Bringing Crypto Knowledge to School: Examining and Improving Junior High School Students' Security Assumptions About Encrypted Chat Apps

Leonie Schaewitz, Cedric A. Lohmann, Konstantin Fischer, and M. Angela Sasse

Ruhr-Universität Bochum, Universitätsstraße 150, 44780 Bochum, Germany
{leonie.schaewitz, konstantin.fischer, cedric.lohmann, martina.sasse}@rub.de

Abstract. End-to-end encryption (E2EE) of everyday communication plays an essential role in protecting citizens from mass surveillance. The especially vulnerable group of children and young adolescents move quickly between chat apps and use them frequently and intensively. Yet they have had the least time to learn about online security compared to other age groups. In a two-part study conducted with four classes at a junior high school ($N = 86$ students, ages 12-16), we examined perceptions of security and privacy threats related to chat apps and understanding of E2EE using a questionnaire. A pre-post measure allowed us to examine how a short instruction video shown in class to explain the concept of E2EE and how it works in chat apps affected students' security understanding and threat perceptions. Our results show that students are aware of a variety of online threats but they are not familiar with the term E2EE. After the instruction, students gained confidence in explaining the concept of encryption and their understanding of the security features of E2EE improved. Our results also show that explanation of threats and E2EE can shift the intention of some participants towards tools that offer more protection.

Keywords: End-to-end encryption · secure communication · secure messaging · security knowledge · threat perceptions.

1 Introduction

Digital communication tools such as chat apps and social networking sites play an important role in teenagers' everyday lives. A PEW research study from 2018 [4] has shown that already then 95% of U.S. teens ages 13 to 17 owned a smartphone or had access to one and regularly used various communication services such as Facebook, Snapchat, and Instagram. These tools help them build and maintain relationships by facilitating communication with friends and family and contact with new people. However, there are also risks associated with using chat apps.

Although most digital communication today is encrypted in some way, many popular chat apps do not employ End-to-End-Encryption (E2EE) by default – and some do not offer it at all. Well implemented E2EE removes the provider's capability to passively eavesdrop on any messages sent between users. This greatly helps preventing undesired breaches of confidentiality for purposes like mass surveillance in the setting of a compromised provider, or for gathering data for targeted advertising in the setting of the "honest-but-curious" threat model [20].

For these reasons, security advocates have long been promoting E2E encrypted communication tools. There are many different tools available, with quite different security models, and understanding their security properties, and how they protect against specific risks is challenging. Previous research with adult users of chat apps has shown that many users do not understand the security properties of (E2E)encryption and different communication tools [7,1,21,8].

In 2020, Lindmeier and Mühling investigated K-12 students' understanding of cryptography and proposed that "students first and foremost lack a clear understanding of networked communication [which] may subsequently prevent them from forming correct mental models about cybersecurity" [18, p. 1]. We argue that trying to get students to develop a sufficiently complex mental model of networked communication would not be the most efficient, or even the most effective way of increasing their day-to-day security. Instead, building on existing mental models and transforming them into functional understanding of E2EE seems more promising.

Hence, to designing effective interventions and education programs, we consider it important to investigate students' beliefs about threats and protections in secure communications as a first step. We therefore report an attempt to capture and improve this type of knowledge. In this paper, we elicited from a group of students (12-16 years old) their perceptions of secure communications in general, and message encryption in particular. The students then got to watch an instructional video about E2EE in chat apps, and their security perceptions and understanding of threats was elicited again.

With this study, we want to explore whether conveying intentionally simplified functional mental models [10] – as simple as "If a chat app uses E2EE, the provider cannot read along" – can pose a feasible solution to our overarching goal of improving students' online security.

This paper addresses the following research questions:

- RQ1 What risks do teenagers perceive when using chat apps, and what are their security needs?
- RQ2 How do teenagers judge the security of chat apps?
- RQ3 What are teenagers' threat models when using chat apps?
- RQ4 What do they know about (end-to-end) encryption (in chat apps)?
- RQ5 Can an instruction video about E2EE change teenagers' perceptions about security threats and their understanding of secure communication?

Our results show that most students did have intuitive knowledge about security goals like confidentiality when using chat apps – for most students a

secure chat app is one that does not allow third parties to read along. Even though most students have heard of the term E2EE before, they struggled to pinpoint the actual security benefits it offers and were understandably unsure about its effectiveness. We see similar assumptions about E2EE, such as that E2EE cannot protect against messages being read along by the app provider, suggesting that misconceptions may form quite early. Our results also show that explanation of threats and E2EE can shift the intention of some participants towards tools that offer more protection.

2 Background and Related Work

2.1 The effectiveness of E2EE in chat apps

Chat apps that implement E2EE like to claim that they protect their users from a range of threats, even a malicious operator. This is not strictly true. Users still rely on the chat app operator's honesty to distribute the correct encryption keys for their contacts. When Alice wants to chat with Bob, she has to query the chat app operator for Bob's encryption keys. A malicious or compromised operator would simply return Mallory's key material, and reroute all of Alice's messages intended for Bob to Mallory, who then re-encrypts the messages for Bob.

If this is done both ways, Alice and Bob will think they are chatting end-to-end encrypted – unknowing that Mallory can happily read along, or manipulate message content. To prevent such key-swapping attacks, multiple modern chat apps support *authentication ceremonies*: Users can check whether they are using the correct encryption secret for a given contact by scanning a QR-Code, or by comparing key fingerprints by hand. However, research has shown that most users struggle to understand the need for authenticated encryption, and are not able to use it correctly [21,23,15].

In the "honest-but-curious" attacker threat model, we assume that the chat app operator does not actively try to compromise message security. Paverd et al. [20, p. 2] define the "honest-but-curious" adversary as follows: "The honest-but-curious adversary is a legitimate participant in a communication protocol who will not deviate from the defined protocol but will attempt to learn all possible information from legitimately received messages."

This threat model translates to the reality of chat apps quite well: The authors, in the context of smart grid energy suppliers, list various factors limiting an operator's capability to mount active attacks – factors that also apply to chat app operators: Regulatory oversight, external audits, and the desire to maintain reputation. We thus argue that even opportunistic[1] E2EE in chat apps can offer desirable security benefits for users, and is something they should look for.

