Investigating youths' learning of online safety and privacy from others: A discussion of study design and statistical analysis considerations

Kerrianne Buchanan, *National Institute of Standards and Technology*
Yee-Yin Choong, *National Institute of Standards and Technology*
Olivia Murphy, *University of Maryland*[1]

Abstract

An important factor for investigating youth's online safety, security, and privacy is to understand how and from where they learn their online behaviors and knowledge. Although research has shown that people within youths' environments (e.g., parents) and the environments themselves (e.g., schools) are important influences, studying the interrelationships between youth and these influences presents many challenges for researchers. This paper describes the challenges inherent in this type of research through a study we conducted to examine parents' involvement with their children's password practices. We discuss challenges in data collection and statistical analysis, and lessons learned that may address these challenges in similar studies of youth.

1. Introduction

Youth (i.e., children under 18) are going online at younger ages and using the internet for many activities [5]. Because youth are putting their personal information online and taking online risks [4, 19], it is important to identify effective practices to help youth stay safe online and protect their privacy, as well as to investigate whether and how youth learn to stay safe online. Several theories support that youth's knowledge, learning, and understanding result from people in their environment [1] (e.g., parents) and the contexts they are embedded within [2] (e.g., families, schools, peer networks). Exploring how these influences shape youth's knowledge and behavior is an important area of study to understand more about youths' online safety, security, and privacy.

The purpose of this position paper is to discuss the challenges of studying the complex interrelationships between youth and the people and factors within their environments. Specifically, we discuss challenges from our work studying parents' involvement with their children's password practices. Passwords are important as they are often used to keep online information protected and provide access control over online accounts. Although in general, youth (like adults) seem to understand that effective passwords are important, research suggests they inconsistently enact effective password behavior [12, 18]. Because parents are an especially important source of children's learning and are often involved in their children's online activities, we sought to extend the work on parents' involvement with children's password behaviors [18, 20] by studying parents' own behaviors and their involvement with password creation and maintenance for their children.

We discuss two key challenges we have faced in this study: 1) collecting data on parents' involvement in each of their children's password practices, and 2) analyzing data to answer research questions about if and how qualities of parents and qualities of each of their children were related to parents' involvement in their children's password practices. We share lessons learned in facing these challenges and highlight methods [8, 14, 15] that can be applied to our study and others that examine how people and environmental factors shape youths' online safety, security, and privacy.

2. Study Methodology

This study is part of a broader research effort to understand youth online security and privacy from the perspectives of children, their parents, and teachers/educators. The goal of this study was to examine the password practices of parents and their involvement with their children's passwords practices. We developed a survey with three sections (see the Appendix for example survey questions). In the first section, parents answered seven question blocks (each with at least one and up to 13 sub-questions) about their technology use, experience, and adoption, and their password practices. The second section asked parents to answer six question blocks (each with at least one and up to nine sub-questions) about each of their children within the target grade range of kindergarten to 12th grade (K-12) about online activities and their involvement in their children's password practices. The final block asked general thoughts and demographics. Survey items were created based on prior research, reviewed by

Workshop on Kids' Online Privacy and Safety (KOPS), co-located with USENIX Symposium on Usable Privacy and Security (SOUPS) 2022, August 7 - 9, 2022, Boston, MA, USA.

[1] This material is based upon work supported by the UMD and NIST Professional Research Experience Program (PREP) under Award Number 70NANB18H165.

content experts and survey design experts, and piloted by parents. The survey was refined based on the reviews and fielded as an online survey through a contracting firm to a panel of participants in the United States. A survey participant was eligible to participate if they were a parent of at least one child in the target grade range. The final dataset included 265 parents who completed the survey answering questions about a total of 448 children. The study was approved by the NIST Research Protections Office (RPO).

