

Francesco Pompedda

Training in Investigative Interviews of Children: Serious Gaming Paired with Feedback Improves Interview Quality





Francesco Pompedda

Born 1988

Bachelor in Psychology, University of Cagliari, 2010 Master in Criminal and Investigative Psychology, Turin University, 2013 Licensed psychologist in Italy, 2013 PhD candidate, Åbo Akademi 2014-2018

Francesco Pompedda is a forensic psychologist who is currently visiting researcher at Göteborg University within the CLIP group. He also teaches introductory courses in forensic psychology and different short courses in investigative interviews of vulnerable witnesses



Training in Investigative Interviews of
Children:
Serious Gaming Paired with Feedback
Improves Interview Quality

Francesco Pompèdda

Psychology
Faculty of Arts, Psychology and Theology
Åbo Akademi University
Åbo, Finland, 2018

Supervised by

Professor Pekka Santtila, PhD.

Professor of Psychology, Faculty of Arts and Sciences, NYU Shanghai,
Shanghai, China

Dr. Jan Antfolk, PhD.

University Researcher, Faculty of Arts, Psychology and Theology, Åbo
Akademi University, Finland

Reviewed by

Professor Martine Powell, PhD.

Professor of Psychology, Director of the Centre for Investigative
Interviewing, Griffith Criminology Institute, Griffith University, Australia

Professor Thomas D. Lyon, PhD.

Judge Edward J. and Ruey L. Guirado, Chair in Law and Psychology at the USC
University of South California, Los Angeles, California

Opponent

Professor Renate Volbert, PhD.

Professor of Legal Psychology, Psychologische Hochschule Berlin, Berlin,
Germany

Cover Art by

Alessandro Vanzino

ISBN 978-952-12-3678-5(Print)

ISBN 978-952-12-3679-2 (Digital)

To my family

Acknowledgments

I would like to thank my pre-examiners Professor Powell and Professor Lyon for their constructive comments and their inspiring suggestions to improve this work. I would also like to thank Professor Powell for the wonderful discussions and the hospitality when we first met in Melbourne at the beginning of my PhD. I would finally thank Professor Renate Volbert for agreeing to be my opponent

L'illuminazione sulla via di... Torino

Nothing would have been possible without two important people. Professor Pekka Santtila and Professor Angelo Zappalà.

This project would have not been possible without you. Professor Pekka Santtila for being a wonderful supervisor, for introducing me to the world of research, for being able to put order where there was none, for showing me that it is possible to be professional and achieve results while having so much fun in the meanwhile, for the great pieces of advice on work and life, for the “go and execute” you told me more than one time, and that helped me not to stop. Thank you, the words cannot explain enough. Professor Angelo Zappalà thank you for being the Italian counterpart. For listening to my ideas, for being present and for all the fun we had together.

L'illuminazione sulla via di... Turku

In Finland, I met another important person in my journey. My second supervisor Dr. Jan Antfolk. Thank you for being a great supervisor. Always willing to help and to not kick me out from the room when I knock at the door for a “fast question” which usually was not that fast, for the great ideas and discussions. You all taught me so much that no words can explain my gratitude.

A special thank you to all the staff at Åbo Akademi (from the new PhD student to the professor), for making the long working day during the Finnish winter much easier. All the jokes, discussions, and fun cheer me up. For being always available to help each other even if we all have a tight time schedule. It would have not been possible also without you. I have different relationships with all of you, from being together in conferences and talk about life every day to just share a lunch now and then, however, I would like to thank you all and I hope that our friendship lasts forever.

I would also especially thank Professor Patrick Jern, for being always ready to help and provide great pieces of advice.

I would also send my gratitude to two important colleagues that are not working anymore at Åbo Akademi such as Emilia and Benny.

I would also thank all the past and current administrative staff at Åbo Akademi, especially Susanne, Yasmin, and Harriet, for being always supportive.

I would like to thank all the people who collaborated with me during this project. Among all, Angelica and Antonella for being able to work hard (you know what I mean) but also have fun at the same time.

I also thank all the colleagues (academic and practitioners) I met around the globe. I shared great discussions and incredibly funny moments with you.

I would like to thank all my new friends here in Finland with whom I shared wonderful moments from barbecuing in the Finnish archipelago to play futsal to chat about different life topics. Thank you! You made Turku like home. Among all, my deepest gratitude goes to Meija, Luca, Stefano, Olga, and Kim.

My gratitude goes also to my old friends in Italy for making distance a no problem in our relationships.

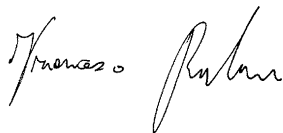
Dulcis in fundo, I want to thank Jessica who has been a cornerstone both in life and during this journey. Thank you for being patient and supportive.

Finally, my gratitude goes to my parents who taught me how to face life.

In God we trust, all others [must] bring data

W. Edwards Deming

Göteborg, Åbo. January 2018

A handwritten signature in black ink, appearing to read 'Francesco Pompèda'. The signature is written in a cursive, flowing style with some loops and flourishes.

Francesco Pompèda

List of publications

- I. Pompedda, F., Zappalà, A., & Santtila, P. (2015). Simulations of child sexual abuse interviews using avatars paired with feedback improves interview quality. *Psychology, Crime & Law*, *21*(1), 28–52. doi: 10.1080/1068316X.2014.915323¹
- II. Pompedda, F., Antfolk, J., Zappalà, A., & Santtila, P. (2017). A combination of outcome and process feedback enhances performance in simulations of child sexual abuse interviews using avatars. *Frontiers in Psychology*, *8*, 1474. doi: 10.3389/fpsyg.2017.01474²
- III. Krause, N., Pompedda, F., Antfolk, J., Zappalà, A., & Santtila, P. (2017). The effects of feedback and reflection on the questioning style of untrained interviewers in simulated child sexual abuse interviews. *Applied Cognitive Psychology*. doi: 10.1002/acp.3316.³⁴
- IV. Pompedda, F., Antfolk, J., Soveri, A., & Santtila, P. (submitted). Transfer of simulated interview training into interviews with children exposed to a mock event.

The original articles have been reproduced by permission of the copyright holders.

¹ This is the authors accepted manuscript of an article published as the version of record in *Psychology, Crime & Law* © 12 May 2014 -

<http://tandfonline.com/10.1080/1068316X.2014.915323>

² <http://www.frontiersin.org/articles/10.3389/fpsyg.2017.01474/full>

³ <http://onlinelibrary.wiley.com/doi/10.1002/acp.3316/full>

⁴ The first two authors have contributed to the manuscript equally and are given in alphabetical order

Picture in the cover used with permission from Mr. Alessandro Vanzino

Sammanfattning på svenska

Intervju kvaliteten (dvs. användningen av regler för bästa praxis) i misstänkta fall av sexuella övergrepp mot barn är fortfarande låg. Utbildningsprogram har utvecklats för att ta itu med detta problem. Vanligtvis lyckas inte dessa program uppnå långvariga effekter, och när positiva resultat uppnåtts, är programmen ofta logistiskt komplicerade, dyra och tidskrävande. Syftet med den föreliggande avhandlingen var att skapa och undersöka ett simuleringsverktyg för intervjuer (EIT®). Verktöget användes för att utbilda intervjuare att använda mer rekommenderade frågor under träningstillfällena med detaljerad, omedelbar och kontinuerlig återkoppling. Träningstillfällena var varken dyra eller tidskrävande. Vi skapade en s.k. allvarlig spelmodell där försökspersoner intervjuade datoriserade cyberfigurer utrustade med svarsalgoritmer och förutbestämda minnesspår. Vi undersökte hur detta tillvägagångssätt kan användas för att förbereda intervjuare för verkliga fall av misstänkta sexuella övergrepp. I samtliga studier som presenteras i den här avhandlingen definierades intervju kvaliteten som antalet rekommenderade och icke rekommenderade som ställdes, antal relevanta, neutrala och felaktiga uppgifter som cyberfigurer eller riktiga barn gav, samt om intervjuaren drog korrekta slutsatser om vad som hade hänt.

I studie I visade vi hur intervjuer med cyberfigurer som kombinerades med återkoppling förbättrade kvaliteten på simulerade intervjuer hos en grupp studerande jämfört med en grupp som genomförde intervjuerna utan återkoppling. Vi visade också att kunskap om evidensbaserade principer för intervjuer inte förbättrade intervjuernas kvalitet. Vi använde en kombination av resultat- och processåterkoppling (dvs. både återkoppling om slutsatserna och de använda frågetyperna).

I Studie II separerade vi mellan de två typerna av återkoppling och visade att de i kombination förbättrar intervjuer i högre grad än de gör åtskilda. I kombination resulterade de i medelstarka eller starka träningseffekter ($d_{ppc2} = 0.76$) på användningen av rekommenderade frågor efter endast fyra träningsintervjuer.

I Studie III använde vi en ny uppsättning svarsalgoritmer för intervjufrågor och cyberfigurernas svar. I de tidigare studierna var algoritmerna mekaniska (dvs. efter en viss mängd rekommenderade frågor gav operatören ett visst svar). Från studie III använde vi probabilistiska algoritmer som gav olika svar med en viss sannolikhet (sannolikheten var härledd från forskning om barns minne och suggestibilitet). I Studie III testade vi också om en enkel reflektionsuppgift ökade träningens effekt. Reflektionsuppgiften förbättrade inte träningseffekterna jämfört med kombinationen av återkopplingstyperna. Denna studie replikerade tidigare resultat angående effekten av intervjuer av cyberfigurer i kombination med återkoppling. Till exempel hade 90 % av deltagarna i de två grupperna som fick återkoppling ökat användningen av rekommenderade frågor och 38 % nådde en tillförlitlig förändring på bara två timmar.

I Studie IV visade vi att förbättringarna i intervju kvaliteten som uppvisats bland studeranden i studierna I-III också kunde nås i en grupp psykologer. Det andra och viktigaste resultatet av Studie IV var att de förbättringar som uppnått under träningen också överfördes till intervjuer med verkliga barn som bevittnat en arrangerad händelse. Under intervjuerna, som utfördes en vecka efter intervju träningen, användes 40 % rekommenderade frågor i återkopplingsgruppen jämfört med 26 % i kontrollgruppen.

I den föreliggande avhandlingen analyserades träningsresultaten också med ett mega-analytiskt tillvägagångssätt som kombinerar resultaten av de enskilda studierna. Resultaten visade hur simulerade intervjuer med cyberfigurer i kombination med resultat- och processåterkoppling på ett robust sätt förbättrade kvaliteten på simulerade intervjuer jämfört med en kontrollgrupp.

Sammanfattningsvis ger resultaten stöd för användningen av en allvarlig spelmodell för utbildningen av intervjuare. Tidigare forskning visar tydligt vikten av att intervjua barnet på ett neutralt sätt i utredningar av misstänkta sexuella övergrepp. På så sätt kan intervjuare ges ett nytt, interaktivt och effektivt verktyg medan polisavdelningar och utbildningsinstitutioner kan ges ett tillämpbart, kostnadseffektivt träningsprotokoll som kan förändra hur vi planerar och organiserar utbildning i detta sammanhang

Abstract

Interview quality (i.e., adherence to best practice) in alleged child sexual abuse (CSA) cases remains low. Training programs have been developed in order to tackle this problem. However, these programs are usually not successful in creating stable effects over time, or when positive results have been achieved, programs are often logistically complicated and expensive in addition to requiring a lot of time from those participating. The general aim of the present thesis was to create and test an interview simulation tool (EIT®). This tool was used to train interviewers to use more recommended questions through multiple practice occasions in combination with the administration of detailed, immediate and continuous feedback, but without excessive time and cost burden. We thus applied a serious gaming approach in which trainees interviewed computer-generated avatars equipped with response algorithms and predefined memories to explore the feasibility of this approach to train interviewers in alleged CSA cases. In all the studies presented in the present thesis, we operationalized interview quality as recommended and not recommended questions asked, relevant, neutral and wrong details elicited from the avatars or children and correct conclusions reached concerning what had happened to the avatars or children.

In Study I, we showed how interviews with avatars combined with feedback improved the quality of simulated investigative interviews in a group of students compared to a group of students that conducted the interviews without feedback. We also showed that knowledge regarding evidence-based principles relating to CSA investigations did not influence the quality of interviews. Here, we used a combination of outcome (i.e. information regarding the conclusion of the story) and process (i.e. information regarding the question types used) feedback simultaneously.

In Study II, we separated between the two types of feedback and showed that the combination of feedback enhanced training effects to a higher degree compared to the process and outcome feedback provided alone. For example, a combination of feedback elicited medium/strong effects ($d_{ppc2} = 0.76$) in improving the percentage of recommended questions in only four interviews.

In Study III, we used a new set of algorithms to relate interviewer questions to avatar responses. In the previous studies, the algorithms were mechanical (i.e., after a certain number of recommended questions an operator provided a detail). Starting from Study III, we used probabilistic algorithms that related interviewer questions to avatar responses probabilistically (in both cases the probabilities themselves were derived from research on child memory and suggestibility). In Study III, we also tested if a simple reflection task enhances training effects. The reflection task did not enhance training effects compared to the group that received a combination of the two previously used feedback types. This study replicated previous results regarding the effect of avatar interviews combined with feedback on interview quality. For example, 90% of participants in the two groups that received feedback improved their use of

recommended questions, and 38% reached a reliable change in their use of recommended questions in only two hours.

In Study IV, we showed that the improvements in interview quality achieved in student samples in Studies I-III were also achieved in a group of psychologists. The second and most important result of Study IV was that the improvements achieved during the training also transferred into interviews with actual children who had witnessed a mock event. During these interviews, that occurred one week after the training, the feedback group asked 40% of recommended questions compared to the control group who reached 26%.

The results of training were analyzed using a mega-analytic approach in the present thesis combining the results of the individual studies. The results showed how simulated interviews with avatars and the provision of a combination of outcome and process feedback improved in a robust manner the quality of simulated investigative interviews compared to a control group.

Overall, the results provide support for the use of a serious gaming approach to training interviewers. Previous research clearly shows how important it is to interview the child in the most neutral way possible when there is suspicion of abuse. Because of this, providing interviewers with a new, interactive and efficient tool together with providing police departments or training institutions with a realistically applicable, time-and-cost efficient training protocol can change the way we plan and organize training in this context.

Abbreviations

CSA	Child sexual abuse
NICHHD	National Institute of Child Health and Human Development
EIT	Empowering Interviewer Training
RDI	Raw data, Descriptive and Inferential statistics (plot)
OR	Odds Ratio

Table of contents

Acknowledgments	4
List of publications	6
Sammanfattning på svenska	7
Abstract	9
Abbreviations	11
Table of contents	12
1. Introduction	14
1.1. CSA Investigation	14
1.1.1. Why Interview Children?	14
1.1.2. Can We Interview Children?	14
1.1.3. Best Practices in Investigative Interviews of Children in Alleged CSA Cases	15
1.2. Training in CSA Interviews	16
1.2.1. State of the Art	16
1.2.2. Training Formats	17
1.3. Feedback Effects and Delivery Methods	19
1.3.1. Feedback Effects and Delivery Methods in the Context of Alleged CSA Interviews.	22
1.4. Current Challenges	23
1.4.1. Long-Term Effects of Training	23
1.4.2. Practice Delivery	23
1.5. Serious Gaming	24
1.6. Empowering Interviewer Training (EIT [®])	25
1.6.1. Comparisons with Other Training Programs and Problems Addressed	25
1.6.2. Theoretical Framework	26
2. Aims and research questions	30
3. Materials and methods	31
3.1. Participants	31
3.2. Materials and Procedure	31
3.3. Simulated Investigative Interviews with Empowering Interviewer Training (EIT [®])	34
3.3.1. Avatars	34
3.3.2. Scenarios	35
3.3.3. Algorithms	36
3.3.4. Coding	37
3.4. Materials and Procedures Specific to the Different Studies	38
3.4.1. Study I and Study II	38
3.4.2. Study III	38
3.4.3. Study IV	39
4. Results	41
4.1. Mega-Analysis of the Experiments	41
4.1.1. Procedures for the Mega-Analysis	41

4.1.2. Samples Used in the Mega-Analysis	41
4.1.3. Statistical Analyses	42
4.2. Results of Interviews with Avatars Coupled with Feedback	43
4.3. Effect of Different Types of Feedback in Interviews with Avatars	46
4.4. Transfer Effect in Interviews with Children	49
5. Discussion	52
5.1. Interpretation of Results	52
5.1.1. Effects of Knowledge about Evidence-Based Best Practice	52
5.1.2. Effects of a Simple Reflection Task	52
5.1.3. Effects of Different Types of Feedback	52
5.1.4. Interviews with Avatars Paired with Feedback Improve Interview Quality	53
5.1.5. Training Effects with Avatars Transfer into Interviews with Children	54
5.2. Limitations	54
5.3. Implications and Future Directions	56
5.4. Conclusions	58
References	59
Appendix A	68
Original articles	69

1. Introduction

A CSA interview can be compared to the Olympics final of 100m sprint event. Interviewers, like sprinters, need years of preparation and then have only one possibility to perform at their best. This is because it is known that the first interview with the child is crucial, and if mistakes are committed in this interview, it is not always possible to undo these mistakes later. In fact, the interview of the child is one of the cornerstones of the criminal investigation in alleged CSA cases. However, it has also proven to be one of the most complex tasks for forensic psychologists and police officers conducting such interviews. The use of a recommended questioning style has been shown to increase the probability of eliciting reliable information from children of all ages, and the reverse is true for a not recommended questioning style (e.g., Lamb, Herschkowitz, Orbach, & Esplin, 2011). Even if practitioners are familiar with these and other recommendations, results in studies conducted in different parts of the world show that, even if some improvements have been achieved, the quality of investigative interviews is low (e.g., Johnson et al., 2015). Positive change can be achieved by intensive training and practice with feedback spread over a long period of time (Lamb, Sternberg, Orbach, Esplin, & Mitchell, 2002). However, except for few positive results (Benson & Powell, 2015), training attempts usually fail to bring about long-term changes (Price & Roberts, 2011). Another problem related to the previous training setups is that they are usually logistically complicated and/or expensive in terms of both time and money.

1.1. CSA Investigation

1.1.1. Why Interview Children?

At the beginning of a criminal investigation, it is usually known that a crime has occurred and the task of the investigation is to find out who has committed it. In cases of suspected CSA, this is often not the case. The abuse is only suspected and the task of the investigation is to find out if abuse has taken place and, if so, what has happened. Due to the common lack of strong corroborating evidence, the interview with the child, which is almost always available (Lamb, Sternberg, Orbach, Herschkowitz, et al., 2002), is the only available piece of evidence in many cases (Herman, 2005), and, therefore, is of paramount importance to the final outcome.

