Published in final edited form as:

J Adolesc Health. 2011 July ; 49(1): 29–35. doi:10.1016/j.jadohealth.2011.04.015.

Development of a researcher codebook for use in evaluating social networking site profiles

Megan A Moreno, MD, MEd, MPH¹, Katie G Egan¹, and Libby Brockman, BS^{2,3}

¹University of Wisconsin, Madison, WI

²University of Washington, Seattle, WA

³Seattle Children's Hospital, Seattle, WA

Abstract

Introduction—Social networking sites (SNSs) are immensely popular and allow for display of personal information, including references to health behaviors. Evaluating displayed content on a SNS for research purposes requires a systematic approach and a precise data collection instrument. The purpose of this paper is to describe one approach to the development of a research codebook so that others may develop and test their own codebooks for use in SNS research.

Methods—Our SNS research codebook began with a basis in health behavior theory and clinical criteria. Key elements in the codebook developmental process included an iterative team approach and an emphasis on confidentiality.

Results—Codebook successes include consistently high interrater reliability. Challenges include time investment in coder training and SNS server changes.

Conclusion—It is our hope that this paper will provide detailed information about one systematic approach to codebook development so that other researchers may use this structure to develop and test their own codebooks for use in SNS research.

INTRODUCTION

Social networking sites and adolescent health

Social networking sites (SNSs) are extremely popular, particularly among adolescents and young adults [1]. It is estimated that over 90% of adolescents have internet access; approximately 70% of adolescents and more than 90% of college students maintain a SNS profile [1–3]. Facebook is currently the most popular SNS and recently surpassed Google for highest total number of weekly hits to any website [4, 5]. The popularity of SNSs may be explained in part by their ability to combine multiple functions into a single web site. SNSs allow members to build a personal web profile, to communicate with others and to build a

© 2011 Society for Adolescent Medicine. Published by Elsevier Inc. All rights reserved.

Corresponding author: Megan A. Moreno, MD, MEd, MPH, University of Wisconsin, Department of Pediatrics, Adolescent Medicine Section, 2870 University Ave, Suite 200, Madison, WI 53705, p: (608) 263-3710 f: (608) 263-0503, mamoreno@pediatrics.wisc.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

The authors would like to recognize the contributions of the following people to the development of the codebook described in this manuscript: Dimitri Christakis, MD, MPH; Malcolm Parks, PhD; Laura Richardson, MD, MPH, Michael Fleming, MD, MPH and Deborah Dickstein, MSW.

social network. Profiles users may incorporate text, photos, graphics, audio and video to build a personal representation of their identity.

In creating a SNS profile adolescents may give researchers and health care providers insight into aspects of their behavior that are not always apparent in offline life. Adolescents report that they often disclose more about themselves on SNSs than they do in person [6]. These disclosures may include references to personal or stigmatizing information, such as references to health risk behaviors including substance use or sexual behaviors. Previous work has found that adolescents' SNS profiles frequently include references to such behaviors through text descriptions and photographs [7–9].

These online displays of health risk behaviors are important as they give researchers and providers new insight into adolescent behaviors that are clearly linked to morbidity and mortality among this population. The three major causes of mortality among adolescents are accidents, homicide and suicide; all of these have been associated with health risk behaviors as precursors [10–12]. Identification of adolescents who at risk for or are engaging in health risk behaviors remains challenging. Many adolescents do not visit health care providers routinely; among the adolescents who do visit health care providers, many are not asked about health risk behaviors [13, 14]. Thus, complementary approaches for identification of health risk behaviors in the adolescent population are needed. SNSs may represent a new approach to enhance our understanding and identification of health risk behaviors in adolescents. In order to establish SNSs as a credible approach to evaluating health risk behavior displays, a framework for consistency and reliability in evaluation and measurement of these web profiles is essential. The purpose of this paper is to describe an approach to the development and use of a research codebook to evaluate adolescents' displayed references to health risk behaviors on SNSs. Our intention is for this paper to provide a framework on which researchers or clinicians may further develop and adapt their own codebooks for use with SNS research. Given that SNSs may rise and fall in popularity, our goal is to provide a framework that is usable and adaptable to researchers as new technologies and new health topics of interest emerge over time.

