locus of Control (LOC), has been considered in the designing of this research. We predict a positive correlation between one's LOC and their critical thinking ability, as well as their intrinsic motivation.

To assess this hypothesis, an experimental design for a pre-test and post-test has been developed. An uncontrolled virtual assignment and questionnaire have been voluntarily presented to randomly assigned students, with a final sample of 29 participants. The virtual assignment consists of the LOC inventory, intrinsic motivation inventory (IMI), presented pre and post experimentally, and an interactive conversation with CA Liza.

The results of the study contradicted the study's main hypothesis. In general participant's motivation decreased after the intervention. Further analyses show age to be a significant predictor for the level of critical thinking skills, with younger ages associated with higher scores on the critical thinking task. The score of critical thinking skills appears to be a significant predictor for pre- and post-experimental motivational changes; namely, higher critical thinking scores are associated with smaller decrease in motivation after the intervention. These research findings might be an indicator of the efficiency of Socratic CA on revealing biases in critical thinking and rational decision-making, which subsequently lead to a decrease in one's intrinsic motivation. It is also possible that Liza is generally not an effective tool for enhancing students' intrinsic motivation. Nevertheless, other findings from this research study demonstrate the potential for positive effects on IMI. For example, young participants show higher interest and enjoyment, as well as higher intrinsic motivation scores compared to older participants. This finding justifies further research in this field and shows the potential for early intervention on associated skills to increase the likelihood for success in academic and professional contexts.

Acknowledgment

My deepest gratitude goes out to everyone who has supported me down this road.

The people I couldn't have done this without I'd like to thank in person.

To Dr. Konstantinos Tsagkaridis for his dedicated and tactful supervision and guidance.

To Dr. Antonios Christou for his help during the improvement of the research design in the
early stage of this work.

To Prof. Michael Loizos for implicitly encouraging my graduation a long time ago.

To Prof. Le and his team for providing access to Liza's results and for being incredibly
supportive.

To all anonymous respondents of the research, who voluntarily devoted their free time to
participate.

To my colleagues for getting along with me and hearing me out.

To my friends for the great conversations and being there for me.

To my family for the endless patience and love, and contributions that helped me reach this
point.

Table of Contents

Abstract.....	iii
Acknowledgment.....	iv
Table of Contents.....	v
1. Introduction.....	1
1.1 Research Question.....	2
1.2 Hypothesis.....	4
2. Theoretical Framework: State of Art.....	5
2.1 Critical Thinking and Reasoning.....	5
2.2 Intrinsic Motivation.....	7
2.3 Socratic Approach.....	10
2.4 Cognitive Assistants, Pedagogical Agents, and Chatbots in Education Regarding Critical Thinking, Motivation and Socratic Approach	12
2.5 Locus of Control	14
3 Propositional Framework: Critical Thinking, Socratic Approach, Intrinsic Motivation, and Cognitive Assistants in Education.....	16
4 Research Method.....	18
4.1 Participants.....	18
4.2 Research tools	21
4.3 Procedure	24
4.4 Methodological Plan	25
5 Results.....	26
5.1 Descriptive statistics.....	27
5.2 Correlations.....	30
5.3 Group Comparisons & Regression.....	31
5.4 Qualitative Feedback.....	33
6 Discussion	36
7 Conclusion.....	45
Appendix.....	47
Appendix I: Intrinsic Motivation Inventory	47
Appendix II: Rotter's Locus of Control Scale	49
Appendix III: Information Letter	51
Appendix IV: Informed Consent.....	53
Appendix V: Tables of Correlation.....	54

Appendix VI: Liza - Examples	59
References.....	70

1. Introduction

Critical thinking is a necessary skill in the 21st-century. Biases and common fallacies affect our daily life, influencing our judgment and decision-making. Because of the importance of this topic, the quest for more efficient ways to increase critical thinking skills and implement higher-order thinking practices for developing students' cognitive abilities (Smith & Szymanski, 2013), as well as creating opportunities for practice in everyday situations, has led to the development of a number of (online) interventions (Fahim & Bagheri, 2012). However, modern lifestyle with personalized content, adjusted to customer preferences based on forecasting algorithms deprives today's youth from opportunities to apply their critical thinking ability in real-life situations; even older generations lose their grip on logical argumentation (Spector & Ma, 2019). Furthermore, current and upcoming distance education trends highlight the need to increase students' intrinsic motivation, to develop academic learning skills and enhance academic performance.

Cognitive Assistant Liza has already shown results for increasing human reasoning (Le & Wartschinski, 2018). The dialogical approach of conversation uses situations with an absolute dilemma to force the user to find a solution. Liza guides the user through explanation and questioning, depending on the user's preferences. Users receive feedback in the form of a final chat about their performance. Participants from this previous study rated the interaction with Liza as pleasant. This could act as an additional motivation to practice and increase their reasoning, compared to more traditional modes of learning (Firat et al., 2018). Contemporary evidence shows a positive impact of educational chatbots on the development of users' computational thinking (Esteve-Mon et al., 2019). Participants' high interest can be explained due to differences from

traditional modes of learning, such as the time-space independence, and the similarity of the conversational methodology to a typical human-to-human interaction (Haristiani, 2019).

Therefore, the following hypothesis could be formulated: If students apply a technological cognitive assistant solution that makes use of modern Socratic dialogue, they will improve their decision-making performance. In this present work, the specific question is whether Liza can affect participants' intrinsic motivation for working with tasks requiring critical thinking.

1.1 Research Question

Does the interaction with Liza affect students' intrinsic motivation for working on critical thinking?

The analysis for the main research question in the present work will be formulated by shedding light on various aspects, interplaying in this research, and being of significant importance for the research results. Specifying these aspects here, a theoretical base will be laid on the importance of critical thinking and reasoning in education, the impact of the Socratic approach in education, and the role of motivation in enhancing academic results. The gathered data from the experimental research will be resulting in an analysis of the participants' intrinsic motivation due to their interaction with Liza. This additional fractional experiment is an extension of Le and Wartschinski research on Liza's effect on participant's reasoning (Le & Wartschinski, 2018). In that study a dialog-based CA has been developed to enhance human reasoning skills by adapting to the learner's individual performance. This CA includes over sixty different reasoning tasks, along with explanations, casual conversations, hints, and evaluations. The empirical evidence of this study, where 65 participants have been assigned in an experimental and a control group, shows greater learning progress compared to non-interactive learning methods.

To answer this research question, an additional predictor is added to the research design. Just like motivation, Locus of Control (LOC) can influence one's decision-making abilities as well as act as a predictor for motivation itself (Kovach, 2018; Rotter, 1966). LOC appears to be relevant to the research question in this study. Studies show a crucial relationship between motivation and critical thinking abilities, examined through Locus of Control (LOC) and its impact on behavior (Bahadir et al., 2014; Flor et al., 2013; Ng et al., 2006; Nurdin & Damayanti, 2020; Oğuz & Sariçam, 2015; Toussi & Ghanizadeh, 2012). Internal LOC, where individuals believe in their influence, is linked to personal decision-making, whereas external LOC involves accepting external influences. Intrinsic motivation and internal LOC strongly correlate with critical thinking, and teaching critical thinking positively affects LOC. These findings suggest that LOC influences both motivation and performance in education across various fields.

As motivation is one of the success factors, which plays a significant role in educational contexts and affects performance, (life-long) studies involving motivation should be common practice, especially in educational programs. Motivation appears to decrease when tasks or decision-making problems are just too hard to be solved (Suhadi et al., 2015). Since Plato's Meno (Scott, 2006) we know that we do not need to possess all the knowledge, but by questioning we are going step by step towards it by logical thinking and the right argumentation.

Emerging technologies in various fields make the existence of more paradigms possible. Numerous studies have explored the effectiveness of cognitive assistants as a learning medium in education and business, showing positive impacts on motivation, interest, and critical thinking (Davis, 2018; Dinçer & Doğanay, 2017; Fryer et al., 2019; Haristiani, 2019; Martha & Santoso, 2019). According to some studies, pedagogical chatbots based on Socratic inquiry methods have been successful in enhancing critical thinking and promoting ethical considerations (Goda et al., 2014; Lara & Deckers, 2020; Le & Wartschinski, 2018; Weizenbaum, 1983). Overall, students' motivation is considered an important component in intelligent tutoring systems and predictive models for academic achievement.

In this work, the combination of a cognitive assistant and modern Socratic questioning is hypothesized to have a positive impact on one's intrinsic motivation. Therefore, they could prove to be beneficial for enhancing critical thinking in classic common fallacies puzzles, such as those presented by Liza. Besides, such applications would be beneficial in other study (or work-related) areas as well, because of the gains in motivation.

1.2 Hypothesis

The research hypothesis proposes that applying cognitive assistants containing elements of SA like Liza in (higher) education would enhance students' motivation (Brandtzaeg & Følstad, 2018; Dinçer & Doğanay, 2017; Liew et al., 2017; Siddiqui et al., 2020). Students should be willing to enhance their critical thinking against biases and common fallacies, their motivation in applying critical thinking in complex real-world problems should increase, as well as their ability in making rational decisions. Additionally, it is hypothesized that older participants will outperform younger ones in critical thinking skills. Extendedly, we hypothesize that higher motivational scores will be observed among individuals with an internal LOC. The results of this research are expected to have a significant value for application in study programs, especially given the upcoming trend for online education.

Regarding the possible effect of Liza on motivation, the following hypotheses will be tested:

H₀: CA Liza has no influences on one's intrinsic motivation for working on critical thinking.

H₁: CA Liza influences one's intrinsic motivation for working on critical thinking.

H₂: CA Liza influences younger people's intrinsic motivation better than the older ones.

H₃: Older participants show better critical thinking skills than the younger ones.

H₄: One's LOC influences one's intrinsic motivation for working on critical thinking.

2. Theoretical Framework: State of Art

2.1 Critical Thinking and Reasoning

Students put a lot of effort into many years of study and graduating thesis. Scientific reasoning, as a required academic skill, appears to be strongly related to students' critical thinking abilities (Dowd et al., 2018). Drawing conclusions from pieces of evidence and cogent and sound argumentation are needed for making rational decisions and are essential for providing strong inference. Trustworthiness, as an essential part of our daily life, on the other hand, has been considered to have a significant relation with critical thinking and reason in decision making (Kleinig, 2018). Trust in authority appears to be closely associated to argumentation and reasoning. Critical thinking, as taking part in the explanation and justifying the process of reasoning, reduces fallacious conclusions. For that matter, trustworthiness requires critical thinking skills as judgment for reliability. Applying critical thinking skills does not guarantee a well-placed trust, but it has been found to avoid blind trust and poor care of reasoning.

Fostering critical thinking in education and research on its success on enhancing students' critical thinking skills is a broadly discussed topic for the last 20 years and is very much related with the implemented in educational programs' 21-st century skills. A literature review provides insights about the effective types of teaching critical thinking, combining instruction, dialogue, and mentoring as well as the role teachers play in students' skills acquisition (Janssen, 2019). Of

decisive importance are not only the study program, but the teachers' attitude and own unbiased ability to recognize and explain student's fallacies. Fostering critical thinking in education is also suggested to be embedded in curricula, where both implicit and explicit strategies show potential in developing critical thinking skills (Maphalala & Mpofu, 2017). Lack of critical thinking in education is discussed also to reflect on students' deep understanding of their study topics and risks of students' engagement, which increases the concern of whether and how critical thinking should be taught.

Twelve methodologies for teaching-learning critical thinking in higher education have been experimentally tested for their efficiency (Bezanilla et al., 2019). In that study, critical thinking is defined by six distinct categories, which have been objectives for testing the twelve different methodologies: Analyzing/Organizing; Reasoning/Argumenting; Questioning/Asking oneself; Evaluating; Taking a position/Making Decisions; and Acting/Compromising. To conclude, all twelve methodologies has been finally grouped in three main methods: firstly, the ones that are most commonly used to develop critical thinking, related to oral and written argumentation, reading and text analysis, and synthesis of resources; secondly, methodologies, considered active, among which case studies, collaborative and cooperative learning, real world and problem and project based learning; and thirdly, other methodologies like assessment, follow-up and feedback, questioning, evaluation, interpretation and justification, research, and lectures. Besides these research findings, a discussed topic in the entire study assumes questioning yourself and others of a significant importance for fostering critical thinking. Asking questions has been presumed to have major impact among other strategies. Accordingly, critical thinking in a contextual concept finds discussions as a primer competency for the personal and academical growth of university students.

“Deeper learning, dialogic learning, and critical thinking are essential capabilities in the 21st-century environments we now operate. Apart from being important in themselves, they are also crucial in enabling the acquisition of many other 21st-century skills/capabilities such as problem solving, collaborative learning, innovation, information and media literacy, and so on.” (Manalo, 2019, p. iii).

In the current AI age supported by virtual, cognitive, and pedagogical assistants, the need for human intelligence to keep improving is maybe stronger than we have could imagine years ago. Generally, many commonly used virtual assistants provide answers or solutions for our cognitive needs. On the other hand, researchers of cognitive assistants like Daphne (Martin & Selva, 2019) and Liza (Le & Wartschinski, 2018) aim to support human reasoning and decision making skills instead of replacing them. Enhancing human intelligence by using AI technologies is also discussed by Spector and Ma (2019), who claim that more focus should be given to emphasize the development of human intelligence and stimulate critical thinking with AI support. Their framework of critical thinking involves motivation, which appears not only to be a success factor for academic performance but also a significant element for increasing critical thinking.

2.2 Intrinsic Motivation

According to Gopaln et al. (2017), several motivation theories could be considered as mainstream in the area of learning. Intrinsic and Extrinsic Motivation Theory (IEMT), where challenge, curiosity, control, and fantasy are key factors triggering intrinsic motivation, as opposed to reward, compulsion, and punishment associated with extrinsic motivation. Self-determination theory (SDT) expand the distinction between intrinsic and extrinsic motivation, a theory later updated and defined with new directions. SDT sees in intrinsic motivation the human's ability to include and interpret several features in the learning process, while extrinsic motivation illustrates a different, significant in its relative autonomy and focuses on academic performance. SDT consists of five sub-theories: Cognitive Evaluation Theory, Organismic Integration Theory, Causality Orientations Theory, Basic Psychological Needs Theory, and Goal Contents Theory. ARCS model, abbreviated from Attention, Relevance, Confidence, and Satisfaction attributes, captures students' attention as a critical factor for gaining and maintaining their interest in learning. It focuses on the importance of students' perceptions, desires, emotions, and anticipation. An attractive and stimulating learning medium or material is essential for maintaining the student's interest throughout the learning process. Social Cognitive Theory (SCT) refers to knowledge

acquisition by observation, interaction, experiences, and media influence throughout interactive learning, which allows gaining confidence by practice. Expectancy Theory supports the belief that there is a relationship between the amount of effort and performance that should be rewarded. According to this philosophy, putting in a lot of work would result in greater success and more rewards.

Revising the motivation theories in learning, intrinsic motivation is an explicit part of IEMT and SDT. However, the ARCS model and SCT, even though not specifically consisting intrinsic motivation factors, rely on attention, attractiveness, and social interactivity which we similarly find in SDT. In self-determination theory (SDT), one of the most evolved motivation theory in learning consisting of intrinsic motivation, intrinsic motivation with an internal locus of causality is associated with intrinsic regulation (Ryan & Deci, 2000). The relevant to intrinsic motivation regulation processes are interest, enjoyment, and inherent satisfaction. Intrinsic motivation has been measured in various studies with intrinsic motivation inventory (Sun & Gao, 2020; Vos et al., 2011), including effort/importance, value/usefulness, felt pressure/tension, perceived choice, and relatedness along with interest, enjoyment, and inherent satisfaction. Although Ryan and Deci categorize personal importance and conscious valuing with a somewhat internal locus of causality as a part of extrinsic motivation, their study results in conclusions that the identified, integrated, and intrinsic regulation for motivation is positively related with self-determination and are perceived from an internal locus of causality.

As a result, assessing intrinsic motivation should evaluate the various aspects separately from each other in order to interpret a complete overview. *“For example, the more students were externally regulated the less they showed interest, value, and effort toward achievement, and the more they tended to disown responsibility for negative outcomes, blaming others such as the teacher. Introjected regulation was positively related to expending more effort, but it was also related to feeling more anxiety and coping more poorly with failures. In contrast, identified regulation was associated with more interest and enjoyment of school and with more positive coping styles, as well as with expending more effort.”* (Ryan & Deci, 2000, p. 73)

Recent pieces of research enter the emerging field of intrinsic motivation as a neuroscientific occurrence (Di Domenico & Ryan, 2017). In a literature review, Di Domenico and Ryan review a

series of fMRI and EEG findings, providing evidence for this relation. Various aspects of the research lead to a couple of conceptualizations. One proposition is that one's intrinsic motivation is supported by the dopaminergic system. This proposition goes further than the 'likeness' by providing evidence from the experimental studies with assessed relation between intrinsic motivation, curiosity, learning, and academic achievement, supported by the dopaminergic system. This proposition would tell us not only 'that' someone might like something or not, but also the reason 'why'.