Correctly implemented, E2EE removes the provider's capability to passively eavesdrop on any messages sent between users. This goes a long way towards

[1] Opportunistic E2EE: A system where users do not verify the correctness of their encryption keys for a given contact. The key server has to be trusted by users.

preventing undesired breaches of confidentiality for purposes like mass surveillance in the case of a compromised provider, or for data-gathering for targeted advertising in the case of the "honest-but-curious" threat model.

2.2 Security perceptions and understanding of E2EE

Previous research has shown that users' understanding of E2EE is limited [2,1,8]. In 2018, Abu-Salma et al. [1, p. 2] asked adults to evaluate the security of the hypothetical E2E encrypted chat app *Soteria*, based on a short textual description: "Soteria communications (messages, phone calls, and video calls) are E2E encrypted." They found that only few participants felt confident explaining E2EE, and that many rated the E2E encrypted chat app's security lower than SMS or phone calls. Participants believed that the provider, government employees, and people with technical knowledge could access the messages sent via the app. Several studies have identified similar worries about E2EE, like the belief that encryption is futile because any encryption could be broken by capable attackers, such as hackers or governmental organizations [2,7,17,26]. While it might be possible that these attackers find ways to circumvent encryption, e.g., by compromising the endpoints where the messages are stored, it is a misconception that modern encryption can be broken and would thus be futile.

Gerber et al. [14] and Dechand et al. [7] found that WhatsApp users were unaware of E2EE, did not understand its associated security features, or did not trust the protection offered. Moreover, the mental models that users have of encryption are generally quite sparse. Wu and Zappala found that users' mental models of encryption can often be described as "a functional abstraction of restrictive access control" [26, p. 395] and Lindmeier and Mühling [18] identified similar models in K-12 students.

2.3 Communicating threat models to end users

Successfully conveying the nuanced differences between a malicious operator and an "honest-but-curious" operator to all chat app users is not feasible – only a dedicated amateur would invest the effort required to acquire expert knowledge in form of a structural mental model [10] and maintaining it (knowledge stored in memory that is not frequently accessed fades). In this paper, we start from the position that E2E encrypted chat apps overall offer security benefits to our target group (junior high school students), because using them reduces the potential risk of a privacy breach at the operator's servers, and that a brief, simplified, but convincing explanation of those benefits can shift at least some of them to consider adoption.

A small number of studies have tested interventions to help users gain a functional understanding of the concept of E2EE. Demjaha et al. [8] tested a metaphor-based approach to convey functional understanding of E2EE but found that none of the different metaphors tested were able to evoke a correct mental model of E2EE in participants. Bai et al. [5] found that a tutorial to

teach "high-level" information about E2EE was able to improve users' understanding of E2EE. However, some misconceptions remained, e.g., several users remained unconvinced that encryption cannot be broken, and still found concepts like integrity and authenticity difficult to grasp. Akgul et al. [3] designed brief educational messages that informed readers about the key principles of E2EE and demonstrated their effectiveness in improving users' understanding of E2EE, using an online questionnaire study. However, when the same messages were tested in a realistic use case, embedded in an actual messaging app, no improvement in comprehension was observed.

One study that examined approaches to teach cryptography to school children used a virtual reality setting that built on a medieval love story where letters are encrypted and decrypted using magic potions [9]. This setting provided the opportunity to use metaphorical descriptions for explaining the complex concept of asymmetric encryption in an immersive way. The study found that presence was a key predictor of learning outcomes. However, the VR environment also poses challenges for the teacher and does not necessarily lead to better learning outcomes than other forms of instruction.

In this paper, we test whether a relatively easy-to-implement 20-minute instruction video explaining the basic security features of E2EE in chat apps can have an impact on students' security perceptions and self-reported behaviors.

3 Method

To answer the research questions, we conducted a two-part study with 86 high school students (ages 12-16). The study consisted of a pre- and post-test design and a teaching unit in the form of an instruction video. The preliminary questionnaire was used to obtain baseline measurements against which the results from the second questionnaire could be compared to gauge whether the students' understanding of E2EE had improved, and what changes they intended to make as a result of the intervention.

3.1 Procedure

The study was conducted during the students' normal classroom time on two days, with a one-day break between questionnaires. On the first day, the students first completed the pre-questionnaire with 8-10 open and 12 closed questions (depending on filter questions) about their security and privacy perceptions in the context of chat apps (see subsection 3.3). Following this, they were shown an instruction video explaining the concept of E2EE in chat apps (see subsection 3.2). Two days later, the students filled out the post-video questionnaire, which included mainly the same questions plus some additional questions about behavioral intentions (see subsection 3.3). Both questionnaires were completed online. The grades 7-9 were in online distance learning at the time of the study due to the COVID-19 pandemic and, hence, filled out the questionnaires from their homes while connected to their teacher and classmates remotely. The 10th grade students had on-site lessons.

3.2 Instruction video

The instruction video was developed in an iterative procedure. An initial version was created by one of the authors, which was discussed with the other authors as well as additional researchers and refined multiple times. We also consulted other web resources on how to explain E2EE to non-experts, such as the Surveillance Self-defense website by the EFF [11].

The final video was about 20 minutes long. In the first part (ca. 3 min.), the basic concept of encryption was explained by the example of the Caesar Cipher, which was already taught in history lessons and known to the students. The second part (ca. 11 min.) focused on secure communication with chat apps and explained the concept of E2EE by means of a fictional narrative in which Bob wants to confess his love to Alice via a chat app without anyone else reading along. The concepts of public and private keys were introduced and a simplified key exchange between Bob and Alice was illustrated. The video then gave an overview about which chat apps currently provide E2EE by default, which allow users to opt-in to E2EE, and which do not offer E2EE at all. The video also instructed viewers on how to check whether the E2EE operates correctly by explaining the meaning of the security number and QR code. The third part (ca. 6 min) summarized why encryption is important (i.e., to keep personal data, such as calls, messages, or pictures private) and explained which types of data are typically not protected by E2EE (e.g., different types of metadata) and which types of attackers E2EE can and cannot protect against.

3.3 Measures

The survey included the following measures, which were used in both the pre- (t1) and post-questionnaire (t2) unless otherwise noted.