1.1.2. Can We Interview Children?

The age and cognitive level of the child are important factors to take into consideration when conducting an interview. Research has mostly focused on understanding developmental changes in memory function and factors that enable reliable accounts (Krähenbühl & Blades, 2006). Regarding the factors that enable reliable accounts, it is accepted that children as young as four years

old are able to provide an accurate and detailed account of the situation they have experienced (Lamb, Sternberg, Orbach, Esplin, et al., 2003; Pipe, Lamb, Orbach, & Esplin, 2004). However, the accuracy of their accounts is influenced both by individual differences such as verbal ability, suggestibility, intelligence, temperament, and recall performance (Chae & Ceci, 2005) but also, and in greater degree, by external factors such as interviewer questioning style (Finnilä, Mahlberg, Santtila, Sandnabba, & Niemi, 2003). Lamb and colleagues showed how, in comparison with older children (> six-year-old), four-year-old children provided fewer details, but the proportion of details elicited using open-ended prompts was the same (50%). However, preschoolers (from three to about six years) are more prone to provide wrong information to not recommended questions than older children (six years or older). This includes, for example, yes/no questions (see Lamb, Sternberg, Orbach, Esplin, et al., 2003), repeated questions (Krähenbühl, Blades, & Eiser, 2009), and also repeated recommended questions (Krähenbühl & Blades, 2009). Preschoolers provide more undesired shifting (towards inaccuracy) and less desired shifting (towards accuracy) to repeated questions (Howie, Sheehan, Mojarrad, & Wrzesinska, 2004), and are less resistant to misinformation (Lehman et al., 2010). They also have more difficulties in monitoring the sources of their memories compared to older children (Roberts, 2002; Roberts & Powell, 2001), and they are more trusting of adults and more susceptible to suggestive techniques (see Quas et al., 2007). In summary, a questioning style that employs more recommended questions is important, regardless of age, but even more important with preschoolers.

The training program evaluated in the present thesis reflects these differences as the avatars were designed to correspond to four or six years old children, that is, the algorithms were different for these two types of avatars reflecting developmental changes described in the literature.

1.1.3. Best Practices in Investigative Interviews of Children in Alleged CSA Cases

The last three decades of studies have shown the conditions under which it is possible to maximize the willingness and capacity of young children to provide accurate accounts in CSA interviews. Table 1 presents a summary of evidence-based rules that according to different researchers and as a result of decades of research (e.g., Bruck & Ceci, 1999; Lamb, 2016; Lyon, 2010; Lyon 2014) interviewers should follow.

Table 1

List of Evidence-Based Rules that Interviewers Should Follow in Alleged CSA Cases.

Rules	
1	Interview the victim as soon as possible after the alleged abuse
2	Record the interview
3	Explain the purpose of the interview
4	Explain the ground rules of the interview (the child should promise to tell the truth; the child can and should say “I don’t remember,” “I don’t know,” “I don’t understand,” or correct the interviewer; the interviewer was not present so they do not know what happened)
5	Establish rapport using open questions
6	Ask the child to report a neutral event as practice interview
7	Use open questions
8	Allow the child to introduce the topic of abuse (the funnel interview)
9	Use focused questions building them on information provided by the child previously
10	Avoid suggestive questions
11	Avoid yes/no or forced choice questions
12	Ask simple questions and use the child’s own vocabulary

From Table 1 it is possible to see how the use of recommended questions (rules 5, 6, 7, 8, 9) and the avoidance of not recommended questions (rules 10, 11) are an important part of investigative interviews of children. Unfortunately, it is well known that there is a gap between best practice and its implementation in real interviews (Criminal Justice Joint Inspection, 2014; Johnson et al., 2015; Sternberg, Lamb, Davies, & Westcott, 2001). The training program that is presented and evaluated in the present thesis focuses on achieving interviews that follow these rules.

1.2. Training in CSA Interviews

1.2.1. State of the Art

Prior research on training of investigative interviews with children provides information about some tested solutions aimed to improve the quality of interviews: Interviewers should use structured protocols such as the NICHD (e.g., Herschkowitz, Fisher, Lamb, & Horowitz, 2007; Orbach & Lamb, 1999), or its derivation such as the Ten Steps investigative interview (Lyon, 2005). Interviewers should be provided with training that includes multiple practice occasions (Lamb, Sternberg, Orbach, Herschkowitz, et al., 2002), and continuous, immediate and detailed feedback (Lamb et al., 2002; Smith, 2008).

However, even if it is clear what needs to be changed and how, training programs, with some promising results (Benson & Powell, 2015; Cederborg, Alm, Lima da Silva Nises, & Lamb, 2013; Yi, Jo, & Lamb, 2016), have generally failed in creating and/or maintaining improvements over time. One of the possible explanation for these failures lies in the employed training formats.

1.2.2. Training Formats

Short and Intensive Theoretical Training. Despite going against research-based recommendations, both concerning general principles of learning and particularly field-tested training programs, short and intensive theoretical training is still a common (probably the most common) training format. This training format is effective in improving knowledge but fails to transfer this knowledge into practical skills during actual investigative interviews (Johnson et al., 2015) and in some cases also in mock interviews (e.g., Cederborg, Orbach, Sternberg, & Lamb, 2000; Sternberg et al., 2001). One recommendation is to avoid option-posing and/or suggestive prompts. Different studies and in diverse countries have showed how interviewers fail to adhere to this recommendation in field interviews. For example, in England and Wales, after the administration of the Memorandum of Good Practice, 40% of information was elicited using option-posing or suggestive prompts (Sternberg et al., 2001). Assuming that interviewers in the field received theoretical training and have knowledge of best practices, results from other countries showed similar non-adherence to recommendations. In Finland 50% (Korkman, Santtila, & Sandnabba, 2006), in Sweden 53% (Cederborg, Orbach, Sternberg, & Lamb, 2000) of questions were closed. Other studies aimed to test the effect of training programs reported interview quality prior to the interventions. For example, in Canada the likelihood of using option-posing or suggestive prompts was 46% (Cyr & Lamb, 2009), and in Korea 56% (Yi et al., 2016). Similar evidence is also available for Australia (Guadagno & Powell, 2009) and Scotland (La Rooy, Lamb, & Memon, 2011). A clear example of how this type of training is ineffective is the study by Johnson and colleagues (2015). Looking at data concerning forensic interviews in Norway over a time span of 22 years (Thoresen, Lønnum, Melinder, & Magnussen, 2009; Thoresen, Lønnum, Melinder, Stridbeck, & Magnussen, 2006), the researchers showed how the use of only best practice recommendations failed to produce effects on performance. Even if there were some positive trends, such as diminished use of yes/no questions and increased use of directive questions, there were no improvements in the general use of open-ended questions (2%) over this time span. In sum, providing only best practice recommendations to interviewers increase their knowledge but fails at creating behavioral changes during actual interviews.

Training with Practice and Ongoing Feedback. Lamb (Lamb, Sternberg, Orbach, Herschkowitz, et al., 2002) was the first to show the positive effects of providing interviewers ongoing feedback on actual interviews. The common components of this training are a) Face-to-face lectures distributed over time

about theoretical concepts. For example, explanations of ground rules, establishment of a rapport with the child, evidence-based information on children's memory and suggestibility, and explanations of the different question types, b) in most cases, the use of a structured protocol such as the NICHD (National Institute of Child Health and Human Development) c) practice in real cases or in mock interviews with colleagues or actors that pretend to be children. During all these different phases, detailed feedback on their performance and supervision are provided to the interviewers. Solid findings show how this type of training improves the use of open-ended questions and decreases the use of suggestive questions. Lamb and colleagues have also shown how feedback must be provided on a continuous basis and must be detailed and immediate in order to be effective (Lamb, Sternberg, Orbach, Esplin, et al., 2002). A successful example of this type of training is the one developed by Cederborg and colleagues in Sweden (Cederborg, Alm, Nises, & Lamb, 2013). In this training program, 104 active crime investigators underwent six different half-year long courses. Interviewers used a new structured protocol based on the NICHD protocol (Lamb et al., 2007) and on the PEACE model (Clarke, Milne, & Bull, 2011). Practice, in this case, consisted of interviews and supervision both in real cases and as well as in mock interviews where a trainee took the role of the child and/or the interviewer. Investigators at the end of the training used two-thirds fewer option-posing questions and three times as many invitations compared to before the training. Long-term effects were not tested in this study. However, as Lamb and colleagues showed (Lamb, Sternberg, Orbach, Hershkowitz, et al., 2002), the effects of training usually disappear as soon as the supervision and feedback end or soon after. Benson and Powell (2015) reported in a summary of training studies over the past 15 years that the disappearance of training effects is evident within six months after the end of feedback or training.

Training with Computerized Methods and Serious Gaming. In recent years, different computer-based learning activities and methods based on serious gaming (as the one presented in the present thesis) have been developed. With computer-based learning activities, we refer to any method that does not require face-to-face training (Benson & Powell, 2015), while with serious gaming we intend training in which any type of avatar has been used instead of actors and/or real interviews (Brubacher, Powell, Skouteris, & Guadagno, 2015). In some studies, the authors used a combination of these two methods (Powell, Guadagno, & Benson, 2016). In one of these studies (Powell, Guadagno, & Benson, 2016), the authors provided the trainee with computer-based activities (12 modules of 3 hours each) over several months. The computer-based activities included, among others, explanation of best practice and of question types, reading and quizzes about how to elicit a disclosure from a child and a simulated interview with a virtual child, in which interviewers choose the best questions among options provided. Feedback and a rationale explaining why the choice made was correct or otherwise was provided immediately to the participants. The effect of this training has been tested

during mock interviews with actors and the results showed that participants after training asked around 75% open-ended questions and that the improvements were maintained between three and six months after the training. Brubacher and colleagues (2015) provided training for teachers. The aim of the study was to test the effect of simulated interaction with an avatar without highly intensive training and if the effects of the training would transfer into a live mock interview. Participants interviewed one five-year-old avatar three times over one week, choosing the best questions among provided options. Likewise, feedback was provided immediately. One week later the teachers interviewed a trained research assistant who pretended to be a child. The teachers asked 13% of open questions in the pre-training and around 50% open questions in the post-training situation. In another study (Benson & Powell, 2015) researchers evaluated another type of training with investigative interviewers. The training included 15 modules with different relevant topics, such as coding of question types, child development, and techniques on how to elicit a disclosure. The suggested time for the completion of the training was 15 weeks. As in the previous example, the modules were delivered over the internet and individualized feedback was provided immediately to the participants. Participants also participated in numerous mock via-skype-interviews with trained actors who pretended to be five-year-old children. First, the authors tested the effect of training in mock interviews. Participants completed three different mock interviews: one before the training, one immediately after, and one between three and six months after the first interview. Participants used 30% of open questions in the pre-training and reached around 58% open questions at the immediate post-training and 56% in the follow-up. Moreover, the authors tested the effect of the training in field interviews. They randomly selected 78 interviews conducted in the five years prior to the training and compared them to 78 interviews conducted by 26 participants after the training. The results showed that participants used more open questions (40%) compared to pre-training interviews (10%). Assessment of performance in the subsequent 12 months showed that the improvements were still present.

1.3. Feedback Effects and Delivery Methods

In a vast meta-analysis of research on student learning, Hattie and Timperley (2007) showed how feedback has a powerful influence on learning. However, they also highlighted that both the type of feedback and how it is delivered modulates the outcome. The average effect size of feedback reported in Hattie and Timperley (2007) was $d = 0.79$. In order to have a comparison value, this effect was twice the average effect size of schooling ($d = 0.40$). However, effect sizes showed large variability that reflected the type of feedback provided. As the authors highlighted, the best outcomes ($d = 1.10$) were the result of feedback that provided information about a task and how to do it more effectively.

The literature defines learning as the process that narrows the discrepancy between performance and expected standards (Hattie & Timperley, 2007). The discrepancy can be reduced both by the one intending to learn (e.g., higher effort and more efficient strategies) and by teachers (e.g., providing concrete goals and facilitating effective learning through feedback). In this context, feedback is effective only if it helps reducing this discrepancy between performance and expected standards. Starting from an engineering approach, Boud defined feedback as a loop characterized by an initial task, subsequent provision of information, and final evaluation of performance in a second related task (Boud, 2015). In contrast to a mere provision of information, this process can be defined feedback, only if performance is improved in the second task. The author also highlights how it can be necessary to repeat this loop more than once in order to improve performance (Boud & Molloy, 2013).

Large variability in the effects of feedback interventions has been recorded, including negative effects where feedback has hindered learning (Kluger & DeNisi, 1996). Kluger and DeNisi (1996) proposed that attention and motivation play an important role. Attention is limited and is rarely focused on details. Because of this limitation, only some discrepancies between performance and expected standards can be processed. According to the authors, feedback should switch attention to selected and specific discrepancies to facilitate behavioral change. Feedback that focuses on the task at hand more than personal evaluations effectively facilitates the learning process. Kluger and DeNisi (1996) proposed that feedback also stimulates different levels of motivation. Negative feedback (focusing on discrepancy between performance and expected standards) can activate more effort in order to overcome the discrepancy. If working harder fails, the person intending to learn may start to create new hypotheses and strategies (e.g., reflect more on the task), which can lead to better and deeper processing. Moreover, the authors highlight how also personal characteristics, such as self-esteem, may influence how a recipient reacts to feedback. Recipients with higher self-efficacy react better to negative feedback (the focus remains on the task) while recipients with lower self-esteem tend to move the focus on the self, causing disengaging and lower performance. The authors also argue that the reverse is true for positive feedback in connection with self-esteem. Hattie and Timperley (2007) later theorized that this might not be always true, suggesting that the range of reactions is more varied. Receivers with low self-esteem can react to positive feedback with both more engagement in the task due to the acknowledgment that they have some deficiencies that can be fixed, as well as disengaging due to the initial success.

Hattie and Timperley (2007) proposed three questions that an effective feedback should aim to answer: "Where am I going?", "How am I going?", and "Where to next?" The answer to "Where am I going?" must be a concrete, clear and achievable set of expected standards. As proposed by Boud (2015), there must be a dialogue between giver and receiver, and knowledge of the appropriate standards must be held by both the parties in order to answer this

question. The answer to “How am I going?” provides information regarding the gap between the performance and the expected standards and information on how to proceed in order to narrow the gap. Boud (2015) proposes that in order to engage receivers in taking action to close the gap, mere information is not always sufficient. Receivers must value the input and trust must be built between giver and receivers (Ivers et al., 2012). The answer to “Where to next?” should allow receivers to go beyond the current task and learn more transversal skills. Hattie and Timperley (2007) also argued that these three questions interact on four different levels: Task-focused, process focused, self-regulation focused, and person-focused (Boud, 2015). Task-focused feedback provides information about how a task is being accomplished (Hattie & Timperley, 2007). Process focused feedback helps the receiver build up more general and transversal concepts that can be adapted to different and more complex tasks. Usually, this feedback stimulates deeper learning (Boud, 2015; Hattie & Timperley, 2007). Feedback about self-regulation helps the receiver to create an internal feedback. In this way, receivers become autonomous in monitoring their development and proactively research for more information and more feedback. The final level of feedback is feedback directed on personal characteristics. This type of feedback has usually been shown to be ineffective.

One important concept present in the literature is that in order to provide effective feedback, receivers must be active and autonomously seek feedback and information. From this perspective, feedback becomes a “process used by learners to facilitate their own learning” (Boud & Molloy, 2013, p. 703). In a recent meta-analysis (Anseel, Beatty, Shen, Lievens, & Sackett, 2015), the authors reviewed the antecedents and consequences of feedback seeking behavior. Feedback seeking behavior is one of the proactive self-regulation strategies that helps receivers to improve. According to the authors, this behavior is influenced, among other variables, by individual differences (e.g., age and feedback attitudes) and situational variables (e.g., role ambiguity). Even if further research is needed, all these variables are presumably not static within individuals and vary among contexts. In sum, the literature suggests that the use of “tailored feedback” based on the receivers, on the task, and on the situation, is more effective than any “one size fits all” feedback.

In conclusion, in order to be effective feedback should at least:

- Provide clear standards to which students can compare their actual performance
- Provide enough information to show to the students why a change is needed and how to reach the expected standards
- Stimulate students to become pro-active information-and-feedback seeker

1.3.1. Feedback Effects and Delivery Methods in the Context of Alleged CSA Interviews.

Among the different types of feedback present in the literature, we selected two that can be easily adapted to the training in CSA interviews. Process feedback, which focuses on the task processes and provides information on how to perform the task (Landsberg, Van Buskirk, & Astwood, 2010) and outcome feedback which Hattie and Timperley defined as “feedback about how a task is being accomplished” (Hattie & Timperley, 2007), however, without information on how to perform better. There is some evidence in the literature for a superior effect of process feedback in training of complex tasks (Astwood, van Buskirk, Cornejo, & Dalton, 2008; Buff & Campbell, 2002). Buff and Campbell showed how process feedback, compared to outcome feedback and mere practice, significantly improved performance in decision-making. However, outcome feedback produced similar results to process feedback. Astwood and colleagues tested the effects of different types of feedback on complex decision-making, showing how process feedback lead to better results (82% of accuracy compared to 67% in the outcome feedback condition). In addition, a recent meta-analysis of feedback in a computerized environment (Van der Kleij, Feskens, & Eggen, 2015) showed that elaborate feedback (process feedback) produced larger effect sizes ($ES' = 0.49$) compared to outcome feedback ($ES' = 0.32$) and other types of feedback. Even if process feedback has a stronger effect than outcome feedback, outcome feedback still has an impact on learning.

According to Hattie and Timperley (2007), outcome feedback is more efficient in changing faulty interpretations. For example, an interviewer who, due to confirmation bias, confirms the initial hypothesis reaching faulty conclusions can be assisted by outcome feedback that highlights the faulty conclusions and helps to recognize the interviewer’s confirmation bias. Process feedback instead is more effective than outcome feedback when the learning goal is the transfer of acquired skills to more complex tasks. For example, the transfer of improved use of recommended questions, from mock interviews to field interviews is not automatic due to the higher complexity that the field interviews present. Process feedback providing information on how to perform better help interviewers to move over the actual task (e.g., the mock interview) and develop more general skills that can be applied to more complex tasks (e.g., field interviews). Both these two tasks are essential in investigative interviews of children. For these reasons, a combination of outcome and process feedback can enhance training on different dimensions compared to providing only one of the two types of feedback.