3. Challenge 1: Data Collection

When designing the survey, we faced the challenge of accounting for the possibility that parents with multiple children might have differing involvement with each of their individual children's password practices. To address this, we designed the survey to present parents with the same set of six questions repeated for each child depending on the number of children each parent had. Although this allowed for collecting information about all of each parent's children individually, the repetitive nature of this process could be burdensome and time consuming to parents completing the survey. To reduce this burden, we added a question when a parent had multiple children, "We will ask you the same questions about your children's online activities and password usage. For the next child, are your answers the same as another child you have given answers for already?" Parents could select if their answers for the next child were the same as or different from a child they had previously entered. If answers were the same, the entire set of questions was skipped for that child. Parents selected this option for 143 of the children (45.83% of children of parents with two or more children).

This approach can be useful for discovering if parents report changing their behaviors for each child. However, this approach may be less sensitive to identifying differences when parents do change their involvement across children, as the data will have lower variability from repeating responses across children. An alternative approach could have been to have participants answer a smaller number of questions for each child so that data could be collected for each child. Another limitation is that this type of data collection relied solely on parent self-report. While this was the goal of the study, to examine parent perceptions and behaviors with their children, a gap in this type of data collection is the perceptions of the children themselves. Future work could benefit from examining responses of parents and their children to perform data analysis of the two as a dyad [10].

4. Challenge 2: Data Analysis

Another challenge we faced was in analyzing the survey data to answer research questions about how qualities of parents and of each of their children were related to parents' involvement in children's password practices. We next discuss the data structure and statistical analysis considerations contributing to this challenge in Sec. 4.1 and Sec. 4.2, respectively.

4.1 Data Structure Considerations

The initial dataset was structured with parents as the rows, or observations, and each survey question as the columns, or variables. Parents' responses for each of their children were repeated across several columns. As displayed in Table 1, this results in a collection of columns for child 1, columns for child 2, and so on. This is referred to as a "wide" data structure.

Table 1. Parents as rows hypothetical data structure

Parent ID	Importance Password Strength	Number of Children	Child 1 Grade	Child 1 Password Creation Help	…	Child 2 Grade	Child 2 Password Creation Help
1	3	2	2	3		11	1
…							
265	3	1	5	2		NA	NA

It is difficult to quickly obtain a summary of results across children, given our dataset structure. For example, examining the frequency of parent responses (*Always, Sometimes, Never*) for the question, "Do you help this child create passwords?" is a challenge given that this variable is found in multiple columns – there is a column for parents' responses across child 1, a column for parents' responses across child 2, and so on. This challenge may result in the decision to analyze results for only one child. This is perfectly reasonable in some scenarios: researchers may be most interested in results for one child or parents of only one child, or they may want to get an initial overview of the results. However, analyzing data in this way negates the ability to examine research questions about effects within families, as this would result in the loss of information about any additional children.

A first step in solving this problem is to restructure the data from the "wide" format into a "long" format (Table 2) such that each child becomes a row nested within their parent. This can be done in popular statistical analysis programs such as R Programming Environment. Rather than repeating child questions as columns for each child, now each child has a row with responses for that child in columns for each question. In this format, the parent survey responses are repeated across each row for their children, as their own demographics and responses are stable and do not differ from child to child. In this way, parents' responses for each of their children is nested within the parent's data. Now analyzing the password creation help item across children can be done, though there are additional considerations for statistical analyses of data in this structure.

Table 2. Child data as rows hypothetical data structure

Parent ID	Importance Password Strength	Number Children	Child	Grade	Password Creation Help
1	3	2	1	2	3
1	3	2	2	11	1
…					
265	3	1	1	5	2

4.2 Statistical Analysis Considerations

Even when data are properly structured, statistical analysis is still a challenge to answer research questions about family effects, as many traditional statistical methods (e.g., ANOVA, multiple regression, chi-square test) assume observations in the data are independent [9, 10, 13]. Our data does not meet this assumption, as the experiences of parents are related to those of their children, as are responses for each child of one parent. Moreover, traditional techniques are often not sufficient to understand unique effects and interrelationships of parent-level (e.g., parents' perceived importance of password strength) and child-level (e.g., child's grade) variables in the data. Statistically partitioning effects due to child-level, parent-level, and the interaction of the two is not only needed to fully account for the nested structure of the data, but may "be of central interest in many studies" [15].