Frequency of chat app use (only asked at t1). For each of the following chat apps, we asked participants how frequently they use it on a 5-point scale (1 = never, 2 = less than once a week, 3 = once a week, 4 = multiple times a week, 5 = multiple times a day): Snapchat, Facetime/iMessage, Telegram, Instagram Messenger, Facebook Messenger, Skype, Signal, WhatsApp, Threema. We also asked via free-text field which chat app they use most frequently.

Perception of secure communication, security needs, and perceived risks. We used free-text fields to receive free-text answers to the following questions: "What does secure communication with the smartphone mean to you?", "If you were to communicate with others over the Internet, what would you like to protect in your communication?", "What are the risks of using a chat app to communicate with other people?"

Perceived security of chat apps. For each chat app, we asked: "How secure do you think it is to send a private message using this service?" Answers were rated on a 7-point scale from 1 = very insecure to 7 = very secure; optional answer: "I do not know the app."

Perception of threats. We then asked the participants to imagine sending a private message to a friend using the chat app they used most frequently and

asked if they thought anyone other than their friend could read this message (yes/no). If answered "yes", we asked to specify who might read the messages and how (free-text responses). If answered "no," we asked why they thought no one else could read the messages. Moreover, we asked them to describe how they can make sure they are communicating with the right person.

Understanding of encryption and E2EE. Participants indicated on a 5-point scale (1 = very unsure to 5 = very sure) how sure they felt explaining the term "encryption", and to describe what it means to them (free-text response). Then, we asked if the term E2EE means anything to them (yes/no; filter question, only at t1) and if so, to provide a brief free-text description. To determine whether participants understood that E2EE protects the content of their messages from third-party access (correct response option from the list), but that other metadata are typically not protected, we asked participants to select from the following list all data that is protected by E2EE (time and duration of conversation; message content; location; how data was transmitted; sender and receiver). In addition, we presented a list of attacks, three with physical access to the phone that E2EE does not protect against (mobile phone theft, friends or parents with access to the phone) and four without physical access and protected by E2EE (blackmail by hackers, government surveillance, messaging app provider, advertising companies). Participants were asked to tick all those that E2EE can protect against.

Knowledge about which chat app uses E2EE and perception of security notifications. For each chat app, participants indicated whether they thought it uses E2EE by default or not. Since WhatsApp is the most popular and frequently used chat app in Germany, we also asked (at t1) if they ever saw WhatsApp's notifications about using E2EE or the change of a security number (yes/no), what these messages mean (free-text response), and how helpful they are (1 = not helpful at all to 7 = very helpful). At t2, we asked how they handle the notification informing them of a security number change.

Behavioral intentions (only at t2). Participants indicated how likely they are to pay attention to whether a chat app encrypts their messages E2E when writing messages in the future (1 = not at all likely to 7 = very likely). We also asked if they will activate E2EE whenever possible (yes/no), how important it is to them that their chats are E2E encrypted (1 = not at all important to 7 = very important), which chat apps they are going to use in the future, and whether they intend to change anything about their messaging behavior. Moreover, we asked whether they told their parents, friends, or relatives about what they have learned about encryption, whether they feel to have a better understanding of what to look for to communicate securely via chat apps, whether they have checked the security number of a contact by scanning the QR code, and to evaluate the video.

3.4 Data analysis

Qualitative data. We coded participants' free-text answers in a data-driven and iterative procedure using the software MAXQDA. The coding was performed

by two researchers – one with a background in IT security, one with a background in cognitive science and psychology. In a first step, both researchers coded the open responses independently from a randomly selected sub-sample of 30 participants to create an initial codebook for each question. Then, the coders discussed and refined their codes and established a final codebook, which was validated by both coders independently coding answers of another set of 20 participants (ca. 23% of the data, which is in the typical range for determining coder agreement [19]). As a measure for intercoder reliability, ReCal2 [13] was used to compute Cohen's Kappa for each code, of which we report a weighted mean for each question that takes into account the frequency of each code. Codes that did not occur in the subsample for which intercoder reliability was calculated were not included in this calculation. The remaining sets of answers were then coded by one coder. An overview of all codes, frequencies, and intercoder reliability is provided in Table 3 in the Appendix.

Quantitative data. To test for significant effects of the video, we conducted repeated-measures analyses, such as the paired-samples t-test to compare differences in means for a single dependent variable at two time points, the repeated-measures multivariate analysis of variance (MANOVA) to detect differences in multiple dependent variables over time, and the McNemar test for paired nominal data. We use $p = 0.05\%$ as the significance level for the statistical tests.

3.5 Research Ethics

Our university department where the study was conducted did not have an institutional review board that time. Instead, our study followed best practices in human subjects research and data protection policies that were reviewed and approved by our institution's data protection authority. The procedure of the study was developed in close consultation with the school's administration and was carried out in the presence of the class teachers. Students' participation was voluntary, without negative consequences. The school informed the students' parents or guardians about our study, and only students who brought a consent form signed by both the student and a parent or guardian, were allowed to take part in our study. We did not collect personal identifiable information about our participants. Participant IDs were distributed randomly among students by their teachers, and at no point did we know which student had given which answer.

3.6 Sample

A total of four classes (one class each from grades 7 to 10) from a German junior high school took part in the study. 100 students took part in the first questionnaire. Of these, 86 also completed the second questionnaire. We suspect that the drop-out of 14 students can be explained by the pandemic situation and distance learning, as only students who attended school lessons from home dropped out. Our final sample consists of 86 students who completed both questionnaires ($n = 36$ female, $n = 45$ male, 2 "diverse", 3 did not want to indicate their gender). These were distributed among the four classes as follows: $n = 22$

class 7, $n = 13$ class 8, $n = 22$ class 9, $n = 29$ class 10. 16 students had practical experience in IT (e.g., internship in IT or computer science) and 17 had a family member or someone close to them who works in IT.

4 Results

4.1 Perception of secure communication, perceived risks, and security needs (RQ1)

For most participants, secure communication means that the messages they exchange with others are private and not accessible to others outside the communication ($n_{t1} = 31$; $n_{t2} = 56$) or that their personal data is protected (e.g., not accessibly by third parties, not forwarded, securely stored, protected from hacking etc.; $n_{t1} = 24$; $n_{t2} = 15$).