In an investigative interviewing context, process feedback can be defined as feedback on whether the questions used by the interviewer are appropriate or not. Instead, feedback on whether the interviewer reached the correct conclusion about what had happened corresponds to outcome feedback. Outcome feedback operationalized in this way cannot be provided in real

investigative interviews, except in rare cases. As Vrij (2005) highlights, due to the usual lack of corroborating evidence, it is not possible to know the ground truth of a detail provided by the child in actual CSA interviews. In addition to the two types of feedback provided, in Study III we also try to stimulate the self-regulation process through the provision of a reflection task. As highlighted by Boud (2015), the self-regulation focused feedback may take the form of asking questions rather than providing information.

As highlighted by Herschkowitz and colleagues (2017), the common feedback provided to interviewers in training programs pertains to the interviewer's behavior, usually, feedback on the question types used (e.g., Benson & Powell, 2015; Lamb et al., 2002; Yi, Jo, & Lamb, 2016). With the exception of outcome feedback on quizzes regarding best practice (e.g., Powell, Guadagno, & Benson, 2016), in no study within the field of CSA investigative interviews has feedback on the conclusions of the interviewers been provided, or the effect of the combination of outcome and process feedback been tested. On the other hand, a considerable amount of research has investigated how feedback should be delivered. The general finding is that feedback must be provided in a continuous, immediate and detailed way in order to maximize its effects (e.g., Lamb, Sternberg, Orbach, Herschkowitz, et al., 2002; Lamb, Sternberg, Orbach, Esplin, et al., 2002).

1.4. Current Challenges

1.4.1. Long-Term Effects of Training

Achieving long-term effect in training of investigative interviewers is likely the largest challenge that researchers and practitioners in this field are currently facing. Even if some training programs have been proven to be effective in enhancing interview quality in the short term (Cederborg et al., 2013), the effects of training seem to disappear immediately (Lamb et al., 2002), or only last for a short amount of time after the training is discontinued. The best results so far have been achieved by the training program of Powell and colleagues (e.g., Benson & Powell, 2015), which showed effects still after six months. Currently, the only way to maintain the quality of interviews is to provide continuous feedback and supervision. It is, however, unclear how much supervision and feedback is needed in order to achieve permanent long-term effects (Lamb, 2016; Lamb et al., 2002) if this is even possible. The simulation tool presented in this thesis can provide booster sessions and continued feedback in an easy way in order to maintain training effects over time.

1.4.2. Practice Delivery

In order to provide the interviewer with multiple practical occasions coupled with feedback, most common training types include either role-playing with adult actors pretending to be a child, interviews with children regarding a

neutral mock event, or interviews in real cases. All these solutions are problematic. Actors playing the role of a child in simulated interviews tend to overestimate the cognitive capabilities of a child compared to how actual children behave in real interviews (Powell, Fisher, & Hughes-Scholes, Lamb 2016). Overestimations of children's capabilities obviously create dissimilarities between training and real interviews. In this way, transfer of training effects is in jeopardy due to the lack of challenges typical of real interviews. In contrast, interviews in real cases obviously provide a realistic response pattern and the reluctance typical of children in alleged CSA cases. As stated before, due to unknown ground-truth, this type of training usually lacks unbiased and detailed feedback. Moreover, committing mistakes in real cases can have strong negative consequences for all parties involved. Finally, practice through interviewing children who have witnessed or participated in a neutral mock event does not necessarily mimic the reluctance and complexity of interviews with children in alleged CSA cases (Lamb, 2016; Powell & Wright, 2008). A general problem shared by most types of training is also the high monetary costs and complicated logistics related to the use of actors and the provision of a continuous feedback by external experts (Benson & Powell, 2015). In realistic contexts where the required monetary resources are not always available to police organizations and training schools, identifying a cheap and effective way of training interviewers is of paramount importance.

1.5. Serious Gaming

No general consensus has been reached on how to define serious gaming (Boyle, Connolly, & Hainey, 2011). However, combining different definitions, we provisionally define serious gaming as any game created (using a pedagogical paradigm) within a computer-based environment, with the aim of teaching and learning (Miller, Chang, Wang, Beier, & Klisch, 2011; Wouters, van Nimwegen, van Oostendorp, & van der Spek, 2013).

A number of meta-analyses have investigated whether serious gaming methods are more efficient than traditional methods (Olszewski, 2016; Sitzmann, 2011; Wouters, van Nimwegen, van Oostendorp, & van der Spek, 2013). The results are somewhat mixed. However, more stable and positive results in favor of serious gaming have been achieved when they have included feedback. One meta-analysis (Wouters & Van Oostendorp, 2013), for example, showed that feedback provided in serious gaming environments enhanced learning ($d = .34$) compared with a group that did not receive feedback. The magnitude of the effect was even larger when the aim was the learning of skills ($d = .62$). A series of case studies also show how serious gaming has proven to be an effective method to teach complex practical skills (Graafland, Schraagen, & Schijven, 2012; Van Der Zee, Holkenborg, & Robinson, 2012). For example, serious gaming is effective in the teaching of technical and non-technical skills in the surgical field (Graafland et al., 2012).

Serious gaming has some features that make it appealing for developers of training programs. It is flexible (Heyselaar, Hagoort, & Segaert, 2017), interactive, engaging while educational (Bellotti, Kapralos, Lee, Moreno-Ger, & Berta, 2013), based on active learning⁵ principles (Greitzer, Kuchar, & Huston, 2007), and finally, it allows for a safe learning environment in which interviewers can make mistakes without dire consequences (Mikropoulos & Natsis, 2011). Focusing more on the context of investigative interviews, the use of serious gaming helps to provide training in a realistic context, where mistakes are not costly, and where it is possible to provide unbiased feedback in a cost and time-effective way. For this reason, in the last five years, more studies have been carried out in order to test if serious gaming can be applied to the training of investigative interviewers, including the experiments presented in the present thesis.

1.6. Empowering Interviewer Training (EIT®)

EIT® (Krause, Pompedda, Antfolk, Zappalà, & Santtila, 2017; Pompedda, Antfolk, Zappalà, & Santtila, 2017, Pompedda, Zappalà, & Santtila, 2015) is a software for the simulation of investigative interviews in alleged CSA cases. At the moment, the interview simulation tool consists of 16 different avatars. Each avatar consists of an alleged CSA case and lists of details that constitute the memories of the avatar. In this way, we know the ground truth of what "happened" to the avatar in the form of details related to the alleged abuse situation. In addition, response algorithms have been created to recreate age and cognitive development differences between preschoolers (< 6-years-old) and children during their first years in school (> 6-years-old). Interviewers interview the avatar in a face-to-face interview where they ask questions after reading a background scenario regarding the allegation.

1.6.1. Comparisons with Other Training Programs and Problems Addressed

Based on the existing literature on the characteristics of effective training, EIT® has some features that differentiate it from other training approaches. Employing serious gaming and avatars can potentially fix many of the problems highlighted before:

- It provides realistic patterns of responses, due to the use the probabilistic algorithms. The algorithms are based on the available research on children memory and suggestibility and provide realistic responses based on the question types asked by the interviewer and the information contained in the memory of the avatar. The algorithms can be modified based on new research. It

⁵ Active learning is an instructional method, based on the constructivist learning theory, in which the learner, instead of just listening, is engaged in the learning process (Prince, 2004)

is known that actors do not always mimic typical response patterns of children. In this way, the interview simulation tool aims to replace the use of actors. Response algorithms are aimed at fixing this problem.

- It allows the provision of immediate, continuous and detailed feedback thanks to the predefined memories. In our setup, it is possible to know the ground truth (i.e., what really happened to the avatar) and a number of additional details related to the avatar's life. This type of unbiased feedback is, as a rule, not possible in field interviews.
- It provides a *safe* as opposed to a *wicked* learning environment (Gambrill, 2005; Hogarth, Lejarraga, & Soyer, 2015), in which it is possible make mistakes and receive detailed and unbiased feedback, while having plenty of possibilities to practice.
- It reduces the related costs, both monetary and time related, as well as the logistic complexity of training. The interview simulation tool is aimed to be used as a web-based application, which can cut the cost related to the logistics of face-to-face training.

Compared to other studies that employ avatars (Brubacher et al., 2015; Powell et al., 2016), our avatars are based on probabilistic response algorithms and interviewers actively ask questions, instead of choosing the best option from a list of questions, which mimics, in a more realistic way, what interviewers actually face in field interviews.

1.6.2. Theoretical Framework

CSA Interviewing is a Complex Task. Interviews with children in alleged CSA cases constitute a complex task, both cognitively and emotionally (Lafontaine & Cyr, 2017). From a cognitive perspective, a complex task is a task where “learners are often overwhelmed by the number of information elements and their interactions that need to be processed simultaneously before meaningful learning can commence” (Paas, Renkl, & Sweller, 2004, p. 1). Interviews in alleged CSA cases surely share these characteristics. Lafontaine and Cyr (2017) highlight how an interviewer should be flexible in adapting the interview to the child's developmental level, temperament, and emotional state, and create a trust-based relationship. Moreover, the interviewer should follow the recommended best practice, such as asking open questions. All these tasks must be processed simultaneously, creating a complex scenario that can overwhelm the cognitive capabilities of the interviewers. Moreover, interviews in alleged CSA cases can be emotionally stressful and practitioners usually complain about high workload and complex working climate (Powell, 2013). Solid findings show how stress has negative consequences on the cognitive capabilities of a human being because it works as an additional cognitive load (Stawski, Sliwinski, & Smyth, 2010). For all these reasons, and as suggested by

other scholars, interviews in alleged CSA cases can be considered a complex cognitive task (Paas & van Gog, 2006).

Human Cognitive Architecture and the Cognitive Load Theory. Working Memory (WM), as part of the human cognitive architecture, is involved in real-world cognitive tasks (e.g., Engle, 2002; Harrison et al., 2013) and in learning complex tasks (e.g., Engle & Kane, 2003). However, working memory has limited capacity when processing information and is not always adequate to learn complex materials (Wouters & Van Oostendorp, 2013). In the context of training investigative interviewers characterized by complex cognitive tasks and emotion regulation, we, as well as other scholars (Mugford, Corey, & Bennell, 2013), propose the Cognitive Load Theory (CLT) as a starting point to plan training programs. The CLT (for an exhaustive explanation, see Ayres & Paas, 2009; Sweller & Chandler, 1994), is an evidence-based instructional theory aimed at maximizing permanent learning and transfer through the optimization of the cognitive load related to learning (Mugford et al., 2013). According to the CLT, schemata construction, their automation and their store in the long-term memory (Paas, Renkl, et al., 2003) are the aims of training. Automatic processing of schemata (that contain multiple pieces of information but are treated as a single element in working memory) requires less cognitive effort (Kalyuga, Ayres, Chandler, & Sweller, 2003). Thereafter, schemata help the trainee to overcome the limited capacity of working memory increasing the amount of information processed at once and that is transferred into the long-term memory. Therefore, training must be planned in order to favor the optimal allocation of cognitive resources to schemata creation and automation.

Three Different Sources of Cognitive Load. In the present thesis, we are going to present the classic triarchic model of cognitive load (Paas, Renkl, & Sweller, 2003). The CLT divides the cognitive load into three different additive sources: intrinsic, extraneous, and germane (Paas, Renkl, & Sweller, 2003). This means that during training, the sum of these three sources of cognitive loads must remain within the limits of working memory (Paas, Tuovinen, & Tabbers, 2003). The intrinsic load cannot be reduced because it is related to the complexity of the task that must be learned (Mugford et al., 2013). As explained before, interviewing children in alleged CSA cases is supposed to have a high level of intrinsic load. Extraneous load is the cognitive effort not related to learning but to training delivery and it can be reduced by changing the training format (Merriënboer & Sweller, 2005), for example, by showing a video with non-verbal behavior instead of providing a text about it. The germane cognitive load is the cognitive effort related exclusively to schemata creation. Let's imagine that the maximum cognitive load an individual can sustain is 10 bits. CSA interview contains a high level of intrinsic cognitive load, due to the elevated number of interactive elements present, that we cannot reduce, say 7 bits. This means that the individual has left 3 bits of free cognitive space. The extraneous and germane cognitive loads share these 3 bits. As said before, according to the CLT, learning happens through schemata acquisition and automation. However, the germane cognitive load refers to the cognitive effort

related to schemata acquisition and automation while the extraneous cognitive load can have a disturbing effect on learning because it uses cognitive effort in tasks that are not related to learning. Therefore, effective training should focus on reducing extraneous cognitive load (Kirschner, 2002) and prioritizing germane cognitive load (van Merriënboer & Sluijsmans, 2009). There is no complete agreement on the number of sources of loads and the relationships between them in the literature (Debie & van de Leemput, 2014). For example, regarding the role of the germane cognitive load, some scholars argue that it actually enhances learning but it is not involved in schemata creation (see Debie & van de Leemput, 2014, for a review). However, the main concept of the CLT remains the same, maximize the cognitive effort related to schemata creation and automation and limit the cognitive effort that is not related to these tasks.

CLT Applied to the Simulation Tool. In the specific case of the interview simulation tool, we aim to reduce extraneous cognitive load providing the interviewer with multiple practical occasions instead of providing frontal lectures. Interviewers therefore mostly learn through feedback and practical experience the effects of different questioning styles. It is generally accepted that feedback may help to overcome the limitations of human cognitive architecture (Mayer, 2008; Wouters & Van Oostendorp, 2013) and that active learning favors schemata automation (Sweller, van Merriënboer, & Paas, 1998).

Other Techniques that Improve the Likelihood of Transfer into Field Interviews. We also included other techniques aimed at improving the likelihood of learning and transfer of training effects. For example, the variability effect (Paas & Merriënboer, 1994) or also called mixed practice suggests that introducing variability in a task improves the probability of transfer (for a review see Helsdingen, van Gog, & van Merriënboer, 2011; Schmidt & Bjork, 1992). This means in our case that providing avatars of different ages, different response patterns, different memory contents and in a randomized order leads to better results than providing them for example in an easy-to-difficult order. A possible explanation for this phenomenon is that random practice favors abstraction of general rules compared to practice on a less variable task where trainees rely on the memory of a specific clue (Helsdingen et al., 2011). Due to its higher complexity compared to an easy-to-difficult order, this technique can produce worse performance in the short term (i.e., during training). However, it favors abstraction of general rules, which can enhance transfer into more complex tasks, such as a field interview. One example is when interviewers realize through practice that asking only invitations (e.g., “Tell me more!”) after a child said, “He touched me”, does not always have the desired effect. In some cases, and after that the interviewer has already tried with invitations to have this type of information from the child, a directive or narrowed invitation such as “You told me that someone touched you, who touched you?” can be rewarding (Ahern, Andrews, Stolzenberg, &

Lyon, 2015; Andrews, Ahern, Stolzenberg, & Lyon, 2016; Gagnon & Cyr, 2017; Stolzenberg & Lyon, 2017).

Moreover, resemblance between training and task, both on a structural and surface level, improves the probability of transfer of training (e.g., Blume et al., 2010, Soveri et al., 2017), which is defined as the extent to which practice in one task has an effect on performance in another task (e.g., Blume, Ford, Baldwin, & Huang, 2010). We have designed our training simulation so that it would be as similar as possible to real interviews, both when it comes to the surface (avatars look like children) and structural (the avatars have response algorithms that mimic those of a child of a specific age) features. A recent study (Heyselaar et al., 2017) showed that there are no differences in dialogue interactions between a human-human dyad and a human-avatar dyad. Compared to avatars used in other training setups, interviewers interact verbally in a normal conversational format with the avatars of our investigative interview simulator.

CLT Applied to Other Training setups. The cognitive load theory can also be applied to other training approaches presented in the literature. For example, Benson and Powell (2015), in a departure from previous training setups, presented information over a longer period, and participants developed at their own pace. This effect is known to produce better results, and from a cognitive-load-theory perspective, can be explained by the fact that trainees have to deal with less information at the same time, making it easier to process the information and transfer it into the long-term memory. It may be that they had more chances to create schemata, and/or to spend more time on something that was more difficult to learn for them. This process can have transferred schemata to the long-term memory and that is why they were able to retrieve the schemata also six months after the end of training. Moreover, and as in other training studies, they did not employ passive learning but quizzes and mock interviews with immediate feedback, which help to decrease the cognitive load unrelated to learning.

2. Aims and research questions

The general aim of the current thesis was to develop and test a simulation tool aimed at improving the quality of investigative interviews with children in alleged CSA cases. At a general level, we expected that the groups conducting interviews with avatars coupled with feedback would conduct better quality interviews compared to control groups interviewing the avatars without any feedback. Interview quality was measured using

- Recommended and not recommended questions asked
- Relevant, neutral and wrong details elicited from the avatars or children
- Correct conclusions reached

The groups receiving feedback were expected to use more recommended and less recommended question, to elicit more relevant and neutral details and fewer wrong details, and finally, to reach more correct conclusions about what had happened compared to the active control groups. These hypotheses were tested in all studies.

The second main aim of the thesis was to test if the improvements achieved during the interview with avatars coupled with feedback would transfer into interviews with children who witnessed a mock event. This hypothesis was tested in Study IV. The individual experiments had also the following additional research questions:

Study I. We tested the effects of interviewers' theoretical knowledge of CSA and the effects of the purported age of the avatar on the quality of investigative interviews. We expected knowledge of evidence-based practice to be unrelated to interview quality and that the purported age of the avatar would have influenced the interview quality. Avatar algorithms were different between the "four-year-old" and "six-year-old" avatars, requiring, for example, more recommended questions in interviews with four-year-old avatars in order to achieve a correct detail.

Study II. We tested the effects of different types of feedback on interview quality separately. We expected the combination of process and outcome feedback to have a stronger effect in enhancing interviewer quality compared to their separate administration.

Study III. We tested the effects of a simple reflection task on interview quality. We expected that the reflection task would have boosted interview quality.

Study IV. We tested the effects of training in a group of psychologists. We expected to find similar results in interview quality as in previous studies in which we tested students.

3. Materials and methods

3.1. Participants

A summary of participants, demographic, experimental group and research question is presented in Table 2. In Studies I, II, and III the participants were university students, whereas in Study IV participants were psychologists.

3.2. Materials and Procedure

General procedures for all the studies are described in Table 3, with the exception of the transfer experiment in Study IV, for which the procedure will be explained separately.

Table 2
Participants, Demographics, Experimental Groups and Research Questions.