Fortunately, existing statistical methodology can facilitate analysis of data in this structure. Multi-level modeling [3, 6, 16] is one such technique that accounts for nested data structures and provides estimates of the effects of variables from different levels. These models are conceptually similar to multiple regression, as predictor variables are input into the model to predict effects on an outcome variable of interest. In addition to providing insights into interrelationships between family factors, these models can also handle a variety of types of data, including categorical, ordinal, binary, and continuous predictors and outcomes [6].

In our study, this model can be used to answer questions about the role of child-level variables or parent-level variables in predicting parent involvement in password creation. Analyses can be conducted across all children in the dataset and could even be used to examine the effect of family size on parents' involvement in children's password practices.

Multi-level modeling concepts have also been applied to other statistical analyses. For example, researchers may be interested in subtypes, or clusters, of participants based on responses to survey questions. In this way, researchers can use results to create profiles or personas of behavior. One method to perform this analysis is latent class analysis (LCA), in which researchers can use theory and statistical modeling parameters to determine how many clusters exist in the data. Multi-level latent class analysis (MLCA) can be used to perform the clustering while accounting for the nested structure of the data [7]. For example, this type of analysis could tell us if there are subtypes of parent involvement with their children's password creation. We may expect to find clusters of parents who help with all aspects of passwords, clusters of parents who don't help at all, and clusters that help inconsistently or just with certain aspects. By entering parent involvement variables into the MLCA analysis, the model can produce a number of clusters and also how participants in that cluster responded to different variables. Such an analysis could be important for identifying what behaviors tend to occur together when developing guidance and education for parents and their children.

While we have conducted preliminary analyses using these concepts, we found that in practice there are challenges to implementing these models, as they often require deep understanding of statistical techniques and statistical software packages. Moreover, when models have many variables and when variables are categorical, interpretation is more difficult. Although many reference materials exist to implement these models, advanced knowledge is still required to understand and use these materials effectively. We believe our study and others will benefit from working with statisticians and data scientists who are familiar with these concepts and challenges to correctly implement and interpret these models. Additionally, the field would benefit from identifying best practices in analyses and statistical tools for researchers to easily conduct these methods.

5. Applications and Future Directions

The considerations described have potential for expanding the types of questions researchers can answer about youths' online safety. Data collection and analysis methods can apply to any type of context, such as schools or teachers. There are also data collection and/or data analysis techniques for examining specific types of relationships among members of a group (e.g., family, peer network). When data are collected from youth and from their families or peer groups, other techniques such as the Social Relations Model can model the effects of different influences, environments, and individual-level factors on outcomes [11, 17]. This could be especially beneficial for studies of multiple members of a family in order to examine family, individual, and reciprocal effects of siblings, mothers, fathers, and/or extended family.

6. References

1. Alfred Bandura and Richard H. Walters. *Social learning theory*. Vol. 1. Prentice Hall, 1977.
2. Urie Bronfenbrenner. Ecological systems theory. In *Six theories of child development: Revised formulations and current issues.* 187-249. Jessica Kingsley Publishers, 1992.
3. Anthony S. Bryk and Stephen W. Raudenbush. *Hierarchical linear models: Applications and data analysis methods*. Sage Publications, Inc, 1992.
4. Aiman El Asam and Adrienne Katz. Vulnerable Young People and Their Experience of Online Risks. *Human-Computer Interaction*, 33(4): 281-304, 2018.
5. EU Kids Online. Kids Online–Findings, methods, recommendations.