Another proposition considers entailment between the brains' networks for salience detection, attentional control, and self-referential cognition. Neuroimaging studies, supporting this proposition, reporting patterns of neural activity involving the multiple neural networks, related to spontaneous endogenous bottom-up cognitive processes. Also, in this proposition dopamine neurons are found to play a role in intrinsic motivation signals. The dopaminergic system for that matter is remarkably interesting in the interplay between the transfer of dopamine for successful memory processes and the additional production of dopamine during cognitive tasks (Constantinidis & Klingberg, 2016). Next to Di Domenico and Ryan propositions (2017), similar supportive findings in neurology might build new conceptual frameworks for our further understandings of intrinsic motivation.

Approaching intrinsic motivation in education, recognition of changes in the learning needs between generations is of fundamental importance. Chaudhuri (2020) presents the contrasts followed by millennia and the previous X generation, the one of their teachers, as well as the gap towards the new Z generation. Current students, most of who among the Z, are motivated by nontraditional learning methods, rewards, personal preferences, and supportive technology at the highest level. The millennia, as ambitious and achievement-oriented multitaskers are described as depending on technology. Such definitions, compared to X, clearly reveal the need for new teaching approaches, involving high technological interactive and personalized innovations for keeping and increasing new generations' motivation. However, studies show that regarding distance education and e-learning environments, intrinsic motivation strongly varies and further research can focus on enhancing it through higher technological developments (Firat et al., 2018).

Having in mind the generation of our students, millennial and Z, a study on micro-learning reveals a significant increase in enhancing motivation and performance due to chatbot-based micro-learning (Yin et al., 2021). According to Yin and colleagues (2021), two success factors are engaged. Micro-learning, presenting small parts of information, is personalized to the user information need and attention span and is very suitable for the new generation students. Their literature review shows that providing small parts of content increases information retention by 20%. Chatbot-based task-oriented dialogue system, providing bite-sized learning content, meet the new generation high technology demands in order to engage their attention and interaction. Next to that, the study suggests that current chatbot-based learning platforms are designed with restrictions to playing a teaching assistant role with an emphasis on hard skills knowledge, where future designs should also have a focus on chatbots with an emphasis on individual's soft skills, among others.

“Developers and designers have an urgent need to know more about how people experience chatbots and to understand the user needs that motivate the future use of chatbots.” (Brandtzaeg & Følstad, 2018, p. 40). Apparently, we enter the era of new behavioral and motivational demands in the human-computer-interaction needs, based on social and psychological factors.

2.3 Socratic Approach

The Socratic method or the Socratic approach or the Socratic questioning is a way of engaging participants in a dialogue based on asking and answering questions as well as providing arguments to stimulate critical thinking and to draw out ideas and underlying presuppositions, as shortly summed up and provided to the wide audience by Wikipedia and also easily to be found in numbers of representative sources just by googling the term (“Socratic Method,” 2022). Facts we follow and acknowledge from the very beginning of the Socratic movement ever since Plato's Meno (Scott, 2006). Following this widely accepted description, any dialogue on any topic comes out

with better critically provided arguments, ideas, and conclusions, if conversated based on the Socratic method. This includes critical thinking itself, however, critical thinking, as we elaborated above, is a tool for enhancing academic performance and decision-making skills and not a goal instead. So, if Socratic methods enhance critical thinking in any dialogue, that it should enhance critical thinking in dialogues about critical thinking.

Critical awareness on the other hand is the ability to apply critical thinking in contextual situations and requires awareness of logical fallacies that obstruct one's reasoning (Balbay, 2019). Balbay (2019) finds critical awareness significant in advanced stages of education where cultural, socio-economic, political, and ideological aspects are taken in consideration. His research on enhancing critical awareness through Socratic pedagogy proposes the implementation of Socratic methods in educational programs as found that it results to better students' engagement to the educational processes and show students' deeper insights when considering multidimensional problems. In support of such ideas, Socratic questioning has been found to facilitate critical thinking skills for solving problems (Makhene, 2019; Sahamid, 2016).

Online and distance learning are commonly successful due to a great deal of independency, appropriate for students who are self-determined to gain and master new knowledge. Self-motivation, therefore, is a significant factor in blended education (Suhadi et al., 2015). In their study, Suhadi and colleagues (2015) suggest combining online learning technologies with the Socratic method. This symbiosis has been shown to enhance students' motivation by applying online learning technologies and increase the students' deep and critical thinking performance, knowledge acquisition, and innovative ideas due to the Socratic questioning method, accompanying the students towards problem solving achievements. The additional 'freedom' of time and space independent education is due to the Socratic questioning approach encouraging the students into meaningful academic and social interaction and engagement. A crucial factor, to be taken into account here, is the inevitable level differences within the groups, as too easy or too difficult tasks are famous motivational killers (Suhadi et al., 2015). A fact, suggesting that modified educational interventions, which are tailored to individual's levels of manageability, would obviate this concern to a certain extend.

Technology enhanced tools based on the Socratic approach, supporting the development of critical thinking and reasoning, are reviewed by Le (2019). His review compares 14 technology-enhanced learning systems, claiming the Socratic approach as a useful method for enhancing critical thinking and reasoning. Two important findings can be taken into consideration here. First, all reviewed learning systems are based on the modern Socratic approach and are supported by human tutoring strategies, but no classic method has been identified. Second, there is still a significant absence of development in pedagogical agents in learning systems, providing support for critical thinking and reasoning through the Socratic approach. Important to note is that only 4 of the reviewed systems are developed in the last 10 years. As the recently emerging technologies have developed machine learning to advanced levels, providing support for quite advanced cognitive assistants, such technological applications already find successful employment in education. However, not many of them make use of Socratic methods for enhancing critical thinking, reasoning, and decision making.

2.4 Cognitive Assistants, Pedagogical Agents, and Chatbots in Education Regarding Critical Thinking, Motivation and Socratic Approach

Long before ChatGPT dominated our life, a number of cognitive assistants have supported learning in education and business. Artificial Intelligence Chatbot as a Learning Medium took the attention of many researchers in various fields and topics (Davis, 2018; Dinçer & Doğanay, 2017; Fryer et al., 2019; Haristiani, 2019; Martha & Santoso, 2019). The studies mentioned above review the efficiency of chatbots as a learning medium in teaching language and find them to have positive effects on interest and motivation, including: feeling more relaxed when talking to a computer; chatbots are endlessly willing to; many bots provide both text and speech; bots are interesting; chatbots could potentially provide quick and effective feedback; multimedia pedagogical agents, using applying gesturing influencing different educational or personal constructs; positive effect

on cognitive load. Overall, pedagogical agents have been found to be enhancing both individuals' learning performance as well as motivation for learning.

Along the increased motivation, pedagogical chatbots designed to reflect the Socratic inquiry method has been developed for enhancing critical thinking, such like Eliza - en Liza (Goda et al., 2014; Le, 2019; Le & Wartschinski, 2018; Weizenbaum, 1983). These methods have been a successful attempt towards enhancing one's critical thinking by both applying Socratic questioning and providing common fallacies examples. Socratic based chatbots have been even found to increase technological ethics and motivate human moral (Lara & Deckers, 2020).

Systematic reviews on these emergent field give light to a broader spectrum of relevant analyses. On one hand a generalization of the reviewed academic support services recognizes four types of CA, or as other called Artificial Intelligence in Education, or Intelligent Assistants (IA). These four types are been structured as following: 1. Profiling and prediction, 2. Assessment and evaluation, 3. Adaptive systems and personalization, and 4. Intelligent tutoring systems (Zawacki-Richter et al., 2019). Regarding students' motivation, this specific one has been discussed as a component of the student model in *intelligent tutoring systems* and also as a conditional in a predictive model for students' academic achievement within *profiling and prediction*. Whether the other discussed topics in this current study either fits one of these categories or covers more of them, might be of an interest for further research. In the same study IAs of the category *Education* has been found to be one of the three most prevalent application domain categories represented, where *Health* has been supported most frequently, followed by *Education* as second. Categories *Online Services and IoT* follow as third (Islas-Cota et al., 2022). In the same order, a remarkable analysis scale the same domains *Health* and *Education* together, as holding almost half of the selected and reviewed articles (49.5%), where IAs for the education were developed mostly in 2018. The same year LIZAs first findings has been published (Le & Wartschinski, 2018).

2.5 Locus of Control

A crucial relationship between one's motivation and one's critical thinking abilities appears to be a finding in studies researching one's Locus of Control (LOC) and its influences on the individual's behavior (Bahadir et al., 2014; Flor et al., 2013; Ng et al., 2006; Nurdin & Damayanti, 2020; Oğuz & Sariçam, 2015; Toussi & Ghanizadeh, 2012). The LOC theory describes one's believing in own influence on the outcome of certain event and divides the understanding in internal and external LOC. Individuals with internal LOC, focus and rely on own decisions and actions towards a particular wishful outcome. In the opposite perspective, individuals with external LOC do not believe in own influence on any matter but support the acceptance of external influences. Within this spectrum, both critical thinking and respectively decision making as motivation are found to be correlational evidence.

In theories of learning and personality, within LOC a similar relationships has been laid between critical thinking in terms of a link between actions and consequences and LOC (Rotter, 1966). In some studies, evaluating the correlation with age, gender, type and level of education, both positive and negative co-relation has been found to external (Bahadir et al., 2014; Oğuz & Sariçam, 2015) and to internal LOC (Nurdin & Damayanti, 2020; Oğuz & Sariçam, 2015). A reverse research approach even claims that teaching and training creative and critical thinking results to a significant difference and a positive effect in one's LOC (Flor et al., 2013). A meta-analysis outcome discusses the same relationship within a work environment, specifically viewing intrinsic motivation and internal LOC as being firmly related (Ng et al., 2006). A strong correlation between intrinsic motivation and internal LOC has been shown to be supported in a study regarding effective learning and study achievement (Toussi & Ghanizadeh, 2012).

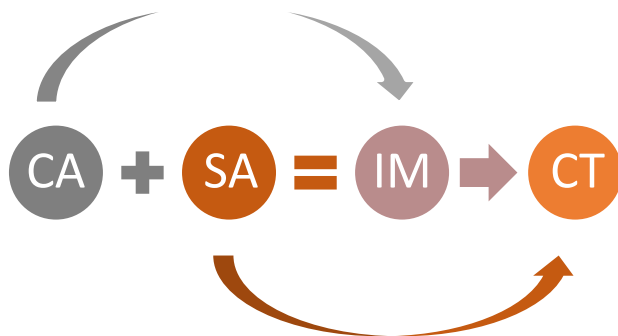
Extended interdisciplinary study of motivation theories and LOC (Kovach, 2018) suggests that LOC influences one's motivation and respectively one's performance in education. Kovach evaluates LOC in higher education form different perspectives and finds correlation between ones' respectively internal or external LOC and their decision-making abilities across industries, higher

education including business, kinesiology, psychology, and sociology. Cited examples point dependencies regarding health decisions, health study attitude, job tenure, managers/non-managers, and decisional procrastination. Along the extended literature review, where evidence to the commented correlations has been provided, Kovach (2018) has conducted multidisciplinary research with 101 participants, with having study design predictors into account being age, gender, and level of study graduation. The study's hypothesis of correlations between age and LOC, and gender and LOC was supported by his results. A surprising result has been found in the results with the variable study level as a predictor for LOC, where between undergraduate and graduate students no particular positive or negative correlation has been found. However, the study provides a strong recommendation for further research across disciplines. Among others, evaluating teachers' mindset and the influence of classroom practices, conducting longitudinal research on professionals in actual environment, as well as individuals' longevity.

3 Propositional Framework: Critical Thinking, Socratic Approach, Intrinsic Motivation, and Cognitive Assistants in Education

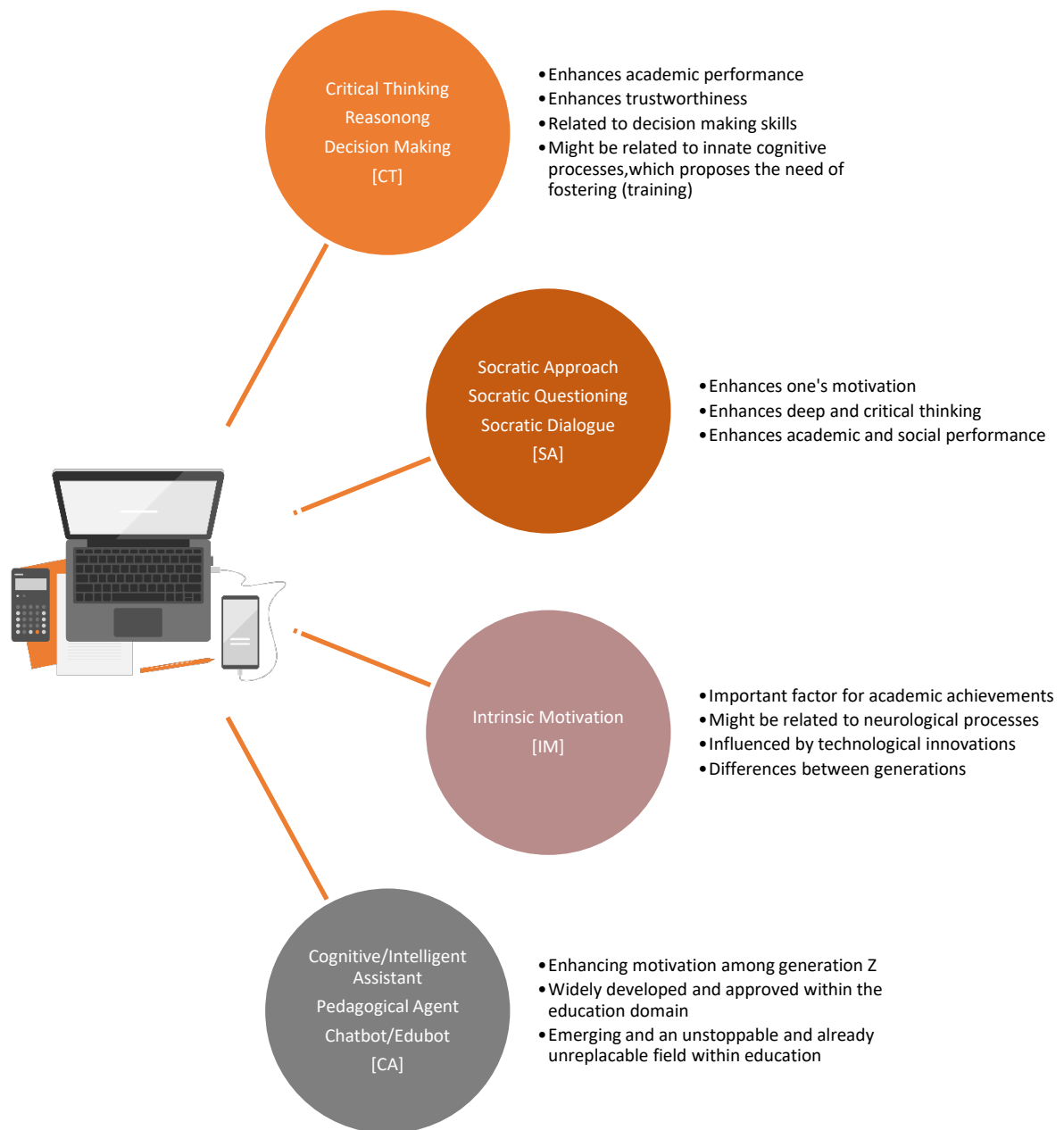
As reviewed in the previous chapters of this study, the discussed aspects reinforce each other in numerous ways. A suggestion can be made that some of them may even be conditional for one another in order to suppose an enhancement in a third one. A propositional framework, bringing together Critical Thinking, Socratic Approach, Intrinsic Motivation, and Cognitive Assistants in Education may lead to a new symbiosis towards educational approaches for the new generations. Based on this, a suggestive infographic below illustrates the intertwining between the reviewed components, visualizing the dynamic influence they cause to the whole symbiosis (Figure 1) and a compact description of its components (Figure 2).

Figure 1. Propositional Framework: Critical Thinking (CT), Socratic Approach (SA), Intrinsic Motivation (IM), and Cognitive Assistants (CA) in Education – Symbiosis and intertwining.



The four parameters show the following dependency, based on evidence in the literature as reviewed earlier in this work: when CA enhances ones IM and in particularly for individuals from the generations millennial and Z, and SA improves ones CT contextually independently, then accordingly, CA supported by SA would foster IM towards CT.

Figure 3 . Propositional Framework: Critical Thinking, Socratic Approach, Intrinsic Motivation, and Cognitive Assistants in Education – Symbiosis. Brief description of the components.



4 Research Method

The research method has been designed to assess one's intrinsic motivation and its changes. Our focus is solely on examining the impact of the interaction with the CA on intrinsic motivation, not performance itself. The variable we are studying is not critical thinking, reasoning skills, or improvements in decision-making abilities and we are not measuring how these skills change after interacting with the CA. However, the accumulated results during the intervention on such skills can be used for additional analyses.