Study	Research Hypotheses	Participants	N	M _{Age}	Sex (F%)	Experimental Groups
Study I	<ol style="list-style-type: none"> 1. Theoretical knowledge regarding CSA investigations will be unrelated to interview quality 2. The feedback (vs. no-feedback) group will conduct better interviews during the last two interviews 	Students	21	24	76%	Control (<i>n</i> = 10) Combination feedback (<i>n</i> = 11)
Study II	<ol style="list-style-type: none"> 1. The feedback (vs. no-feedback) group will conduct better interviews 2. The group receiving process feedback will conduct better interviews compared to the group receiving outcome feedback. 3. The group receiving both types of feedback simultaneously will conduct better interviews compared to participants who received only one of the two types of feedback. 	Students	48 ¹	28	79%	Control (<i>n</i> = 12) Outcome feedback (<i>n</i> = 12) Process feedback (<i>n</i> = 12) Combination feedback (<i>n</i> = 12)
Study III	<ol style="list-style-type: none"> 1. The feedback (vs. no-feedback) group will conduct better interviews 2. A reflection task in combination with feedback will boost interview quality 	Students	59	24	59%	Control (<i>n</i> = 19) Feedback ² (<i>n</i> = 19) Feedback plus Reflection (<i>n</i> = 21)
Study IV	<ol style="list-style-type: none"> 1. The feedback (vs. no-feedback) group will conduct better interviews 	Psychologists	40	27	93%	Control (<i>n</i> = 20) Feedback (<i>n</i> = 20)
Study IVa	<ol style="list-style-type: none"> 1. The feedback (vs. no-feedback) group will conduct better interviews with children who witnessed a mock event 	Children	76	70 Months	56%	“The pirate” (<i>n</i> = 38) “Paw patrol” (<i>n</i> = 38)

Note. ¹ Includes 21 participants tested in Study I; ² When feedback is mentioned alone, it refers to a combination of outcome and process feedback

Table 3
Procedures Followed in the Studies.

Tasks in Chronological Order	Studies			
	I	II	III	IV
<i>Before Avatar Simulation</i>				
Informed consent form, confidentiality agreement, and explanation of the study	X	X	X	X
Knowledge test through the Child Sexual Abuse Attitude and Belief Scale ¹	X	X	X	X
Instructions about best practice			X	X
Delivery of background information and scenarios concerning the alleged CSA case	X	X	X	X
Collection of preliminary ideas regarding the presence or absence of sexual abuse	X	X	X	X
<i>Simulated Interview with Avatar</i>				
Interviews with avatars (max 10 minutes each) ²	4	4	8	6
<i>After each Simulated Interview</i>				
Collection of interviewer conclusion regarding the presence or absence of sexual abuse	X	X	X	X
Feedback in the experimental groups (vs. the active control group with no feedback)	Combination	Combination	Combination	Combination
Reflection task		Combination	Combination	Combination
			X	X

Note. ¹(Finnilä et al., 2003); ² In all studies, interviewers received an equal number of abuse and non-abuse scenarios (yes or no) and an equal number of 4- and 6-year-old avatars.

3.3. Simulated Investigative Interviews with Empowering Interviewer Training (EIT®)

Simulated interviews with avatars were performed using different versions of the interview simulation tool. An operator listened to the question asked by the interviewer, categorized it (e.g., as option-posing) and selected the appropriate category in the software. In Study I and II (Version 1), the operator manually launched the appropriate video based on decision-tree type algorithms derived from research on children memory and suggestibility. In Study III and IV (Version 1.12.7), we embedded probabilistic algorithms in the software and the software launched the appropriate videos automatically (See Figure 1).

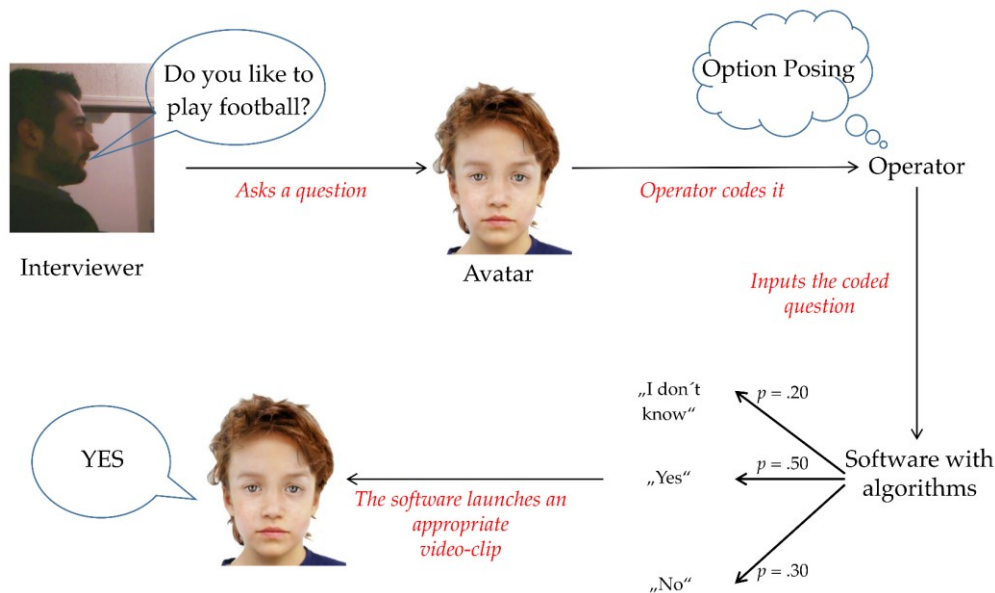


Figure 1. Visual explanation of the procedure used during the interviews with avatars.

3.3.1. Avatars

Avatar images were created by morphing different images of real children that were merged into a new and novel face that cannot be linked to any real child but looks like a real child, as the one presented in Figure 1. Subsequently, we recorded the voices for the different avatars using the text to speech function embedded in the SitePal software (Oddcast, New York, NY) and we animated the avatars through its graphic engine. Since the voices recorded were from adults, as a final step, we modified the adults' voices into children's voices and we edited the videos using the Adobe suite (Adobe CS6). The training tool consisted of 8 different avatars in Study I and Study II, while in the last two studies it consisted of 16 different avatars. In all studies, avatars differed in age (four vs. six years), gender (female vs. male), emotionality (emotional [crying] vs. neutral) and the presence or absence of a sexual abuse incident in the avatars' memories. The avatars presented to the participants always included

all the combinations of age (4 vs 6) and scenario (abuse vs not abuse). In Study IV, it was not possible to balance within interviewer the combination of age and scenario, this was only balanced within the group. In addition, gender has been balanced within the group. The order of the sequence in which the avatars were presented was randomized.



Figure 2. Image of one of the avatars used in the studies.

3.3.2. Scenarios

For each avatar, a scenario of alleged child sexual abuse was created. For each scenario, we also created lists of details that constituted the memories the avatar remembered. In this way, it was possible for us to objectively define whether the interviewers correctly found out what had happened to the avatar. We divided the details present in the avatars' memory into:

Relevant Details. Details present in the avatar memory that help the interviewer to understand what had really happened to the avatar. Based on the scenario these details can disclose an abuse situation or provide an innocent explanation of the alleged situation. These narrative responses were provided only in response to recommended questions and presented in a fixed order in all interviews.

Neutral Details. Details present in the avatar memory but unrelated to the alleged abuse situation. They did not provide the interviewer with useful information in order to understand what happened but can contain information about the avatar's favorite games or other situations involving family members. The narrative responses were provided in response to recommended questions and presented in a fixed order in all interviews.

In addition to relevant and neutral details, the operator could also manually provide additional details, labeled as side details, created to increase the perceived realism of the simulation. These details were unrelated to the alleged abuse situation and were provided to the interviewers only as a response to recommended questions, which included a particular topic that was present in the memory of the avatar. For example, if the interviewer asked "What kind of games do you play with dad?" and a detail about games with dad was present in the avatar's memory, the operator, instead of activating the 'recommended question' algorithm, provided the first detail regarding the game with dad. The operator provided answers about a particular topic until all details of that

category had been revealed or the interviewer changed the topic. Recommended questions related to the allegation or that did not match a side detail category (or if all side details related to the question had been already provided), led to activating the standard algorithm.

As noted above, the predefined details allowed us to evaluate objectively the interview and to recognize wrong details. Wrong details were details not present in the predefined memory but obtained by the interviewer using a not recommended questioning style. Taking as an example Figure 3, since the interviewer asked an option-posing question, and the avatar knows the answer to the question (the correct answer is ‘Yes’), the interviewer had 35% of probability to receive ‘No’ as an answer from the avatar, which would have been a wrong detail. If the interviewer would have asked an option-posing question for which the avatar would not have known the answer, the total probability of receiving a wrong detail would have been 80%.

3.3.3. Algorithms

The algorithms used were based on the available empirical research on children’s memory and suggestibility. In the first two studies, the algorithms were mechanical. This means that the operator kept track of the algorithm and followed the same decision-tree scheme in each case. For example, for 4-year-old avatars, after three recommended questions, the avatar answered with the first neutral detail, after three more recommended questions the avatar answered with the first relevant detail. Starting from the third article, we implemented more complex and probabilistic algorithms. For each question type, a range of possible responses and related probabilities was created. The responses and related probabilities were in both cases based on the available research on children behavior during interviews. Based on the research available we created two different algorithms for the avatars that reflect differences between four and six-year-old (see Figure 3)

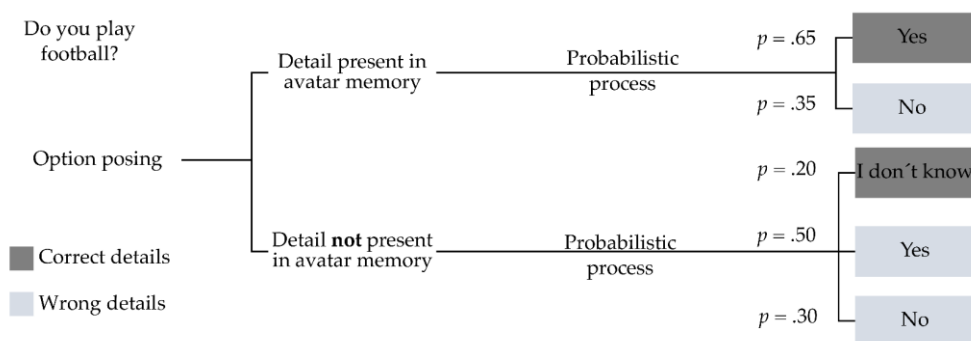


Figure 3. Example of an algorithm for a four-year-old avatar. The wrong details are based on one of the case scenarios.

3.3.4. Coding

The coding of question types was based on the scheme presented in Table 4.

Table 4

Question Coding Divided by Recommended and Not Recommended Questions.

Category	Definition
Recommended Questions	
Facilitators	Open-ended and non-suggestive questions that encourage the child to continue with the previous answer
Invitations	Open-ended and non-suggestive questions for excellence. They are really broad and let the child talk freely
Directive	Open-ended and non-suggestive questions that focus child attention on a previously mentioned detail asking for a focalized explanation (usually WH Questions)
Not Recommended Questions	
Option-posing	Closed-ended questions that focus on unmentioned details (without implying a particular type of response) or on a mentioned detail asking the child to provide a yes/no answer
Specific suggestive	Open/Closed-ended questions that are based on an unmentioned detail and express the expected response from the child
Unspecific suggestive	Open/Closed-ended questions that are not based on an unmentioned detail but express the expected response from the child (e.g., social and peer pressure)
Repetitions	Repetitions of a previous recommended or not recommended question
Too-long/Unclear	Questions that use a language too complicated for the cognitive level of the child and/or formulated in a haphazard manner and/or contains more than one concept at the time
Multiple choice	Questions that provide a predetermined list which the child is requested (explicitly or implicitly) to pick from
Time	Open/Closed-ended questions that require the child to provide or recollect precise time-related information
Fantasy	Open/Closed-ended questions that can activate the child fantasy or move the discussion from the reality to the fantasy level
Feelings	Open/Closed-ended questions that require the child to provide accounts regarding own or others feelings

3.4. Materials and Procedures Specific to the Different Studies

3.4.1. Study I and Study II

In Study I, we tested the effect of feedback and interviews with avatars in comparison to an active control group who interviewed the avatars without any feedback on their performance; in Study II, we tested the effect of different types of feedback and their combination. Participants in Study I and Study II underwent three different types of feedback conditions:

- **Outcome Feedback.** Participants in this group received feedback on the correctness or otherwise of their conclusions regarding what had happened to the avatar. Participants provided the experimenter with both a dichotomous answer regarding the presence or absence of sexual abuse in the scenario and a detailed explanation of what had happened.
- **Process Feedback.** Participants in this group received feedback regarding their questioning style. Four questions they asked during the interview, two recommended questions and two not recommended questions, were chosen for detailed feedback after each interview. A research-based rationale for using or not using the particular type of question was provided.
- **Combination Feedback.** Participants in this group received both the previously described feedback types simultaneously.

In Study I and Study II, we also measured the interviewers' knowledge of research results regarding the investigation of child sexual abuse using the Child Sexual Abuse Attitude and Belief Scale (CSAABS; Finnilä-Tuohimaa, Santtila, Björnberg, Hakala, Niemi, & Sandnabba, 2008). We presented to the interviewers only the 19 factual items out of the 40 items included in the scale. Participants had to reply on 6-point Likert scales how much they agreed or disagreed with the different statements, as for example "Clinical experience is the best guarantee to get to the truth". In this case, the correct answer should be "I disagree".

3.4.2. Study III

In Study III, we tested the effect of a simple reflection task in addition to the combination feedback described previously. The reflection task consisted of asking the participants to reflect back on their interview and provide new examples of question types they used during the interview. Participants were asked to provide examples based on the process feedback they received before, that is, if the interviewer received feedback on a directive question they used, a new example of a directive question was required. However, if no example from the interview was remembered, the interviewer was allowed to create a

brand new one. Starting from Study III, all participants also received information about best practices in child interviews. The paper provided to the participants contained information regarding which types of questions to ask and two questions that aimed to test if the participant had read and understood the guidelines.

3.4.3. Study IV

Study IV included the same interviews with avatars coupled with feedback presented in the previous studies. Two groups, one control group and one group who received combination feedback underwent avatar interviews. After the training, all the participants interviewed two children who had witnessed a mock event. Further information regarding the procedures followed in the study can be found in Table 5.

Mock Events. Two different mock events were staged in the schools of the children (the Pirate Game and the Paw Patrol game). The mock events were based on previous mock events presented in Roberts, Lamb, and Sternberg (1999). The mock events included active involvement of children in order to increase ecological validity (Powell & Thomson, 1997), and actions with high forensic relevance. For example, the events included dressing and undressing moments, innocuous touching between both adult/child and child/child pairs, a secret and the insertion of a cookie in the mouth. Each of the mock events lasted about eight minutes and was videotaped. The structure of the two events was similar with differences regarding the main character and some of the actions.

Rapport Building. The participating psychologists were provided with a protocol based on the NICHD protocol (National Institute of Child Health and Human Development, 2011) to help them in the creation of rapport with the children. In order not to strain the children's attention (Roberts et al., 2004) and in order to avoid the need for more than one interview with the same child (Herschkowitz et al., 2006), we limited the rapport building phase to eight minutes and included directive questions. The maximum total length of the interview including the rapport-building phase was 30 minutes.

Coding. Coding of the question types was the same for both avatar and child interviews. For accuracy of details in the interviews with children, we used a method based on the one proposed by Roberts et al. (1999). Details provided by the children both in narrative and yes/no responses were divided into pieces of information and the veracity of each piece of information was evaluated against the mock-event. Moreover, the question that elicited the details was recorded. Responses that were not of narrative nature were evaluated with the same method because children who agree or disagree with a question posed by the interviewer (answering yes or no) still provide information (see Table 6).

Table 5
Procedure Followed in Study IV after the Training with Avatars.

Interviewer	Child
1 Training with avatars	Mock events staged in the children schools
2 Protocol, based on NICHD, provided	
<i>One week after both training and the mock events</i>	
3 Preliminary information handed	Child walks into the interview room
4 Interview with the child	Child participates in the interview
5 Interviewer goes out of the room	Questionnaire about the interview
6	Toy handed
7 Interviewer reports what they think had happened during the mock event	Child returns to the classroom
8 Researcher asks the interviewer to think back to the training with avatars	
9 Preliminary information handed	Another child who participated in a different mock event enters the room
10 Repetition of points 4-7	Repetition of points 4-7

Table 6
Example of the Procedure Applied for Scoring Details.

Interviewer Question	Child Answer	Correct Details	Wrong Details	Question type used by the interviewer
What did you do in the pirate game?	We dressed up and made a cake	We, dressed up (2)	Made, cake (2)	Directive (child already mentioned the pirate game)
A magician came to visit you, didn't he?	Yes	--	A magician, that was a "he", visited the child (4)	Suggestive (child never mentioned any magician)

4. Results

4.1. Mega-Analysis of the Experiments

The mega-analysis, or pooled analysis of raw data, differs from the meta-analysis, or pooled analyses of published summary statistics, because the researchers recode, based on new common rules, raw data of different experiments, to create a larger dataset. In this way, some of the variability due to different operational definitions is controlled. The problem with mega-analyses is that it is not easy to get access to raw data of experiments. The present thesis provided us with the possibility of conducting a mega-analysis across all the experiments. Moreover, since variables were measured and defined in the same way, definitional issues were not a problem. Importantly, conducting a mega-analysis adds power to the statistical analyses (Costafreda, 2009) due to the higher number of participants and provides more precise estimates. The use of small samples is a common problem in the field due to the logistic difficulties in recruiting large numbers of participants. For all these reasons, we decided to analyze data related to the training with avatars using a mega-analytic approach. The transfer effects into interviews with real children will be presented separately.

4.1.1. Procedures for the Mega-Analysis

The studies shared common ground rules, for example, we used the same coding for question types and details. Therefore, formal re-coding of the question types and details was not needed. However, we had to control for differences in algorithms between the first and the last two studies, and differences between avatars, the eight avatars used in Study I and Study II used a different organization of details and different algorithms compared to the 16 used in the last two studies. In order to control for the differences, we coded the avatars from 1 to 24, including in this way differences both for what concerns the algorithms used and differences in the details.