The risks most frequently mentioned by participants were that their messages could be read by others ($n_{t1} = 38$; $n_{t2} = 54$), that they could be hacked ($n_{t1} = 8$; $n_{t2} = 12$), that the other person forwards their messages ($n_{t1} = 9$; $n_{t2} = 7$), data misuse (e.g., that data is sold, $n_{t1} = 9$; $n_{t2} = 2$), that personal information about them was publicly revealed ($n_{t1} = 6$; $n_{t2} = 4$), or that the person they communicate with was pretending to be someone else ($n_{t1} = 5$; $n_{t2} = 5$).

When asked what they want to protect when communicating online, most respondents mentioned their messages ($n_{t1} = 27$; $n_{t2} = 39$), general private data ($n_{t1} = 29$; $n_{t2} = 33$), photos/videos ($n_{t1} = 17$; $n_{t2} = 21$), location data ($n_{t1} = 11$; $n_{t2} = 9$), contacts and numbers ($n_{t1} = 4$; $n_{t2} = 8$), account information/passwords ($n_{t1} = 7$; $n_{t2} = 1$), or everything ($n_{t1} = 5$; $n_{t2} = 6$).

4.2 Perceived security of chat apps (RQ2)

The three most often used chat apps in our sample are WhatsApp (multiple times a day: 87.2%; never: 1.2%), Snapchat (multiple times a day: 66.3%; never: 18.6%) and the Instagram Messenger (multiple times a day: 45.3%; never: 22.1%). All other chat apps were only used by a small proportion of participants or very infrequently, thus, we focus on these three chat apps when presenting the results.

When we asked students to rate the security of the different chat apps they use (on a 7-point scale) at baseline (t1), the perceived security of sending private messages via the chat apps was at a medium level for WhatsApp ($M = 4.35$, $SD = 1.70$, $n = 86$), Snapchat ($M = 3.91$, $SD = 1,55$, $n = 80$), and Instagram Messenger ($M = 3.56$, $SD = 1.51$, $n = 80$). The Facebook Messenger was rated as least secure ($M = 2.81$, $SD = 1.44$; $n = 52$) and Threema as most secure ($M = 5.20$, $SD = 2.68$; however, only $n = 5$ people rated Threema, the rest did not know the service). All means and standard deviations of students' security ratings of the different apps can be seen in subsection 4.2.

$N = 74$ of our participants (all WhatsApp users except one) knew or suspected that WhatsApp offers E2EE by default, $n = 52$ correctly indicated that Snapchat does not have E2EE by default (37 Snapchat users, 15 non-users), and

Perceived Security of Chat Apps						
	Before Video ($t1$)			After Video ($t2$)		
	N	M	SD	N	M	SD
WhatsApp	86	4.35	1.700	85	5.24	1.593
Snapchat	80	3.91	1.552	81	3.06	1.495
Instagram Messenger	80	3.56	1.508	82	3.00	1.491
Skype	63	4.21	1.427	68	4.38	1.446
iMessage	58	4.62	1.705	66	4.79	1.524
Facebook Messenger	52	2.81	1.442	67	3.39	1.487
Telegram	26	3.65	1.810	45	4.27	1.514
Signal	6	4.83	1.472	32	5.53	1.545
Threema	5	5.20	2.683	29	5.17	1.416

Table 1. Means (M) and standard deviations (SD) of students' assessments of the security of chat apps from 1(very insecure) to 7(very secure) before (t1) and after (t2) watching the video. N: Number of students that knew the app and gave an answer.

$n = 44$ correctly indicated that the Instagram Messenger does not have E2EE by default (32 Instagram users, 12 non-users).

$N = 75$ have seen the information message provided in WhatsApp to explain E2EE, which was rated rather helpful ($M = 5.11$, $SD = 1.97$; $n = 76$; scale: 1 = not very helpful, 7 = very helpful).

Moreover, $N = 54$ saw the information message about the change of a security number in WhatsApp. The perceived helpfulness of this message was on a medium level ($M = 4.30$, $SD = 2.28$; $n = 63$). Most participants guessed that the message meant that their contact's number ($n = 27$), security number ($n = 14$), or cell phone ($n = 7$) had changed. $N = 3$ said that the app had been reinstalled, 4 associated security with the message, 16 did not know, and 21 gave other responses, including three claiming the chat was now no longer secure.

4.3 Perception of threats: attackers, methods, & protections (RQ3)

Prior to the lecture, the majority of participants ($n = 63$) believed that if they sent a private message to a friend via the chat app service they used most often, someone other than their friend could also read this message ($n = 53$ used an E2E encrypted chat app, such as WhatsApp, $n = 10$ used chat apps without E2EE, such as Snapchat or Instagram most frequently). Most of them believed that the provider of the app ($n = 35$), hackers ($n = 17$), friends ($n = 7$), government or intelligence ($n = 7$), family members ($n = 4$), or others ($n = 12$; e.g., persons to whom the message is forwarded or shown, others with access to the account, companies who buy this data) could read their messages.

When asked how others could read their messages, they mention hacking ($n = 13$), access via the receiver ($n = 10$; e.g., that the friend forwards/shows the message to someone else), that the provider has access ($n = 9$), via physical

access to the device ($n = 6$), via logging into the chat ($n = 6$), or via access to the server ($n = 5$) as potential methods.

$N = 23$ did not believe that a person other than the receiver could read their messages ($n = 22$ of them used E2E encrypted chat apps). As reasons, they mentioned (E2E)-encryption of their messages ($n = 5$), their perceptions about the protection of their data (by the app; $n = 7$), or other, such as that their messages were not interesting, or that they send the message only to a specific person or number.

When asked how they could check if they were writing to the right person, only one person mentioned to verify the person by scanning the QR code. Most participants answered with different strategies, such as asking about something personal ($n = 22$), talking to or calling the person (n = 16), checking the name ($n = 11$) or number ($n = 10$), asking them to send a photo of themselves ($n = 7$), checking the profile picture ($n = 3$) or writing style ($n = 5$). 14 people were not sure, and 24 mentioned other aspects, such as knowing or trusting the person, or gave unclear answers.

