

Examining Parent versus Child Reviews of Parental Control Apps on Google Play

Turki Alelyani¹, Arup Kumar Ghosh², Larry Moralez², Shion Guha³,
Pamela Wisniewski²

¹ *Stevens Institute of Technology, Hoboken, New Jersey 07030, USA*
talelyan@stevens.edu

² *University of Central Florida, Orlando, FL 32816, USA*
arupkumar.ghosh@ucf.edu, larrymoralez@knights.ucf.edu,
pamela.wisniewski@ucf.edu

³ *Marquette University Milwaukee, WI 53233, USA*
shion.guha@marquette.edu

Abstract. Mobile devices have become a ubiquitous means for teens and younger children to access the internet and social media. Such pervasive access affords many benefits but also exposes children to potential online risks, including cyberbullying, exposure to explicit content, and sexual solicitations. Parents who are concerned about their children’s online safety may use parental control apps to monitor, manage, and curate their children’s online access and mobile activities. This creates tension between the privacy rights and interests of children versus the legal, emotional, and moral imperatives of parents seeking to protect their children from online risks. To better understand the unique perspectives of parents and children, we conducted an analysis of 29,272 reviews of 52 different parental control apps from the Google Play store. We found that reviews written by parents differed statistically from those written by children such that it is possible to computationally automate the process of differentiating between them. Furthermore, latent themes emerged from the reviews that revealed the complexities and tensions in parent-child relationships as mediated by parental control app use. Natural Language Processing (NLP) revealed that the underlying themes within the reviews went beyond a description of the app, its features or performance and more towards an expression of the relationship between parents and teens as mediated through parental control apps. These insights can be used to improve parental control app design, and therefore the user experience of both parents and children.

Keywords: Privacy, Parental Control Apps, User Reviews, Computational Analysis, Classification, Parent-Child Relationships, Google Play

1 Introduction

With the proliferation of smartphones among youth, online safety has become a considerable concern within families [1, 2]. This is especially true because mobile smart devices have become the norm for teenagers [3], providing constant access to the internet that is often not monitored by their parents. However, parents have a legal and

emotional duty to ensure safety for their children in online contexts [4]. To do this, parents use a wide array of strategies to monitor their teens' technology use, including 16% of parents, according to a Pew Research, who install parental control applications apps on their teens' mobile devices to filter and block inappropriate online activities [3]. An analysis of 75 Google Play parental control apps found that the features of these apps may be too clumsy and privacy invasive for families that value open communication, trust, and a teen's desire to gain independence from his or her parents [5]. Ghosh et al. confirmed this claim from the perspective of teens and younger children by qualitatively analyzing online reviews posted from the vantage point of child users [6]. However, a key limitation of these studies is that researchers used qualitative methods on a relatively small sample of child reviews and were unable to conduct a comparative study of parent versus child reviews.

We build upon this work by conducting the first large-scale analysis of 29,272 reviews for 52 parental control apps to understand the unique perspectives of parents and children. We conducted a quantitative examination of the online reviews for parental control apps to understand whether parents and teens rate and write about parental control apps differently in their online reviews. We also examine the interpersonal relationships between parents and children through the lens of online privacy and surveillance. Specifically, we pose the following research questions:

RQ1: *Can we use computational methods to accurately distinguish between online reviews written by parents versus those written by children?*

RQ2: *Does the content of online reviews differ depending on whether the user is a parent or child? If so, how?*

To answer these questions, we scraped and analyzed publicly posted online reviews for 52 parental control apps available for download on the Google Play store. We first analyzed the reviews by applying Topic Modeling and N-Grams techniques to extract our linguistic prediction drivers. We then evaluated six predictive models including Naïve Bayes, Support Vector Machines, Neural Network, Logistic Regression, K-Nearest Neighbors, and Classification and Regression Trees. We compared the results of N-grams and Topic Modeling as different techniques for features extraction. We then generated topics based on parent versus child and high versus low rated reviews (Low: 1-3 ratings; High: 4-5 ratings) to understand the key differences in these reviews.

Our paper makes two unique contributions. First, we show that it is possible to build computational models that accurately predict the origin of online reviews (parents or children) using linguistic indicators. We compared and contrasted six common machine learning algorithms to highlight their performance in such classification tasks. Second, we reveal that latent themes expressed within online app reviews reveal more insights than just the strengths and weakness of the app. They express a multitude of emotions and a manifestation of the complex tensions that exist in parent-teen relationships, specifically those around privacy rights and parental control through surveillance tactics. These findings have important implications for the analysis of online reviews that extend beyond the context of adolescent online safety and serve as an important lens for future social computational research.

2 Background

2.1 Teen Technology Use and Parental Relationships

Technology use among teens and parental mediation have become an important research topic [7–12]. Yet, the majority of research in this space derives from the social sciences with little contribution from a social computational perspective. For instance, several researchers have conducted interview-based studies to highlight the tensions between parents and children when it comes to rule-setting and ensuring the online safety of youth [1, 13]. Others found that teens desire privacy as they are in the process of individuating and establishing their identities online [14, 15].

2.2 “Practical Obscurity” versus “Parental Stalking”

Teens are often forced to disclose personal information to their parents, as parents want more transparency into their teens’ online activities for the purpose of ensuring their online safety [11]. Yet, according to privacy theories, everyone should have some level of authority to decide how their personal information is disclosed to others [16, 17]. Blackwell et al. studied how “practical obscurity” (i.e., the limited visibility) of mobile devices makes it harder for parents to know their children’s online activities and, as a consequence, parents often misjudge the frequency and nature of their teens’ technology use [1]. For instance, they under-estimate how often their teens use social media apps or even which apps their children use.

To increase access to their teens online mobile activities, parents can install parental control apps on the teens’ smartphone that allow them to monitor and restrict various functions, including calls, text messaging, web browsing, and installations [18]. In general, parental control apps are a way for parents to control their children’s behavior as a means to protect them, as opposed to helping teens self-regulate and protect themselves [5]. Recent research has shown that teens equate such parental control apps to a form of “parental stalking” [6]. Others have argued that these apps engender an incongruity with the core values (e.g., privacy, autonomy) important to different families and may negatively impact parent-teen relationships [19] and shown that the use of currently available apps was associated with children experiencing more (not fewer) online risks [20]. Human-Computer Interaction (HCI) researchers have recommended and conceptualized that more collaborative approaches be used to manage these tensions [6, 8, 21, 22].