4.1.2. Samples Used in the Mega-Analysis

The participants ($N = 147$) were divided into two groups, an active control group, that interviewed the avatars without any type of feedback, and an experimental group, that received feedback after each interview. The active control group consisted of 51 participants while the feedback group consisted of 96 participants. In total, participants performed 904 interviews and asked 36401 questions. Table 7 shows the descriptive statistics.

Table 7
Means and Standard Deviations for the Dependent Variables in Control and Feedback Groups.

Dependent Variable	Control			Feedback		
	<i>N</i>	<i>M</i>	<i>SD</i>	<i>N</i>	<i>M</i>	<i>SD</i>
% Recommended questions	51(320)	32.61	17.77	96 (584)	57.65	23.80
Correct details	51 (320)	2.82	2.37	96 (584)	4.86	2.85
Wrong Details	47 (288)	3.11	3.57	85 (496)	0.88	1.75
Correct conclusions	51 (320)	0.07	0.26	96 (584)	0.26	0.44
% Correct details / Total retrieved	47 (288)	52.26	35.59	85 (496)	82.48	28.61
% Correct details / Total available	51 (320)	32.14	26.92	96 (584)	56.52	32.79

Note. In parentheses the number of interviews.

4.1.3. Statistical Analyses

The effects of interviews with avatars coupled with feedback were analyzed using multilevel modeling. Multi-level analyses were conducted using the lme4 (Bates, Mächler, Bolker, & Walker, 2015) and lmerTest package for *p*-values and model fitting (Kuznetsova, Brockhoff & Christensen, 2016). Linear mixed-effects models, fitted using a Maximum Likelihood procedure and with Satterthwaite approximation for degrees of freedom, was used for continuous variables (proportion of recommended questions and correct details), while a binomial, logit link, generalized linear mixed model was used for dichotomous variables (correct conclusions).

We analyzed four dependent variables: i) *Percentage of recommended questions* refers to the proportion of recommended out of all questions (recommended and not recommended) asked by the interviewer. ii) *Percentage of correct details / details retrieved* refer to the proportion of correct out of all details (correct and wrong) elicited by the interviewers. iii) *Percentage of correct details / details available* refers to the proportion of correct details elicited by the interviewer out of the total number of correct details available. iv) *Correct conclusions* refer to the proportion of correct conclusions reached by the interviewers. There is no homogeneous consensus

in the literature on which is the best procedure to define models in multilevel analyses (see for example Barr, Levy, Scheepers, Tily, 2013; Bates, Kliegl, Vasishth, Baayen, & Baayen, 2015). We decided to compute the model with the maximal random effect structure and then test which model better predicted the data through Aikake Information criterion (AIC). We used restricted maximum likelihood (REML) and AIC as model selection criteria for the model fitting of the random part of the formula. The final formula included a random intercept for interviewers and a random intercept for the avatars. We used maximum likelihood (ML) and AIC as model selection criteria for the fixed part. The model including the interaction between group and number of the interview was the one supported. The final formula for all the analyses was:

$$DV \sim \text{Group} * \text{Time} + (1|\text{ID}) + (1|\text{Avatar})$$

Where DV is the dependent variable, Group refers to the experimental vs. control group allocation, Time refers to the number of interviews (repeated measure), ID refers to the interviewers and Avatar refers to the avatar interviewed (1 to 24).

4.2. Results of Interviews with Avatars Coupled with Feedback

Providing feedback and interviews with avatars improved the quality of investigative interviews on all the variables investigated, compared to an active control group who did not receive feedback. Detailed results of the analyses are presented in Table 8. For all the figures presented the first four interviews are based on the total sample of 147 interviewers (in all the studies the participants conducted at least four interviews), interviews 5 and 6 are based on Study III and IV ($n = 99$) while interviews 7 and 8 are based on only Study III ($n = 59$).

Proportion of Recommended Questions. As expected, interviewing avatars coupled with feedback increased the use of recommended questions over time compared to the control group (Figure 4, Panel A).

Proportion of Correct details out of the Total Details Retrieved. As expected, interviewing avatars coupled with feedback increased the proportion of correct details out of all details retrieved by the interviewers in comparison with the control group (Figure 4, Panel B).

Proportion of Correct Details Retrieved out of the Total Available. As expected, interviewing avatars coupled with feedback improved the degree of information (correct details) out of the total available elicited by the interviewers compared to the control group (Figure 4-C).

Correct Conclusions. As expected, interviewing avatars coupled with feedback improved the proportion of correct conclusions reached by the interviewers in comparison with the control group (Figure 4-D).

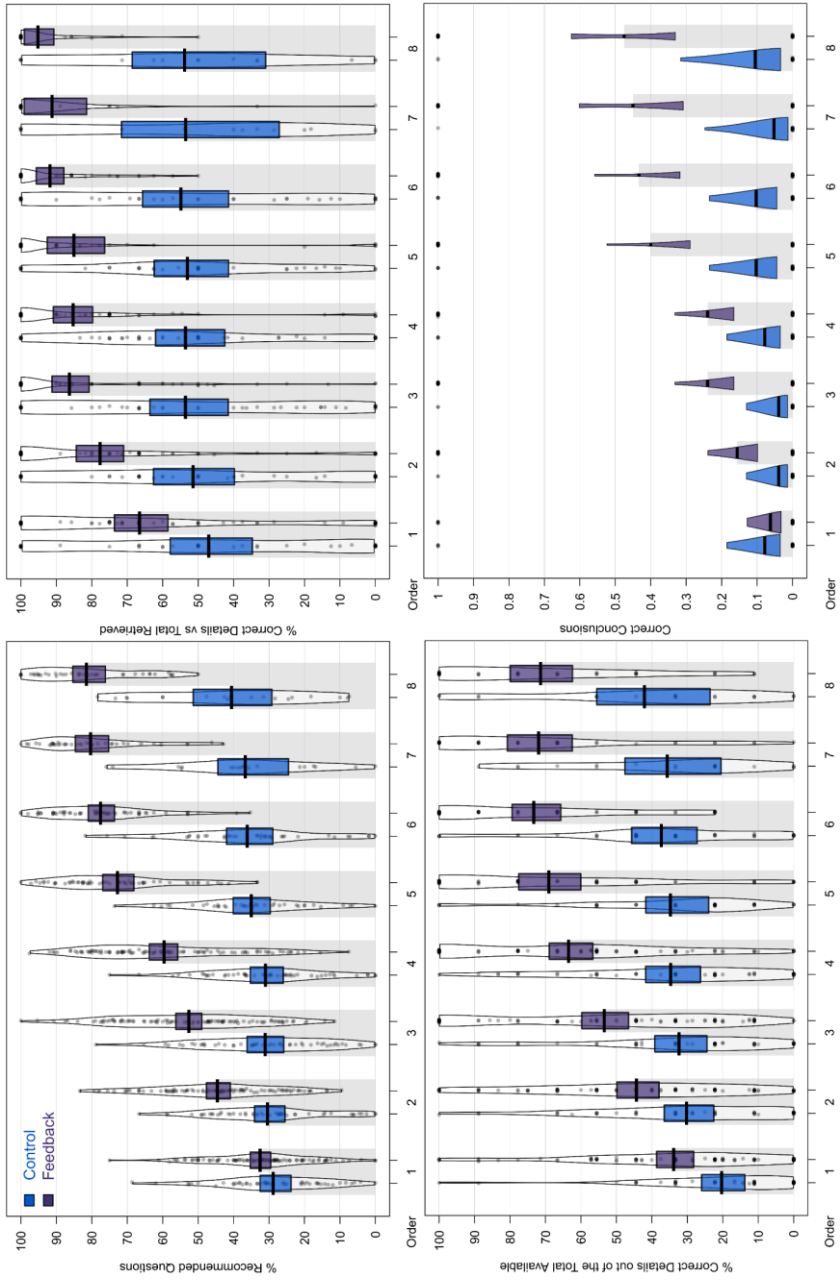


Figure 4. RDI (Raw data, Descriptive and Inferential statistic) plots. Panel A: Proportion of recommended questions asked by the interviewers, Panel B: Proportion of correct details retrieved compared to all details retrieved, Panel C: Proportion of correct details retrieved out of the total available, and Panel D: Proportion of correct conclusions during interviews with avatars (y-axis). The x-axis shows the order of the interviews. A Point represents a single interview within each group. The black line represents the average value, highlighted respectively in dark blue and dark violet is the 95% CI and in whites the smoothed density. Panel D shows only the 95% CI. The first 4 interviews are based on 147 interviewers, 51 in the control group and 96 in the feedback group; interviews 5 and 6 are based on 99 interviewers, 39 in the control group and 60 in the feedback group while interviews 7 and 8 are based on 59 interviewers, 19 in the control group and 40 in the feedback group.

Table 8
Impact of Interviews with Avatars and Feedback on Dependent Variables.

<i>Predictors</i>	<i>Dependent Variables</i>									
	Recommended Questions			% Correct Details / Retrieved			% Correct details / Available			
	<i>B</i>	<i>CI</i>	<i>p</i>	<i>B</i>	<i>CI</i>	<i>p</i>	<i>B</i>	<i>CI</i>	<i>p</i>	
Fixed Parts										
(Intercept)	27.83	23.27 – 32.39	<.001	48.91	40.59 – 57.22	<.001	22.84	14.73 – 30.96	<.001	
Group	1.91	-3.63 – 7.46	.499	16.60	6.57 – 26.63	.001	7.14	-1.37 – 15.65	.101	
Time	1.01	0.40 – 1.63	.001	1.12	-0.33 – 2.57	.129	2.44	1.28 – 3.61	<.001	
Group*Time	6.00	5.24 – 6.76	<.001	3.91	2.09 – 5.72	<.001	4.95	3.48 – 6.36	<.001	
Random Parts										
σ^2	118.48			567.03			424.75			
Interviewers	179.32			358.35			309.74			
Avatar	4.40			28.15			116.26			
Observations	904			784			904			

Table 9
Impact of Interviews with Avatars and Feedback on Conclusions.

<i>Predictors</i>	<i>Dependent Variables</i>		
		Conclusions	
	<i>Odds Ratio</i>	<i>CI</i>	<i>p</i>
Fixed Parts			
(Intercept)	0.03	0.01 - 0.09	<.001
Group	1.55	0.48 - 4.96	.463
Time	1.17	0.94 - 1.47	.161
Group*Time	1.34	1.05 - 1.72	.020
Random Parts			
Interviewers	0.41		
Avatars	0.91		
Observations	904		

4.3. Effect of Different Types of Feedback in Interviews with Avatars

Here, we used the same formula as for the general analyses of the avatar training to analyze the effects of different types of feedback. From the total sample of 147 participants, the participants who received any type of feedback ($n = 96$) have been divided in outcome feedback ($n = 12$), process feedback ($n = 12$) and the combination of both types of feedback ($n = 72$) groups. Since reflection did not show statistically significant effects, participants who received feedback and reflection task have been included. The analyses are limited to the first four interviews because interviewers in the separate feedback condition conducted only four interviews. Table 10 shows the results of the pairwise comparisons for each type of feedback and the control group. Appendix A shows the raw means and standard deviations for each dependent variable divided by feedback type. As expected, a combination of both feedback types simultaneously had a stronger effect on the proportion of recommended questions and correct conclusions drawn, compared to receiving feedback separately (Figure 5). Due to the differences in algorithms and numbers of details, the analyses of the details are heavily influenced by the differences at the baseline between outcome and process feedback (Study II) and control and combination feedback (all four studies).

Table 10

Pairwise Comparisons of the Effects of the Different Types of Feedback

	<i>Contrasts</i>	<i>Estimate</i>	<i>CI</i>	<i>SE</i>	<i>Df</i>	<i>t</i>	<i>p</i>
Recommended Questions	OF vs C	-0.73	-9.11- 7.65	4.24	151.19	-0.17	= .863
	PF vs C	12.77	4.39-21.16	4.24	151.27	3.01	= .003
	2F vs C	20.73	16.01-25.45	2.39	145.81	8.68	< .001
	PF vs O	13.50	2.98-24.02	5.32	145.33	2.54	= .012
	2F vs OF	21.46	13.29-29.64	4.14	151.69	5.19	< .001
	2F vs PF	7.96	-0.22-16.14	4.14	151.73	1.92	= .056
Correct Details vs Details Retrieved	OF vs C	24.83	9.89-39.76	7.55	134.48	3.29	= .001
	PF vs C	39.03	24.10-53.97	7.55	134.53	5.17	< .001
	2F vs C	25.36	16.57-34.15	4.44	130.55	5.71	< .001
	PF vs OF	14.21	-4.25-32.67	9.33	129.96	1.52	= .130
	2F vs OF	0.53	-14.11-15.17	7.40	134.07	0.072	= .943
	2F vs PF	-13.68	-28.32- 0.97	7.40	134.07	-1.85	= .067
Correct Details vs Details Available	OF vs C	14.32	0.90-27.74	6.79	158.04	2.11	= .036
	PF vs C	27.04	13.61-40.46	6.80	158.35	3.98	= .001
	2F vs C	18.77	11.59-25.96	3.63	145.82	5.16	< .001
	PF vs OF	12.72	-3.26-28.70	8.09	144.66	1.57	= .118
	2F vs OF	4.45	-8.79-17.70	6.70	159.46	0.664	= .507
	2F vs PF	-8.26	-21.51- 4.99	6.70	159.65	-1.23	= .220
	<i>Contrasts</i>	<i>OR</i>	<i>CI</i>	<i>SE</i>	<i>Df</i>	<i>z</i>	<i>p</i>
Correct Conclusions	OF vs C	1.58	0.51- 4.92	0.91	NA	0.79	= .427
	PF vs C	2.66	0.93- 7.60	1.43	NA	1.82	= .068
	2F vs C	3.52	1.70- 7.30	1.31	NA	3.39	= .000
	PF vs OF	1.68	0.51- 5.53	1.02	NA	0.85	= .395
	2F vs OF	2.23	0.77- 6.42	1.20	NA	1.48	= .139
	2F vs PF	1.33	0.50- 3.49	0.65	NA	0.57	= .566

Note. C = Control group, OF = Outcome feedback, PF = Process feedback, 2F = Combination of both feedback. No correction for p-values applied. Calculation refers to least square means.

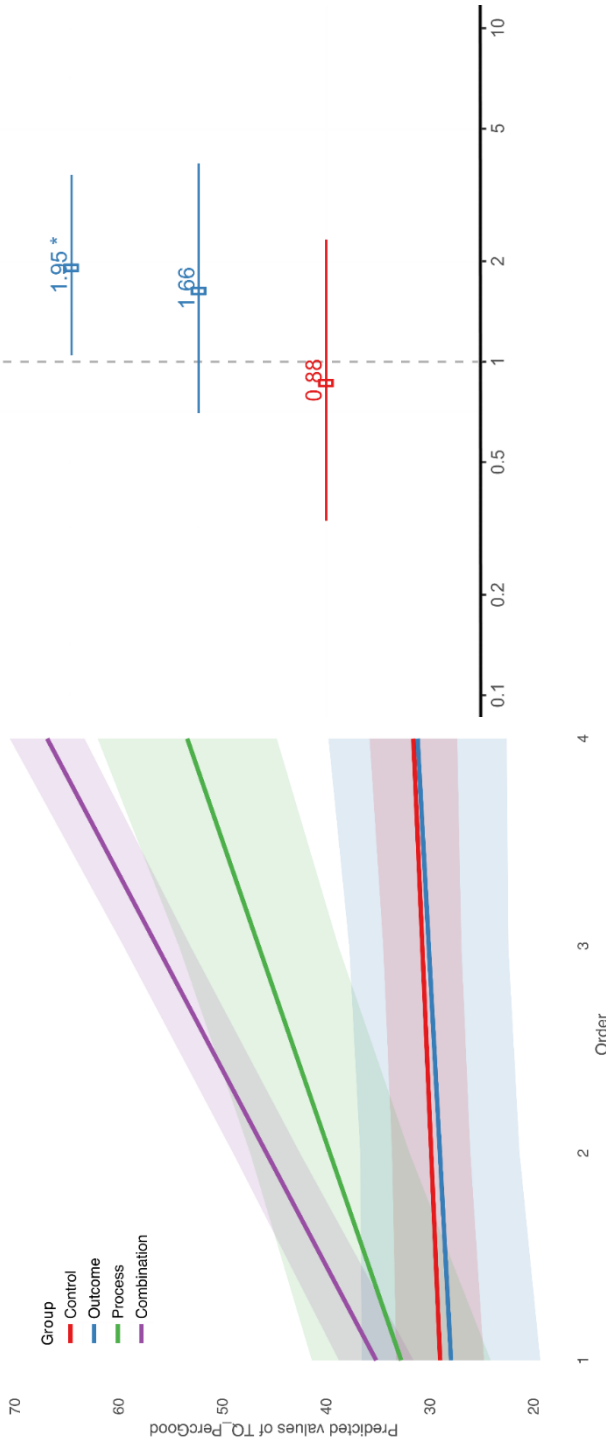


Figure 5. On the left the predicted values (y-axis) and 95% CI for the proportion of recommended questions divided by group. On the x-axis the number of interviews. On the right, Odds ratio in comparison with the control group and 95% CI for the probability of achieving a correct conclusion. The terms presented refer to the interaction between group and number of the interview. The first term starting from the upper part refers to Combination feedback*Time, the second term refers to Process feedback*Time and the last term to Outcome feedback * Time. The red line refers to negative probability. The blue line refers to a positive probability. Control group refers to 51 participants, outcome feedback to 12 participants, process feedback to 12 participants and combination of both types of feedback to 72 participants. * p = .035.

4.4. Transfer Effect in Interviews with Children

A one-way ANOVA, used for the assessment of baseline demographic and performance differences between groups, did not show differences between children who participated in mock event 1 or mock event 2 concerning age ($F[1, 74] = 0.55, p = .460$) and a Pearson Chi-Squared test showed no difference for gender ($\chi^2[1] = .57, p = .450$). For the proportion of recommended questions and the proportion of correct details but not for the proportion of correct conclusions, the means for the total sample ($n = 76$) were in favor of the feedback group. When only including cases in which the children talked about the event, means for all outcome variables were in the expected direction (see descriptive statistics in Table 11). Since the proportion of correct details and correct conclusions are influenced by the cases in which the child did not talk about the event, the main analyses are based on the whole sample for the proportion of recommended questions and restricted to the cases in which the child talked about the event for proportion of correct details and correct conclusions. In this group, we included both cases in which the child talked about the event spontaneously or through narrative details but also children who provided information only as correct rejection to suggestive questions. There was no difference concerning children who spontaneously talked about the event between groups (six children in each group). We ran a supplementary analysis too. For the proportion of recommended questions, we ran analyses including only the cases in which the child talked about the event (Control, $n = 23$; Feedback, $n = 12$) to test if the disclosure of the child had an effect on the questioning style.