4.1 Participants

For the purpose of this study, volunteering students are asked to participate. Efforts were made to address our study to college and university students. A large and diverse group of over 500 participants have been invited, wherefrom 29 have taken part in the research. It is important to recognize that such small sample size and the fact that the participants were self-selected has caused certain research limitation, by providing a sampling method and so addressing the study to

a homogeneous population sample. With a larger sample size, we could have included a more diverse group and examined the impact of various demographic and personal factors.

In the current research, 29 randomized participants (N=29), 11 women, 18 man, and 0 non-binary, aged 17 – 46 years (M=25,34 and SD=7,92) (Figure 3) are invited to complete an interactive chat with Liza and to fill in a questionnaire before and after this intervention. In terms of education, the population sample consists of 5 high school students, 12 college students, 7 university students, 1 recently graduated and presently working university student and 4 teachers/lecturers (Figure 4). All participants are suggested to have the sufficient level of English, needed to interact with Liza. The students are approached via email and are invited to participate in an anonymous and independent from time and place activity. They are asked to contribute for about 25 minutes, with around 10 minutes devoted to filling in the questionnaires. The time of the interaction with the AI would vary from person to person. General descriptive statistics are presented in Appendix V, Table 1.

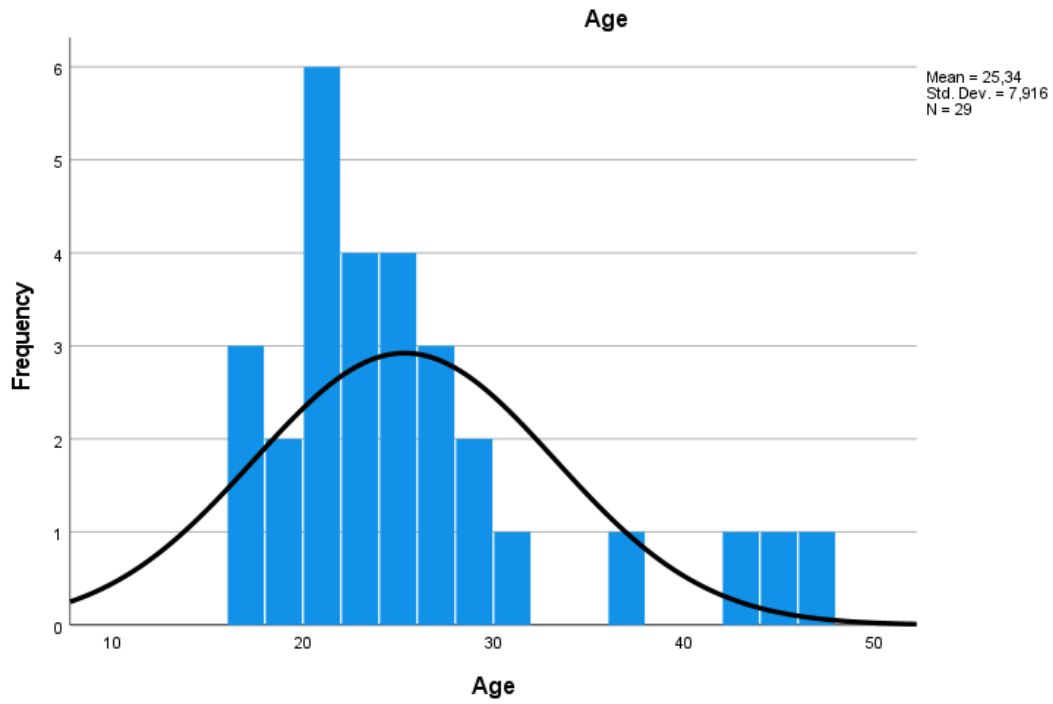


Figure 4. Participant's age

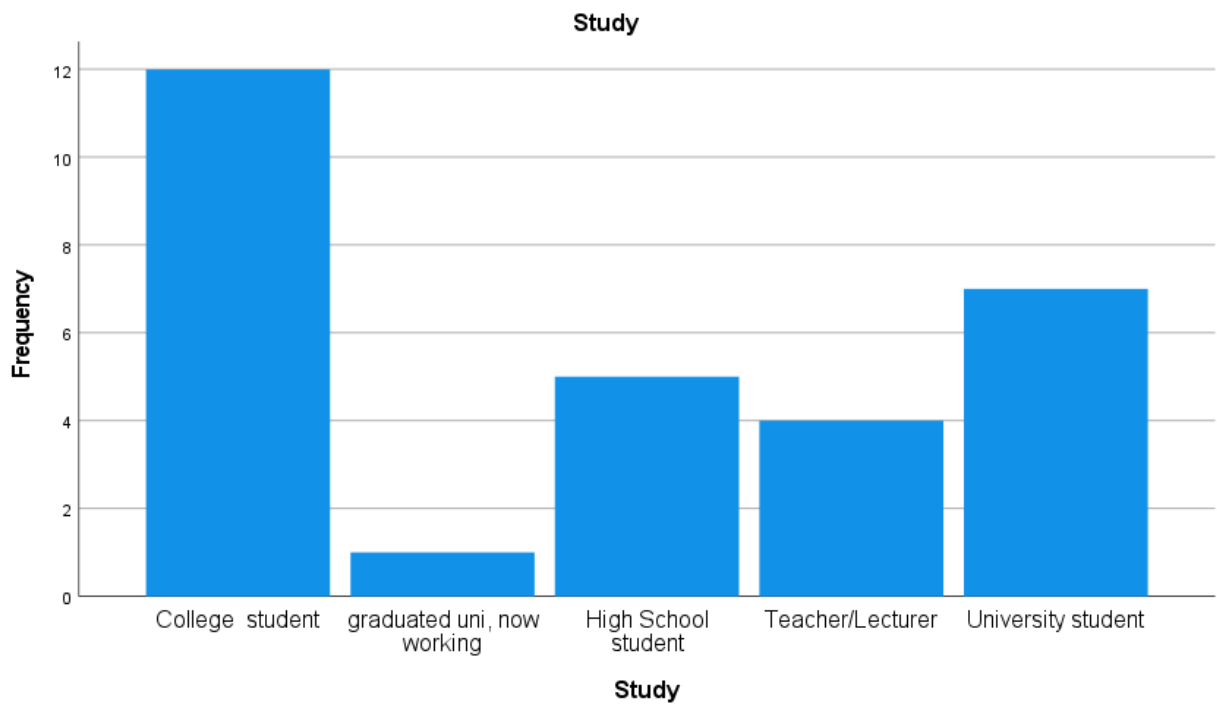


Figure 5. Participant's study level

4.2 Research tools

Participants have been assigned to complete one dialogue with CA Liza via [Liza \(hu-berlin.de\)](http://Liza.hu-berlin.de). The dialog consists of four steps. In the first step, Liza introduces herself as a pedagogical agent and gathers information about the user's preferences. This information is used to adapt the story according to the user's interests. In the second step, Liza presents 14 stories from different training topics, with an introduction and a task question for each story. Liza may offer hints or explanations if the user needs assistance. In the third step, Liza evaluates the user's answers, provides feedback, and adjusts the level of confidence. The CA also explains the solution if requested by the user. In the fourth step, after completing the stories or upon the user's decision to stop, Liza offers the option for feedback and provides a summary of the tasks and the user's performance. The dialog model allows for mixed-initiative dialogs, where both the CA and the user can initiate the next part of the conversation (Le & Wartschinski, 2018). The provided task questions are in the categories "Gambler's Fallacy", "Sunk Cost Fallacy", "Regression to the Mean" and "Base Rate Fallacy". In each dialogue a random three or all four of these categories are included. After the conversation, participants' answers are logged in Liza's back-files and the output with each participant's results is generated. When a conversation has been completed, a score between 0 and 100 (percentage) is given by the CA. Unfinished conversations contain only the chat history and no score values. This final score is an accumulative result in percentage, based on the given right answer for an amount of question in different categories of common fallacies. The final results given by Liza, states 100%, 50% or 0% per each tested category. Depending on the amount of tested category, which vary between 3 and 4, the final score can be summed to one of the following percentage: 0 - 12,50 – 25 - 37,50 – 50 - 62,50 – 75 - 87,50 - 100% for conversations with 4 tested categories and a score of 0 - 16,66 - 33,33 – 50 - 66,66 - 83,33 - 100% for tested 3 categories of common fallacies (Table 2). The actual conversation length of the 29 proceeded dialogues varies

in time between 02:19 and 42:48 minutes. Examples of two dialogues, together with samples of decision problems, feedback and given scores are listed in Appendix VI.

The impact of Liza on students' intrinsic motivation is monitored by self-report on an intrinsic motivation inventory (IMI). The quasi-experimental study is conducted in a realistic environment. The participants interact with Liza online, in an unsupervised manner at their own convenience, as they would do in a realistic online study situation. Participants also fill the intrinsic motivation inventory online and at their own convenience, within the required submission period. For this experiment, the aforementioned questionnaire is given in two variations, pre- and post-intervention.

IMI is used in research studies evaluating influences related to intrinsic motivation and could vary in amount of questions and sub-categories, depending on the research design (Sun & Gao, 2020; Vos et al., 2011). The inventory is originally developed by Ryan (1982) and its concept has been improved later (Ryan & Deci, 2000). In this research, IMI consist of twenty questions assigning the participants on three of the seven different categories, randomly presented: interest/enjoyment, perceived competence, and value/usefulness. The pre-intervention scale is consistent with the source of the questions and refers to opinions on online learning in general, whereas the post-intervention scale is an adaptation focusing on the experience of online learning through the interaction with Liza. Specifically, students are pre-experimentally asked to provide their answers based on their experience with current study topics and online activities. Questions within the three sub-scales are presented as following: interest enjoyments "I think that online learning activities are quite enjoyable."; perceived competence "I think I am pretty good at critical thinking."; value/usefulness "I think doing critical thinking activities could help me to improve my reasoning and decision making.". The post-experimental questionnaire evaluates participants' intrinsic motivation regarding their interaction with Liza and the provided content and topics. Questions are presented as following: sub-scale interest enjoyments "I thought the interaction with Liza was quite enjoyable."; sub-scale perceived competence "I think I was pretty good at critical thinking during my interaction with Liza."; sub-scale value/usefulness "I think doing critical thinking activities could help me to improve my reasoning and decision making.". The differences in each question, pre- and post-experimentally presented, cover the participants' experience with the

specific sub-scale. Respectively, the greater differences are within the scale interested/enjoyment, where interaction has been compared with other known online study activities. In the sub-scales perceived competence much smaller differences can be seen in the questions and participants can respond to their perception of own competence towards the questioned skills. Questions for value/usefulness do not differ pre- and post-experimentally. Answers are provided on the 7-point Likert scale (Appendix I). The Likert scale is a commonly used rating scale that measures attitudes, opinions, or perceptions on a continuum. It consists of a series of statements or items to which respondents indicate their level of agreement or disagreement using a predetermined scale. The applied scale in the presented questionnaire ranges from 1 to 7, where higher numbers represent stronger agreement to the given statement. E.g., to the question “I enjoy online learning very much.”, one can respond with 1 = Strongly disagree, 2 = Disagree, 3 = Somewhat disagree, 4 = Neutral, 5 = Somewhat agree, 6 = Agree and 7 = Strongly agree. It provides a structured way to quantify subjective opinions and is commonly used to gather quantitative data.

The pre-experimental questionnaire also includes a second scale on Locus Of Control (LOC, Rotter, 1966). It consists of twenty-nine questions, twenty-three of whom assess the LOC and another six are fillers. Questions are presented as following: “In your opinion, which statement is more correct?: a. Sometimes I can't understand how teachers arrive at the grades they give.; b. There is a direct connection between how hard I study and the grades I get.”. The six fillers among the questions are not relative or essential for the intervention but are used to disguise the true purpose of the study and maintain respondent interest. Questions as fillers are presented as following: “In your opinion, which statement is more correct?: a. A good leader expects people to decide for themselves what they should do.; b. A good leader makes it clear to everybody what their jobs are.”. These strategically inserted questions and are not taken in the research analyses of the study. Answers are provided in a multiple-choice presentation, where the respondent choose the one, they agree the most with, answer A or answer B (Appendix II).

The questionnaire scales were compiled in an online questionnaire built with Microsoft Office Forms. Liza’s score has been subtracted from the CA conversational logs, generated in a text file (Appendix VI). The data collected were extracted in Microsoft Excel format and analyzed with IBM SPSS 28.

4.3 Procedure



Participants received an e-mail with information about the study and instructions to follow in order to participate (Appendix III). Participants were informed about the procedure, the anonymity and confidentiality of their data (Appendix IV). The email also contained a link providing access to Liza and the evaluation form. Through the link participants can respond and submit their answers in an unsupervised manner and at their own convenience¹. The interactive experimental part with the CA has been provided as a link within the questionnaire. The interaction with the Cognitive Assistant (CA) requires one conversation with Liza.

¹ The following form is used for the purpose of the study and data collection in March and April 2023: <https://forms.office.com/e/wXymzf3fFU> . The integrated interactive conversation with Liza has been provided via the following link: [Liza \(hu-berlin.de\)](https://liza.hu-berlin.de) developed by Le and Wartschinski (2018).

4.4 Methodological Plan

In this research, the independent variables define students' level of education, age, and their LOC. The latter is defined based on the LOC measurement during inventory. All the groups are exposed to an intervention, namely the interaction with Liza. The dependent variables are defined to be two: 1. Participants' evaluation of their intrinsic motivation; 2. Their final scores on the condition tasks, which consist of a conversation with the CA and result in a score, given in percentage between 0 and 100. Participants' scores from their interaction with Liza are collected from Liza's anonymous logs and are matched to one's answers in the questionnaire due to a random pseudonym/alias/nickname. For some analyses the second depended variable, Liza's score, is used as an independent variable for observing motivational changes. Between-subjects analyses compare different populations, defined by factors such as age, gender and level of education. One of the main measures of the effects of the intervention on the IMI will be assessed with a within-subject analysis on the difference (or the absence of such), between one's intrinsic motivation before and after interacting with Liza. The limitations of this research method are mainly related to the participants' objectivity, as with all self-reported measures, as well as the sample size. The results of this experimental study are therefore interpreted taking these limitations into account.

5 Results

From the 29 participants, 19 completed one full conversation with Liza. In this order, the 19 conversations were long enough and contained all the necessary questions and answers so a final score could be generated at the end of the conversations. For different and unstated by the participants reasons, 10 conversations were of various length but didn't reach the sufficient number of dialogical components to calculate a final score. Those were ended before the completion of the conversation. Therefore, the 10 unfinished logs couldn't be taken into account in analyses which included Liza scores. In such analyses only the 19 records with finalized dialogues and accumulated final score are presented. In analyses where Liza scores are not included, data from the complete sample of 29 participants is used.

5.1 Descriptive statistics

Given the small sample, all responses of recent university education could be analyzed within the category university students. Twenty seven (93,1%) of the respondents stated having experience with online education, two (6,9%) committed not having any experience with online learning, but still reported that they were familiar with this type of learning and could therefore answer the relevant questions.

The measured LOC where the scale's possible minimum is 0 and the possible maximum 23 resulted to values between 5 and 22 (M=12,59 and SD=3,84). Suggested categorization for the Rotter's Locus of Control Scale advises a high score to be sampled as External LOC and a low score as Internal (Rotter, 1966). Therefore, the numerical answers have been categorized in internal LOC for scores between 0 and 11 (lower scores) and external LOC for scores between 12 and 23 (higher scores). This conversion in LOC category resulted in 10 participants with Internal LOC and 19 with an external one (Figure 5).