2.3 Online Reviews and Parental Control Apps

Ghosh et al.’s qualitative analysis of online reviews for 37 parental control apps examined what children think about parental control apps’ effectiveness and invasiveness [6]. They found that most children felt that the apps were excessively restrictive and privacy invasive. This previous work focused only on Google Play reviews posted by children. To our knowledge, online reviews have not been used yet

to understand parents' perspectives on these apps, nor how they differ from the perspectives of the children. To fill this gap, we scraped 29,272 reviews for 52 parental control apps to conduct a social computational analysis that differentiates between parent and child reviews, as well as models the different themes expressed within these reviews.

Analyzing online reviews is valuable as they have been shown to effectively help make better products [30–32] and boost profits [33]. For example, Epstein et al. used online app reviews, a survey, and interviews to improve the design of menstrual apps for women [34]. Wang et al. created a framework for product recommendation by leveraging the power of online reviews [26]. In addition, user feedback was also used to understand reasons for disliking apps [35]. Analyzing online reviews is also a common approach among computational social science researchers [23–26] and is a newer approach used within intersectional fields, such as Human-Computer Interaction (HCI) and Natural Language Processing (NLP) [27–29].

3 Study Design

App stores such as Google Play let users review their downloaded apps and assign a numerical rating (i.e. 1-5 stars). Users may highlight specific strengths and weaknesses of the app. Ratings for each app are then aggregated and displayed for the user to view. This data source captures different perspectives regarding aspects such as the app's functionality, benefits, and cost. These reviews can help developers overcome some of their flaws in the development process [30–32], as well as helping consumers make important decisions as to what apps will meet their needs as end users.

Below, we describe our approach to data collection, data cleaning, and analysis. Our methodology consisted of two phases: First, we applied machine learning techniques to identify different features and perspectives mentioned in the user reviews for both teens and parents, as well as the sentiments and opinions associated to these features. Second, we classified these reviews based on the extracted features. Table 1 shows all of the app used in the analysis. For each app reviewed contained more than one review, and the total number of reviews is included in the table as well.

3.1 Data Collection

We scraped publicly available user reviews on Google Play using the app review downloading tool Heedzy¹. Each review had the following attributes: 1) app name, 2) date, 3) user name, 4) review, and 5) rating. Ratings were numerical values (represented as a star) given by the user, ranging from 1 = worst to 5 = best. As shown in Table 1, a total of 29,272 user reviews for 52 apps were collected for this analysis. No users were involved in this study and IRB approval was not obtained. We excluded user names from the exemplar quotations shared in this paper to maintain anonymity.

¹ www.heedzy.com/feedback

Table 1. Summary of App Names and Number of Reviews used in the analysis

App Name	Reviews	App Name	Reviews
Bitdefender	95	Dashboard	117
Cerberus	3000	Board	109
Cybersafe	9	iNetClean	16
ESET	84	Parental Control	145
SecureKids	13	Familoop	38
Funamo	555	Launcher	17
Kakatu	114	PhoneWatcher - Mobile Tracker	314
KIDOZ	2113	Qustodio	996
Kids Place	3000	Ranger Pro Safe Browser	42
Kidslox	50	Remote Control	50
Kids Zone	490	ReThink - Stops Cyberbullying	121
MamaBear	465	Safe Browser	634
McAfee	312	Safe Browsing	222
MMGuardian	1060	Screen Time Companion App	2935
Mobicip	413	Screen Time	3002
Mobile Fence	2103	SecureTeen	2605
Net Nanny	646	Securkin	9
Norton Family	922	ShieldMyTeen	609
NQ Family Guardian	198	TeenSafe Child	132
Land of Kids	23	Trackidz	52
Xooloo	47	SafeKiddo	57
Web Blocker	98	TangTracker	32
Mobile Security	33	SURFIE	11
shieldMyTeen	609	Safe Kids	3
Parental Control and Locator	3	Privacy Camp	3
Block	6	Family Safety	8

3.2 Data Preprocessing

NLTK², a third-party library for Python for natural language processing, was used to remove stop-words and frequently used words from each review. A MALLET list was used to identify stop words [36]. We followed an iterative process to remove frequently used words that would mislead our models by giving additional weight to specific keywords. Many of these words are common in the English language (e.g., “and”,

² www.nltk.org

“this”, “is”, “are”). We also removed words that appeared too frequently (e.g., “app,” “please,” and “fix”). We note that these words suggest that users often post reviews for developers to fix problems within the app, but otherwise, were irrelevant to this research.

3.3 RQ1: Classifying App Review Authors

We employed a rule-based classification technique to extract rules for both parents and teens reviews based on research conducted by Ghosh et al. [6]. This helped in mapping the attributes of a review with a parents/teen label. A rule set consists of multiple rules $R_s = \{R_1, R_2, \dots, R_{1n}\}$. For example, in teen reviews, attributes such as “my parents”, “my mom”, and “my dad” were identified. For parents, “my teen”, “my son”, and “my child” were key attributes. We used these rules to establish ground truth for classifying the authors of these reviews. After classification, we extracted different linguistic features for each group. These features can be represented as collections of words or a set of variables categorizing a specific context. [37]. We then added Term Frequency-Inverse Document Frequency (TF-IDF) vectorization to identify other important features that represent the parent and teen classes. These features served as predictors for the model to classify authors of app reviews.

3.4 RQ2: Understanding Themes in App Reviews

We represented each review as a bag-of-words, using n-grams as features [38]. N-grams can capture groups of words in each review that may represent some patterns or important features. Relevant examples of useful 2-grams include “keep track”, “sucks worst”, and “parents allow.” This enabled us to build a text corpus to test against the full dataset for extracting latent themes. We tested this corpus against six common machine learning algorithms. Table 2-3 show the performance accuracy for both N-grams and Topic Modeling, the mean absolute error, as well as a comparison of the confusion matrices for each of the 5 classifiers.