4.4 Assumptions about encryption and E2EE (RQ4)

When asked at t1 what the term "encryption" means, $n = 32$ referred to the protection of their messages, which are protected from being read by people outside the communication. $N = 14$ had associations with access control, describing encryption as a barrier that protects or keeps something secret, or as a password or mechanism for locking accounts, messages, or devices, and $n = 12$ described it as converting data into another form, such as a (secret) code or something that makes it unreadable. Some described that encryption means that their data is protected ($n = 11$), that their messages are secure ($n = 4$), or they simply explained the term with the term ($n = 5$; e.g., that something is encrypted), mentioned other/unclear aspects ($n = 3$), or to not know the answer ($n = 9$).

Only $n = 25$ indicated that the term E2EE meant anything to them. 13 described E2EE meaning that their messages are not readable by third parties, while the remainder did not refer to the non-readability of their messages but simply recited the term ($n = 5$; e.g., "messages are encrypted from beginning to end"), described that E2EE means that their messages ($n = 2$) or data are secure ($n = 1$), or that they did not know how to explain E2EE ($n = 3$).

Figure 1 shows that, when asked what data E2EE protects when communicating via a chat app, most participants ($n = 67$) correctly selected "message content." However, $n = 60$ also selected "sender and receiver", $n = 49$ "location", $n = 31$ "way of data transmission", and $n = 23$ "time and duration of conversation" – indicating great uncertainty about whether the metadata is protected by E2EE. In terms of potential attackers (see Figure 2), participants seemed to assume that E2EE mainly protects them from hackers ($n = 62$). Only half of the participants ($n = 43$) stated that E2EE can protect their private communications from the app provider.

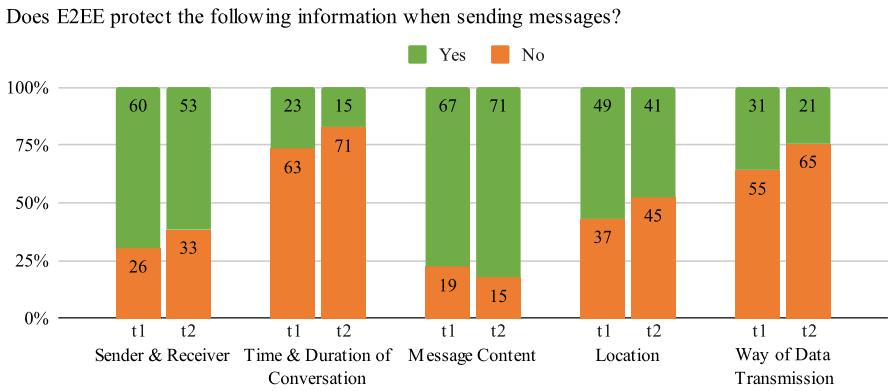

Fig. 1. Students' answers on what data is protected by E2EE. Before the video ($t1$) and after ($t2$).

4.5 Effects of the video (RQ5)

Perception of security, threats, & protections. To investigate whether the perceived security of the three most used chat apps, WhatsApp, Snapchat, and Instagram (dependent variables), had changed after the video (from t1 to t2), we ran a repeated measures MANOVA (using $n = 76$ data sets of participants who rated all three apps at t1 and t2). The analysis revealed a significant difference in the perceived security of the chat apps over time, Wilks's $\lambda = 0.56$, $F(3, 73) = 19.09$, $p < .001$, $\eta_p^2 = .44$. While the perceived security of WhatsApp increased significantly, $F(1, 75) = 19.23$, $p < .001$, $\eta_p^2 = .20$, the perceived security of Snapchat, $F(1, 75) = 26.75$, $p < .001$, $\eta_p^2 = .26$, and the Instagram Messenger, $F(1, 75) = 14.94$, $p < .001$, $\eta_p^2 = .17$, decreased. Converted to r as an effect size (see [12, p. 538]), these findings reflect a medium-sized effect for WhatsApp ($r = 0.45$) and Instagram ($r = 0.41$), and a large effect for Snapchat ($r = 0.51$). See subsection 4.2 on page 10 for means and standard deviations.

The number of participants who correctly stated that WhatsApp uses E2EE by default increased from 74 to 79, and the number of participants who correctly stated that Snapchat [Instagram Messenger] has no E2EE by default increased from 52 to 68 [44 to 64]. See Table 2 for an overview of participants' ratings.

With regard to perceived threats, the number of participants who believed that someone else can read their messages was reduced from $n = 63$ to $n = 46$. Of those, $n = 20$ still mentioned the app provider (with $n = 11$ stating to use only E2E encrypted chat apps most frequently; $n = 6$ mentioned a mixture of apps with and without E2EE; $n = 3$ used apps without E2EE). On a positive note, the number of participants who mentioned E2E(encryption) as the reason they did not think it was possible for others to read their messages increased from 5 to 26.

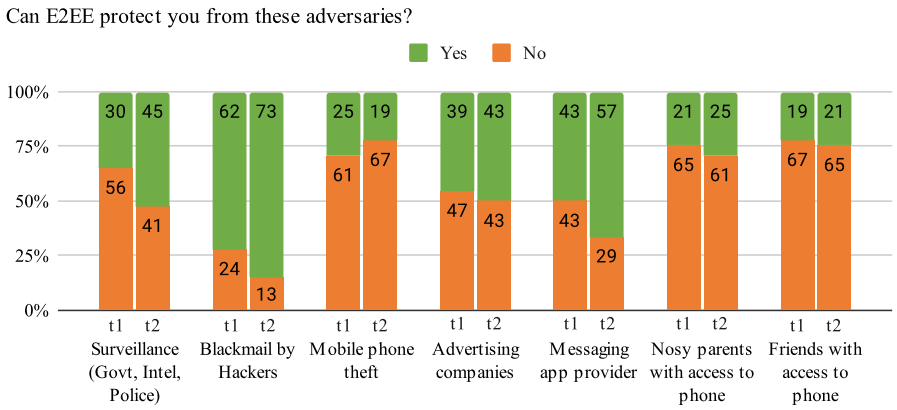

Fig. 2. Students' answers on who E2EE protects them from. Before the video ($t1$) and after ($t2$).