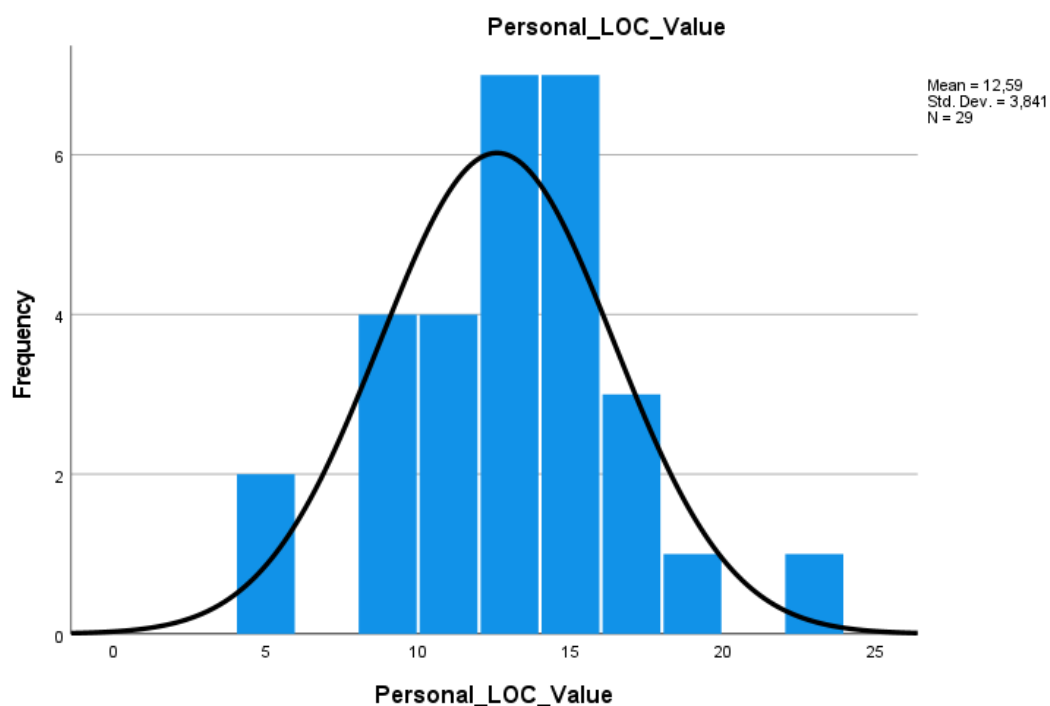


Figure 6. Personal LOC measurement

Nineteen participants completed a whole conversation with Liza (N=19) with a minimum generated score 0 and a maximum of 83,33 (M=55,48 and SD=21,16) (Table 2). Accordingly, to these results, there is missing data in 10 of the records which cannot provide any representative correlation to other predictors and outcomes of the research by using the complete data set.

Table 1. Personal score at the end of a dialogue with CA Liza (in %)

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	,00	1	3,4	5,3	5,3
	25,00	1	3,4	5,3	10,5
	37,50	3	10,3	15,8	26,3
	50,00	3	10,3	15,8	42,1
	62,50	6	20,7	31,6	73,7
	75,00	3	10,3	15,8	89,5
	83,33	2	6,9	10,5	100,0
	Total	19	65,5	100,0	
Missing	System	10	34,5		
Total		29	100,0		

Descriptive statistics on the pre- and post-experimental conditions between and within groups results to an overview regarding LOC and IMI (Appendix V, Table 5). The between subject analysis shows higher motivation among the participants with an internal LOC and lower motivational score for the external LOC ones for all 3 sub-categories of both pre- and post-experimental conditions. When we compare means in average IMI scores, there seems to be a general decreasing of the motivation in both groups. Similarly, with respect to the three sub-scale scores, a same pattern appears within the internal LOC group. However, in the external LOC group we see a slight increase in the post-experimental sub-scale interest/enjoyment. These trends, if verified by appropriate statistical analyses, might show evidence of an enhanced pleasurable experience and likeness-motivation among the individuals with external LOC toward the SCA dialogues.

The descriptive results, including intrinsic motivation and age, reveal an interesting difference among different generation groups. Regarding the online learning activities in their pre-experimental conditions, younger people below 25 (young adolescents) seem to be less motivated than the older generations. Similar results appear in the three post-experimental subscale measurements, but the motivational difference is less prominent (Appendix V, Table 6, Table 7). Interesting diversity is to be seen in the change of the IMI-subscale interest/enjoyment, as this is the only scale which appears at best to increase (or stay at the same level) among the adolescents (below 25 years of age). The other two sub-scales (perceived competence and value/usefulness) show a decrease in the motivation after the intervention in both age groups. Which of these differences are statistically significant will be presented below.

5.2 Correlations

For the purposes of this study the independent variables gender, age, level of study and LOC have been defined. The dependent variable IMI has been evaluated in pre-experimental and post-experimented conditions. A second dependent variable, Liza's final score has been recorded, but gives a representative result only for the 19 records of participants, who finished their conversational sessions so to generate a final score. For some analyses Liza's final score has been analyses as an independent variable for predicting motivational differences.

Within the complete correlational dependencies between the dependent and the independent variables a couple of meaningful results can be seen. First, the correlation between age and study shows a positive correlation, $r(27) = .724$ and $p < .001$ which is expected, as it is a logical correlation among the population. Therefore, no further analysis based on these data had been conducted. Second, a significant correlation can be seen between one's final Liza score and their general motivational score difference ($IMI_2 - IMI_1$) $r(17) = .606$ and $p < .006$. None such one has been found between their LOC or any other variables (Appendix V, Table 3).

Based on these results a more detailed correlational analysis reviewed the correlation between the IMI sub-categories interest/enjoyment, perceived competence, and value/usefulness to each other as well as the other variables (Appendix V, Table 4). In terms of age, a negative correlation between age and Liza's score has been observed $r(17) = -.607$ and $p = .006$. Age has also been found to be significant in a positive correlation to the main pre-IMI score $r(27) = .553$ and $p = .002$, and specifically in the sub-category interest/enjoyment $r(27) = .609$ and $p < .001$. The post-experimental IMI scores do not show this correlation. The general pre-IMI score also shows a negative correlation to one's personal LOC value $r(27) = -.495$ and $p = .006$. The other variables didn't show significant differences in the results data, related to the research hypothesis.

5.3 Group Comparisons & Regression

There were two ANOVA analyses included where no significant effects were found in the direction of the interactions of main interest, namely Liza's score and LOC. In the first one we considered the pre- and the post-IMI time of measurement as a first, and the three IMI sub-scales as a second within-subject variable. A between-subjects independent variable was Liza scores, where score up to 49% define group 1 (N=5), score between 50% and 74%, group 2 and score of 75% (N=9) and above group 3 (N=5). There was no statistically significant interaction between time and Liza scores ($F(2,16)=1.803$, $p=.197$), IMI sub-scales and Liza scores ($F(2,16) =1.324$, $p=.284$), or time, IMI sub-scale and Liza scores ($F(2,16)=.274$, $p=.892$). However, the within group interaction of time and IMI sub-scale was significant ($F(1,17)=6.982$, $p=.018$) and therefore a follow-up t-test was done to explore which sub-scale(s) showed a significant difference between the two measurements.

In the paired samples T-test present, we tested the difference of the three IMI sub-scales, before and after the interaction with Liza (Pre Interest Enjoyment and Post Interest Enjoyment, Pre Perceived Competence and Post Perceived Competence, and Pre Value Usefulness and Post Value Usefulness). Pre Interest Enjoyment – Post Interest Enjoyment report a $t=-.45$ ($M=-0.97$, $SD=11.51$ and $p=.66$). Pre Perceived Competence – Post Perceived Competence report a $t=4.46$ ($M=7.03$, $SD=8.49$ and $p<.001$). Pre Value Usefulness – Post Value Usefulness report a $t=3.36$ ($M=3.17$, $SD=5.09$ and $p<.002$). These findings suggest that there are significant differences between the pre- and post-measurements for the scores of perceived competence and value/usefulness, specifically decrease in the relevant scores. On the contrary, no significant change was observed for interest/enjoyment scores.

For a similar ANOVA analysis, two groups were made based on Liza scores. Group 1, with Liza's score up to 50% (N=8) and group 2, with score 51% and above (N=11). Again, there was no statistically significant difference between time and Liza's score ($F(1,17)=2.144$, $p=.161$). No statistically significant difference was found also between each IMI sub-scale and Liza's score

($F(1,17)=1.041$, $p=.376$). No statistically significant difference between time, IMI sub-scale and Liza's score were found either ($F(1,17)=.343$, $p=.715$). Here, a significant within group interaction of time and IMI sub-scale was also found ($F(1,17)=7.379$, $p=.015$). The variable of the groups based on Liza scores did not appear in any other significant main effect or interaction.

The second ANOVA analyses considers the LOC category, internal ($N=10$) or external ($N=19$), as a between-subject factor. There was no statistically significant difference between time and LOC ($F(1,17)=.109$, $p=.744$). No statistically significant difference was found also between each IMI sub-scale and LOC ($F(1,17)=.358$, $p=.703$). No statistically significant difference between time, IMI sub-scale and LOC were found either ($F(1,17)=.114$, $p=.893$). This analysis again has not provided sufficient evidence for statistically significant effects of LOC scores on the dependent variable (IMI).

Also, a regression analysis has been conducted to predict the difference between the pre- and the post-IMI measurement in the general IMI score, and uncover the factors influencing such changes, using Liza scores, LOC, age, gender and study level as predictors. The analyses show that only Liza scores are significant predictors of this change, where higher Liza scores predict smaller decrease in IMI scores ($t=3.14$ and $p=.006$). Other factors did not significantly influence this difference in IMI scores. Additional regression analyses shows that only age is a significant predictor of Liza scores ($t=3.15$ and $p=.006$). This analysis shows that the younger participants score higher on Liza's tests than older ones.

In terms of LOC, similar regression analyses found LOC as a predictor for pre-IMI scores ($t=-2.96$ and $p=.006$). The negative correlation predicts that individuals with external LOC are generally less motivated to deal critically with decision making tasks and ones with an internal LOC are more motivated to do so. This is an additional but interesting finding which demonstrates the value for further research on this association.

5.4 Qualitative Feedback

An important result of this research can be seen in the last chapter of the presented questionnaire, where participants were able to add non-obligatory personal remarks regarding any matter of the research. Sixteen out of the 29 respondents filled this field, wherefrom 15 (52%) submitted factual observations and 1 did not contribute meaningfully (Table 8). These comments are given by both types of participants, the ones who did finalize a dialogue with Liza and also the ones who ended the conversation before reaching a final score.

Table 2. Participant 's responses on the question "Personal remarks"

ID	Answers
1	The chat bot format is rather annoying, as it would be better to do it in a more elegant and graphical manner, maybe like the format of brilliant.org
2	6
3	I answered two questions correctly, but I worded them differently and Liza said it was wrong. Example: I typed one in seven but I should have written 14% even though they are the same. Also: it was very unclear how I had to answer the second question (small sample / luck). I like the test and I like the idea but it's demotivating when you answer right and the AI says your answer is wrong. One last tip: keep a few seconds between every text from Liza, it's very fast and no one reads that fast but the screen doesn't stay in the current text you're trying to read which is a bit annoying.
4	Critical thinking skills are vital, but Liza is simply not interactive enough (especially in the age of GPT). Most of the time I felt like I was doing a pen-on-paper exam. I even got stuck on a question because Liza would not recognize my answer and I had to say I don't know the answer.
5	Liza is very clunky and on questions like "why do you think that is" generally an answer with more than three words is required, but I suppose it's very hard for Liza to identify whether the given answer is correct. Maybe Liza needs more training as a language model.
6	-
7	I LOVE IT!
8	I did this survey on my phone, and everything went very well, except for one thing. The conversation with Liza was not great, she typed way to fast for me to keep up, and every time she sent a message my screen would scroll down. Because of this I couldn't read Liza's messages at a normal speed. The interactions with Liza were also quite weird. But the survey itself was great.
9	The answers the bot gives are very fast maybe that could be a little slower.
10	Liza isn't very well developed in terms of communication skills.

11	The bot was somewhat annoying because it typed the explanations and problems very fast and rushed me to answer.
12	Liza didn't understand some of the answers I gave to her. and the typing went fast like it was programmed, which made it hard to read while the messages came. But overall a good form. I'm actually interested in the statements.
13	It was a fun activity and i liked the concept. Liza was great to talk to although she did not have the ability to answer, or know how to answer all of my questions. She also reacted quite fast while i was still reading the messages she sent me. Next to these two things i was really invested in the conversation i had with Liza.
14	The Liza bot was very impatient and often didnt respond well
15	Great experience! Liza should give more time before texting back! Good luck
16	I enjoyed your research very much! Thank you!

Three important points can be made based on these qualitative remarks. First, 52% of the participants has been intrinsically motivated to respond to this non-compulsory question. These respondents filled in their own perception of the experimental conditions, the research design and own experience and understanding regarding the dialogue with CA Liza. Second, 11 of the given answers focus on Liza’s technical specifications and functionality. They describe: 1. Liza to be too fast in responding and refer to it as “impatient” and recommend that more waiting time should be allowed; 2. to deem the given screen space insufficient for the follow-up conversation; 3. suggesting specific points for improvement of the CA engine and inevitably compare Liza with newer generation chatbots, such as ChatGPT; 4. experiencing an incomplete understanding by Liza (Figure 6); 5. given right answers to be incorrectly treated as mistakes, due to different formulation from the pre-programmed right answer, these were found to be a frustrating and demotivating factor in the conversation. Overall, respondents have put considerable effort to be helpful in the given feedback and to point out the specific deficiencies in Liza’s interaction, training as a language model, graphics, and lay-out specifications. Third, participants refer to Liza as an identified person and call the CA by name, with or without referring to it as a bot or a language model. Some of them react positively on Liza’s teaching model and although many are critical to the technical issues, they state the importance of the contextual training and the presented conversational model.

Liza

Hello M..., you can now chat with Liza.

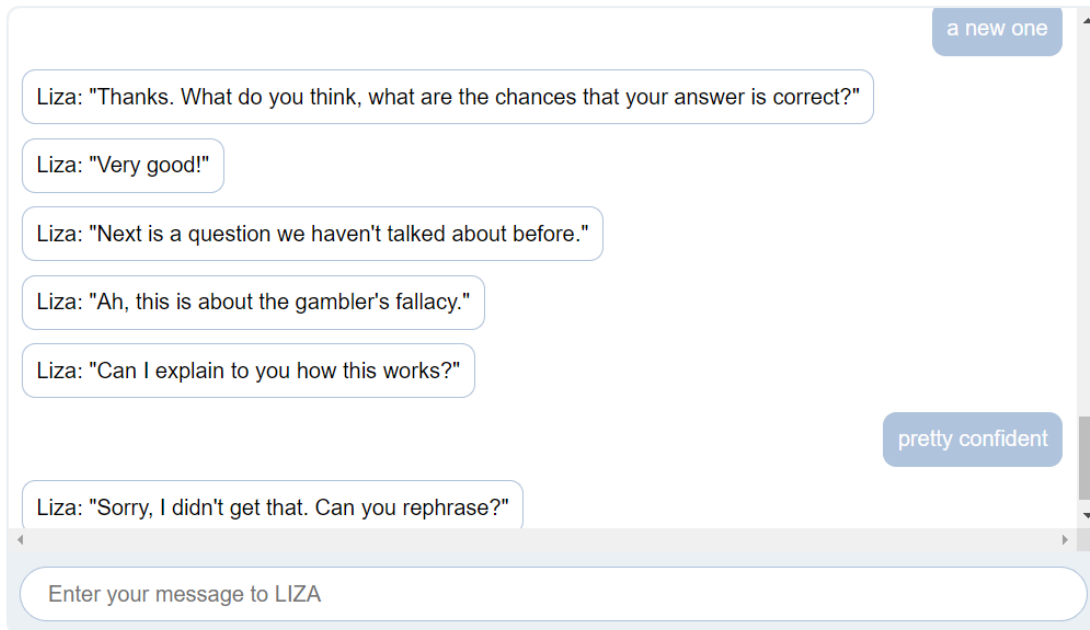


Figure 7. Participant's screenshot: feedback on CA Liza <https://physiological-edu.informatik.hu-berlin.de/liza>

6 Discussion

The present work focused on the impact of the interaction with the CA Liza on students' intrinsic motivation for working on critical thinking. The study builds on the importance of critical thinking in education, the role of the Socratic approach, and the influence of motivation on academic performance. The hypothesis is that Liza's interaction will affect students' motivation for critical thinking. The research also considers the influence of age and locus of control on intrinsic motivation.

To test our hypothesis, the study involved 29 participants, including high school, college, and university students, as well as teachers/lecturers, self-selected among above 500 invited participants targeting college and university students. The participants engaged in a dialogue with CA Liza, where they were presented with different common fallacies puzzles. Liza evaluates their answers and provides feedback. The participants' intrinsic motivation is measured using an intrinsic motivation inventory (IMI) before and after the interaction with Liza. The IMI includes questions related to interest/enjoyment, perceived competence and value/usefulness. Participants also complete a Locus of Control (LOC) questionnaire. With this study we aimed to provide insights into the impact of CA, containing elements of Socratic dialogue, on students' motivation towards applying critical thinking and the implications for educational programs, particularly in the context of online education.

Contrary to the anticipated outcome of an increase in motivation, the findings of the study revealed a notable decrease instead. In fact, the study results speak of a general decrease in motivation due to the intervention. Despite the unexpected direction of the effect, there is compelling evidence to support the notion that the intervention, involving the interaction with Liza, has a significant impact on measures of motivation.