Table 2. Performance accuracy of N-grams and topic modeling

Algorithm	Accuracy		Mean Absolute Error	
	NG	TM	NG	TM
Logistic Regression	0.73	0.59	0.48	1.08
K-Nearest Neighbors	0.67	0.57	0.74	1.08
Classification and Regression Trees	0.64	0.52	0.70	1.51
Naive Bayes	0.73	0.53	0.51	1.33
Support Vector Machines	0.53	0.53	1.35	1.32
Neural Network	0.68	0.63	0.58	0.93

Table 3. Comparison of confusion matrix results

Algorithm	Precision		Recall		F1 - Score	
	NG	TM	NG	TM	NG	TM
Logistic Regression	0.69	0.45	0.73	0.59	0.70	0.48
K-Nearest Neighbors	0.59	0.51	0.67	0.57	0.61	0.53
Classification and Regression Trees	0.61	0.50	0.64	0.52	0.63	0.50
Naive Bayes	0.66	0.28	0.73	0.53	0.68	0.37
Support Vector Machines	0.53	0.51	0.53	0.53	0.37	0.37
Neural Network	0.67	0.49	0.68	0.63	0.67	0.54

Next, we used topic modeling, specifically the latent Dirichlet allocation algorithm (LDA) via MALLET, to extract the hidden semantic structure for both parent and teen reviews [39]. Topics are collections of word tokens which represent the context of the analyzed text. MALLET identifies the most relevant topic for each review by converting collection of text to features.

The LDA algorithm is a generative statistical model often applied to discrete data such as text corpora and is used to categorize texts from a document to a specific category. Textual features are then transformed into numerical representations that can be processed efficiently. HCI research has increasingly begun use of topic models [40–42] to explore and make sense of large-scale text data in conjunction with qualitative inferences from topic models, particularly from online communities. This allows us to understand what influences how parents and teens administer a given rating. We used a common convention of selecting the number of topics that represent 80% of the overall variance to set the number of topics for each group [42] [5]. Tables 4-7 show the extracted topics with respect to the followings:

- (1) An exploratory analysis for both parent and child as well as apps rating, Table 4-6.
- (2) Parent versus child and high versus low rated reviews (Low: 1-3 ratings; High: 4-5) to understand the key differences in these reviews, Table 7.

4 Results

4.1 Examining Apps Reviews by Ratings

Initially, we did an exploratory analysis across three groups of reviews irrespective of if the review was posted by a parent or child. To do so, we classified the extracted reviews into three groups according to their rating. Our team interpreted the results qualitatively based on table topic models. The first group of ratings consists of reviews with rating 1 and 2. The second group consists of reviews with rating 3 and 4. The third group consists of reviews with rating 5.

The groups provided insights into the relationship between apps rating ranges and

the extracted topics. Tables 4-6 outline relevant topics for each category. For instance, Table 4 and 5, which represent reviews with high and medium ratings, reflect some satisfactions with the apps by both teens and parent, along with suggestions for improvements. These include payment issues, user interface, installations and blocking issues. Table 6 shows the first group, which represents the majority of reviews - 8742 - and has the range of occurrences between 436 and 1453.

These topics were mainly reflecting users' dissatisfactions with several apps' features including license, upgrading, installations as well as some compatibility issues. For example, some of the extracted topics may reflect functionality issues as in Topic 2, Topic 3, and Topic 4. Other topics may reflect dissatisfaction with the apps due to other reasons mentioned earlier as in Topics 4-10. Additionally, the reported topics show that there is a relationship between apps with low rating scores and the review themes. For instance, Topic 3 may explain some concerns regarding apps setting or security.

Table 4. Topics on high rating apps reviews

ID	Topic	Occurrence
1	play kid google found pretty turn block	263
2	browser monitor internet mode safe text content sites	338
3	device settings password stars working uninstall issue blocking	395
4	kids add option make feature works nice pay problem	457
5	screen love son year day home button times	489
6	time control set tablet parents limit children parental	490
7	good daughter update games location days show awesome	559
8	free works web phones find stars website trial	650
9	great work features easy android version track install service	745
10	phone child access kids lock back blocked things usage installed	808

Table 5. Topics on medium rating apps reviews

ID	Topic	Occurrence
1	year free worth make download pay blocked trial	589
2	access device android installed lock block put feature monitoring	815
3	kids play games phones things bad hate worry tool	1071
4	love parent son kids great usage limits tablet helps	1063
5	parents children easy control monitor safe parental internet online web	1589

ID	Topic	Occurrence
6	time screen set limit tablets tablet day chores tasks school	1794
7	control devices recommend highly found amazing reviews lot	1559
8	phone child daughter track mind content thing peace location block	2042
9	great good works features work job happy nice needed	2479
10	kid settings back perfect awesome home life mobile	2315

Table 6. Topics on low rating apps reviews

ID	Topic	Occurrence
1	tablet great update son working support year internet	436
2	child block access sites lock parent blocks porn	534
3	work uninstall stupid works deactivate chrome useless	654
4	settings screen browser mode android open home website	811
5	kids hate parents bad install version control games	771
6	good back times uninstalled buy change delete star location	721
7	play google download give features people stars sucks	974
8	time device password account set waste devices reset found problem	1010
9	phone won kid thing daughter locked longer call	1408
10	lifetime free license pay money email make trust trial years	1453

4.2 Distinguishing between Parent versus Child Reviews

To address RQ1, we ran three different classifiers on the data set to determine which worked best to classify parent and teen reviews. Table 2 shows the results of Naïve Bayes (NB), Support Vector Machines (SVM), and Neural Network (NN) to predict whether a review was entered by a teen or parent. Based on the extracted N-Grams features, the output depended upon whether or not the model estimated the right class (parent or teen). There were 10 reviews and each review were associated with the top three topics. The scores represent the weight these topics have within each review, so they can be used later on to build our models. To train our proposed models, we used 80% of the dataset for training and 20% for testing on 29,272 reviews, and we reported the results on 10-fold cross validation. We analyzed the results from the accuracy

measure for each classifier. Naïve Bayes (NB) produced the highest score having correctly classified 75% of the reviews. The Support Vector Machines (SVM), which has been described as an outstanding classifier in the context of text classification, achieved a 72% accuracy measure [43–45]. Neural Network (NN) produced the worst results with an accuracy measure score of 69%.