Theoretical foundation of the codebook

In developing our codebook we first investigated theoretical models to guide its structure so that we could appropriately capture data relevant to adolescents' health risk behaviors. The Media Practice Model describes that adolescents choose and interact with media based on who they are, or who they want to be, at the moment [15]. This model influenced our interest in viewing SNSs as representations of adolescents' actual as well as desired identities. We began to consider that displays on SNSs may represent not just events or behaviors, but also intentions towards behaviors.

A second theoretical model that influenced our codebook was the Social Cognitive Theory (SCT). This theory predicts that adolescents who observe media characters engaging in behaviors such as alcohol use without experiencing negative consequences will be more likely to adopt the behaviors portrayed [16–19]. SCT further explains that media messages that contain relevant motivations, desirable associations and positive consequences are more likely to promote adoption of the referenced behavior. This model was considered as we developed coding schemes to represent adolescents' varied descriptions of behaviors to include references to motivations, associations as well as consequences as depicted on SNSs.

Origins of the codebook

Our codebook development was further influenced by a review of the literature for qualitative methods. As qualitative data can provide “thick description” of experiences or events we wanted to maximize the reliability in collecting and interpreting such data [20]. We followed other investigators’ strategies for interpreting qualitative interview data such as holding recurrent team meetings to discuss and further develop the coding structure as it evolved [21].

Finally, in developing codes for criteria reflecting health risk behaviors we consulted established clinical criteria. Basing the coding scheme on clinical criteria, our intention was to maximize the clinical application of data obtained via SNSs and to promote translation of findings into further clinical research. To promote consistency in coding we wanted to avoid individual coders needing to apply personal or subjective judgment to references. Our goal was to create coding schemes based on clear criteria that coders could frequently reference while they were evaluating SNS profiles. One example of our application of clinical criteria to our coding schemes is the use of the CRAFFT problem drinking criteria in evaluation of alcohol reference coding [22]. The CRAFFT problem drinking criteria provide precise criteria that are used to define problem drinking in the adolescent population, such as drinking and driving, or forgetting what one did while drinking.

CODEBOOK DEVELOPMENT

Confidentiality

In creating the codebook, our first objective was to address ethical concerns that may arise in evaluating data from SNS profiles. We met with our university’s Institutional Review Board (IRB) several times at the start of and during the creation of the codebook. As per recommendations by the IRB, a priority area was to maintain the confidentiality of the data obtained from profiles. We designed the codebook so that names and contact information were not collected as identifiers. In early studies using our codebook, we tracked each profile’s unique URL addresses (i.e. “www.myspace.com/837201738493”) to avoid duplication of coding. As soon as data collection was complete this information was deleted. This attention to confidentiality has also been applied to any papers or presentations that are based on data from our codebooks: when writing papers or creating presentations that include quotes from participants’ profiles, quotes are always slightly altered so that an online search will not match the quote to the participant’s profile. Additionally, when displaying photographic examples in papers or presentations, these are always altered or de-identified so as not to identify the participant. In developing these standards we worked with a pediatric ethicist to promote application of ethical guidelines to our data collection methods. Through our discussions with human subjects committees and a pediatric ethicist, we helped develop a manuscript describing how research on publicly available SNS profiles may be analogous to research on published personal ads in newspapers. In both cases, the information displayed may be quite personal and not intended to be for research purposes. However, in both cases the information has been selected by its owner to be published in a forum accessible to the public.[23]

Profile security

We then determined how we would approach profile privacy settings. SNS profiles may be “publicly available” such that profile content is accessible by the entire online social network, or “privately available” in which profile content is only available with explicit permission of the profile owner. Most of our studies have focused on SNS profiles that are publicly available. As our goals include investigating SNSs as a potential innovative venue for health interventions, we were interested in profiles that could be accessed by any online

web user. Thus, we were interested in adolescents' profiles that would be available to the their friends, as well as teachers, parents or even health care providers in consideration that any future intervention efforts could potentially involve one or more of these groups.