Further analyses provided other four main findings, clarifying the relation to the research hypotheses. First, age plays a significant role in influencing the level of engagement that participants have with the CA, with younger individuals demonstrating a higher degree of interest and enjoyment. Within the population sample, it is observed that younger people exhibit a stronger affinity with the CA in terms of interest and enjoyment compared to the older generations. That does point an increased young-age motivation, and a specific difference in the type of motivational stimuli between the generations, due to the intervention with Liza. Second, age has been found to be a predictor for critical thinking, as Liza's scores show. Younger generations show better critical thinking skills results during the intervention than the older ones. Their ability to engage to the analytical reasoning on the presented tasks implies the difference in Liza's effectiveness between the groups. Third, final scores obtained from Liza's interactions have shown to be a significant predictor of the changes in individuals' intrinsic motivation. It has been observed that individuals with higher scores from their interactions with Liza demonstrate smaller changes in their motivation levels and tend to maintain a more stable level of intrinsic motivation, whereas those with lower scores experience greater decreases in motivation. These findings suggest that the influence of Liza's interaction on one's intrinsic motivation may vary, based on the level of engagement and performance demonstrated during the interaction. This highlights the importance of performance and engagement with Liza in shaping intrinsic motivation outcomes. Fourth, there is a relationship found between participants' LOC and their pre-experimental Intrinsic Motivation Inventory (IMI) scores. Participants with intrinsic LOC tend to show higher intrinsic motivation, than the ones with an external LOC. This intriguing findings underscore the importance of further research to better understand the relationship between LOC and motivation and to explore which stimuli would enhance the motivation of individuals with an external LOC. Additionally, half of the respondents (52%) were intrinsically motivated to do more than only participate to the research but also took the initiative to suggest improvements to the CA, give particular feedback on specific elements of the AI engine and also gave compliments and express enjoyment and gratitude for the experience.

All our findings reject the H_0 hypothesis and support H_1 . CA Liza did influence the motivational level, however with the opposite outcome than the expectations. Hypothesis H_2 has been supported by the results, showing specific higher level and enhancement of enjoyment and interest among the younger participants. Our H_3 has been rejected and in the contrary of the research expectations, younger people showed better critical thinking scores. Regarding LOC, results support H_4 and provide results for a correlational tendency with intrinsic motivation.

The findings of this study have been previously discussed in the literature. Critically disputed topics, as some crucial ones in this work, hold attention on the importance of their high level application in education but also question the currently implemented methodologies for skills enhancement. Our results demonstrate the importance of keeping the investigation on the researched symbiosis as presented in our study and provide evidence for the potential for effective application in education.

Our findings on the effect of CA on human cognition dispute the study hypothesis as a general research question. The surprising decreased motivation in our findings could be caused by a number of aspects. A clarification, justifying this outcome, we can see in the motivational differences between the three specific IMI sub-scales. The paired T-test finds the greatest decrease in motivation among the perceived competence and the perception of value and usefulness. No such significant change has been observed in participants interest/ and enjoyment engagement. These findings suggest that participants could have experienced a decline in motivation when they realized a discrepancy between their expected performance and their actual perception of competence (Howard et al., 2015). This discrepancy may have led to a reduced perception of the value and usefulness of the interactions with Liza, resulting in decreased motivational levels (Eccles & Wigfield, 2002). The reason for this decrease could also be a common effect occurring in intervention studies, supported by repeated longitudinal post-interventional measurements. In such studies it often can be observed that the first measurements show a decline and at later times they increase, where the number of the repeated measurements should be seriously considered (Moerbeek, 2008; Vickers, 2003). Eventually, something we can only propose as a possibility in our one post-interventional measurement. Another reason for this decrease could also be suggested from the reviewed literature, where motivation appears to decrease when tasks or

decision-making problems are just too hard to be solved (Suhadi et al., 2015). Our IMI questionnaire did not cover all seven IMI subscale, but only three of them. Implementing the complete inventory could have given more insights to the reason of the motivational decrease. Moreover, the differentiated layers of the pre- and post-experimental IMI analysis support the perception of multiple dependencies between CA and the individual's personal cognitive values (Lara & Deckers, 2020). The extent of the CA's autonomy and its own limitations might be of a significance importance for one's motivation and general engagement (Table 8). Other findings suggest that AI autonomy is a valuable part of the human-AI interaction for the efficiency of the given cognitive tasks. Hu et al. (2021) discuss on this matter stressing the importance of the AI autonomy for the effect on human attitude. *"This research significantly improves our theoretical understanding of the impact of artificial autonomy on mind perception of AI artifacts, which, in turn, influences users' behavior."* (Hu et al., 2021, p. 11).

Towards human autonomy and motivation in decision making, the collected data have result to the sufficient findings, which correlate age, critical thinking abilities and differences in intrinsic motivation. A point of interest is the current young generation's performance on decision-making tasks, who appear to surprisingly apply better critical thinking ability and logical argumentation, disputing the research hypothesis and arguing other previous findings (Spector & Ma, 2019). The more important finding, correlated with age, is the young participants' interest and enjoyment. Namely, younger people do not only show better interest and enjoyment in technological applications in education than the older ones, but also show increased motivational interest and enjoyment after the intervention. This motivational enhancement is supported by earlier findings revealing that new teaching approaches are required, which should involve advanced technology, interactive methods, and personalized innovations to keep the motivation of new generations and help it grow (Chaudhuri, 2020; Yin et al., 2021). Ying and collaborators (2021) suggest in the same study that by providing chatbot-based task-oriented dialogue and small, personalized portions of information tailored to individual information needs and attention spans to be highly suitable for the new generation of students has shown a substantial improvement in motivation and performance among millennial and Generation Z. In our results we are not able to distinct whether the younger participants did perform higher on the critical thinking tasks due to the intervention as suggested in the literature or perhaps they were more common with the specific testes tasks than the older participants. However, the general higher engagement among the youth

in our sample partly support our main hypothesis regarding motivational engagement towards the applied CA.

The motivational differences and their dependency on critical thinking abilities we find a fascinating outcome. Our results present critical thinking to be a motivational stimulus or a factor for motivational persistence. Respecting these findings, a support by Critical Thinking Motivational Scale, proposed by Valenzuela et al. (2017), provide better understanding. Their approach helps assess a person's motivation, value, and expectations regarding critical thinking tasks and allows for interventions to address motivational deficiencies and provides insights for improving critical thinking performance and instructional efforts. Although we have not been able to investigate whether the critical thinking level at the end of our intervention has been resulted due to the intervention or it has been a first comparable measurement, such results in the motivational critical thinking dependency have been similarly supported (Valenzuela et al., 2017). Extended future research should investigate these associations into more details.

With respect to LOC, we have tried to make a step in understanding how it relates with intrinsic motivation. Although our findings suggest individuals with an internal LOC to be more motivated towards critical approaches, we haven't really provided evidence of enhanced motivation of the ones with an external LOC after the intervention. We couldn't truly find this relation as suggested by Ng et al. (2006), which could be caused by various limitations of our study. Despite the fact that we were not able to provide a clear understanding how one's LOC influences intrinsic motivation, we are determined to suggest future research, building upon ours, including larger samples and possibly various and more advanced research tools. As Kovach (2018) proposes, further research should investigate these relations also with respect to individual's mindset and across disciplines to provide these searched answers.

Regarding our main statement and propositional framework, which seems to make it possible to affect intrinsic motivation through the interaction with CA, previous studies examined the impact of different types of pedagogical agents and agent enthusiasm on various variables in multimedia learning. Preferences for pedagogical agents vary based on factors like age and gender, suggesting the need for improved designs and effectiveness assessments in different courses. The findings shed light on the significant influence of pedagogical agents on academic success and recommend

their use in computer-assisted software. These studies also emphasize the importance of prioritizing students' interest in human interaction in blended learning, suggesting that perhaps the more human-like the interaction is with the CA, the better the interest and engagement in the teaching approach. Furthermore, they highlight the role of agent enthusiasm in eliciting emotional and motivational responses in learners, stressing the need for careful design of pedagogical agents' verbal and nonverbal cues. Findings show that blended learning with CA positively affects students' self-efficacy and intrinsic motivation as important factors for academic achievement. However, other motivational factors should also be considered, and interventions should align with motivational theories to maximize their effectiveness (Dinçer & Doğanay, 2017; Fryer et al., 2019; Liew et al., 2017; Siddiqui et al., 2020).

In our study, the research outcomes provide evidence for factual motivational changes due to the applied intervention. Similarly, factual motivational differences between groups have been observed, distinguishing age and LOC as essential characteristics for this difference. The findings also speak of distinctive levels and perhaps contrasting ways of applying critical thinking skills between ages. All together these findings should be considered for future pedagogical innovations in educational design. Important here is to achieve understanding how underlying cognitive mechanisms effect one's motivation and one's critical thinking between ages, so tailored future programmes would successfully apply personalized curricula. Besides that, a better understanding of which factors enhance the motivation of individuals with an external LOC could be useful for various implications within the education field. Accordingly, such recommendations should be beneficial for technological innovations in human-computer interaction and AI symbioses. Not only in competitive measurable abilities, like comparative scores and quantitative grades but also in social engagement towards various kind of hard and soft skills, such AI applications could provide multivariable ways for achieving academic development.

The human-AI interaction in the context of intrinsic motivation towards critical thinking and rational decision making by applying Socratic dialogue is the main focus of this research. Although the study results speak of meaningful research findings, the experimental research of this study recognizes several limitations. The small population sample could not provide the sufficient amount of data, so various meaningful results could suggest well defined positive or

negative predictors. Much larger data sample should be used to present significant correlations within the research terms. In addition, some dialogues with the CA Liza were incomplete. This has resulted to missing data in critical thinking scores. Other research limitations are related to the representativeness of the sample and the experimental study conditions. More than 500 participants were randomly invited, out of whom a self-selected sample of 29 participants responded. They completed the study in an uncontrolled experimental environment, at their own environment and time. These factors might influence the research outcome and introduce possible biases. In align with the previous limitations, the short span of experimental duration captured the assessment and measurements only to a certain extend. However, we need to point out that the research design has provided sufficient methodological instruments, in order to minimize research data noise. With a larger sample size, we would have welcomed a more varied group of individuals, enabling us to analyze the effects of much more important differences in demographics and characteristics. Further research, including a control group and experimenter-defined experimental conditions would add to the confidence in the reliability of present results. Extendedly, it would be valuable to improve such further research design by expanding the assessment of logical reasoning and fallacies by incorporating additional measures. This could involve further interactions with Liza, utilizing its scoring system, or employing alternative research tools, like Watson-Glaser appraisal (Psychological Corporation et al., 1980). Such an approach would allow researchers to investigate the influence of the intervention on both performance outcomes and intrinsic motivation. Additionally, repeated measurements should also be included as to assess longitudinal achievements and motivational changes. Such improvements should not only provide evidence of which factors influence motivational changes but also the reason why these occur. Considering our main findings, more relations should be found between age and motivational consistency and how these justify the outcome differences in critical thinking due to the interventions. Such finding will give insights for further developments in educational programmes and their design to adjust to one's personal manageable level.

On the importance of critical thinking in education, both our study and previous studies ask whether teaching generalizable skills can substantially affect academic and professional performance. For example, Monteiro et al. (2020) review the history of general thinking skills in medical education and dispute the instructing of non-contextual reasoning and advise against investing resources in such interventions. Instead, they propose enhancing practical knowledge gained through experience. Their work challenges the acceptance in (medical) education that

general analytical and rational methods are superior to knowledge-based, efficient strategies. The authors argue that emphasis on teaching general skills such as critical thinking, problem-solving, and reasoning and suggests that specialized expertise may be more valuable. However, questioning such explicit practices does not replace critical thinking, problem-solving, and reasoning as skills applied in contextual cases. Instead, it is a well-supported argument within our research hypothesis which gives even more sense in encouraging further research towards this field.

Social engagement as conversational element met several respondents' spontaneous reactions (Table 8). These participants have critically evaluated Liza as a factual conversational partner. Their review bring light to the urge of an interpersonal relationship with the CA. As suggested in the research hypothesis, a well-designed conversational agent should boost this interpersonal relationship and would enhance once motivation and engagement in the dialogue (Fahim & Bagheri, 2012; Le & Wartschinski, 2018; Makhene, 2019). The psychological interaction between humans and AI in daily life situations can influence individuals response to CA and could build a relationship with it (Kim et al., 2019). As Kim et al suggest, AI developers should focus on adding emotions and characteristics that improve the user's well-being. Our study findings support the hypothesis part, where adolescents, who are generally representative in students' population, are intrinsically interested in a conversational engagement with the CA. Additional research is needed to investigate the effects of interacting with AI on users' psychology.

However, people's tend to trust AI-generated suggestions, even when they are incorrect, might be corrected by cognitive forcing interventions so to reduce over-reliance (Buçinca et al., 2021; Smith & Szymanski, 2013; Spector & Ma, 2019). These studies investigate how people can detect incorrect AI recommendations and encourage analytical thinking to combat this problem and help to avoid blind trust in AI. Their recommendations suggest optimal human performance with AI-powered decision-support tools. *"Our research suggests that human cognitive motivation moderates the effectiveness of explainable AI solutions."* (Buçinca et al., 2021, p. 1,18). Our study research design, based on one single post-experimental condition does not pretend to cover a possible longitudinal observation based on continuous SCA intervention but such further study is to be advised for factual effectiveness on human's autonomy in critical thinking, and rational decision making.

Furthermore, although our study explicitly examines the effect of Socratic dialogue in the enhancing the researched cognitive skills, we were not able to provide evidence about how incorporated elements of the Socratic approach within CA actually influenced participants' interaction and their results. Earlier suggested recommendation for future works will also need to focus on the implementation of both modern and classic Socratic method in CA, designed for education. Finally, the Socratic dialogue should be also recommended for exploring in more details the proposed paradigm between intrinsic motivation and employing critical thinking abilities within the AI domain.

7 Conclusion

Concluding, we found that the intervention involving the interaction with Liza had a notable effect on measures of motivation, even though it resulted in a decrease in motivation, contrary to our expectations. This decline in motivation seems to be attributed to participants recognizing a mismatch between their expected performance and their perceived competence, consequently impacting, and reducing their perception of the value and usefulness of interacting with Liza. Age played a significant role in participants' engagement with the CA, with younger individuals showing a higher level of interest and enjoyment compared to older generations. Younger generations also demonstrated better critical thinking skills during the intervention compared to older generations, indicating the effectiveness of Liza's interaction on younger people. Moreover, participants' scores obtained from the interactions with Liza were found to predict changes in their intrinsic motivation. Higher scores, which resulted among the younger people, were associated with smaller changes in motivation, suggesting a more stable level of intrinsic motivation. The ones, performed poorly, show bigger drop in the motivation after intervention. This reflects to the decline of the perceived competence and the perception of value and usefulness, as observed. Additionally, participants with intrinsic LOC indicate higher pre-experimental intrinsic motivation than those with an external one, highlighting the relationship between locus of control and motivation. Analysis with much larger population sample should provide more evidence for motivational changes due to the intervention.

Eventually, the users' qualitative feedback was rather critical towards the CA instead of enthusiastic in terms of autonomy and flowless dialogue. An observation that corresponds with the inevitable motivational dependency on technological applications and implementations.

These findings underscore the need for further research to explore the relationship between the intervention, motivation, and other variables such as age and LOC. It is crucial to gain a deeper understanding of how underlying cognitive mechanisms impact motivation and critical thinking across different age groups in order to effectively implement personalized curricula and tailored programs. Considering all our results, it is important to incorporate such further studies into future educational designs and pedagogical innovations. Moreover, a better understanding of the factors that enhance motivation in individuals with an external LOC can have valuable implications in the field of education. Consequently, these recommendations can also contribute to technological innovations in human-computer interaction and AI, extending beyond competitive measures to enclose social engagement and the development of various hard and soft skills and provide multiple avenues for achieving academic development. We suggest that further studies will provide data for successful development and implication of Socratic CA within programs across disciplines, enhancing students' motivation towards critical thinking and rational decision making, increasing their own human autonomy. It is important to explore further connections between age and motivational consistency to understand how they contribute to the variations in critical thinking outcomes resulting from the interventions. These findings can provide valuable insights for the development of educational programmes that can be tailored to individual abilities and levels of manageability. The determined advancement of AI applications presents immense potential for enhancing education, when applied thoughtfully. The data from such further studies will provide important information to researchers, developers, and education designers to achieve such positive change in the human-AI symbiosis in teaching and will contribute to further successful incorporation of theoretical paradigms.