Moreover, ca. 40% of participants ($n = 34$) in t2 indicated that they had to check the security number or QR code of their contact to determine if they were writing with the correct person. In t1 only one person suggested this strategy.

Overall, participants' confidence in being able to explain the term encryption increased significantly from $M = 2.99$ ($SD = 1.02$) to $M = 3.47$ ($SD = 0.97$). A paired-samples t-test showed that this difference was statistically significant, $t(85) = \text{-}4.96, p < .001$, and represented a medium-sized effect, $d_z = -0.54$. And the number of participants who described that encryption means that no one else can read their messages increased from $n = 32$ to $n = 52$.

Figure 1 and Figure 2 show the prevalence of participants' assumptions about what protection E2EE does and does not provide (before and after the video). An exact McNemar's test determined that there was a statistically significant difference in the proportion of participants assuming protection against state surveillance by E2EE before ($n = 30$) and after the instruction video ($n = 45$), $p = .017$. There was also a significant increase in the proportion of participants assuming protection against the app provider, $p = .029$ ($t1 : n = 43$; $t2 : n = 57$), and hackers, $p = .035$ ($t1 : n = 62$; $t2 : n = 73$), after the video. All other comparisons were not significant. Although these changes represent some improvement, many participants still had misconceptions after the video about what threats E2EE can and cannot protect against, and that metadata (e.g., sender and recipient) is not protected by E2EE.

Intentions and actions. $N = 40$ participants indicated an intention to change something about their chat app use, $n = 33$ did not, and $n = 13$ were not yet sure. The likelihood that they will check whether a chat app used E2EE in the future was rated on a medium level ($M = 4.12$, $SD = 1.84$), but the majority ($n = 65$) want to activate E2EE in chat apps whenever possible.

Overall, participants stated that it is rather important to them that their chats are E2E encrypted ($M = 5.19$, $SD = 1.74$). Moreover, $n = 18$ said they

	Students' answers before the video		Students' answers after the video		Ground truth	Students' improvement
	E2EE	No E2EE	E2EE	No E2EE		
WhatsApp	74	12	79	7	E2EE	5.81%
Snapchat	34	52	18	68	No E2EE[a]	18.60%
iMessage	49	37	51	35	E2EE[b]	2.33%
Skype	43	43	42	44	opt-in[c]	1.16%
Telegram	32	54	38	48	opt-in[c]	−6.98%
Threema	32	54	40	46	E2EE	9.30%
Signal	26	60	44	42	E2EE	20.93%
Facebook Messenger	25	61	22	64	opt-in[c]	3.49%
Instagram Messenger	42	44	22	64	No E2EE	23.26%

[a] Snapchat uses E2EE encryption for Images, but not for text messages
[b] iMessage offers E2EE when texting with other iMessage users (replaces SMS)
[c] E2EE has to be enabled by the user for each chat individually

Table 2. Students' answers on whether the listed chat apps use E2EE by default, before and after the video. Ground truth given for all chat apps in July 2021. *Students Improvement* denotes the increase of correct answers after the video was shown.

had tried to verified the security number of one of their contacts by scanning the QR code after the instruction video, n = 35 talked with their parents, friends, or relatives about what they had learned about encryption, and most of them ($n = 73$) had the feeling to know better what to look for when using a chat app to communicate securely.

5 Discussion

5.1 Perceived security of chat apps.

Interestingly, participants rated WhatsApps' security only at a medium level, even though most knew it offers E2EE by default, and had seen the notification from WhatsApp stating that E2EE was being used to secure their messages. One explanation is that participants do not consider WhatsApp secure due to negative experiences unrelated to E2EE, such as cyberbullying or exposure to harmful content, or because it is owned by Faceboook, which monetizes its users' data, and is therefore not trusted – which is in line with prior findings, such as [2,7,14]. Another explanation, supported by our data, is that many participants saw, but did not understand the notification about E2EE in their chats and did not associate E2EE with an effective increase in security. This is likely since the majority of our participants did not know what the term E2EE meant prior to the instruction video.

Users' stance of "not trusting WhatsApp" is understandable, and arguably "safer" than assuming WhatsApp will under no circumstances be able to read

chat messages because it says it uses E2EE. If one does not trust WhatsApp, which belongs to Facebook, a company publicly associated with mishandling user data rather than protecting it [22], it is very hard to convince them that WhatsApp would implement E2EE correctly, not add backdoors, or stray from the role of an "honest-but-curious" provider. As security researchers, we have to agree that it comes down to how much you trust the company not to do this – when our participants do not, that skepticism should be seen as healthy.

That said, we wanted understand that it is not a reason to chose a chat app that doesn't offer E2EE over one that does. While E2EE cannot protect from resourceful targeted attack, current E2EE implementations in chat apps are likely to offer protection against passive eavesdroppers with access to the provider's systems. In the real world, E2EE can hinder mass surveillance and make app operators less tempted to collect private message contents for mischief.

5.2 Understanding of E2EE & perception of threats.

Prior to the video, most participants were unsure about the concept of encryption and reported associations similar to those found in previous research [18,26]: Encryption was described as access control, or as transformation of data into (secret) code.

They intuitively knew that confidentiality is a key security goal when using chat apps, as for most students a secure chat app is one that does not allow third parties to read along. However, many participants believed that their communications with chat apps were not confidential, but that their messages could be read by others outside the communication. They named a number of different actors they suspected of being able to read the messages, most notably the provider of the communication tool – even if they claimed to use E2E encrypted chat apps like WhatsApp.

This clearly shows that our participants – similar to what previous research on adults has shown – did not connect E2EE to the protection of their messages' content from the provider of the chat apps [7,14]. This might have been the case because they connected encryption to the easily breakable Caesar Cipher, which they had learned about previously.

After seeing the instruction video, they seem to have learned to associate E2EE with increased security: They assessed chat apps which do not provide E2EE by default as significantly less secure than before. Moreover, WhatsApp, offering E2EE by default, was assessed as more secure than before. Because we measured these effects not immediately, but two days after showing the video, we assume that these effects can be considered long-term. Moreover, our findings also show that a short instruction video explaining threats and E2EE can shift the intention of some participants towards chat tools that offer more protection.