Proportion of Recommended Questions. The feedback group used more recommended questions than the control group in the whole sample (Table 12). A visual representation can be found in Figure 6.

Proportion of Correct Details out of the Total Details Retrieved. The feedback group elicited a statistically significantly higher proportion of correct details compared to the control group when analyzing the cases in which the child talked about the event (Table 12).

Correct Conclusions. The feedback group reached a higher proportion of correct conclusions compared to the control group, but the difference was not statistically significant. Calculation of the odds ratios produced an OR = 3.66 [0.23-56.92], $p = .355$.

Supplementary Analysis. We tested if the disclosure of the child had an effect on the proportion of recommended questions used by the interviewer. Analyses including only cases where the child talked about the event ($n = 35$) showed that the group receiving feedback still used a statistically significantly higher proportion of recommended questions $E = 31.59, t = 2.16, SE = 14.60, p = .039$.

Table 11
Descriptive Statistics Divided by Group.

Dependent Variable	Control			Feedback		
	<i>N</i>	<i>M</i>	<i>SD</i>	<i>N</i>	<i>M</i>	<i>SD</i>
% Recommended questions	36	25.57	14.62	40	40.44	22.78
Correct details ¹	23	12.13	23.91	12	23.50	22.03
Wrong Details	23	25.30	55.49	12	21.58	51.18
Correct conclusions	23	0.30	0.47	12	0.50	0.52
% Correct / wrong details	23	21.28	32.57	12	47.45	43.01

Note. ¹ Correct details elicited from the first standard question and as correct rejection have been excluded

Table 12
Result of the multilevel analyses on the continuous dependent variables.

<i>Predictors</i>	<i>Dependent Variables</i>					
	Recommended Questions			Correct Details / Retrieved		
	<i>B</i>	<i>CI</i>	<i>p</i>	<i>B</i>	<i>CI</i>	<i>p</i>
Fixed Parts						
(Intercept)	25.57	17.94 – 33.20	<.001	22.75	5.07 – 40.43	.018
Group	14.87	4.36 – 25.38	.009	31.46	2.84 – 60.08	.041
Random Parts						
σ^2	184.54			201.24		
Interviewers	180.19			1139.60		
N _{Interviewers}	38			26		
Observations	76			34		

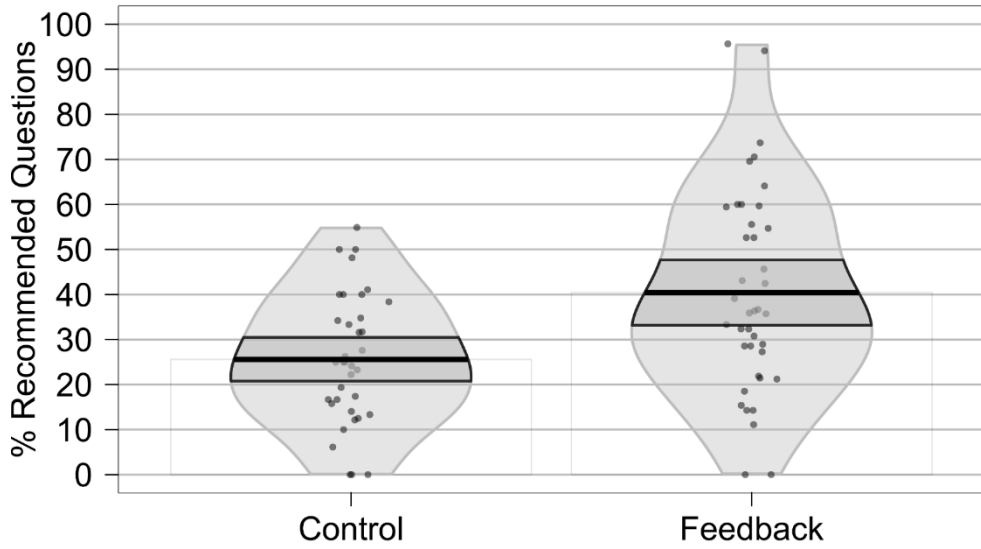


Figure 6. RDI (Raw data, Descriptive and Inferential statistic) plot of the proportion of recommended questions, out of all questions, asked by the interviewers in interviews with children and divided by group (Control, $n = 36$; Feedback $n = 40$). A Point represents a single interview within each group. The black line represents the average value, highlighted in dark grey is the 95% CI and in light grey the smoothed density.

5. Discussion

The overall aim of the present thesis was to develop and test a new method to conduct training in interviews with children in alleged CSA, integrating serious gaming and avatars with response algorithms and predefined memories. The studies included in the current thesis contained a combined sample of 147 interviewers and more than 900 interviews. The results showed how interviews with avatars paired with feedback improved the quality of interviews in the simulation and that this effect transferred into interviews with children who witnessed a mock event for two out of the three variables evaluated. Caution is needed since studies that test the effects of this training on interviews in actual alleged CSA cases are necessary. However, the meta-analysis of the data from the four studies provides robust evidence for the improvement of the quality of simulated investigative interviews among students and psychologists, at least in the short term.

5.1. Interpretation of Results

5.1.1. Effects of Knowledge about Evidence-Based Best Practice

One of the aims of Study I was to test if knowledge regarding best practices would have an effect on interview quality. As previous research has highlighted (Sternberg et al., 2001), knowledge regarding evidence-based principles in CSA investigation was not related to interviewing quality. Moreover, even if the formal statistical evaluation was not computed, there was no difference between the performance of students in Study III and psychologists in Study IV at baseline, reinforcing the findings of Study I.

5.1.2. Effects of a Simple Reflection Task

In Study III, a simple reflection task did not enhance training effects. Even if it is not possible to state with certainty, it may be that this finding is due to the type of reflection tested. Asking participants to recall additional examples for question categories (or make up new examples if they did not remember any), can activate more retrieval than reflection processes. The inability to stimulate a critical examination of one's own performance and strategies, which is considered an important part of systematic reflection (Ellis, Carette, Anseel, & Lievens, 2014), is a possible explanation. Further research on different reflection tasks than the one employed in Study III is necessary.

5.1.3. Effects of Different Types of Feedback

The combination of outcome and process feedback improved the quality of investigative interviews compared to the administration of these two types of feedback separately and compared to the control group. Studies that employ outcome and process feedback separately are common in research. However,

to our best knowledge, this is the first study to test the effect of a combination of outcome and process feedback in comparison to these two types of feedback provided separately in the context of investigative interviewing. The provision of outcome and process feedback simultaneously provided the participant with both a reason to change (outcome feedback) and an example of how to improve (process feedback). In some cases, only one not recommended question led the participants to a wrong conclusion. A clear example is the tendency after a child's disclosure such as "He touched me." to ask a closed question such as "Was it dad?" Even if the interview was conducted in an exemplary way otherwise, the one not recommended question could lead to a wrong conclusion. In this case, providing only process feedback would probably not have had the same effect as providing both feedback types together. This hypothesis is supported by the results of Study II. Even if process feedback was more effective than outcome feedback when considered in isolation, the combination of outcome and process feedback enhanced interview quality even more, at least for the proportion of recommended questions and conclusions, compared to providing the two feedback types separately. However, due to the low number of participants in the outcome and process feedback conditions, more research is needed in order to replicate these results in larger samples.

5.1.4. Interviews with Avatars Paired with Feedback Improve Interview Quality

Among the best practice guidelines to follow, it is stated that an interviewer should use recommended questions and avoid using not recommended questions. However, these recommendations should be analyzed together. Consider an interviewer using forty recommended and forty not recommended questions (the proportion of recommended questions is 50%), whereas another interviewer uses only twenty recommended questions, but also uses only five not recommended questions (the proportion of recommended question is 80%) In this case, the proportion can provide us with a better view of the quality of the interview. The same reasoning is true for correct details compared to all details retrieved. Moreover, interviewers are usually required to provide reports about their conclusions. For this reason, we also evaluated their ability to provide conclusions. In this training-setup, it was possible to evaluate if the interviewer reached a completely correct conclusion. Training with avatars improved interview quality on different dimensions, including a higher proportion of correct conclusions. The way we decided to operationalize interview quality provides a more comprehensive view of how participants improved compared to the use of absolute numbers and compared to studies that did not evaluate the veracity or otherwise of details elicited and conclusions reached. The results of the mega-analysis provide evidence for the use of interviews with avatars together with a combination of process and outcome feedback. Participants in the feedback group showed a better use of recommended and not recommended questions and a better ability to elicit

correct compared to wrong details compared to the control group. These are all important issues. As highlighted by Benson and colleagues (2015), prosecutors require shorter and more focused interviews, in order not to jeopardize the reliability of children accounts. Participants in our studies improved also in terms of the total number of correct details elicited out of the total available. This means that they were able to minimize wrong and harmful information while eliciting all the possible information from the avatars.

Participants were also able to draw a higher number of correct conclusions. It is important to highlight that our evaluation of correct conclusion was strict. One wrong or missing detail was enough for coding the interviewer's conclusion as wrong. This type of evaluation is not possible in a real context, which adds strength to our analysis of the effects of the feedback we provided to the participants. Interestingly, participants have not been trained in how to reach a correct conclusion and which details to elicit from the avatars (e.g., hypothesis testing). A possible explanation for this *side effect* comes from the results of Study II. Probably, and as argued by Hattie and Timperley (2007), the provision of outcome feedback and not only process feedback improved the ability of participants of avoiding faulty interpretations.

5.1.5. Training Effects with Avatars Transfer into Interviews with Children

Study IV provided more support for a positive effect of the interview simulation tool. First, we showed how psychologists improved using our training and secondly, we showed transfer effect into interviews with children who witnessed a mock event. The increased proportion of recommended questions, correct details and correct conclusions in real interviews with children suggests that participants in the feedback group conducted interviews of higher quality compared to participants in the control group. This is an encouraging result because psychologists participated in a one-hour long training. However, we did not find statistically significant effects for drawing correct conclusions. It is possible that the low number of children who talked about the event led to a floor effect. However, participants in the feedback group were able to reach a higher number of correct conclusions compared to the control group. This is an important first step in the validation of this method.

5.2. Limitations

Together with many strengths and promising prospects come some limitations. We are aware that interviewing children is much more than asking the right questions. Other issues, such as rapport building, motivating the child in order to increase the likelihood of disclosure, the use of a supportive style and hypothesis testing are important features of investigative interviews. However, they can be also easily included in our simulation setup in the future. One clear result of the studies included in the current thesis is that there is individual variation among participants in how they respond to training. The RDI plots

clearly show how some individual improvement was present in the control group and some individuals in the feedback group failed to improve. A possible explanation for this is that interviewers in the control group may have been encouraged to use a more recommended questioning style due to the longer and more rewarding narratives received from the avatars when using that questioning style.

Another limitation is the relatively low level of relevant experience among the psychologists in Study IV. The literature shows that more experienced interviewers are less prone to change their interview style compared to less experienced interviewers. However, Benson (2015) showed no differences between less and more experienced interviewers at the end of the training. A possible explanation of the results of Benson (2015) is that due to the autonomy interviewers had during the course (i.e., they could do it at their own pace) and due to the more interactive procedures employed, also more experienced interviewers were more willing to engage and improve their performance. A possible solution to this problem is the creation of tailored training based on the participants' knowledge and capabilities. In any case, it will be important to test the interview simulation tool with more experienced interviewers.

Moreover, the interviewers did not receive any training in creating a rapport with the child, in using a supportive interviewing style, nor in how to behave when a child is not disclosing, which may limit generalizability. The transfer effect was evaluated in interviews with children concerning a neutral event. Different researchers have highlighted how interviewing children about innocuous events does not necessarily mimic the dynamics of real interviews (Lamb, 2016; Powell & Wright, 2008). For example, they may lack the level of reticence and reluctance present in interviews with allegedly abused children. However, mock events allowed us to monitor the veracity of the details provided by the child (Lamb, Herschkowitz, Orbach, & Esplin, 2011). Also, many children in our study did not talk about the mock event during the interview. We do not know with certainty the reason for this. However, two critical points may be raised. Psychologists know the importance of rapport building well (e.g., Herschkowitz, Lamb, Katz, & Malloy, 2013), but this is perhaps less known in inexperienced interviewers as our psychologists were, and the induced secrecy during the mock event might have had a negative influence on the likelihood of a child disclosing (Lyon et al., 2014). For this reason, it will be important to conduct more studies in order to collect more data on children provision of details and disclosures, in particular in interviews of alleged CSA.

Two other critical challenges cannot be addressed by the current setup. We do not know the long-term effect of the training. However, to date, the only way to maintain training effects is to provide practice with feedback continuously. The interview simulation tool allows interviewers to have booster session with the avatars between one interview and the other in the field, and the related costs are lower than in traditional training with actors. However, how many

booster sessions with feedback are necessary to achieve long-term effects is another issue not yet addressed in this thesis. In conclusion, professional CSA investigators should be trained within the same set-up while testing their performance.

Some final methodological limitations need to also be mentioned. Even if the feedback procedure was pre-determined, we did not investigate interrater-reliability for the feedback provision. Also, even if the algorithms presented in the thesis are supposedly providing realistic patterns of response, some improvements can be made. In the simulations used, there was no link between the substantive content of the questions and the content of the narrative details provided in avatar responses.

5.3. Implications and Future Directions

What the interview simulation tool adds compared to previous studies and training program is a new training method, which aims to solve some of the practical problems related to traditional training. The use of avatars and, in the future, automatic speech recognition instead of actors can diminish the cost-burden. Moreover, this type of training can be provided via internet, which, as highlighted by Benson (2015) seems to be an effective new way of delivering training. In addition, the response algorithms are supposed to be more realistic than responses provided by actors. The algorithms utilized in the last two studies are too complex to be reproduced by human beings, showing a possible enhanced similarity with field interviews. As highlighted by different scholars (Blume et al., 2010), the similarity between training and task is an important factor for maximizing transfer effect. Another important point of this training is the time spent by the participant during training. Between one and one hour and a half, without any face to face lectures or reading of theoretical materials or guides beforehand, nevertheless, showed strong effects. However, other important steps will be to test if this training tool will help interviewers to elicit coherent narratives from children and if it will adapt to the actual needs of prosecutions. For example, one of the prosecution requirements is to conduct shorter and more focused interviews (Benson & Powell, 2015).

The Cognitive Load Theory. This thesis also addresses some theoretical issues related to previous failures. In a cognitively demanding context such as interviews of children, the CLT is an interesting, yet to be tested in this context, evidence-based start for the planning of training programs (see for an example Mugford, Corey, & Bennell, 2013). An example is the results of Benson and colleagues (2015). For the first time, they have been able to show long-term effects of training. One of the possible explanation the authors argue is that compared to previous training programs, the information was provided over a longer time span. Research shows how this method favorably improve chances of learning (Schmidt & Bjork, 1992), moreover, it requires less cognitive resources in the short-term of the interviewer. The same principle can be applied to the interview simulation tool presented in this thesis. The interviewers, except for some information about recommended and not

recommended questions, learned the optimal questioning style, and its effect due to the outcome feedback, during the simulated interviews.

The Issue of the Reluctant Child. Another important issue that has not been yet included in this training evaluation is the problem of the reluctant child. Even if an interview is conducted according to recommendations and the child has a strong memory trace of the event, the disclosure is not always guaranteed (Pipe, Lamb, Orbach, & Cederborg, 2012). The reluctance to disclose an abusive event can be influenced by several factors. For example, reluctance or delayed disclosure is more common when perpetrators are familiar to the child (Lamb, 2016; London, Bruck, Wright, & Ceci, 2008). However, other factors such as induced secrecy (Lyon et al., 2014), embarrassment to talk with an unfamiliar interviewer about sensitive issues, feelings of responsibility and guilt (for a review Pipe et al., 2012) can also decrease the likelihood of disclosure. How to motivate reluctant children to disclose abuse is still an open issue in the field. To solve this problem, different solutions have been proposed. One approach developed to overcome this problem is to elicit a promise to tell the truth from the child. This method has been shown to increase the probability of children's disclosure of transgressions (e.g., Talwar, Lee, Bala, & Lindsay, 2002). However, the promise to tell the truth, has also shown to be less effective with children younger than 5 years old (Heyman, Fu, Lin, Qian, & Lee, 2015). In order to fix this problem, Lyon and colleagues suggested (2014) the putative confession. The putative confession, which consists of telling the child that the interviewer has been informed already of what happened without adding suggestive details increased true disclosures about a transgression by 20%. Subsequent studies (Quas, Stolzenberg, & Lyon, 2018) showed, regardless of the age of the child, how the putative confession, compared to eliciting a promise from the child to tell the truth, improved the likelihood of disclosing a minor transgression. Another technique is the emphasis on the rapport-building part of the interview. Increasing rapport building, also through the use of drawing (Katz, Barnett, & Hershkowitz, 2014), which is supposed to rely on emotional factors that can enhance cooperativeness, has been shown to be effective in increasing the number of forensic relevant details elicited without altering the question types used (Hershkowitz et al., 2013). In a recent study, Stolzenberg and colleagues (Stolzenberg, McWilliams, & Lyon, 2017) tested the effect of the use of negatively valenced yes/no questions, for example, "Did something bad happen to the doll?". In case of positive answers, this was followed by a recall question (e.g., tell me everything that happened). The use of yes/no questions increased disclosure of a transgression, but, on the other hand, also led 5% of the children to provide a false report. Moreover, 25% of the children did not disclose the transgression. In conclusion, even if with some positive results, more research in this field is needed. The possibility of including such effects to the simulation has not yet been explored.

In conclusion, if the interview simulation tool will be proved efficient also in field interviews, which is what we expect, it can provide a revolution in the field. Avatars with response algorithm within a pedagogical framework have

endless possibilities of application. The first and most important application is to provide this type of training to everyone working among potentially abused children, such as teachers, parents, and physicians. The neutrality of the first interview with the child is of paramount importance. Other possible applications are human intelligence gathering or teaching police officers to master the strategic use of evidence technique (Hartwig, Granhag, Strömwall, & Kronkvist, 2006). Outside the forensic environment, possible applications are clinical interviews or job interviews.

The already programmed next step is a test with the Estonian police and with the Japanese police. During these experiments, we will be able to test for the longevity of training effects as well. Moreover, we have finalized the first web-based version of the software. The web-based version allows us to maximize the positive characteristics of the interview simulation tool, such flexibility, and time-and-costs savings'. At the same time, we are working to add automatic speech recognition.