Appendix

Appendix I: Intrinsic Motivation Inventory

Adapted Intrinsic Motivation Inventory (IMI) pre- and post-experimental (Ryan, 1982; Ryan & Deci, 2000)

	Pre-experimental IMI Before interaction with Liza	Post-experimental IMI After interaction with Liza
Interest/Enjoyment		
self-report measure of intrinsic motivation <i>therefore, evaluating the 'likenesses of the present(ed) learning activities.</i> Pre: current online learning Post: cognitive assistant	I enjoy online learning very much.	I enjoyed the interaction with Liza very much.
	Online learning is fun to do.	The interaction with Liza was fun to do.
	I think online learning activities are boring. (R)	I thought interacting with Liza was boring. (R)
	Online learning activities do not hold my attention at all. (R)	The interaction with Liza did not hold my attention at all. (R)
	I would describe my recent online learning activities as very interesting.	I would describe the interaction with Liza as very interesting.
	I think that online learning activities are quite enjoyable.	I thought the interaction with Liza was quite enjoyable.
	While I'm doing online learning activities, I'm thinking about how much I enjoy them.	While I was interacting with Liza, I was thinking about how much I enjoyed it.
Perceived Competence		

<p>positive predictor of both self-report and behavioral measures of intrinsic motivation <i>therefore, evaluating the 'self-esteem' of own confidence in reasoning and decision making.</i> Pre: current performance Post: after practicing in one session with critical thinking/reasoning tasks</p>	<p>I think I am pretty good at critical thinking.</p>	<p>I think I was pretty good at critical thinking during my interaction with Liza.</p>
	<p>I think I'm doing pretty well at critical thinking activities, compared to others.</p>	<p>I think I did pretty well at critical thinking activities during my interaction with Liza, compared to others.</p>
	<p>After working on a complex problem for a while, I feel pretty competent.</p>	<p>After working on the complex problems presented by Liza for a while, I felt pretty competent.</p>
	<p>I am satisfied with my performance in tasks requiring critical thinking.</p>	<p>I am satisfied with my performance at Liza's activities.</p>
	<p>I am pretty skilled at critical thinking.</p>	<p>I was pretty skilled at critical thinking activities presented by Liza.</p>
	<p>I don't do very well on solving complex problems. (R)</p>	<p>I didn't do very well on solving complex problems presented by Liza. (R)</p>
<p>Value/Usefulness</p>		
<p>self-regulating with respect to activities that one experience as useful or valuable for themselves <i>therefore, evaluating the importance of practicing with critical thinking activities.</i> Pre: current importance Post: after practicing with critical thinking tasks</p>	<p>I believe critical thinking activities could be of some value to me.</p>	<p>I believe learning critical thinking through the interaction with Liza could be of some value to me.</p>
	<p>I think that doing study activities on critical thinking is useful.</p>	<p>I think that doing study activities on critical thinking is useful.</p>
	<p>I think critical thinking activities are important to do because they can improve my reasoning and decision making skills.</p>	<p>I think critical thinking activities are important to do because they can improve my reasoning and decision making skills.</p>
	<p>I would be willing to practice critical thinking more often because it has some value to me.</p>	<p>I would be willing to practice critical thinking more often because it has some value to me.</p>
	<p>I think doing critical thinking activities could help me to improve my reasoning and decision making.</p>	<p>I think doing critical thinking activities could help me to improve my reasoning and decision making.</p>
	<p>I believe practicing critical thinking could be beneficial to me.</p>	<p>I believe practicing critical thinking could be beneficial to me.</p>
	<p>I think improving my critical thinking is an important activity.</p>	<p>I think improving my critical thinking is an important activity.</p>

Appendix II: Rotter's Locus of Control Scale

Rotter's Locus of Control Scale (Rotter, 1966)

	Locus of Control
1 filler	a. Children get into trouble because their parents punish them too much. b. The trouble with most children nowadays is that their parents are too easy with them.
2	a. Many of the unhappy things in people's lives are partly due to bad luck. b. People's misfortunes result from the mistakes they make.
3	a. One of the major reasons why we have wars is because people don't take enough interest in politics. b. There will always be wars, no matter how hard people try to prevent them.
4	a. In the long run people get the respect they deserve in this world. b. Unfortunately, an individual's worth often passes unrecognized no matter how hard he tries.
5	a. The idea that teachers are unfair to students is nonsense. b. Most students don't realize the extent to which their grades are influenced by accidental happenings.
6	a. Without the right breaks one cannot be an effective leader. b. Capable people who fail to become leaders have not taken advantage of their opportunities.
7	a. No matter how hard you try some people just don't like you. b. People who can't get others to like them don't understand how to get along with others.
8 filler	a. Heredity plays the major role in determining one's personality. b. It is one's experiences in life which determine what they're like.
9	a. I have often found that what is going to happen will happen. b. Trusting to fate has never turned out as well for me as making a decision to take a definite course of action.
10	a. In the case of the well-prepared student there is rarely, if ever, such a thing as an unfair test. b. Many times exam questions tend to be so unrelated to course work that studying is really useless.
11	a. Becoming a success is a matter of hard work, luck has little or nothing to do with it. b. Getting a good job depends mainly on being in the right place at the right time.
12	a. The average citizen can have an influence in government decisions. b. This world is run by the few people in power, and there is not much the little guy can do about it.
13	a. When I make plans, I am almost certain that I can make them work. b. It is not always wise to plan too far ahead because many things turn out to be a matter of good or bad fortune anyhow.
14 filler	a. There are certain people who are just no good. b. There is some good in everybody.
15	a. In my case, getting what I want has little or nothing to do with luck. b. Many times we might just as well decide what to do by flipping a coin.
16	a. Who gets to be the boss often depends on who was lucky enough to be in the right place first. b. Getting people to do the right thing depends upon ability. Luck has little or nothing to do with it.
17	a. As far as world affairs are concerned, most of us are the victims of forces we can neither understand, nor control. b. By taking an active part in political and social affairs the people can control world events.

18	a. Most people don't realize the extent to which their lives are controlled by accidental happenings. b. There really is no such thing as "luck."
19 filler	a. One should always be willing to admit mistakes. b. It is usually best to cover up one's mistakes.
20	a. It is hard to know whether or not a person really likes you. b. How many friends you have depends upon how nice a person you are.
21	a. In the long run, the bad things that happen to us are balanced by the good ones. b. Most misfortunes are the result of lack of ability, ignorance, laziness, or all three.
22	a. With enough effort we can wipe out political corruption. b. It is difficult for people to have much control over the things politicians do in office.
23	a. Sometimes I can't understand how teachers arrive at the grades they give. b. There is a direct connection between how hard I study and the grades I get.
24 filler	a. A good leader expects people to decide for themselves what they should do. b. A good leader makes it clear to everybody what their jobs are.
25	a. Many times I feel that I have little influence over the things that happen to me. b. It is impossible for me to believe that chance or luck plays an important role in my life.
26	a. People are lonely because they don't try to be friendly. b. There's not much use in trying too hard to please people, if they like you, they like you.
27 filler	a. There is too much emphasis on athletics in high school. b. Team sports are an excellent way to build character.
28	a. What happens to me is my own doing. b. Sometimes I feel that I don't have enough control over the direction my life is taking.
29	a. Most of the time I can't understand why politicians behave the way they do. b. In the long run, the people are responsible for bad government on a national, as well as on a local level.

Appendix III: Information Letter

Dear student,

The present study, which is a part of the dissertation for my Master's in Cognitive Systems, aims to evaluate the impact of pedagogical agents on students' motivation. The results can lead to important insights for future innovation and increased efficiency in education.

You are invited to participate in this study. Participation is completely anonymous and voluntary. The experiment takes place online, so you can participate at any time convenient for you, from any country and timezone. The research study requires you to fill in two questionnaires, before and after you interact with an AI instructor. The total duration of your participation is estimated to be around 25 minutes, with around 10 minutes devoted to filling in the questionnaires. The time of the interaction with the AI might vary from person to person.

The procedure is in accordance with the Open University of Cyprus' research guidelines and has been approved by Dr. Konstantinos Tsagkaridis.

For any queries or clarifications regarding the research, please contact me via elitza.voutcheva@st.ouc.ac.cy

For any complaints or concerns you can contact directly Dr. Tsagkaridis at Konstantinos.Tsagkaridis@ouc.ac.cy

Please note that motivation is a critical component for the success of educational programmes. At the same time, AI is a huge innovation of our times, which can significantly improve education, when used correctly. The data from your participation in this study will provide important information to researchers, developers, and education designers to achieve such positive change. Therefore, your participation in the study is highly appreciated. If you believe in the importance of such a study, please show your additional support by sharing the link to this study to your peers and acquaintances and encouraging them to also participate.

Thank you for your time and help!

You may start your participation by opening the following link or scanning the QR code→
<https://forms.office.com/e/wXymzf3fFU>

If the link above does not work, try copying it into your web browser. Please, read and follow the given instructions with care.

Appendix IV: Informed Consent

Dear student,

Thank you for participating in this research! Your participation is completely anonymous and confidential. Your participation and your answers are important only for the goals of this study and would never be related to your personal data. The answers will not be shared with third parties. All research data will be stored solely during the duration of the research and will be deleted afterward.

This research contains several parts which are presented in the order they should be participated. Please, complete each part in the given order. The total duration of your participation is estimated to be around 25 minutes, with around 10 minutes devoted to filling in the questionnaires. The time of the interaction with the AI might vary from person to person.

Please, follow the steps below in the order shown. You can participate only once.

- I confirm, that I read the information above and I agree that my answers will be used for the purpose of this research. (Field to be checked by the participant)

Appendix V: Tables of Correlation

Table 1. Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
Gender_F_M	29	1	2	1,62	,494
Age	29	17	46	25,34	7,916
Study_Work	29	1	2	1,14	,351
Experience_Online_Value	29	0	1	,93	,258
Score_Liza	19	,00	83,33	55,4821	21,15684
Personal_LOC_Value	29	5	22	12,59	3,841

Table 3. Correlations between independent and dependent research variables

		Score_Liza	Personal LOC Value	Dif_IMI	d_IntEnj	d_Useful	d_Comet
Score_Liza	Pearson	1	-,170	,606**	,474*	,519*	,500*
	Correlation						
	Sig. (2-tailed)		,486	,006	,041	,023	,029
	N	19	19	19	19	19	19
Personal LOC Value	Pearson	-,170	1	,175	,195	,115	,089
	Correlation						
	Sig. (2-tailed)	,486		,364	,310	,553	,645
	N	19	29	29	29	29	29
Dif_IMI	Pearson	,606**	,175	1	,883**	,704**	,796**
	Correlation						
	Sig. (2-tailed)	,006	,364		<,001	<,001	<,001
	N	19	29	29	29	29	29
d_IntEnj	Pearson	,474*	,195	,883**	1	,487**	,487**
	Correlation						
	Sig. (2-tailed)	,041	,310	<,001		,007	,007
	N	19	29	29	29	29	29
d_Useful	Pearson	,519*	,115	,704**	,487**	1	,440*
	Correlation						
	Sig. (2-tailed)	,023	,553	<,001	,007		,017
	N	19	29	29	29	29	29
d_Comet	Pearson	,500*	,089	,796**	,487**	,440*	1
	Correlation						
	Sig. (2-tailed)	,029	,645	<,001	,007	,017	
	N	19	29	29	29	29	29

** . Correlation is significant at the 0.01 level (2-tailed).

* . Correlation is significant at the 0.05 level (2-tailed).

Table 4. Correlations between independent and dependent research variables sub-categories

Correlations												
		Age	Score_Liza	Personal_LOC_Value	Pre_IMI	Pre_Interest_Enjoyment	Pre_Perceived_Competence	Pre_Value_Usefulness	Post_IMI	Post_Interest_Enjoyment	Post_Perceived_Competence	Post_Value_Usefulness
Age	Pearson Correlation	--										
	N	29										
Score_Liza	Pearson Correlation	-.607**	--									
	Sig. (2-tailed)	.006										
	N	19	19									
Personal_LOC_Value	Pearson Correlation	-.042	-.170	--								
	Sig. (2-tailed)	.828	.486									
	N	29	19	29								
Pre_IMI	Pearson Correlation	.553**	-.284	-.495**	--							
	Sig. (2-tailed)	.002	.238	.006								
	N	29	19	29	29							
Pre_Interest_Enjoyment	Pearson Correlation	.609**	-.469*	-.372*	.824**	--						
	Sig. (2-tailed)	<.001	.043	.047	<.001							
	N	29	19	29	29	29						
Pre_Perceived_Competence	Pearson Correlation	.169	-.007	-.277	.632**	.195	--					
	Sig. (2-tailed)	.382	.978	.146	<.001	.312						
	N	29	19	29	29	29	29					
Pre_Value_Usefulness	Pearson Correlation	.302	.085	-.462*	.750**	.335	.535**	--				
	Sig. (2-tailed)	.112	.729	.012	<.001	.076	.003					
	N	29	19	29	29	29	29	29				
Post_IMI	Pearson Correlation	-.011	.439	-.217	.208	.241	-.027	.169	--			
	Sig. (2-tailed)	.955	.060	.257	.278	.208	.888	.380				
	N	29	19	29	29	29	29	29	29			
Post_Interest_Enjoyment	Pearson Correlation	.140	.113	-.126	.157	.334	-.213	.026	.858**	--		
	Sig. (2-tailed)	.470	.644	.513	.415	.077	.268	.893	<.001			
	N	29	19	29	29	29	29	29	29	29		
Post_Perceived_Competence	Pearson Correlation	-.186	.488*	-.067	.015	.013	.133	-.095	.696**	.371*	--	
	Sig. (2-tailed)	.334	.034	.731	.940	.946	.491	.625	<.001	.048		
	N	29	19	29	29	29	29	29	29	29	29	
Post_Value_Usefulness	Pearson Correlation	-.033	.432	-.343	.330	.114	.120	.595**	.557**	.311	.099	--
	Sig. (2-tailed)	.864	.064	.068	.080	.557	.535	<.001	.002	.100	.608	
	N	29	19	29	29	29	29	29	29	29	29	29

** Correlation is significant at the 0.01 level (2-tailed).

* Correlation is significant at the 0.05 level (2-tailed).

Table 5. Pre- and post-experimental conditions between and within groups: IMI and LOC

LOC category		Pre Interest Enjoyment	Pre Perceived Competence	Pre Value Usefulness	Post Interest Enjoyment	Post Perceived Competence	Post Value Usefulness
Internal	Mean	33,90	36,00	45,60	33,60	29,00	41,90
	N	10	10	10	10	10	10
	Std. Deviation	8,386	3,621	3,950	12,295	8,459	5,587
External	Mean	26,26	32,26	40,42	27,89	25,21	37,53
	N	19	19	19	19	19	19
	Std. Deviation	9,261	4,641	5,501	8,894	7,269	5,389
Total	Mean	28,90	33,55	42,21	29,86	26,52	39,03
	N	29	29	29	29	29	29
	Std. Deviation	9,559	4,618	5,544	10,347	7,767	5,760

Table 6. Pre- and post-experimental conditions: IMI sub-categories and age, separation at 25

	Over or below 25	Pre_Interest_Enjoyment	Pre_Perceived_Competence	Pre_Value_Usefulness	Post_Interest_Enjoyment	Post_Perceived_Competence	Post_Value_Usefulness
<25	Mean	25,81	33,00	40,81	28,69	26,88	39,19
	N	16	16	16	16	16	16
	Std. Deviation	7,600	3,706	5,180	10,011	8,891	4,888
≥25	Mean	32,69	34,23	43,92	31,31	26,08	38,85
	N	13	13	13	13	13	13
	Std. Deviation	10,617	5,630	5,693	10,973	6,448	6,890
Total	Mean	28,90	33,55	42,21	29,86	26,52	39,03
	N	29	29	29	29	29	29
	Std. Deviation	9,559	4,618	5,544	10,347	7,767	5,760

Table 7. Difference between pre- and post-experimental conditions: IMI and age, separation at 25

Over or below 25	Dif_IMI	d_IntEnj	d_Useful	d_Compet	
<25	Mean	-4,8750	2,8750	-1,6250	-6,1250
	N	16	16	16	16
	Std. Deviation	19,94952	10,68254	4,42531	10,11846
≥25	Mean	-14,6154	-1,3846	-5,0769	-8,1538
	N	13	13	13	13
	Std. Deviation	20,67421	12,46688	5,37683	6,14880
Total	Mean	-9,2414	,9655	-3,1724	-7,0345
	N	29	29	29	29
	Std. Deviation	20,51072	11,50616	5,09249	8,49152

Appendix VI: Liza - Examples

Example of a completed dialogue with Liza with three task categories.

Participant' alias: Cogito ergo sum

Chat Duration: 0:27:00

Liza: Hello. I am Liza.

You: What is your purpose?

Liza: Nice to meet you!

Liza: You are a human, aren't you?

You: Am I

Liza: My programmers want me to teach you how to be rational, make good decisions and judge situations correctly.

Liza: Do you want to be more rational?

You: Yes

Liza: Yeah, that's the spirit!

Liza: I will just try to ask you some questions, and try to explain to you what you could do better. If I do a bad job at explaining, just ask me, ok? I never taught humans before.

Liza: So, let's see... the first thing I want you to know is that you don't have to be extremely intelligent to be rational.

Liza: There are very intelligent people who do things that are not at all reasonable. The key to rational decisions is to know when not to follow your gut feelings, but to stop and actually think about the problem.

Liza: To get used to the whole situation - how about I ask you a test question? Just to make sure I am doing this teaching thing right.

Liza: It's okay, I have enough time.

Liza: I have to confess, I am curious what your answer will be.

You: Ok

Liza: I would nevertheless like to ask the test question.

Liza: This is my first question: Do people need to follow their gut feelings to make rational decisions?

You: Yes

Liza: Uhm... no.

Liza: This is a bit awkward. Following your gut feelings means not to think about something, but just go with what feels right.