The reported findings illustrate that the extracted features by N-Grams technique contributed in identifying parents' reviews from child reviews. From our analyses, parents' reviews were associated with concerns including functionality issues, suggestions for improvement and cost issues. Some of these features include “monitors usage including”, “google play doesn't”, “support unable”, and “app reason rooted.”

Child reviews were mostly expressing frustration toward their parents. For instance, some of the extracted features for teens include negative sentiments regarding the parental control apps installed on their devices explicitly mentioned their parent or parents. Examples include “even stupid parents”, “people creating disgusting”, “hate parents”, and “dislike dad put.”

The coherence of our analysis shows how well the extracted features by N-Grams can be contributed to improving the performance of the proposed models. In other words, parents and teens features may have shared a common theme within each group which led to the increasing of the models' performance accuracy. Additionally, reviews written by either group may reveal that concerns are centered around specific type of issues. A more thorough research of parental control apps can provide an array of clues to providing future strategies for apps designers.

Our findings show that the both models, (NB) and (LR), substantially outperformed the other models. This finding confirmed previous studies' conclusions that NB is an outstanding classifier in text classifications [45]. K-Nearest Neighbors (KNN) and Neural Network (NN) scored 63% and 68%, respectively. Support Vector Machines (SVM) and Classification and Regression Trees (CART) produced the lowest performance in accuracy scoring 53% and 64%, respectively.

On the contrary, we observe low accuracy measure on topic modeling results compared to N-Grams results. For instance, LR and NN scored 59% and 63%, the highest performance with higher MAE 1.08 and 0.93, respectively. A discrepancy between the calculated performance for N-Grams (NG) and Topic modeling (TB) can be explained by the text length where classifiers tend to perform better on shorter text. KNN scored 57% on accuracy for both techniques. NN and LR scored the highest accuracy for Topic modeling. KNN produced 57% in accuracy compared to lower accuracy when it is applied on N-Grams. Finally, CART, SVM, and NB produced the worst accuracy with low variance among each other, 52% and 53%. This finding confirms previous research findings that Naïve Bayes is very sensitive to the dataset [43].

Table 3 shows Precision, recall, and F-measure of each proposed classifier. We compared the results when using N-grams and Topic Modeling as different techniques for features extraction. As explained earlier, N-Grams produce short text containing 2-3 words. In contrast, topic modeling produces different topics where each topic consists of several words. We experienced a high discrepancy between the two results produced by N-gram and topic modeling. In N-Grams, we achieved the highest precision of 69% and highest recall for LR. NN achieved the second highest precision 67% and 68% in

recall. NB performed 66% and 73% in precision and recall. Finally, KNN, CART, and SVM range between 53% and 61% for Precision and between 53% and 57% for recall.

Our N-Grams classifiers performance seems promising given the experienced limitation in the extracted reviews. For instance, teens' reviews tend to be very short compare to parents' reviews which can be hard for classifiers to identify the correct pattern. Additionally, some apps had a larger number of reviews compared to others. Consequently, high variance can be achieved within the dataset which can diminish the classification accuracy. The other category was performed on topic modeling achieved the range between 45% and 59% for precision and recall in the following classifiers: LR, KNN, CART, and SVM. NB reported the lowest precision 28%. Finally, we investigated the misclassification issues in the topic modeling analysis and found that the variability of the used vocabulary by different users can be a significant factor in achieving lower scores in precision and recall. This finding of the low precision and recall in topic modeling is consistent with previous study [46].

4.3 Contrasting Parent versus Child Reviews

To understand the different themes expressed in the reviews by parents and teens (RQ2), we compared the results of N-grams and Topic Modeling as different techniques for features extraction. As shown in Table 7, we then generated topics based on parent versus child and high versus low rated reviews (Low: 1-3 ratings; High: 4-5) to understand the key differences in these reviews.

Latent themes emerged from the data to reveal differences between parent and child reviews. The topics demonstrated a relationship between apps rating scores and the review themes for both parents and children.

High parental ratings accounted for 54% of reviews. Parent reviews tended to range from one complete sentence to more than 5 sentences. Positive reviews focused on the app's ability to protect the online safety of their child. For instance, one positive review explained, "*I can monitor everything my son does.*" Low parental ratings accounted for 17% of reviews. Negative reviews were associated with concerns such as functionality, installation, licensing, and cost. In one example, the parent wrote, "*Keeps crashing after update making my phone unusable because it takes forever to get the program to close and you are locked out of everything.*"

In contrast, child reviews tended to be short sentence fragments emoting anger and frustration towards their parent. High child ratings represented only 5% of reviews. The few positive reviews from children showed that they appreciate some of the app's features. For instance, one child explained, "*I'm 9...with kid search it has kid friendly things that work for my age! Keep up!*" These reviews also suggested that some children understood their parents' concerns regarding their safety and the negative effects of technology overuse. Keywords such as "*safe, help, addicted*" appeared in several topics. Low child ratings comprised 24% of the reviews and included emotional charged words, such as "*Hate it,*" "*F you,*" "*sucks,*" "*stupid,*" "*dumb,*" and "*bad.*" Topics in this group often reflected a child's frustration regarding privacy violations by their parents and limits on their freedom.

Table 7. Parent and child topics under high and low app rating.