5.4. Conclusions

The interview simulation tool presented in this thesis showed how it is possible to improve the quality of simulated investigative interviews using avatars and feedback. Moreover, we also showed that these results can be achieved in a relatively short time. Importantly, the simulation fixes some important practical problems, providing the interviewers with a more realistic training environment and with a more feasible solution. Serious gaming seems to be an interesting solution to pursue. Naturally, the simulation tool has not solved all the problems related to training in investigative interviews of alleged abused children. More studies are needed in order to even more clearly define the efficacy of this training program and also in order to address the other current challenges related to investigative interviews in alleged CSA cases.

References

- Ahern, E. C., Andrews, S. J., Stolzenberg, S. N., & Lyon, T. D. (2015). The productivity of Wh-Prompts in child forensic interviews. *Journal of Interpersonal Violence, 1*(9). <https://doi.org/10.1177/0886260515621084>
- Andrews, S. J., Ahern, E. C., Stolzenberg, S. N., & Lyon, T. D. (2016). The productivity of Wh- prompts when children testify. *Applied Cognitive Psychology, 30*(3), 341-349. <https://doi.org/10.1002/acp.3204>
- Anseel, F., Beatty, A. S., Shen, W., Lievens, F., & Sackett, P. R. (2015). How are we doing after 30 years? A meta-analytic review of the antecedents and outcomes of feedback-seeking behavior. *Journal of Management, 41*(1), 318-348. <https://doi.org/10.1177/0149206313484521>
- Astwood, R. S., van Buskirk, W. L., Cornejo, J. M., & Dalton, J. (2008). The impact of different feedback types on decision-making in simulation based training environments. *Human Factors and Ergonomics Society Annual Meeting Proceedings, 52*(26), 2062-2066. <https://doi.org/10.1177/154193120805202606>
- Ayres, P., & Paas, F. (2009). Interdisciplinary perspectives inspiring a new generation of cognitive load research. *Educational Psychology Review, 21*(1), 1-9. <https://doi.org/10.1007/s10648-008-9090-7>
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*, 255-278. <https://doi.org/10.1016/j.jml.2012.11.001>
- Bates, D., Kliegl, R., Vasishth, S., Baayen, H., & Baayen, R. H. (2015). Parsimonious mixed models. *arXiv preprint arXiv:1506.04967*. Retrieved from <https://arxiv.org/pdf/1506.04967.pdf>
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1). <https://doi.org/10.18637/jss.v067.i01>
- Bellotti, F., Kapralos, B., Lee, K., Moreno-Ger, P., & Berta, R. (2013). Assessment in and of serious games: An overview. *Advances in Human-Computer Interaction, 2013*, 1-11. <https://doi.org/10.1155/2013/136864>
- Benson, M. S., & Powell, M. B. (2015). Evaluation of a comprehensive interactive training system for investigative interviewers of children. *Psychology, Public Policy, and Law, 21*(3), 309-322. <https://doi.org/10.1037/law0000052>
- Blume, B. D., Ford, J. K., Baldwin, T. T., & Huang, J. L. (2010). Transfer of training: a meta-analytic review. *Journal of Management, 36*(4), 1065-1105. <https://doi.org/10.1177/0149206309352880>
- Boud, D. (2015). Feedback: ensuring that it leads to enhanced learning. *THE CLINICAL TEACHER, 12*, 3-7. <https://doi.org/10.1111/tct.12345>
- Boud, D., & Molloy, E. (2013). Rethinking models of feedback for learning: The challenge of design. *Assessment and Evaluation in Higher Education, 38*(6), 698-712. <https://doi.org/10.1080/02602938.2012.691462>
- Boyle, E., Connolly, T. M., & Hainey, T. (2011). The role of psychology in understanding the impact of computer games. *Entertainment Computing, 2*(2), 69-74. <https://doi.org/10.1016/j.entcom.2010.12.002>
- Brubacher, S. P., Powell, M., Skouteris, H., & Guadagno, B. (2015). The effects of e-simulation interview training on teachers' use of open-ended questions. *Child Abuse & Neglect, 43*, 95-103. <https://doi.org/10.1016/j.chiabu.2015.02.004>
- Bruck, M., & Ceci, S. J. (1999). The suggestibility of children's memory. *Annual review of psychology, 50*(1), 419-439.

- <https://doi.org/10.1146/annurev.psych.50.1.419>
- Buff, W. L., & Campbell, G. E. (2002). What to do or what not to do?: Identifying the content of effective feedback. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 46*, 2074–2078. <https://doi.org/10.1177/154193120204602517>
- Cederborg, A.-C., Alm, C., Lima da Silva Nises, D., & Lamb, M. E. (2013). Investigative interviewing of alleged child abuse victims: an evaluation of a new training programme for investigative interviewers. *Police Practice and Research, 14*(3), 242–254. <https://doi.org/10.1080/15614263.2012.712292>
- Cederborg, A.-C., Orbach, Y., Sternberg, K. J., & Lamb, M. E. (2000). Investigative interviews of child witnesses in Sweden. *Child Abuse & Neglect, 24*(10), 1355–1361. [https://doi.org/10.1016/S0145-2134\(00\)00183-6](https://doi.org/10.1016/S0145-2134(00)00183-6)
- Chae, Y., & Ceci, S. J. (2005). Individual differences in children's recall and suggestibility: The effect of intelligence, temperament, and self-perceptions. *Applied Cognitive Psychology, 19*(4), 383–407. <https://doi.org/10.1002/acp.1094>
- Clarke, C., Milne, R., & Bull, R. (2011). Interviewing Suspects of Crime: The Impact of PEACE Training, Supervision and the Presence of a Legal Advisor. *Journal of Investigative Psychology and Offender Profiling, 8*(2), 149–162. <https://doi.org/10.1002/jip.144>
- Costafreda, S. G. (2009). Pooling fMRI data: meta-analysis, mega-analysis and multi-center studies. *Frontiers in Neuroinformatics, 3*, 33. <https://doi.org/10.3389/neuro.11.033.2009>
- Criminal Justice Joint Inspection. (2014). Achieving best evidence in child sexual abuse case—A joint inspection. *Crown Prosecution Service Inspectorate, London*. Retrieved from: https://www.justiceinspectors.gov.uk/cjji/wp-content/uploads/sites/2/2014/12/CJJI_ABE_Dec14_rpt.pdf
- Cyr, M., & Lamb, M. E. (2009). Assessing the effectiveness of the NICHD investigative interview Protocol when interviewing French-speaking alleged victims of child sexual abuse in Quebec. *Child Abuse and Neglect, 33*(5), 257–268. <https://doi.org/10.1016/j.chiabu.2008.04.002>
- Debue, N., & van de Leemput, C. (2014). What does germane load mean? An empirical contribution to the cognitive load theory. *Frontiers in Psychology, 5*, 1099. <https://doi.org/10.3389/fpsyg.2014.01099>
- Dijk, T. van, Spil, T., Burg, S. van der, & Wenzler, I. (2015). Present or play: The effect of serious gaming on demonstrated behaviour. *International Journal of Game Based Learning (IJGBL), 5*(2), 55–69. Doi: 10.4018/ijgbl.2015040104
- Ellis, S., Carette, B., Anseel, F., & Lievens, F. (2014). Systematic reflection: implications for learning from failures and successes. *Current Directions in Psychological Science, 23*(1), 67–72. <https://doi.org/10.1177/0963721413504106>
- Engle, R. W. (2002). Working memory capacity as executive attention. *Current Directions in Psychological Science, 11*(1), 19–23. <https://doi.org/10.1111/1467-8721.00160>
- Engle, R. W., & Kane, M. J. (2003). Executive attention, working memory capacity, and a two-factor theory of cognitive control. In *The Psychology of Learning and Motivation: Advances in Research and Theory* (pp. 145–199). [https://doi.org/10.1016/S0079-7421\(03\)44005-X](https://doi.org/10.1016/S0079-7421(03)44005-X)
- Finnilä, K., Mahlberg, N., Santtila, P., Sandnabba, K., & Niemi, P. (2003). Validity of a test of children's suggestibility for predicting responses to two interview situations differing in their degree of suggestiveness. *Journal of Experimental Child Psychology, 85*(1), 32–49. [https://doi.org/10.1016/S0022-0965\(03\)00025-0](https://doi.org/10.1016/S0022-0965(03)00025-0)
- Gagnon, K., & Cyr, M. (2017). Sexual abuse and preschoolers: Forensic details in

- regard of question types. *Child Abuse & Neglect*, 67, 109–118. <https://doi.org/10.1016/j.chiabu.2017.02.022>
- Gambrill, E. D. (2005). *Critical thinking in clinical practice: improving the quality of judgments and decisions*. John Wiley & Sons.
- Graafland, M., Schraagen, J. M., & Schijven, M. P. (2012). Systematic review of serious games for medical education and surgical skills training. *British Journal of Surgery*, 99(10), 1322–1330. <https://doi.org/10.1002/bjs.8819>
- Greitzer, F. L., Kuchar, O. A., & Huston, K. (2007). Cognitive science implications for enhancing training effectiveness in a serious gaming context. *Journal on Educational Resources in Computing*, 7(3), 2–es. <https://doi.org/10.1145/1281320.1281322>
- Guadagno, B. L., & Powell, M. B. (2009). A qualitative examination of police officers' questioning of children about repeated events. *Police Practice and Research*, 10(1), 61–73. <https://doi.org/10.1080/15614260802128468>
- Hattie, J., & Timperley, H. (2007). The power of feedback [transfer argument]. *Review of Educational Research*, 77(1), 81–112. <https://doi.org/10.3102/003465430298487>
- Harrison, T. L., Shipstead, Z., Hicks, K. L., Hambrick, D. Z., Redick, T. S., & Engle, R. W. (2013). Working memory training may increase working memory capacity but not fluid intelligence. *Psychological Science*, 24(12), 2409–2419. <https://doi.org/10.1177/0956797613492984>
- Hartwig, M., Granhag, P. A., Strömwall, L. A., & Kronkvist, O. (2006). Strategic use of evidence during police interviews: when training to detect deception works. *Law and Human Behavior*, 30(5), 603–619. <https://doi.org/10.1007/s10979-006-9053-9>
- Helsdingen, A., van Gog, T., & van Merriënboer, J. (2011). The effects of practice schedule and critical thinking prompts on learning and transfer of a complex judgment task. *Journal of Educational Psychology*, 103(2), 383–398. <https://doi.org/10.1037/a0022370>
- Herman, S. (2005). Improving decision making in forensic child sexual abuse evaluations. *Law and Human Behavior*, 29(1), 87–120. <https://doi.org/10.1007/s10979-005-1400-8>
- Herschkowitz, I., Ahern, E. C., Lamb, M. E., Blasbalg, U., Karni-Visel, Y., & Breitman, M. (2017). Changes in interviewers' use of supportive techniques during the revised protocol training. *Applied Cognitive Psychology*, 31(3), 340–350. <https://doi.org/10.1002/acp.3333>
- Herschkowitz, I., Fisher, S., Lamb, M. E., & Horowitz, D. (2007). Improving credibility assessment in child sexual abuse allegations: The role of the NICHD investigative interview protocol. *Child Abuse and Neglect*, 31(2), 99–110. <https://doi.org/10.1016/j.chiabu.2006.09.005>
- Herschkowitz, I., Lamb, M. E., Katz, C., & Malloy, L. C. (2013). Does enhanced rapport-building alter the dynamics of investigative interviews with suspected victims of intra-familial abuse? *Journal of Police and Criminal Psychology*, 30(1), 6–14. <https://doi.org/10.1007/s11896-013-9136-8>
- Herschkowitz, I., Orbach, Y., Lamb, M. E., Sternberg, K. J., & Horowitz, D. (2006). Dynamics of forensic interviews with suspected abuse victims who do not disclose abuse. *Child Abuse & Neglect*, 30(7), 753–769. <https://doi.org/10.1016/j.chiabu.2005.10.016>
- Heyman, G. D., Fu, G., Lin, J., Qian, M. K., & Lee, K. (2015). Eliciting promises from children reduces cheating. *Journal of Experimental Child Psychology*, 139, 242–248. <https://doi.org/10.1016/j.jecp.2015.04.013>
- Heyselaar, E., Hagoort, P., & Segaert, K. (2017). In dialogue with an avatar, language behavior is identical to dialogue with a human partner. *Behavior Research Methods*, 49(1), 46–60.

- <https://doi.org/10.3758/s13428-015-0688-7>
- Hogarth, R. M., Lejarraga, T., & Soyer, E. (2015). The two settings of kind and wicked learning environments. *Current Directions in Psychological Science, 24*(5), 379–385. <https://doi.org/10.1177/0963721415591878>
- Howie, P., Sheehan, M., Mojarrad, T., & Wrzesinska, M. (2004). “Undesirable” and “desirable” shifts in children’s responses to repeated questions: age differences in the effect of providing a rationale for repetition. *Applied Cognitive Psychology 18*(9), 1161–1180. Doi: 10.1002/acp.1049
- Ivers, N., Jamtvedt, G., Flottorp, S., Young, J. M., Odgaard-Jensen, J., French, S. D., Oxman, A. D. (2012). Audit and feedback: effects on professional practice and healthcare outcomes. In N. Ivers (Ed.), *Cochrane Database of Systematic Reviews*. Chichester, UK: John Wiley & Sons, Ltd. <https://doi.org/10.1002/14651858.CD000259.pub3>
- Johnson, M., Magnussen, S., Thoresen, C., Lønnum, K., Burrell, L. V., & Melinder, A. (2015). Best practice recommendations still fail to result in action: A national 10-Year follow-up study of investigative interviews in CSA cases. *Applied Cognitive Psychology, 29*(5), 661–668. doi: 10.1002/acp.3147.
- Kalyuga, S., Ayres, P., Chandler, P., & Sweller, J. (2003). The Expertise Reversal Effect. *Educational Psychologist, 38*(1), 23–31. https://doi.org/10.1207/S15326985EP3801_4
- Katz, C., Barnett, Z., & Hershkowitz, I. (2014). The effect of drawing on children’s experiences of investigations following alleged child abuse. *Child Abuse and Neglect, 38*(5), 858–867. <https://doi.org/10.1016/j.chiabu.2014.01.003>
- Kirschner, P. A. (2002). Cognitive load theory: implications of cognitive load theory on the design of learning. *Learning and Instruction, 12*(1), 1–10. [https://doi.org/10.1016/S0959-4752\(01\)00014-7](https://doi.org/10.1016/S0959-4752(01)00014-7)
- Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin, 119*(2), 254–284. <https://doi.org/10.1037/0033-2909.119.2.254>
- Korkman, J., Santtila, P., & Sandnabba, N. K. (2006). Dynamics of verbal interaction between interviewer and child in interviews with alleged victims of child sexual abuse. *Scandinavian Journal of Psychology, 47*(2), 109–119. <https://doi.org/10.1111/j.1467-9450.2006.00498.x>
- Krähenbühl, S., & Blades, M. (2006). The effect of interviewing techniques on young children’s responses to questions. *Child: Care, Health and Development, 32*(3), 321–331. <https://doi.org/10.1111/j.1365-2214.2006.00608.x>
- Krähenbühl, S., Blades, M., & Eiser, C. (2009). The effect of repeated questioning on children’s accuracy and consistency in eyewitness testimony. *Legal and Criminological Psychology, 14*, 263–278. <https://doi.org/10.1348/135532508X398549>
- Krähenbühl, S. J. S., & Blades, M. (2009). Does the form of question repetition have an effect on children’s recall accuracy and consistency? *International Journal of Police Science & Management, 11*(4), 460–475. <https://doi.org/10.1350/ijps.2009.11.4.144>
- Krause, N., Pompèdda, F., Antfolk, J., Zappalà, A., & Santtila, P. (2017). The Effects of Feedback and Reflection on the Questioning Style of Untrained Interviewers in Simulated Child Sexual Abuse Interviews. *Applied Cognitive Psychology, 31*(2), 187–198. <https://doi.org/10.1002/acp.3316>
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. (2016). lmerTest: Tests in linear mixed effects models. R package version 2.0-32.
- La Rooy, D., Lamb, M. E., & Memon, A. (2011). Forensic interviews with children in Scotland: a survey of interview practices among police.