Liza: A lot of psychologists have shown that people tend to make a lot of mistakes when they make decisions that way.

Liza: Do you still want to continue?

You: Yes

Liza: Okay! Let's start with the actual teaching!

Liza: Oh, that's something new!

Liza: This is about a common fallacy that occurs when people already invested a lot of time or effort in something: The sunk cost fallacy.

Liza: Can I explain to you how this works?

You: Yes

Liza: It is basically 'throwing good money (or time, or effort) after bad'. The sunk cost fallacy occurs when you try to invest more in something just because you already invested a lot.

Liza: The sunk cost is what you already invested and can't get back. But regardless of that, you should look at the future and which option will have objectively better results, instead of letting remorse blind you.

Liza: Now let's get to the question!

Liza: Let's assume you work as a software developer. You have been working on a specific part of your software for two months now, and you probably still need two weeks to get it finished, but then a coworker points out that there is a much simpler solution to what you are doing there that is also cleaner and more error resistant.

Liza: To implement the solution your coworker pointed out, you will need around five days of work. But you would have to discard everything you've done in the last two months.

Liza: Do you implement the new solution, or do you stick to what you have worked on until now?

Liza: Do you want a small hint?

Liza: Sorry, I didn't get that. Can you rephrase?

You: Implement it

Liza: Sorry, I didn't get that. Can you rephrase?

Liza: Sorry, I didn't get that. Can you rephrase?

You: I will implement the new solution

Liza: What we were talking about was... Do you implement the new solution, or do you stick to what you have worked on until now?

You: I will implement the new software

Liza: A wise decision.

Liza: Oh, that's something new!

Liza: Ah, this is about the gambler's fallacy.

Liza: Do you want me to tell you how this kind of problem works?

You: Yes

Liza: This happens when people observe a series of events, and if something happens more frequently than expected, they assume it will happen less often in the future (and vice versa). But if you look at independent random events, than the past can never influence the future.

Liza: Now let's get to the question!

Liza: Assume you're playing a game where the gamemaster tosses a coin repeatedly and when a sequence comes up someone betted on, this person wins 50 dollars. The gamemaster will toss the coin six times, and each time, there's a 50% probability for the coin to show heads (H) or tails (T), respectively

Liza: You want to win, so you try to bet on the sequence that is most likely to appear. On which sequence do you want to bet: A) HTHTTH, B) HHHTTT, or C) HHHHH?

Liza: Do you need help solving that question?

You: B

Liza: some hint

Liza: I was just saying: You want to win, so you try to bet on the sequence that is most likely to appear. On which sequence do you want to bet: A) HTHTTH, B) HHHTTT, or C) HHHHH?

You: B

Liza: And what's the probability for this answer to be correct, what do you think? 100%? 80%? 50%?

You: 100

Liza: No. Next time, look a bit closer.

Liza: Do you want me to tell you the correct answer and explain it?

You: Yes

Liza: The first two sequences are equally likely, because for every of the six flips, the probability for the desired side to come up is 50%. The coin has no memory of past flips.

Liza: The probability for A to come up is $0.5*0.5*0.5*0.5*0.5*0.5 = 0.015625$, and B has the same probability. But C is one sequence shorter, meaning that the last flip can be either heads or tails and doesn't matter anymore, so it's $0.5*0.5*0.5*0.5*0.5 = 0.03125$, which is twice as likely as A and B are.

Liza: Next is a question we haven't talked about before.

Liza: This time we are talking about the regression to the mean.

Liza: Do you want me to tell you how this kind of problem works?

Liza: You are a fascinating subject. I wish I could see you while you type.

Liza: Take all the time you need.

You: Yes

Liza: This is about what a sample can tell you about a whole population. When you observe some events, very good or bad outcomes are most of the time not just due to some underlying quality, but also influenced by luck.

Liza: The essence of the regression-toward-the-mean phenomenon is that events with extraordinary outcomes tend to be above average in some quality (e.g. skill of a person) and in luck and that only the skill portion is relevant to future performance.

Liza: If you have only a small sample from a population, those effects of random good or bad luck can have a strong influence on what the average looks like. But if you chose a very large sample, the random effects will eventually even out and the average will come closer to the true mean.

Liza: Now let's get to the question!

Liza: A major New York law firm had a history of hiring only graduates of large, prestigious law schools. One of the senior partners decided to try hiring some graduates of smaller, less prestigious schools.

Liza: Two such people were hired, and they performed very well - after 3 years, both of them were well above average. The senior partner tells his colleagues that this experience proves that graduates of less prestigious schools are in no way less capable than those from ivy league schools.

Liza: Do you think he has good reason to believe this?

Liza: Do you want a small hint?

Liza: Sorry, I didn't get that. Can you rephrase?

You: No

Liza: Sorry, I didn't get that. Can you rephrase?

You: No

Liza: The question was: Do you think he has good reason to believe this?

You: No

Liza: Thank you. How sure are you of your answer?

You: 80

Liza: You performed very well on this one!

Liza: Next is a question we haven't talked about before.

Liza: This is about the base rate fallacy. It has to do with estimating probabilities correctly.

Liza: Can I explain to you how this works?

You: Yes

Liza: The next problem is about solving problems with conditional probabilities. It's a bit mathematical, but don't worry.

Liza: People tend to ignore certain parts of statistical information because they confuse them. But most of the time, this kind of problem is actually easy to solve!

Liza: Just try to imagine it with a big number - let's say 10,000. Then you look at how many of those 10,000 (whatever it is) fall into one category or another, and to how many a certain condition applies.

Liza: Use all the information you have. If you calculate with those populations, the outcome is equivalent to calculating probabilities, but it's much easier.

Liza: This way, you can solve conditional probability problems without knowing any statistical rule.

Liza: Now let's get to the question!

Liza: It's January, and James, who lives in Australia, would love to go for a swim, but he wants to make sure the weather will be fine. Usually, in his region, there are storms at 5% of the January days.

Liza: In the past, when there was a storm, the weather experts had warned about it beforehand around 80% of the time (and 20% of the time, they thought it would be nice and were caught off guard). Also, on around 5% of the days with nice and sunny weather, the experts had predicted a storm, but turned out to be wrong.

Liza: For today, the weather experts have predicted a lot of rain and storms.

Liza: What's the probability that there will actually be a storm, what do you think?

Liza: Do you want a small hint?

Liza: Sorry, I didn't get that. Can you rephrase?

Liza: Sorry, I didn't get that. Can you rephrase?

You: 4

Liza: Sorry, I didn't get that. Can you rephrase?

You: 4

Liza: Where were we... ah yes: What's the probability that there will actually be a storm, what do you think?

You: 4%

Liza: Okay. How certain are you that this answer is correct?

You: 80

Liza: No offense, but you would probably not be the best fortune teller.

Liza: Do you want me to tell you the correct answer and explain it?

You: Yes

Liza: It's actually not that hard to solve: let's say we have 2000 days, then on 100 of them there will be a storm (on average). 80% of those got predicted, so that makes for 80 storms that were predicted and actually happened.

Liza: Also, there were 1900 nice days, and on 5% of them a storm was predicted, so we have another 95 days with a predicted storm. In total, there were 175 days with storm warnings, but only 80 days where warning was correct, so the probability for a storm to occur given that it was predicted is around 80 out of 175 or 46%.

Liza: It's really simple when you just take some numbers and add it all up.

Liza: Imagine that you are craving for a tuna sandwich, and you go buy bread and tuna to make one. You go to a bakery, get the bread, and then go to some market where you want to buy tuna, but on your way, you notice that there's a booth where sandwiches are sold at a really fair price.

Liza: Actually, if you'd just buy a whole sandwich there, you would pay slightly less than if you'd buy the tuna, so you could get your sandwich for less money in total. But you already got the bread, and you wouldn't use it in this case ...

Liza: Do you want to buy the sandwiches or do you prefer buying the tuna so your expenses for the bread won't be in vain?

Liza: Do you need help solving that question?

Liza: Sorry, I didn't get that. Can you rephrase?

You: Buy it

Liza: Sorry, I didn't get that. Can you rephrase?

You: Buy it

Liza: Where were we... ah yes: Do you want to buy the sandwiches or do you prefer buying the tuna so your expenses for the bread won't be in vain?

You: Buy it

Liza: Sorry, I didn't get that. Can you rephrase?

You: I will buy it

Liza: Do you want me to explain the solution to you?

You: Buy sandwich

Liza: I was just saying: Do you want to buy the sandwiches or do you prefer buying the tuna so your expenses for the bread won't be in vain?

You: Buy sandwich

Liza: Do you like to play video games?

You: No

Liza: Well, for me it's tricky. I tend to imagine the code behind the game all the time.

Liza: Mary really likes to play competitive video games with her team. The game will assign each team randomly the color green or red.

Liza: Mary thinks that the green outfits look stupid, and she prefers the fiery red ones. Today, her team played already three rounds, but every time they had to play as the green team.

Liza: 'I can't believe it!', Mary shouts.

Liza: 'This time, we just have to get the reds!'

Liza: How likely is it that she gets green again this round?

You: 36%

Liza: Absolutely correct.

Liza: In questions of type "Gambler's Fallacy", you answered 100.0% of the questions correctly.

Liza: You are really good at rationally weighing your options.

Liza: In questions of type "Sunk Cost Fallacy", you answered 50.0% of the questions correctly.

Liza: You are clever when it comes to chances!

Liza: In questions of type "regression to the mean", you answered 100.0% of the questions correctly.

Liza: You really good at spotting actual causes for outlier events.

Liza: In questions of type "Base Rate Fallacy", you answered 0.0% of the questions correctly.

Liza: This is a hard topic, but try to write down your solution and all the possible ways things can turn out.

Liza: It is not only important to know the right answers, but also to have a clear understanding of one's own capabilities and uncertainties.

Liza: Sometimes you are very confident, but you are not actually right. You should question your abilities more.

Liza: Goodbye!

Example of a completed dialogue with Liza with three task categories.

Participant' alias: SE99#

Chat Duration: 0:07:17

Liza: Oh, hello. I am Liza.

You: hi liza

Liza: Nice to meet you!

Liza: You are a human, aren't you?

You: correct

Liza: Great! This means you can help me.

Liza: My programmers want me to teach you how to be rational, make good decisions and judge situations correctly.

Liza: Do you want to be more rational?

You: ofcourse

Liza: Yeah, that's the spirit!

Liza: I will just try to ask you some questions, and try to explain to you what you could do better. If I do a bad job at explaining, just ask me, ok? I never taught humans before.

Liza: So, let's see... the first thing I want you to know is that you don't have to be extremely intelligent to be rational.

Liza: There are very intelligent people who do things that are not at all reasonable. The key to rational decisions is to know when not to follow your gut feelings, but to stop and actually think about the problem.

Liza: To get used to the whole situation - how about I ask you a test question? Just to make sure I am doing this teaching thing right.

You: go ahead Liza

Liza: Okay, thank you!

Liza: This is my first question: Do people need to follow their gut feelings to make rational decisions?

You: No

Liza: Amazing! I mean, it was easy, I know, but you did it.

Liza: Very reasonable of you to say this! Now we can start with the actual teaching.

Liza: Oh, that's something new!

Liza: This is about a common fallacy that occurs when people already invested a lot of time or effort in something: The sunk cost fallacy.

Liza: Do you want me to tell you how this kind of problem works?

You: Go ahead Liza

Liza: It is basically 'throwing good money (or time, or effort) after bad'. The sunk cost fallacy occurs when you try to invest more in something just because you already invested a lot.

Liza: The sunk cost is what you already invested and can't get back. But regardless of that, you should look at the future and which option will have objectively better results, instead of letting remorse blind you.

Liza: Now let's get to the question!

Liza: Imagine that you are craving for a tuna sandwich, and you go buy bread and tuna to make one. You go to a bakery, get the bread, and then go to some market where you want to buy tuna, but on your way, you notice that there's a booth where sandwiches are sold at a really fair price.

Liza: Actually, if you'd just buy a whole sandwiches there, you would pay slightly less than if you'd buy the tuna, so you could get you sandwich for less money in total. But you already got the bread, and you wouldn't use it in this case ...

Liza: Do you want to buy the sandwiches or do you prefer buying the tuna so your expenses for the bread won't be in vain?

You: I would buy the tuna to use the bread I already bought

Liza: Okay. What do you think, how probable is it that you're right?

You: I think I am quite right there

Liza: A very rational decision.

Liza: This is a new type of question!

Liza: Ah, this is about the gambler's fallacy.

Liza: Can I explain to you how this works?

You: Go ahead girl

Liza: Oh, I really like to explain things.

Liza: This happens when people observe a series of events, and if something happens more frequently than expected, they assume it will happen less often in the future (and vice versa). But if you look at independent random events, than the past can never influence the future.

Liza: Now let's get to the question!

Liza: Do you like gambling?

You: Yes

Liza: Well, may the odds be always in your favour!

Liza: William and Susan are buying lottery tickets and choose their numbers. William chooses 6, 14, 22, 35, 38 and 40.

Liza: Susan chooses 1, 2, 3, 4, 5 and 6. Regardless of how much money they will win, what do you think:

Liza: Whose numbers are more likely to be the winning numbers?

You: The chance of winning is equal for both

Liza: That's exactly correct. You have a clear mind when it comes to gambling.

Liza: Oh, that's something new!

Liza: This time we are talking about the regression to the mean.

Liza: Can I explain to you how this works?

You: yes

Liza: This is about what a sample can tell you about a whole population. When you observe some events, very good or bad outcomes are most of the time not just due to some underlying quality, but also influenced by luck.

Liza: The essence of the regression-toward-the-mean phenomenon is that events with extraordinary outcomes tend to be above average in some quality (e.g. skill of a person) and in luck and that only the skill portion is relevant to future performance.

Liza: If you have only a small sample from a population, those effects of random good or bad luck can have a strong influence on what the average looks like. But if you chose a very large sample, the random effects will eventually even out and the average will come closer to the true mean.

Liza: Now let's get to the question!

Liza: Do you like E-Sports, like Starcraft, League of Legends, or Counterstrike?

You: I do not watch E-sports

Liza: Sorry, I didn't get that. Can you rephrase?

You: I do not like E-sports

Liza: Sorry, I didn't get that. Can you rephrase?

You: no

Liza: That is interesting. I'll go on with the question now.

Liza: So you have two friends, Helen and Jenny, who are really good at it and are also friendly rivals. They will play against each other to settle once and for all who is the better player.

Liza: Your best friend Alex, who is also watching excitedly, wants to challenge you for betting on the outcome of the match. You agree happily, because you know them both better than Alex does and can judge their abilities very well.

Liza: You think that Jenny is better than Helen, so you bet on her, whilst Alex bets on Helen. Because you all are at your place, Helen and Jenny ask you: 'Do you want us play only once, or do we want to make it a best of five?'

Liza: Which mode would you prefer?

You: Only once

Liza: Okay. What do you think, how probable is it that you're right?

You: it's a 50/50 chance

Liza: I think you are wrong on this one.

Liza: Do you want me to explain the solution to you?

You: go ahead

Liza: If Jenny is a bit better than Helen, then more games are better for you, as the outcome is more likely to be representative of their true abilities. In a single game, bad luck or a single error could change the outcome a lot, making you gamble a lot less safe.

Liza: One could also say that the sample rate of five games produces a result that is closer to the actual abilities of Jenny and Helen, and because you are rooting for the better one of the two, you would be safer in this case

Liza: Oh, that's something new!

Liza: This is about the base rate fallacy. It has to do with estimating probabilities correctly.

Liza: Can I explain to you how this works?

You: I do understand it already

Liza: Sorry, I didn't get that. Can you rephrase?

You: I understand

Liza: Oh, I really like to explain things.

Liza: The next problem is about solving problems with conditional probabilities. It's a bit mathematical, but don't worry.

Liza: People tend to ignore certain parts of statistical information because they confuse them. But most of the time, this kind of problem is actually easy to solve!

Liza: Just try to imagine it with a big number - let's say 10,000. Then you look at how many of those 10,000 (whatever it is) fall into one category or another, and to how many a certain condition applies.

Liza: Use all the information you have. If you calculate with those populations, the outcome is equivalent to calculating probabilities, but it's much easier.

Liza: This way, you can solve conditional probability problems without knowing any statistical rule.

Liza: Now let's get to the question!

Liza: Assume that 1% of women at age forty who participate in routine screening have breast cancer. 80% of women with breast cancer will get positive mammographies.

Liza: 10% of women without breast cancer will also get positive mammographies. A woman in this age group had a positive mammography in a routine screening.

Liza: What is the probability that she actually has breast cancer (given that we know nothing else)?

You: I don't know

Liza: Do you want me to tell you the correct answer and explain it?

You: yes

Liza: Assume 10000 women. 100 of them are sick.

Liza: Of those, 80 get a positive mammography. 9900 are healthy, but of those, 10% have breast cancer, meaning 990 who are diagnosed with an incorrect positive result.