Parents: High Ratings	Child: High Ratings
<ol style="list-style-type: none"> 1. Phone, app, games, chores, times, earn, daughters, knowing, downloaded, amount 2. Time, limit, usage, tasks, limits, helps, track, extra, helped, helpful 3. Apps, access, devices, give, year, lock, tablets, web, content, ability 4. Daughter, good, works, found, free, home, find, check, perfect 5. Son, tablet, app, love, mind, online, settings, manage, peace, keeping 6. Phone, child, safe, play, don, things, children, block, phones, worry 7. App, children, internet, recommend, highly, activity, problem, happy, add, service 8. App, child, great, parents, parent, device, features, feel, android, protect 9. Control, monitor, easy, great, work, make, installed, parental, feature, love 10. Kids, screen, set, ve, day, school, long, hours, back, put 	<ol style="list-style-type: none"> 1. Works, hate, screen, games, day, year, bad, isn, website, sad 2. Dad, installed, things, doesn, happy, im, date, helps, lol, block, 3. App, recommend, lot, homework, likes, mind, step, helped, likes, heck, understand 4. Tablet, play, limit, make, playing, life brother, glad, delete, quot 5. Love, kids, years, kid, device, find, teen nice, downloaded, control, 6. Phone, kid, stop, thing, found, hope, won, addicted, cousin, face 7. Good, mom, great, loves, apps, spend dang, fix, world, back 8. Parents, don, put, pretty, settings, usage, didn, airplane, setting, review 9. Time, safe, made, hour, track, mobile, control, change, manage, tab 10. App, child, work, password, deleted, give, blocked, mad, hey, feature
Parents: Low Ratings	Child: Low Ratings
<ol style="list-style-type: none"> 1. Time, son, good, day, useless, hours, password, application, change, show 2. App, doesn, games, parent, mode, buy, allowed, blocking, sites, kid 3. Child, great, blocked, home, place, open, site, videos, button, received 4. Device, uninstall, times, back, google, trial, don, message, log, find 5. App, update, working, free, data, year, features, school, monitoring, stopped 6. Son, work, location, give, settings, days, norton, pay, android ll 7. Apps, tablet, works, won, screen, control, play, account, phones, fix 8. Children, access, installed, kids, block, worked, monitor, stars, make, problem 9. Phone, put, don, locked, service, call, thing, email, long, issue 10. Daughter, quot, set, lock, support, ve, version, paid, didn, samsung 	<ol style="list-style-type: none"> 1. Parents, don, sucks, people, school, friends, makes, anymore, youtube, talk 2. Life, child, privacy, thing, children, feel, hour, parent, app, text 3. Hate, download, stuff, horrible, person, net, wont, great, forever, times 4. Kids, apps, blocked, good, control, screen, freedom, block, things, stop 5. Time, dad, stupid, doesn, made, uninstall, work, worst, easy, game 6. Phone, put, parents, day, hours, making, gonna, dumb, unlock, set 7. App, quot, trust, play, install, dont, delete, lot, teen, ve 8. Kid, tablet, im, games, won, make, bad, downloaded fix, internet 9. Mom, device, settings, blocks, google, mode, safe, watch, teens, volume 10. App, password, installed, kid, give, minutes, didn, star, stalking, back

These quotes highlight how teens are not satisfied with the apps being installed on their devices. On the other hand, parents expressed satisfaction or positive feedback. For instance, “safe online remote” and “Good app children” may explain a positive experience with an app’s features. Table 4 contains examples of the extracted topics using LDA for common parent and teen topics from high and low rating reviews.

5 Discussion

5.1 Parents and Children Write Reviews in Different Ways

Our analysis addresses the parent and teen communities' perspectives on parental control app reviews which range from enjoyment and satisfaction to sadness and displeasure. Parents reviews are largely found to be long and complete, varying in the range of one complete sentence to more than five sentences. For instance, one complete sentence may explain an app's feature, "I can monitor everything my son does." Complete reviews of an app with a 5-star may highlight elements that the developer has designed well, for instance:

"I can now let my son uses my phone without worrying if he is going to get into something he shouldn't! I also love how easy the app was to set up! I cannot recommend this highly enough for anyone that has children or works around children." – Parent, Parental Control by Familoop, 2016

This parent praised the apps ability to alleviate their worries about what their son was looking at on his phone. Furthermore, it indicates that the app was easy to set up. The parent is happy with the effectiveness of the app and its initial usability. Thus, effectiveness and ease of use are elements that will engender a positive experience in parents and should be noted by developers.

Parental reviews with a 1-star rating often remark on their dissatisfaction with the app. These types of reviews tend to be longer given that the parent may want to justify their rating, for example:

"I have had sooooooo many issues with this application! It has week days/ends mixed up, the timer doesn't work properly with games, it's a day behind in its reporting, etc. Those issues I have come to live with because at least it blocks inappropriate apps. NOPE! The last straw was when I found out today that my son has FULL access to the Internet even though I have it all blocked with this app. I'm talking FULL ACCESS! PORN GALORE! Do NOT trust this application!" – Parent, ESET Parental Control, 2016

This example demonstrates the types of frustrations a parent may have using parental control apps. Simple UI elements like the calendar and timers are malfunctioning. This may be indicative of two scenarios. In one, the app developer lacked adequate quality controls and shipped a product that is malfunctioning. In the other, the apps usability may not be intuitive or learnable enough for parents of various technological backgrounds. Distinguishing between parent and teen reviews may help inform developers on how to design effective UI elements for both parent and teen users. New designs can then be user tested by parents and teens separately to ensure that the needs of both user groups are being met.

While these reviews suggest that parents are eager to share their positive and negative experiences, they tend to not share their teens frustration or displeasure.

Positive reviews by teens accounted for only 5% of the total reviews, compared to 54% by parents. This suggests that teens are having fewer positive experiences with the parental control apps. Indeed, teen reviews often feature expressions of anger related to restrictive features. Some examples include short descriptions such as “Hate it”, “F you”, and “Cuz I am child”. However, teens also admit that control apps can be helpful, but some features should be improved:

“I am a kid. I used to be on my phone all the time but this app got me up and out. Now though, it glitches and says I've been on my phone for 11 hours when I only play on it on the bus, which is an hour max. It also doesn't let me respond to texts when time is used up. I also cannot get on contacts without having my parents unblock it. Still is a great app though. Hope this glitch will be fixed soon”. – Teen, Screen Time Companion App, 2015

In addition to expressing their frustrations with the control app's restrictions, reviews by teens were found to be shorter than those of their parents. Despite their shortened length, however, these shortened reviews may reveal additional security concerns not initially considered by the developers. For instance, in the quote below, the teen highlights the possibility of their parent's phones being stolen. Criminals with access to the parent's phone may also have access to critical information regarding their teen.