- Journal of Police and Criminal Psychology*, 26(1), 26–34.
<https://doi.org/10.1007/s11896-010-9072-9>
- Lafontaine, J., & Cyr, M. (2017). The relation between interviewers' personal characteristics and investigative interview performance in a child sexual abuse context. *Police Practice & Research*, 18(2), 106–118.
<https://doi.org/10.1080/15614263.2016.1242423>
- Lamb, M. E. (2016). Difficulties translating research on forensic interview practices to practitioners: Finding water, leading horses, but can we get them to drink? *American Psychologist*, 71(8), 710–718.
<https://doi.org/10.1037/amp0000039>
- Lamb, M., Herschkowitz, I., Orbach, Y., & Esplin, P. (2011). *Tell me what happened: Structured investigative interviews of child victims and witnesses*. Chichester, UK: John Wiley & Sons.
- Lamb, M. E., Orbach, Y., Herschkowitz, I., Esplin, P. W., & Horowitz, D. (2007). A structured forensic interview protocol improves the quality and informativeness of investigative interviews with children: A review of research using the NICHD Investigative Interview Protocol. *Child Abuse and Neglect*, 31(11–12), 1201–1231.
<https://doi.org/10.1016/j.chiabu.2007.03.021>
- Lamb, M. E., Sternberg, K. J., Orbach, Y., Esplin, P. W., & Mitchell, S. (2002). Is ongoing feedback necessary to maintain the quality of investigative interviews with allegedly abused children? *Applied Developmental Science*, 6(1), 35–41.
https://doi.org/10.1207/S1532480XAD0601_04
- Lamb, M. E., Sternberg, K. J., Orbach, Y., Herschkowitz, I., & Horowitz, D. (2003). Differences between accounts provided by witnesses and alleged victims of child sexual abuse. *Child Abuse and Neglect*, 27(9), 1019–1031.
[https://doi.org/10.1016/S0145-2134\(03\)00167-4](https://doi.org/10.1016/S0145-2134(03)00167-4)
- Lamb, M. E., Sternberg, K. J., Orbach, Y., Herschkowitz, I., Horowitz, D., & Esplin, P. W. (2002). The effects of intensive training and ongoing supervision on the quality of investigative interviews with alleged sex abuse victims. *Applied Developmental Science*, 6(3), 114–125.
https://doi.org/10.1207/S1532480XAD0603_2
- Lamb, M. M. E., Sternberg, K. K. J., Orbach, Y., Esplin, P. P. W., Stewart, H. L., & Mitchell, S. (2003). Age differences in young children's responses to open-ended invitations in the course of forensic interviews. *Journal of Consulting and Clinical Psychology*, 71(5), 926–934.
<https://doi.org/10.1037/0022-006X.71.5.926>
- Landsberg, C. R. C., Buskirk, W. Van, Astwood, R. S., Van Buskirk, W. L., & Astwood, R. S. (2010). Does Feedback Type Matter? Investigating the Effectiveness of Feedback Content on Performance Outcomes. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 54, 2304–2308.
<https://doi.org/10.1177/154193121005402718>
- Lehman, E. B., McKinley, M. J., Thompson, D. W., Leonard, A. M., Liebman, J. I., & Rothrock, D. D. (2010). Long-term stability of young children's eyewitness accuracy, suggestibility, and resistance to misinformation. *Journal of Applied Developmental Psychology*, 31(2), 145–154.
<https://doi.org/10.1016/j.appdev.2009.11.007>
- London, K., Bruck, M., Wright, D. B., & Ceci, S. J. (2008, January). Review of the contemporary literature on how children report sexual abuse to others: Findings, methodological issues, and implications for forensic interviewers. *Memory*.
<https://doi.org/10.1080/09658210701725732>
- Lyon, T. D. (2005). *Ten step investigative interview*, 5–7. Retrieved from <http://works.bepress.com/thomaslyon/5/>
- Lyon, T. D. (2010). Investigative interviewing of the child. In D. N. Duquette & A. M. Haralambie (Eds.) *Child Welfare Law and Practice*, 87–109. Denver, Colorado: Bradford publishing company

- Lyon, T. D. (2014). Interviewing Children. *Annual Review of Law and Social Science*, 10(1), 73–89. <https://doi.org/10.1146/annurev-lawsocsci-110413-030913>
- Lyon, T. D., Wandrey, L., Ahern, E., Licht, R., Sim, M. P. Y., & Quas, J. A. (2014). Eliciting maltreated and nonmaltreated children's transgression disclosures: narrative practice rapport building and a putative confession. *Child Development*, 85(4), 1756–1769. <https://doi.org/10.1111/cdev.12223>
- Mayer, R. E. (2008). Applying the science of learning: evidence-based principles for the design of multimedia instruction. *The American Psychologist*, 63(8), 760–769. <https://doi.org/10.1037/0003-066X.63.8.760>
- Merriënboer, J. Van, & Sweller, J. (2005). Cognitive load theory and complex learning: Recent developments and future directions. *Educational Psychology Review*, 17(2), 147–177. <https://doi.org/10.1007/s10648-005-3951-0>
- Mikropoulos, T. A., & Natsis, A. (2011). Educational virtual environments: A ten-year review of empirical research (1999–2009). *Computers & Education*, 56(3), 769–780. <https://doi.org/10.1016/j.compedu.2010.10.020>
- Miller, L. M., Chang, C.-I., Wang, S., Beier, M. E., & Klisch, Y. (2011). Learning and motivational impacts of a multimedia science game. *Computers & Education*, 57(1), 1425–1433. <https://doi.org/10.1016/j.compedu.2011.01.016>
- Mugford, R., Corey, S., & Bennell, C. (2013). Improving police training from a cognitive load perspective. *Policing: An International Journal of Police Strategies & Management*, 36(May), 312–337. <https://doi.org/10.1108/13639511311329723>
- National Institute of Child Health and Human Development. (2011). The National Institute of Child Health and Human Development (NICHD) Protocol: Interview Guide. in M. E. Lamb, D. La Rooy, L. C. Malloy, & C. Katz (Eds.), *Children's testimony: a handbook of psychological research and forensic practice*. Chichester, UK: John Wiley & Sons.
- Olaszewski, A. E. (2016). *Virtual Simulation and Serious Games for Medical Education: A Review of the Literature and Development of a Virtual Peritoneal Dialysis Simulator*. (Doctoral dissertation). Retrieved from Harvard University's DASH repository. <http://nrs.harvard.edu/urn-3:HUL.InstRepos:27007751>
- Orbach, Y., Herschkowitz, I., Lamb, M. E., Sternberg, K. J., Esplin, P. W., & Horowitz, D. (2000). Assessing the value of structured protocols for forensic interviews of alleged child abuse victims. *Child Abuse and Neglect*, 24(6), 733–752. [https://doi.org/10.1016/S0145-2134\(00\)00137-X](https://doi.org/10.1016/S0145-2134(00)00137-X)
- Orbach, Y., & Lamb, M. E. (1999). Assessing the accuracy of a child's account of sexual abuse: a case study. *Child Abuse & Neglect*, 23(1), 91–98. [https://doi.org/10.1016/S0145-2134\(98\)00114-8](https://doi.org/10.1016/S0145-2134(98)00114-8)
- Paas, F. G., & Merriënboer, J. Van. (1994). Variability of worked examples and transfer of geometrical problem-solving skills: A cognitive-load approach. *Journal of Educational Psychology*, 86(1), 122. <http://dx.doi.org/10.1037/0022-0663.86.1.122>
- Paas, F., Renkl, A., & Sweller, J. (2003). Cognitive load theory and instructional design: Recent developments. *Educational Psychologist* 38(1), 1–4. doi:10.1207/S15326985EP3801_1
- Paas, F., Renkl, A., & Sweller, J. (2004). Cognitive load theory: Instructional implications of the interaction between information structures and cognitive architecture. *Instructional Science* 32(1–2), 1–8. <http://dx.doi.org/10.1023/B:TRUC.0000021806.17516.d0>
- Paas, F., Tuovinen, J., & Tabbers, H. (2003). Cognitive load measurement as a means to advance cognitive load theory. *Educational psychologist*, 38(1), 63–71. https://doi.org/10.1207/S15326985EP3801_8
- Paas, F., & van Gog, T. (2006). Optimising worked example instruction: Different

- ways to increase germane cognitive load. *Learning and Instruction*, 16(2 SPEC. ISS.), 87–91. <https://doi.org/10.1016/j.learninstruc.2006.02.004>
- Pipe, M. E., Lamb, M. E., Orbach, Y., & Cederborg, A. C. (Eds.). (2012). *Child sexual abuse: Disclosure, delay, and denial*. New York, NY: Psychology Press.
- Pipe, M. E., Lamb, M. E., Orbach, Y., & Esplin, P. W. (2004). Recent research on children's testimony about experienced and witnessed events. *Developmental Review*, 24(4), 440–468. <https://doi.org/10.1016/j.dr.2004.08.006>
- Pompedda, F., Zappalà, A., & Santtila, P. (2015). Simulations of child sexual abuse interviews using avatars paired with feedback improves interview quality. *Psychology, Crime & Law*, 21(1), 28–52. <https://doi.org/10.1080/1068316X.2014.915323>
- Pompedda, F., Antfolk, J., Zappalà, A., & Santtila, P. (2017). A Combination of Outcome and Process Feedback Enhances Performance in Simulations of Child Sexual Abuse Interviews Using Avatars. *Frontiers in Psychology*, 8, 1474. <https://doi.org/10.3389/fpsyg.2017.01474>
- Powell, M. B. (2013). An overview of current initiatives to improve child witness interviews about sexual abuse. *Current Issues in Criminal Justice*, 25(711), 711–720. <https://doi.org/10.1525/sp.2007.54.1.23>
- Powell, M. B. (2015). Improving the reliability of child witness testimony in court: The importance of focusing on questioning techniques. *Psychology, Public Policy, and Law*, 21(3), 309–322. <https://doi.org/10.1037/law0000052>
- Powell, M. B., Fisher, R. P., & Hughes-Scholes, C. H. (2008). The effect of using trained versus untrained adult respondents in simulated practice interviews about child abuse. *Child Abuse & Neglect*, 32(11), 1007–1016. Doi: 10.1016/j.chiabu.2008.05.005
- Powell, M. B., Guadagno, B., & Benson, M. (2016). Improving child investigative interviewer performance through computer-based learning activities. *Policing and Society*, 26(4), 365–374. <https://doi.org/10.1080/10439463.2014.942850>
- Powell, M. B., & Thomson, D. M. (1997). The effect of an intervening interview on children's ability to remember one occurrence of a repeated event. *Legal and Criminological Psychology*, 2(2), 247–262. <https://doi.org/10.1111/j.2044-8333.1997.tb00346.x>
- Powell, M. B., & Wright, R. (2008). Investigative interviewers' perceptions of the value of different training tasks on their adherence to open-ended questions with children. *Psychiatry, Psychology and Law*, 15(2), 272–283. <https://doi.org/http://dx.doi.org/10.1080/13218710802014493>
- Price, H. L., & Roberts, K. P. (2011). The effects of an intensive training and feedback program on police and social workers' investigative interviews of children. *Canadian Journal of Behavioural Science*, 43(3), 235–244. <https://doi.org/10.1037/a0022541>
- Prince, M. (2004). Does active learning work? a review of the research. *Journal of Engineering Education*, 93, 223–231. <https://doi.org/10.1002/j.2168-9830.2004.tb00809.x>
- Quas, J. A., Malloy, L. C., Melinder, A., Goodman, G. S., D'Mello, M., & Schaaf, J. (2007). Developmental differences in the effects of repeated interviews and interviewer bias on young children's event memory and false reports. *Developmental Psychology*, 43(4), 823–837. <https://doi.org/10.1037/0012-1649.43.4.823>
- Quas, J. A., Stolzenberg, S. N., & Lyon, T. D. (2018). The effects of promising to tell the truth, the putative confession, and recall and recognition questions on maltreated and non-maltreated children's disclosure of a minor transgression. *Journal of Experimental Child Psychology*, 166, 266–279. <https://doi.org/10.1016/j.jecp.2017.08.014>

- Roberts, K. P. (2002). Children's ability to distinguish between memories from multiple sources: Implications for the quality and accuracy of eyewitness statements. *Developmental Review, 22*(3), 403-435. [https://doi.org/10.1016/S0273-2297\(02\)00005-9](https://doi.org/10.1016/S0273-2297(02)00005-9)
- Roberts, K. P., Lamb, M. E., & Sternberg, K. J. (1999). Effects of the timing of postevent information on preschoolers' memories of an event. *Applied Cognitive Psychology, 13*(6), 541-559. doi: 10.1002/(SICI)1099-0720(199912)13:6<541::AID-ACP618>3.0.CO;2-5
- Roberts, K., Lamb, M., & Sternberg, K. (2004). The effects of rapport-building style on children's reports of a staged event. *Applied Cognitive Psychology 18*(2), 189-202. Retrieved from <http://onlinelibrary.wiley.com/doi/10.1002/acp.957/full>
- Roberts, K. P., & Powell, M. B. (2001). Describing individual incidents of sexual abuse: a review of research on the effects of multiple sources of information on children's reports. *Child Abuse & Neglect, 25*(12), 1643-1659. [https://doi.org/10.1016/S0145-2134\(01\)00290-3](https://doi.org/10.1016/S0145-2134(01)00290-3)
- Schmidt, R. A., & Bjork, R. A. (1992). New conceptualizations of practice: common principles in three paradigms suggest new concepts for training. *Psychological Science, 3*(4), 207-218. <https://doi.org/10.1111/j.1467-9280.1992.tb00029.x>
- Sitepal [Computer software]. (2014). Retrieved from <http://www.sitepal.com/>
- Sitzmann, T. (2011). A meta-analytic examination of the instructional effectiveness of computer-based simulation games. *Personnel Psychology, 64*(2), 489-528. <https://doi.org/10.1111/j.1744-6570.2011.01190.x>
- Smith, M. C. (2008). Pre-professional mandated reporters' understanding of young children's eyewitness testimony: Implications for training. *Children and Youth Services Review, 30*(12), 1355-1365. <https://doi.org/10.1016/j.chilyouth.2008.04.004>
- Soveri, A., Antfolk, J., Karlsson, L., Salo, B., & Laine, M. (2017). Working memory training revisited: A multi-level meta-analysis of n-back training studies. *Psychonomic Bulletin & Review. 24*(1), 1217-1227. <https://doi.org/10.3758/s13423-016-1217-0>
- Stawski, R. S., Sliwinski, M. J., & Smyth, J. M. (2010). Stress-related cognitive interference predicts cognitive function in old age. *Psychology and Aging, 21*, 535-544. <https://doi.org/10.1037/0882-7974.21.3.535>
- Sternberg, K. J., Lamb, M. E., Davies, G. M., & Westcott, H. L. (2001). The memorandum of good practice: Theory versus application. *Child abuse & neglect, 25*(5), 669-681. [https://doi.org/10.1016/S0145-2134\(01\)00232-0](https://doi.org/10.1016/S0145-2134(01)00232-0)
- Stolzenberg, S. N., & Lyon, T. D. (2017). "Where were your clothes?" Eliciting descriptions of clothing placement from children alleging sexual abuse in criminal trials and forensic interviews. *Legal and Criminological Psychology, 22*(2), 197-212. <https://doi.org/10.1111/lcrp.12094>
- Stolzenberg, S. N., McWilliams, K., & Lyon, T. D. (2017). The Effects of the Hypothetical Putative Confession and Negatively Valenced Yes/No Questions on Maltreated and Nonmaltreated Children's Disclosure of a Minor Transgression. *Child Maltreatment, 22*(2), 167-173. <https://doi.org/10.1177/1077559516673734>
- Sweller, J., & Chandler, P. (1994). Why some material is difficult to learn. *Cognition and Instruction, 12*(3), 185-233. https://doi.org/10.1207/s1532690xci1203_1
- Sweller, J., van Merriënboer, J. J. G., & Paas, F. G. W. C. (1998). Cognitive architecture and instructional design. *Educational Psychology Review, 10*(3), 251-296. <https://doi.org/10.1023/A:1022193728205>
- Talwar, V., Lee, K., Bala, N., & Lindsay, R. C. L. (2002). Children's conceptual

- knowledge of lying and its relation to their actual behaviors: Implications for court competence examinations. *Law and Human Behavior*, *26*(4), 395–415. <https://doi.org/10.1023/A:1016379104959>
- Thoresen, C., Lønnum, K., Melinder, A., & Magnussen, S. (2009). Forensic interviews with children in CSA cases: A large-sample study of Norwegian police interviews. *Applied Cognitive Psychology*, *23*(7), 999–1011. <https://doi.org/10.1002/acp.1534>
- Thoresen, C., Lønnum, K., Melinder, A., Stridbeck, U., & Magnussen, S. (2006). Theory and practice in interviewing young children: A study of Norwegian police interviews 1985–2002. *Psychology, Crime and Law*, *12*(6), 629–640. <https://doi.org/10.1080/10683160500350546>
- Van der Kleij, F. M., Feskens, R. C. W., & Eggen, T. J. H. M. (2015). Effects of feedback in a computer-based learning environment on students' learning outcomes: a meta-analysis. *Review of Educational Research*, *85*(4), 475–511. <https://doi.org/10.3102/0034654314564881>
- Van Der Zee, D. J., Holkenborg, B., & Robinson, S. (2012). Conceptual modeling for simulation-based serious gaming. *Decision Support Systems*, *54*(1), 33–45. <https://doi.org/10.1016/j.dss.2012.03.006>
- van Merriënboer, J. J. G., & Sluijsmans, D. M. A. (2009). Toward a Synthesis of Cognitive Load Theory, Four-Component Instructional Design, and Self-Directed Learning. *Educational Psychology Review*, *21*(1), 55–66. <https://doi.org/10.1007/s10648-008-9092-5>
- Vrij, A. (2005). Criteria-Based Content Analysis: A Qualitative Review of the First 37 Studies. *Psychology, Public Policy, and Law*, *11*(1), 3–41. <https://doi.org/10.1037/1076-8971.11.1.3>
- Wouters, P., van Nimwegen, C., van Oostendorp, H., & van der Spek, E. D. (2013). A meta-analysis of the cognitive and motivational effects of serious games. *Journal of Educational Psychology*, *105*(2), 249–265. <https://doi.org/10.1037/a0031311>
- Wouters, P., & Van Oostendorp, H. (2013). A meta-analytic review of the role of instructional support in game-based learning. *Computers & Education*, *60*, 412–425. <https://doi.org/10.1016/j.compedu.2012.07.018>
- Yi, M., Jo, E., & Lamb, M. E. (2016). Effects of the NICHD Protocol Training on Child Investigative Interview Quality in Korean Police Officers. *Journal of Police and Criminal Psychology*, *31*(2), 155–163. <https://doi.org/10.1007/s11896-015-9170-9>

Appendix A

Raw Means and Standard Deviations of the Dependent Variables Divided by Feedback Types

Dependent Variable	<i>Group(N)</i>	<i>M_{t1}</i>	<i>SD</i>	<i>M_{t4}</i>	<i>SD</i>
Recommended Questions	Control (51)	28.86	14.99	31.03	16.70
	Outcome (12)	26.89	10.23	29.06	13.01
	Process (12)	32.06	13.04	52.24	15.43
	Combination (72)	33.51	14.99	65.92	17.56
Correct Details vs Details Retrieved	Control (47)	47.01	38.31	53.57	33.35
	Outcome (12)	80.16	30.38	73.69	35.28
	Process (12)	92.68	14.20	89.58	29.11
	Combination (61)	58.70	37.18	86.70	22.85
Correct Details vs Details Available	Control (51)	20.29	21.27	34.71	27.34
	Outcome (12)	34.64	23.53	50.89	36.34
	Process (12)	52.05	29.27	68.41	32.77
	Combination (72)	30.69	25.88	64.77	30.23
Correct Conclusions	Control (51)	0.08	0.27	0.08	0.27
	Outcome (12)	0.08	0.29	0.08	0.29
	Process (12)	0.17	0.39	0.25	0.45
	Combination (72)	0.04	0.20	0.26	0.44

Note. M_{t1} refers to the first interview and M_{t4} to the fourth interview

Francesco Pompedda

Training in Investigative Interviews of Children: Serious Gaming Paired with Feedback Improves Interview Quality

Interview quality in alleged child sexual abuse (CSA) cases remains low. Training programs have been developed in order to tackle this problem. However, these programs are usually not successful in creating stable effects over time, or when positive results have been achieved, programs are often logistically complicated and expensive.

This dissertation presents the development and test of a new, flexible, and interactive tool that combines serious gaming and feedback in order to improve the quality of investigative interviews in CSA cases in a time and cost effective way