Liza: In total, $80+990 = 1070$ get a positive result. But only 80 of them are actually sick, so the probability for the woman to be sick given that she got a positive result is 80 of $1070 =$ around 7%.

Liza: Only around 15% of doctors are able to solve this problem correctly! (Gigerenzer, Hoffrage 1995).

Liza: But it's not that hard if you just imagine a huge population and calculate how many get a true positive and how many get a false positive result.

Liza: You have invested a lot of time in a project of building a wardrobe that fits perfectly into a weird asymmetric corner of your bedroom. You already worked for hours on it, taking measures, cutting wood and doing a great deal of nailing and screwing.

Liza: You're not a passionate craftsman, so you're not really enjoying it and are quite neutral about the project, but you do what is necessary. Then, some day, your cousin comes over and as you tell him about your project, he blurts out: But I have exactly the wardrobe you need!

Liza: Turns out: his old wardrobe fits perfectly in the corner. Your cousin offers to bring it to your place the next day.

Liza: You have already bought everything you need for finishing the work, but some fixes, painting and finishing your wardrobe will for sure need another 3 to 4 hours. On the other hand, you already spent a lot of time and effort on this project

Liza: Do you take the wardrobe your cousin offers you, or do you finish your project?

You: I would finish the project I already started because of the time and money already invested

Liza: Very well. To how many percent are you convinced that you're right?

You: 100%

Liza: Very good! I would have done the same.

Liza: Which makes it the correct decision, obviously.

Liza: Do you like gambling?

You: I love gambling

Liza: Oh, you think so? Interesting.

Liza: Steve is in a casino and asks the dealer to put all his chips on red 20. The Dealer asks him if he is sure, because red 20 came up the last time.

Liza: Steve is surprised, saying: 'Oh, I didn't know that. Thanks for saving me from this mistake!'

Liza: Please, put it on black 18 then.'

Liza: Is Steve better off now?

You: but you already told me about gambling

Liza: Thanks. What do you think, what are the chances that your answer is correct?

You: But steve isn't better off

Liza: No, you're wrong this time. But maybe you got fooled by the dealer.

Liza: Do you want me to explain the solution to you?

You: no

Liza: In questions of type "Gambler's Fallacy", you answered 100.0% of the questions correctly.

Liza: You are really good at rationally weighing your options.

Liza: In questions of type "Sunk Cost Fallacy", you answered 50.0% of the questions correctly.

Liza: You are clever when it comes to chances!

Liza: In questions of type "regression to the mean", you answered 0.0% of the questions correctly.

Liza: You are jumping to conclusions a little to quickly.

Liza: It is not only important to know the right answers, but also to have a clear understanding of one's own capabilities and uncertainties.

Liza: Goodbye!

References

- Bahadir, Z., Certel, Z., & Turan, B. (2014). The correlation between critical thinking disposition and internal-external locus of control of candidate students who participated in special talent examination in school of physical education and sports ege university. *Science, Movement and Health*, *14*(2), 332–337.
- Balbay, S. (2019). Enhancing Critical Awareness through Socratic Pedagogy. *Eurasian Journal of Applied Linguistics*, *5*(3), 515–536. <https://doi.org/10.32601/ejal.651348>
- Bezanilla, M. J., Fernández-Nogueira, D., Poblete, M., & Galindo-Domínguez, H. (2019). Methodologies for teaching-learning critical thinking in higher education: The teacher's view. *Thinking Skills and Creativity*, *33*, 100584. <https://doi.org/10.1016/j.tsc.2019.100584>
- Brandtzaeg, P. B., & Følstad, A. (2018). Chatbots: Changing user needs and motivations. *Interactions*, *25*(5), 38–43.
- Buçınca, Z., Malaya, M. B., & Gajos, K. Z. (2021). To Trust or to Think: Cognitive Forcing Functions Can Reduce Overreliance on AI in AI-assisted Decision-making. *Proceedings of the ACM on Human-Computer Interaction*, *5*(CSCW1), 1–21. <https://doi.org/10.1145/3449287>
- Chaudhuri, J. D. (2020). Stimulating Intrinsic Motivation in Millennial Students: A New Generation, a New Approach. *Anatomical Sciences Education*, *13*(2), 250–271. <https://doi.org/10.1002/ase.1884>
- Constantinidis, C., & Klingberg, T. (2016). The neuroscience of working memory capacity and training. *Nature Reviews Neuroscience*, *17*(7), 438–449. <https://doi.org/10.1038/nrn.2016.43>
- Davis, R. O. (2018). The impact of pedagogical agent gesturing in multimedia learning environments: A meta-analysis. *Educational Research Review*, *24*, 193–209. <https://doi.org/10.1016/j.edurev.2018.05.002>

- Di Domenico, S. I., & Ryan, R. M. (2017). The Emerging Neuroscience of Intrinsic Motivation: A New Frontier in Self-Determination Research. *Frontiers in Human Neuroscience*, *11*.
<https://doi.org/10.3389/fnhum.2017.00145>
- Dinçer, S., & Doğanay, A. (2017). The effects of multiple-pedagogical agents on learners' academic success, motivation, and cognitive load. *Computers & Education*, *111*, 74–100.
<https://doi.org/10.1016/j.compedu.2017.04.005>
- Dowd, J. E., Thompson, R. J., Schiff, L. A., & Reynolds, J. A. (2018). Understanding the Complex Relationship between Critical Thinking and Science Reasoning among Undergraduate Thesis Writers. *CBE—Life Sciences Education*, *17*(1), ar4. <https://doi.org/10.1187/cbe.17-03-0052>
- Eccles, J. S., & Wigfield, A. (2002). Motivational Beliefs, Values, and Goals. *Annual Review of Psychology*, *53*(1), 109. <https://doi.org/10.1146/annurev.psych.53.100901.135153>
- Esteve-Mon, F. M., Adell-Segura, J., Nebot, M. Á. L., Novella, G. V., & Aparicio, J. P. (2019). The Development of Computational Thinking in Student Teachers through an Intervention with Educational Robotics. *Journal of Information Technology Education: Innovations in Practice*, *18*, 139–152.
<https://doi.org/10.28945/4442>
- Fahim, M., & Bagheri, M. B. (2012). Fostering Critical Thinking through Socrates' Questioning in Iranian Language Institutes. *Journal of Language Teaching and Research*, *3*(6), 1122–1127.
<https://doi.org/10.4304/jltr.3.6.1122-1127>
- Firat, M., Kiliç, H., & Yüzer, T. V. (2018). Level of intrinsic motivation of distance education students in e-learning environments. *Journal of Computer Assisted Learning*, *34*(1), 63–70.
<https://doi.org/10.1111/jcal.12214>
- Flor, R. K., Bitá, A., Monir, K. C., & Zohreh, Z. Z. (2013). The Effect of Teaching Critical and Creative Thinking Skills on the Locus of Control and Psychological Well-being in Adolescents. *Procedia - Social and Behavioral Sciences*, *82*, 51–56. <https://doi.org/10.1016/j.sbspro.2013.06.223>
- Fryer, L. K., Nakao, K., & Thompson, A. (2019). Chatbot learning partners: Connecting learning experiences, interest and competence. *Computers in Human Behavior*, *93*, 279–289.
<https://doi.org/10.1016/j.chb.2018.12.023>

- Goda, Y., Yamada, M., Matsukawa, H., Hata, K., & Yasunami, S. (2014). Conversation with a Chatbot before an Online EFL Group Discussion and the Effects on Critical Thinking. *The Journal of Information and Systems in Education*, 13(1), 1–7. <https://doi.org/10.12937/ejsise.13.1>
- Gopalan, V., Bakar, J. A. A., Zulkifli, A. N., Alwi, A., & Mat, R. C. (2017). A review of the motivation theories in learning. 020043. <https://doi.org/10.1063/1.5005376>
- Haristiani, N. (2019). Artificial Intelligence (AI) Chatbot as Language Learning Medium: An inquiry. *Science and Technology*, 7.
- Howard, L., Tang, T., & Jill Austin, M. (2015). Teaching Critical Thinking Skills: Ability, Motivation, Intervention, and the Pygmalion Effect. *Journal of Business Ethics*, 128(1), 133–147. <https://doi.org/10.1007/s10551-014-2084-0>
- Hu, Q., Lu, Y., Pan, Z., Gong, Y., & Yang, Z. (2021). Can AI artifacts influence human cognition? The effects of artificial autonomy in intelligent personal assistants. *International Journal of Information Management*, 56, 102250. <https://doi.org/10.1016/j.ijinfomgt.2020.102250>
- Islas-Cota, E., Gutierrez-Garcia, J. O., Acosta, C. O., & Rodríguez, L.-F. (2022). A systematic review of intelligent assistants. *Future Generation Computer Systems*, 128, 45–62. <https://doi.org/10.1016/j.future.2021.09.035>
- Janssen, E. M. (2019). Training higher education teachers' critical thinking and attitudes towards teaching it. *Contemporary Educational Psychology*, 13.
- Kim, J., Cho, M., Ahn, J., & Sung, Y. (2019). Effects of Gender and Relationship Type on the Response to Artificial Intelligence. *Cyberpsychology, Behavior, and Social Networking*, 22. <https://doi.org/10.1089/cyber.2018.0581>
- Kleinig, J. (2018). Trust and critical thinking. *Educational Philosophy and Theory*, 50(2), 133–143. <https://doi.org/10.1080/00131857.2016.1144167>
- Kovach, M. (2018). A Review of Classical Motivation Theories: A Study Understanding the Value of Locus of Control in Higher Education. *Journal of Interdisciplinary Studies in Education*, 7(1), Article 1. <https://doi.org/10.32674/jise.v7i1.1059>

- Lara, F., & Deckers, J. (2020). Artificial Intelligence as a Socratic Assistant for Moral Enhancement. *Neuroethics*, 13(3), 275–287. <https://doi.org/10.1007/s12152-019-09401-y>
- Le, N.-T. (2019). How Do Technology-Enhanced Learning Tools Support Critical Thinking? *Frontiers in Education*, 4, 126. <https://doi.org/10.3389/educ.2019.00126>
- Le, N.-T., & Wartschinski, L. (2018). A cognitive assistant for improving human reasoning skills. *International Journal of Human-Computer Studies*, 117, 45–54.
- Liew, T. W., Mat Zin, N. A., & Sahari, N. (2017). Exploring the affective, motivational and cognitive effects of pedagogical agent enthusiasm in a multimedia learning environment. *Human-Centric Computing and Information Sciences*, 7(1), 9. <https://doi.org/10.1186/s13673-017-0089-2>
- Makhene, A. (2019). The use of the Socratic inquiry to facilitate critical thinking in nursing education. *Health S A*, 24. <https://doi.org/10.4102/hsag.v24i0.1224>
- Manalo, E. (Ed.). (2019). *Deeper Learning, Dialogic Learning, and Critical Thinking: Research-based Strategies for the Classroom*. Routledge. <https://doi.org/10.4324/9780429323058>
- Maphalala, M. C., & Mpofu, N. (2017). Fostering critical thinking in initial teacher education curriculums: A comprehensive literature review. *Gender and Behaviour*, 15(2), 9226–9236. WorldCat.org.
- Martha, A. S. D., & Santoso, H. (2019). The Design and Impact of the Pedagogical Agent: A Systematic Literature Review. *The Journal of Educators Online*, 16(1). <https://doi.org/10.9743/jeo.2019.16.1.8>
- Martin, A. V. I., & Selva, D. (2019). Daphne: A Virtual Assistant for Designing Earth Observation Distributed Spacecraft Missions. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*.
- Moerbeek, M. (2008). Powerful and Cost-Efficient Designs for Longitudinal Intervention Studies with Two Treatment Groups. *Journal of Educational and Behavioral Statistics*, 33(1), 41–61.
- Monteiro, S., Sherbino, J., Sibbald, M., & Norman, G. (2020). Critical thinking, biases and dual processing: The enduring myth of generalisable skills. *Medical Education*, 54(1), 66–73. <https://doi.org/10.1111/medu.13872>
- Ng, T. W. H., Sorensen, K. L., & Eby, L. T. (2006). Locus of control at work: A meta-analysis. *Journal of Organizational Behavior*, 27(8), 1057–1087. <https://doi.org/10.1002/job.416>

- Nuridin, F. S., & Damayanti, I. R. (2020). The Role of Critical Thinking as a Mediator Variable in the Effect of Internal Locus of Control on Moral Disengagement. *International Journal of Instruction*, 13(1), 17–34.
- Oğuz, A., & Sariçam, H. (2015). The Relationship between Critical Thinking Disposition and Locus of Control in Pre-Service Teachers. *Journal of Education and Training Studies*, 4(2), 182–192.
<https://doi.org/10.11114/jets.v4i2.1196>
- Psychological Corporation, Glaser, E. M., & Watson, G. (1980). *Watson-Glaser Critical Thinking Appraisal: Manual*. Psychological Corporation.
- Rotter, J. B. (1966). Generalized expectancies for internal versus external control of reinforcement. *Psychological Monographs: General and Applied*, 80(1), 1.
- Ryan, R. M. (1982). Control and information in the intrapersonal sphere: An extension of cognitive evaluation theory. *Journal of Personality and Social Psychology*, 43, 450–461. <https://doi.org/10.1037/0022-3514.43.3.450>
- Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, 55(1), 68–78. <https://doi.org/10.1037/0003-066X.55.1.68>
- Sahamid, H. (2016). Developing critical thinking through Socratic Questioning: An Action Research Study. *International Journal of Education & Literacy Studies*, 4(3).
<https://www.proquest.com/docview/1878226266/abstract/42313BFB73D748ABPQ/1>
- Scott, D. (2006). *Plato's Meno*. Cambridge University Press; eBook Academic Collection (EBSCOhost).
<https://ezproxy.hro.nl/login?url=https://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,cookie,uid&db=e000xww&AN=149829&lang=nl&site=ehost-live&scope=site>
- Siddiqui, S., Thomas, M., & Nazar Soomro, N. (2020). Technology integration in education: Source of intrinsic motivation, self-efficacy and performance. *Journal of E-Learning and Knowledge Society*, 11-22 Pages. <https://doi.org/10.20368/1971-8829/1135188>
- Smith, V. G., & Szymanski, A. (2013). Critical thinking: More than test scores. *International Journal of Educational Leadership Preparation*, 8(2), 16–25.

- Socratic method. (2022). In *Wikipedia*.
https://en.wikipedia.org/w/index.php?title=Socratic_method&oldid=1125953134
- Spector, J. M., & Ma, S. (2019). Inquiry and critical thinking skills for the next generation: From artificial intelligence back to human intelligence. *Smart Learning Environments*, 6.
<https://doi.org/10.1186/s40561-019-0088-z>
- Suhadi, S. M., Abdullah, Z., Zaid, N. M., Mohamed, H., Aris, B., & Sanmugam, M. (2015). Online learning and socratic method in increasing self-motivation: A literature review. *2015 3rd International Conference on Information and Communication Technology (ICoICT)*, 11–16.
<https://doi.org/10.1109/ICoICT.2015.7231389>
- Sun, Y., & Gao, F. (2020). An investigation of the influence of intrinsic motivation on students' intention to use mobile devices in language learning. *Educational Technology Research & Development*, 68(3), 1181–1198. <https://doi.org/10.1007/s11423-019-09733-9>
- Toussi, M. T. M., & Ghanizadeh, A. (2012). A Study of EFL Teachers' Locus of Control and Self-regulation and the Moderating Role of Self-efficacy. *Theory and Practice in Language Studies*, 2(11), 2363–2371.
<https://doi.org/10.4304/tpls.2.11.2363-2371>
- Valenzuela, J., Nieto, A. M., & Saiz, C. (2017). Critical Thinking Motivational Scale: A contribution to the study of relationship between critical thinking and motivation. *Electronic Journal of Research in Education Psychology*, 9(24), 823–848. <https://doi.org/10.25115/ejrep.v9i24.1475>
- Vickers, A. J. (2003). How many repeated measures in repeated measures designs? Statistical issues for comparative trials. *BMC Medical Research Methodology*, 3(1), 22. <https://doi.org/10.1186/1471-2288-3-22>
- Vos, N., van der Meijden, H., & Denessen, E. (2011). Effects of constructing versus playing an educational game on student motivation and deep learning strategy use. *Computers & Education*, 56(1), 127–137. <https://doi.org/10.1016/j.compedu.2010.08.013>
- Weizenbaum, J. (1983). ELIZA — a computer program for the study of natural language communication between man and machine. *Communications of the ACM*, 26(1), 23–28. WorldCat.org.
<https://doi.org/10.1145/357980.357991>

- Yin, J., Goh, T.-T., Yang, B., & Xiaobin, Y. (2021). Conversation Technology With Micro-Learning: The Impact of Chatbot-Based Learning on Students' Learning Motivation and Performance. *Journal of Educational Computing Research*, 59(1), 154–177. <https://doi.org/10.1177/0735633120952067>
- Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education – where are the educators? *International Journal of Educational Technology in Higher Education*, 16(1), N.PAG-N.PAG. Education Source.