“Freaking hate this. It's bullshit. My parents are hacking me. No one get this all. It's more safe without it. Imagine if some one got hold of their phones. It's bullshit.” –Teen, Secure Teen Parental Control, 2015

One review revealed that parental control apps can contribute to the increased toxic relationship between teens and their parents, while also exacerbating other social issues:

“Im 15. my dad got this app just to limit time on my phone. I have no problem with that and i agree that i use my phone too often. but how you can restrict apps is the worst. i could have a really nice conversation with a new person i met at school. not anymore. i have a social problem and texting helps me talk to people. well now im screwed. my friends dont want to text me anymore because they know my dad can see my messages. I am not even gonna start on not having a wifi signal because its such bullshit...” –Teen, Screen Time Companion App, 2015

This review suggests that teens may be understanding of the parent's desire to control their mobile phone usage but disagree to the extent to which their behaviors are restricted. This not only creates tension between the teen and the parent, but also limits the teen's ability to socialize according to current conventions of their age group. The latter may lead to a sense of alienation. Understanding the needs and desires of teen mobile users could potentially avoid this conflict by way of curating restrictions based on the varying interpersonal dynamics of parents and teens.

In general, our results show that teens were open to communicate and share their frustrations where it seems like there is a lack of communication with their parents

when it comes to privacy issues. Teens demand privacy and more autonomy as they feel more restricted and disclosed by installing these apps.

5.2 Reviews Reveal Relational Tensions between Parents and Children

Topic modeling revealed additional insights into the relationship between the extracted features and app rating. The three groups in topic modeling, tables 5-7, show different patterns for low, medium, and high rating apps. For instance, low rating apps tend to be mostly negative and include keywords such as mom, dad, block, hate, privacy, horrible, stupid, and ruin. Many of these keywords represent teens expressing their anger and irritation regarding the apps. Some of these keywords such as 'block' or 'blocked' occur in low rating reviews by both parents and teens.

However, in light of the explicit quotes examined in section 5.1, it is likely that these words are being used by each group differently. That is, parents are going to use the word blocked in a negative review if the app failed at blocking the teen's mobile usage. Whereas a teen is likely to use it in a negative review when it successfully blocks their access. Topics in high rating apps are similar between both user groups with keywords such as 'help' or 'helped' and 'safe.' While tensions are likely to occur between parents and teens, in many cases the app was able to help the family solve problems regarding their safety and that these safety concerns were understood by both parties. It is important, then, for developers to search for common needs that overlap between the two user groups to design effective solutions.

These findings have implications beyond classifying parent and teen reviews based on their linguistic factors. In many cases, topic modeling revealed that the underlying themes within the reviews went beyond a description of the app, its features, or its performance. Instead, reviews were often an expression of the relationship between parents and teens as mediated through parental control apps. Thus, the written component of a review appears far more important than a quantitative rating of app usability, and more, a valuable signal of the underlying parent-teen relationship. Future studies should focus on review content as an important indicator of understanding these relationships.

Our work is consistent with previous studies where N-Grams outperformed other techniques due to the length of the extracted text [46]. Topic modeling and N-Grams helped to generate some labels related to different domains including design, privacy, license, and app costs. These types of analyses can be used to inspire designers to embrace new communication strategies so users can be pro-active in sharing their experience. Finally, our study found that both teens and parents are willing to explain the reasoning behind their rating. This can be demonstrated in the three groups as each one may represent different categories. One implication of this finding is that both teens and parents are encouraged to communicate and share their thoughts.

These analyses are an important source of information for apps developers to improve the quality of the developed apps. The applied techniques and generated features assessed the model to improve the performance accuracy for the six machine learning classifiers.

5.3 Implications for the Future Design of Parental Control Apps

Key insights that arose from our results may help us shape the future design of parental control apps. Our results showed that parents and children liked and disliked the currently available apps for different reasons. Although most parents were generally positive about the apps, they were mainly concerned about apps cost, license, bugs, and functionality issues. So, app designers need to make sure that: (1) their apps offer free and low-cost versions, (2) apps are bug-free, and (3) they provide tech support for parents who may not be tech savvy or intuitive help documentation.

Teens, on the other hand, posted positive reviews when they felt that the apps helped them break addictive patterns and better manage their screen-time. However, they more often left low ratings because of how the parental control app negatively changed the relationship dynamic between themselves and their parents. Stalking, restriction, and privacy were common themes among the low-rated child reviews. This finding raises the question of how parental control apps might be designed in a way that is more supportive of nurturing positive parent-teen relationships while still ensuring a teen's online safety? To do this, developers and researchers should pro-actively embrace direct interactions with teen users for more feedback. Doing so will provide additional clarification regarding the teen's concerns. Teens need to have their voice heard being a major stakeholder in the design process. What's more, giving teens a voice in the design process will allow for the development of parental control apps that respect a teen's need for autonomy and privacy, while providing security that parents seek. This can benefit the teens' mobile experience while also facilitating more positive relationships between teens and their parents.

5.4 Limitations and Future Research

Several limitations should be considered while interpreting the reported results. First, our topic modeling Parameter K were set to be 10, based on common convention derived from our observation of each group's size. This result can change in the case of different parameters. Second, our analysis was based on 52 parental control apps found on Google Play with large variance of the number of reviews for each app. Finally, the extracted reviews for teens were small compare to the parents' reviews, so results could differ in future studies with more teen reviews. Therefore, we suggest that future research consider validating the generalizability of our findings across different platforms (i.e., iOS) and a wider range of adolescent online safety apps to see if the patterns we uncovered hold in these new contexts. We also encourage social computational researchers to work with qualitative researchers to find synergistic ways to meaningfully analyze large data sets from the strengths of both perspectives.