5.3 Limitations

The study was conducted at a single German junior high school. Although we were able to include classes from four different grades, the sample's diversity

is limited. Due to the COVID-19 pandemic, the students from grades 7-9 were in online distance learning and hence, did not receive the lecture in presence as originally planned. Instead, we used a pre-recorded video for all students. Because we used only one version of the video in the study, the results may not be generalizable. An advantage of using a pre-recorded video, however, is that every class had received exactly the same information, which enhances the comparability of results between students.

5.4 Future Work

Future research could extend the intervention to include different versions of the instruction video or even different educational approaches. A between-subjects design could be used to compare the effectiveness of different approaches, for example, our narrative-based approach against a more fact-based approach, or an approach that works with metaphorical explanations. Moreover, future work could examine differences across age groups and test whether the effects of different educational approaches differ for students of different ages.

Finally, future research should also examine teachers' security perceptions and understanding of E2EE. This was out of scope for our research, but teachers are not IT security specialists, so it is important to develop appropriate training for them as well.

6 Conclusion

Our results show that an educational intervention explaining what E2EE is, and what protection a chat app that includes it can offer, had an effect with a good portion of our participants. After the instruction video, more students understood that – while it does not protect from all threats – choosing a chat app with E2EE offers protection against specific ones. We conclude that including threat models and security attributes of different technologies, such as E2EE, in the school curriculum would be beneficial – especially when it is directly related to the digital communication tools the students use.

Increasing students' general awareness of threats to digital technology and the benefits of different protection mechanisms is of course beneficial. But using concrete examples of how threats apply to their day-to-day activities, and that there are choices they can make to protect them, increases motivation and facilitates learning. After all, we know from mental models research that reasoning about concrete, familiar constructs significantly increases the chances of reaching correct conclusions [16]. Discussing them with their peers and applying the knowledge to their communication with each other should lead to repeated use, creating new habits and normalizing the use of security [25]. Our participants gained confidence in explaining the concept of encryption. This is a positive finding, because it could lead to an increase in self-efficacy [6], which in turn is a key predictor of secure behavioral practices. And increasing confidence in their ability to use IT security at a young age, is vital. From adults, security

researchers today still hear the futility argument, "Hackers can always find a way in," first reported by Weirich and Sasse [24, p. 139].

We think that conveying the strength of modern encryption is important to fight negative consequences that come with the futility mindset. School is a promising place to instill this mindset. Educational interventions should not try to impart structural knowledge about encryption, but convey simple functional models that help students make the right decisions. Knowing that E2EE is not futile, and gaining confidence, provide a basis for secure habits they can build on as they progress through life.

Acknowledgment

Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany´s Excellence Strategy - EXC 2092 CASA – 390781972.

References

1. Abu-Salma, R., Redmiles, E.M., Ur, B., Wei, M.: Exploring user mental models of end-to-end encrypted communication tools. In: 8th USENIX Workshop on Free and Open Communications on the Internet (FOCI 18) (2018)
2. Abu-Salma, R., Sasse, M.A., Bonneau, J., Danilova, A., Naiakshina, A., Smith, M.: Obstacles to the adoption of secure communication tools. In: 2017 IEEE Symposium on Security and Privacy (SP). pp. 137–153. IEEE (2017)
3. Akgul, O., Bai, W., Das, S., Mazurek, M.L.: Evaluating in-workflow messages for improving mental models of end-to-end encryption. In: 30th USENIX Security Symposium (USENIX Security 21). USENIX Association (Aug 2021)
4. Anderson, M., Jiang, J.: Teens, social media and technology 2018. Pew Research Center **31**, 2018 (2018)
5. Bai, W., Pearson, M., Kelley, P.G., Mazurek, M.L.: Improving non-experts' understanding of end-to-end encryption: An exploratory study. In: 2020 IEEE European Symposium on Security and Privacy Workshops (EuroS&P). pp. 210–219 (2020). https://doi.org/10.1109/EuroSPW51379.2020.00036
6. Bandura, A.: Self-efficacy: Toward a unifying theory of behavioral change. Psychological Review **84**, 191–215 (1977). https://doi.org/10.1037//0033-295x.84.2.191
7. Dechand, S., Naiakshina, A., Danilova, A., Smith, M.: In encryption we don't trust: The effect of end-to-end encryption to the masses on user perception. In: 2019 IEEE European Symposium on Security and Privacy (EuroS&P). pp. 401–415. IEEE (2019)
8. Demjaha, A., Spring, J.M., Becker, I., Parkin, S., Sasse, M.A.: Metaphors considered harmful? An exploratory study of the effectiveness of functional metaphors for end-to-end encryption. In: Proceedings 2018 Workshop on Usable Security. vol. 2018. Internet Society (2018)
9. Dengel, A.: Public-private-key encryption in virtual reality: Predictors of students' learning outcomes for teaching the idea of asymmetric encryption. CoolThink@ JC p. 41 (2020)
10. diSessa, A.: Models of computation. In: Norman, D.A., Draper, S.W. (eds.) User Centered System Design: New Perspectives on Human-Computer Interaction, pp. 201–218. Lawrence Erlbaum Associates, Hillsdale, New Jersey (1986)