https://eprints.lse.ac.uk/60512/1/EU%20Kids%20online%20III%20.pdf: London. 2014.
6. W.Holmes Finch, Jocelyn E. Bolin, and Ken Kelley. *Multilevel Modeling Using R*. CRC Press, 2014.
7. Kimberly L. Henry and Bengt Muthén. Multilevel Latent Class Analysis: An Application of Adolescent Smoking Typologies With Individual and Contextual Predictors. *Structural Equation Modeling: A Multidisciplinary Journal*, 17(2): 193-215, 2010.
8. J. M. Jenkins, J. Rasbash, and T. G. O'Connor. The role of the shared family context in differential parenting. *Developmental Psychology*, 39(1): 99-113, 2003.
9. David A. Kenny and Charles M. Judd. Consequences of violating the independence assumption in analysis of variance. *Psychological Bulletin*, 99(3): 422-431, 1986.
10. David A. Kenny, Deborah A. Kashy, and William L. Cook. *Dyadic data analysis*. Dyadic data analysis. Guilford Press, 2006.
11. David A. Kenny and Lawrence La Voie. The Social Relations Model. In *Advances in Experimental Social Psychology*, L. Berkowitz, Editor. 141-182. Academic Press, 1984.
12. Priya Kumar, et al. 'No Telling Passcodes Out Because They're Private'. 2017.
13. Mary L. McHugh. The chi-square test of independence. *Biochemia medica*, 23(2): 143-149, 2013.
14. Thomas. G. O'Connor, et al. Family settings and children's adjustment: differential adjustment within and across families. *British Journal of Psychiatry*, 179: 110-5, 2001.
15. Stephen W. Raudenbush, Robert T. Brennan, and Rosalind C. Barnett. A multivariate hierarchical model for studying psychological change within married couples. *Journal of Family Psychology*, 9(2): 161-174, 1995.
16. Tom A. B. Snijders. Multilevel Models for Family Data. In *Advances in Family Research*, J.M.A.M. Janssens, J.J.F. ter Laak, and L.W.C. Tavecchio, Editors. 193-208. Thesis, 1995.
17. Tom A.B. Snijders and David A. Kenny. The social relations model for family data: A multilevel approach. *Personal Relationships*, 6(4): 471-486, 1999.
18. Mary Theofanos, Yee-Yin Choong, and Olivia Murphy. 'Passwords Keep Me Safe' – Understanding What Children Think about Passwords.
19. Tena Velki, et al. Empirical study on the risky behavior and security awareness among secondary school pupils - validation and preliminary results. In *2017 40th International Convention on Information and Communication Technology, Electronics and Microelectronics*, 1280-1284, 2017.
20. Leah Zhang-Kennedy, et al. From nosy little brothers to stranger-danger: Children and parents' perception of mobile threats. Association for Computing Machinery, Inc. 2016.

7. Appendix

Subset of Survey Questions

A. Parents' Password Practices
1. Password creation:
   a. Importance of considerations (*easy to remember, easy to type, strong–hard to crack, same as other passwords*)
   b. Whether password generators are used (*Always, Sometimes, Never but know about it, Never and don't know about it*)
   c. Create a password for a hypothetical account on family doctor's website
2. Password tracking and maintenance:
   a. Methods used to keep track of personal passwords (*memorize, browser/device saved and auto-filled, use mnemonics, someone else remembers, write on paper, save in files, save in emails, password manager, do not track*)
   b. Frequency of changing personal passwords (*30, 31-60, 61-90, 91-120, 121-180 days, change only when necessary, change depending on accounts*)

B. Questions about Children
1. Do you help this child create passwords? (*Always, Sometimes, Never*)
   a. If *Always* or *Sometimes*, how (check all apply)?
      - *I create passwords for this child.*
      - *This child and I work together to create his/her passwords.*
      - *I only give this child guidance, but he/she creates the passwords.*
   b. If *Always* or *Sometimes*, rate the importance of considerations when helping this child with password creation (*easy to remember, easy to type, strong–hard to crack, same as other passwords*)
2. Do you help this child keep track of passwords? (*Always, Sometimes, Never*)
   a. If *Always* or *Sometimes*, how (check all apply)?
      - *I have a list (paper or electronic) of this child's passwords.*
      - *I memorize this child's passwords.*
      - *I have this child create a list of passwords and he/she is responsible for keeping the list.*
      - *I give this child guidelines on how he/she should keep track of the passwords.*
3. Has helping your children with their passwords changed your own password practices? (*Yes–why, No–why not?*)