6 Conclusion

Our N-Grams and Topic Modeling analyses revealed new insights into the relationship and tensions. These analyses are an important source of information for analysts and apps developers to improve the quality of the developed apps between parents and

children by applying computational methods to parental control app reviews. A key contribution of this work is that we integrated domain knowledge into computational models for empirical validation at a reasonable scale. Yet, these findings have implications beyond classifying parent and child reviews based on their linguistic factors. In many cases, topic modeling revealed that the underlying themes within the reviews went beyond a description of the app and its features or performance, and more towards an expression of the relationship between parents and teens as mediated through parental control apps. Thus, reviews seem to be far more important than a quantitative rating of app usability and more, a valuable signal of the underlying parent-teen relationship. These insights can be used to improve parental control app design, and therefore the user experience of both parents and children.

References

1. Blackwell, L., Gardiner, E., Schoenebeck, S.: Managing Expectations: Technology Tensions Among Parents and Teens. In: Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing. pp. 1390–1401. ACM, New York, NY, USA (2016a)
2. Boyd, D.: It's Complicated: The Social Lives of Networked Teens. Yale University Press, New Haven (2014)
3. Anderson, M.: Parents, Teens and Digital Monitoring, <http://www.pewinternet.org/2016/01/07/parents-teens-and-digital-monitoring/>, (2016)
4. Baumrind, D.: A developmental perspective on adolescent risk taking in contemporary America. *New Directions for Child and Adolescent Development*. 1987, 93–125 (1987). doi:10.1002/cd.23219873706
5. Wisniewski, P., Ghosh, A.K., Rosson, M.B., Xu, H., Carroll, J.M.: Parental Control vs. Teen Self-Regulation: Is there a middle ground for mobile online safety? In: Proceedings of the 20th ACM Conference on Computer Supported Cooperative Work & Social Computing. ACM, Portland, OR, USA (2017)
6. Ghosh, A.K., Badillo-Urquiola, K., Guha, S., LaViola Jr., J., Wisniewski, P.: Safety vs. Surveillance: What Children Have to Say about Mobile Apps for Parental Control. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. ACM, New York, NY, USA (2018)
7. Ashktorab, Z., Vitak, J.: Designing Cyberbullying Mitigation and Prevention Solutions Through Participatory Design With Teenagers. In: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems. pp. 3895–3905. ACM, New York, NY, USA (2016)
8. Hashish, Y., Bunt, A., Young, J.E.: Involving Children in Content Control: A Collaborative and Education-oriented Content Filtering Approach. In: Proceedings of the 32Nd Annual ACM Conference on Human Factors in Computing Systems. pp. 1797–1806. ACM, New York, NY, USA (2014)
9. Wisniewski, P., Jia, H., Wang, N., Zheng, S., Xu, H., Rosson, M.B., Carroll, J.M.: Resilience Mitigates the Negative Effects of Adolescent Internet Addiction and Online Risk Exposure. In: Proceedings of the 33rd Annual ACM Conference on

- Human Factors in Computing Systems. pp. 4029–4038. ACM, New York, NY, USA (2015)
10. Wisniewski, P., Xu, H., Rosson, M.B., Perkins, D.F., Carroll, J.M.: Dear Diary: Teens Reflect on Their Weekly Online Risk Experiences. In: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems. pp. 3919–3930. ACM, New York, NY, USA (2016)
 11. Yardi, S., Bruckman, A.: Social and Technical Challenges in Parenting Teens' Social Media Use. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. pp. 3237–3246. ACM, New York, NY, USA (2011)
 12. Yardi, S., Bruckman, A.: Income, Race, and Class: Exploring Socioeconomic Differences in Family Technology Use. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. pp. 3041–3050. ACM, New York, NY, USA (2012)
 13. Hiniker, A., Schoenebeck, S.Y., Kientz, J.A.: Not at the Dinner Table: Parents' and Children's Perspectives on Family Technology Rules. In: Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing. pp. 1376–1389. ACM, New York, NY, USA (2016)
 14. Cranor, L.F., Durity, A.L., Marsh, A., Ur, B.: Parents' and Teens' Perspectives on Privacy In a Technology-Filled World. In: Proceedings of the Tenth Symposium On Usable Privacy and Security. USENIX, Menlo Park, CA (2014)
 15. Livingstone, S.: Taking risky opportunities in youthful content creation: Teenagers' use of social networking sites for intimacy, privacy, and self-expression. *New Media & Society*. 10, 393–411 (2008)
 16. Nissenbaum, H.: PRIVACY AS CONTEXTUAL INTEGRITY. *Washington Law Review*. 79, (2004)
 17. Petronio, S.S.: *Boundaries of Privacy: Dialects of Disclosure*. SUNY Press (2002)
 18. Zaman, B., Nouwen, M.: Parental controls: advice for parents, researchers and industry, <http://www.lse.ac.uk/media@lse/research/EUKidsOnline/Home.aspx>
 19. Czeskis, A., Dermendjieva, I., Yapit, H., Borning, A., Friedman, B., Gill, B., Kohno, T.: Parenting from the Pocket: Value Tensions and Technical Directions for Secure and Private Parent-teen Mobile Safety. In: Proceedings of the Sixth Symposium on Usable Privacy and Security. pp. 15:1–15:15. ACM, New York, NY, USA (2010)
 20. Ghosh, A.K., Badillo-Urquiola, K., Rosson, M.B., Xu, H., Carroll, J., Wisniewski, P.: A Matter of Control or Safety? Examining Parental Use of Technical Monitoring Apps on Teens' Mobile Devices. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. ACM, New York, NY, USA (2018)
 21. Amato, G., Bolettieri, P., Costa, G., Torre, F., Martinelli, F.: Detection of Images with Adult Content for Parental Control on Mobile Devices? In: Proceedings of the 6th International Conference on Mobile Technology, Application & Systems. pp. 35:1–35:5. ACM, New York, NY, USA (2009)
 22. Ko, M., Choi, S., Yang, S., Lee, J., Lee, U.: FamiLync: Facilitating Participatory Parental Mediation of Adolescents' Smartphone Use. In: Proceedings of the 2015