11. Electronic Frontier Foundation: Surveillance self-defense: Tips, tools and how-tos for safer online communications, https://ssd.eff.org/en
12. Field, A.: Discovering statistics using IBM SPSS statistics (4th ed.). Sage, London, UK (2013)
13. Freelon, D.G.: Recal: Intercoder reliability calculation as a web service. International Journal of Internet Science **5**(1), 20–33 (2010)
14. Gerber, N., Zimmermann, V., Henhapl, B., Emeröz, S., Volkamer, M.: Finally johnny can encrypt: But does this make him feel more secure? In: Proceedings of the 13th International Conference on Availability, Reliability and Security. pp. 1–10 (2018)
15. Herzberg, A., Leibowitz, H.: Can johnny finally encrypt? evaluating e2e-encryption in popular im applications. In: Proceedings of the 6th Workshop on Socio-Technical Aspects in Security and Trust. pp. 17–28 (2016)
16. Johnson-Laird, P.N.: Mental models: Towards a cognitive science of language, inference, and consciousness. No. 6, Harvard University Press (1983)
17. Krombholz, K., Busse, K., Pfeffer, K., Smith, M., von Zezschwitz, E.: "If HTTPS were secure, I wouldn't need 2FA"- End user and administrator mental models of HTTPS. In: 2019 IEEE Symposium on Security and Privacy (SP). pp. 246–263. IEEE (2019)
18. Lindmeier, A., Mühling, A.: Keeping secrets: K-12 students' understanding of cryptography. In: Proceedings of the 15th Workshop on Primary and Secondary Computing Education. WiPSCE '20, Association for Computing Machinery, New York, NY, USA (2020). https://doi.org/10.1145/3421590.3421630
19. O'Connor, C., Joffe, H.: Intercoder reliability in qualitative research: Debates and practical guidelines. International Journal of Qualitative Methods **19**, 1–13 (2020). https://doi.org/10.1177/1609406919899220
20. Paverd, A., Martin, A., Brown, I.: Modelling and automatically analysing privacy properties for honest-but-curious adversaries. Tech. Rep (2014)
21. Schröder, S., Huber, M., Wind, D., Rottermanner, C.: When signal hits the fan: On the usability and security of state-of-the-art secure mobile messaging. In: European Workshop on Usable Security. IEEE. pp. 1–7 (2016)
22. Team Guild: A timeline of trouble: Facebook's privacy record (August 2021), https://guild.co/blog/complete-list-timeline-of-facebook-scandals/, [Online; posted 04-August-2012]
23. Vaziripour, E., Wu, J., O'Neill, M., Whitehead, J., Heidbrink, S., Seamons, K., Zappala, D.: Is that you, Alice? A usability study of the authentication ceremony of secure messaging applications. In: Thirteenth Symposium on Usable Privacy and Security (SOUPS 2017). pp. 29–47 (2017)
24. Weirich, D., Sasse, M.A.: Pretty good persuasion: a first step towards effective password security in the real world. In: Proceedings of the 2001 workshop on New security paradigms. pp. 137–143 (2001)
25. Wenger, E.: Communities of practice and social learning systems: the career of a concept. In: Social learning systems and communities of practice, pp. 179–198. Springer (2010)
26. Wu, J., Zappala, D.: When is a tree really a truck? Exploring mental models of encryption. In: Fourteenth Symposium on Usable Privacy and Security (SOUPS 2018). pp. 395–409. USENIX Association, Baltimore, MD (2018)

Appendix

Code Frequencies for Free-Text Answers Before (t1) and After (t2) the Instruction Video

Def. Secure Communication ($\kappa = 0.92$)

Code	**t1**	**t2**
Messages not accessible by 3rd parties	31	56
Data protection	24	15
Communication with friends/familiy	7	2
Encryption	3	6
Important	8	2
Not (very) important	2	0
Dont know	1	0
Other/unclear	16	16

Security Needs ($\kappa = 0.92$)

Code	**t1**	**t2**
General (private) data	29	33
Accounts and passwords	7	1
Location data	11	9
Contacts and numbers	4	8
Messages	27	39
Pictures	17	21
Everything	5	6
Nothing	3	2
Don't know	2	0
Other/unclear	9	4

Risks and Threats ($\kappa = 0.77$)

Code	**t1**	**t2**
Messages readable by third parties	38	54
Hacking	8	12
Other person forwards messages	9	7
Data misuse	9	2
Personal information revealed	6	4
Person pretends to be someone else	5	5
Miscommunication	4	2
Unwanted sceenshots	3	2
Data stored on servers	2	3
Surveillance	2	1
None	1	1
Dont know	7	2
Other/unclear	4	10

Who can read your messages ($\kappa = 0.93$)

Code	**t1**	**t2**
Developer/provider	35	20
Hacker	17	16
Government/Intelligence	7	8
Friends	7	6
Family	4	7
Other/unclear	13	11

How can they read messages ($\kappa = 0.86$)

Code	**t1**	**t2**
Hacking	13	13
Access via person	10	6
Provider has access	9	4
Physical access to device	6	3
Login to chat	6	3
Access via server	5	3
Not encrypted	1	4
Dont know	9	6
Other/unclear	10	14

Why can't they read messages ($\kappa = 0.82$)

Code	**t1**	**t2**
Encryption/E2EE	5	26
Data protection	7	9
Dont know	1	1
Other/unclear	10	5

Authentication Strategy ($\kappa = 0.84$)

Code	**t1**	**t2**
Check security number	1	34
Ask something personal	22	12
Call/talk to the person	16	13
Check name	11	8
Check number	10	7
Send photo	7	3
Check profile/ profile picture	3	5
Writing style	5	4
Encryption/E2EE	0	5
Dont know	14	7
Other/unclear	24	7

Encryption Meaning ($\kappa = 0.84$)		
Code	**t1**	**t2**
Messages not readable by third parties	32	52
Transformation into (secret) code	12	14
Access control	13	7
Data protection	11	3
Messages are secure	4	5
Something is encrypted	5	3
Dont know	9	2
Other/unclear	3	5
E2EE Meaning ($\kappa = 0.89$)		
Code	**t1**	**t2**
Messages not readable by third parties	13	52
Messages are secure	2	5
Data protection	1	2
From start to end encrypted	5	3
Transformation into (secret) code	0	6
Dont know	3	9
Other/unclear	1	17
WhatsApp notification: E2EE ($\kappa = 0.94$)		
Code	**t1**	
Messages not readable by third parties	49	
Messages/data are secure	18	
Dont know	6	
Other/unclear	14	
WhatsApp notification: Security Number ($\kappa = 0.94$)		
Code	**t1**	**t2**
Change of number	27	32
Security number changed	14	15
Change of phone/device	7	15
App was reinstalled	3	4
Offers protection/security	4	4
Dont know	16	9
Other/unclear	21	24
Behavioral Change (self report) ($\kappa = 1.00$)		
Code		**t2**
No		40
Yes		33
Maybe		13

Table 3. Code frequencies for free-text answers before (t1) and after (t2) the instruction video. κ denotes Cohen's Kappa for each code group, weighted by code frequency.