- ACM International Joint Conference on Pervasive and Ubiquitous Computing. pp. 867–878. ACM, New York, NY, USA (2015)
23. Kim, S., Zhang, J., Chen, Z., Oh, A., Liu, S.: A Hierarchical Aspect-sentiment Model for Online Reviews. In: Proceedings of the Twenty-Seventh AAAI Conference on Artificial Intelligence. pp. 526–533. AAAI Press, Bellevue, Washington (2013)
 24. Mason, R., Gaska, B., Durme, B.V., Choudhury, P., Hart, T., Dolan, B., Toutanova, K., Mitchell, M.: Microsummarization of Online Reviews: An Experimental Study. In: Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence. pp. 3015–3021. AAAI Press, Phoenix, Arizona (2016)
 25. Michael, L., Otterbacher, J.: Write Like I Write: Herding in the Language of Online Reviews. In: Proceedings of the 8th International Conference on Weblogs and Social Media, ICWSM 2014. pp. 356–365 (2014)
 26. Wang, J., Zhao, W.X., He, Y., Li, X.: Leveraging Product Adopter Information from Online Re-views for Product Recommendation. In: Ninth International AAAI Conference on Web and Social Media. AAAI Publications (2015)
 27. Ida, Y., Nakamura, T., Matsumoto, T.: Domain-dependent/Independent Topic Switching Model for Online Reviews with Numerical Ratings. In: Proceedings of the 22Nd ACM International Conference on Information & Knowledge Management. pp. 229–238. ACM, New York, NY, USA (2013)
 28. Zhao, W.X., Wang, J., He, Y., Wen, J.-R., Chang, E.Y., Li, X.: Mining Product Adopter Information from Online Reviews for Improving Product Recommendation. *ACM Trans. Knowl. Discov. Data.* 10, 29:1–29:23 (2016). doi:10.1145/2842629
 29. Moghaddam, S., Jamali, M., Ester, M.: Review Recommendation: Personalized Prediction of the Quality of Online Reviews. In: Proceedings of the 20th ACM International Conference on Information and Knowledge Management. pp. 2249–2252. ACM, New York, NY, USA (2011)
 30. Jawecki, G., Fuller, J.: How to use the innovative potential of online communities? Netnography – an unobtrusive research method to absorb the knowledge and creativity of online communities. *International Journal of Business Process Integration and Management.* 3, 248–255 (2008). doi:10.1504/IJBPIIM.2008.024982
 31. Korfiatis, N., García-Bariocanal, E., Sánchez-Alonso, S.: Evaluating content quality and helpfulness of online product reviews: The interplay of review helpfulness vs. review content. *Electronic Commerce Research and Applications.* 11, 205–217 (2012). doi:10.1016/j.elerap.2011.10.003
 32. Vasa, R., Hoon, L., Mouzakis, K., Noguchi, A.: A Preliminary Analysis of Mobile App User Reviews. In: Proceedings of the 24th Australian Computer-Human Interaction Conference. pp. 241–244. ACM, New York, NY, USA (2012)
 33. Duan, W., Gu, B., Whinston, A.B.: Do online reviews matter? — An empirical investigation of panel data. *Decision Support Systems.* 45, 1007–1016 (2008). doi:10.1016/j.dss.2008.04.001
 34. Epstein, D.A., Lee, N.B., Kang, J.H., Agapie, E., Schroeder, J., Pina, L.R., Fogarty, J., Kientz, J.A., Munson, S.: Examining Menstrual Tracking to Inform

- the Design of Personal Informatics Tools. In: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems. pp. 6876–6888. ACM, New York, NY, USA (2017)
35. Fu, B., Lin, J., Li, L., Faloutsos, C., Hong, J., Sadeh, N.: Why People Hate Your App: Making Sense of User Feedback in a Mobile App Store. In: Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. pp. 1276–1284. ACM, New York, NY, USA (2013)
 36. McCallum, A.K.: MALLET: A Machine Learning for Language Toolkit. (2002)
 37. Guyon, I., Elisseeff, A.: An Introduction to Feature Extraction. In: Feature Extraction. pp. 1–25. Springer, Berlin, Heidelberg (2006)
 38. Brown, P.F., deSouza, P.V., Mercer, R.L., Pietra, V.J.D., Lai, J.C.: Class-Based n-gram Models of Natural Language. Presented at the Computational Linguistics (1992)
 39. Blei, D.M.: Latent Dirichlet Allocation. 30
 40. Guha, S., Baumer, E.P.S., Gay, G.K.: Regrets, I've Had a Few: When Regretful Experiences Do (and Don'T) Compel Users to Leave Facebook. In: Proceedings of the 2018 ACM Conference on Supporting Groupwork. pp. 166–177. ACM, New York, NY, USA (2018)
 41. Muller, M., Guha, S., Baumer, E.P.S., Mimno, D., Shami, N.S.: Machine Learning and Grounded Theory Method: Convergence, Divergence, and Combination. In: Proceedings of the 19th International Conference on Supporting Group Work. pp. 3–8. ACM, New York, NY, USA (2016)
 42. Baumer, E.P.S., Mimno, D., Guha, S., Quan, E., Gay, G.K.: Comparing grounded theory and topic modeling: Extreme divergence or unlikely convergence? *Journal of the Association for Information Science and Technology*. 68, 1397–1410 (2017). doi:10.1002/asi.23786
 43. Alelyani, T., Mao, K., Yang, Y.: Context-Centric Pricing: Early Pricing Models for Software Crowdsourcing Tasks. In: Proceedings of the 13th International Conference on Predictive Models and Data Analytics in Software Engineering. pp. 63–72. ACM, New York, NY, USA (2017)
 44. Friedman, J.H.: On Bias, Variance, 0/1—Loss, and the Curse-of-Dimensionality. *Data Min. Knowl. Discov.* 1, 55–77 (1997). doi:10.1023/A:1009778005914
 45. Friedman, N., Geiger, D., Goldszmidt, M.: Bayesian Network Classifiers. *Machine Learning*. 29, 131–163 (1997). doi:10.1023/A:1007465528199
 46. Ciurumelea, A., Schaufelbühl, A., Panichella, S., Gall, H.C.: Analyzing reviews and code of mobile apps for better release planning. In: 2017 IEEE 24th International Conference on Software Analysis, Evolution and Reengineering (SANER). pp. 91–102 (2017)