Inclusion and exclusion criteria

The next consideration was to define what criteria would guide profile inclusion and exclusion, and determined how the codebook could accurately track profile exclusions. The first inclusion criteria considered was the availability of data on the profile. In designing our initial codebook, and in most of our studies to date, we have included only publicly available profiles as described above [9]. As privacy settings grow increasingly complex, our codebooks now include guidelines on which sections of information must be accessible on the profile for the profile to be included in the study. Second, our codebooks were designed to target profiles that are still actively used by their owners. Thus, codebooks were designed to include an evaluation of a specific timeframe during which evidence of profile use must be present in order to consider the profile as still actively maintained by its owner. Other inclusion criteria pertain to information displayed on the profile regarding age, gender, language, home location or other demographic factors. Once inclusion categories were determined, a method of tracking exclusions was needed within the codebook. We designed the codebook to allow for tracking numbers of exclusions, categories of exclusions and to prevent duplications of exclusions using the URL as a unique identifier. An example of an inclusion/exclusion flow chart is included as Figure 1.

Coding variables of interest

We then defined our variables of interest and began developing coding schemes for these variables. This was done in an iterative process. The first step was establishing whether clinical criteria were present to operationally define a given variable for the codebook. For example, the CRAFFT criteria, used to identify problem alcohol use among adolescents, were directly applied to SNS profiles to code "problem alcohol use"[22]. In situations in which clinical criteria were not available, such as in coding for "alcohol use" we consulted dictionaries, medical textbooks and thesauruses to develop initial coding keywords and phrases.

The second step in developing these coding schemes involved applying the codebook to a sample of SNS profiles. Often this coding resulted in identification of further keywords and synonyms linked to the variable of interest. One example was identification of the words "wasted" or "drunk" as additional representations of the variable "intoxication". The third step involved team discussions of both our findings from profile content, as well as notes on the evaluation process itself. If modifications were suggested, the process began again with the coding of a practice set of profiles and then presenting findings and notes on the process to our research group. We worked intensively with a Communication professor with background in Internet studies to develop codes during this stage [24, 25].

Following this iterative process the codebook included the following for each variable: a description of the variable including any relevant clinical criteria, a list of example terms that represent the variable and a list of example terms that might be considered but that do not represent the variable. One example is in evaluation of profile photographs for alcohol use, Figure 2 is an example Facebook profile which features commonly displayed references to alcohol (this profile has been altered to protect the profile owner's identity). Photographs that clearly label the profile owner holding a marked beverage identifiable as alcohol (i.e. a beer bottle) would meet criteria as an alcohol reference. Photographs that feature the profile owner holding a red plastic cup (often used when drinking from a keg) would not be considered an alcohol reference in our codebook because it is unclear what that cup is

holding. Our codebook was designed to be conservative and always give the benefit of the doubt to the profile owner.

We then established further details of the codebook. These included in what time period the profile should be evaluated. For example, some profile owners have had a Facebook profile for several years; investigators using the codebook would need to know how far back in time to conduct a profile evaluation. The codebook also included instructions regarding which comments on the SNS profile should be evaluated (comments from the profile owner only versus comments from other profile viewers). The codebook also needed to address which personal photographs should be evaluated; Facebook includes several photograph categories such as profile pictures, photo albums and tagged pictures. As Facebook profiles can include thousands of photographs, clear criteria for how many individual photographs should be viewed, and how to ensure that the photograph is representing the profile owner, were key considerations in our codebook development. For example, in a recent study we only evaluated the “tagged photographs” section in which all photographs are clearly labeled with names attached to each person in the photograph. In doing so, we could ensure that the photograph we were evaluating was indeed one of the profile owner. After all criteria were developed, we established the process of coding profiles. The codebook was designed so that sections of the profiles are evaluated in the same order each time. Variables are presented in the codebook are presented in the same prescribed order that the coder follows during the profile evaluation process.

Codebook layout

We designed our codebook in Microsoft Excel®. This has been a useful tool as it allowed us to separate several layers of the codebook onto several separate excel sheets within one excel file. The layers may include the following: 1) the full codebook and coding criteria, 2) the data collection sheet with variables on the horizontal axis, allowing for participant data to be entered across each individual row, 3) the exclusion tracking form, and 4) the form for coders to track questionable references that they wish to discuss with the team, as well as the outcome of the team’s discussion regarding those references. Table 1 includes an example section of a data collection sheet including demographic information and references to alcohol use.

Adaptation of the codebook

Our initial codebooks were designed to be applied to MySpace profiles. Over time, our research interests have included evaluation of Facebook profiles. Our research interests have also grown to include new health variables of interest, such as mental health and physical activity. For each adaptation of the codebook, the same iterative process is applied including application of relevant clinical criteria, development of an initial set of keywords to apply to profiles, evaluation of a practice dataset, and discussion with the team to further refine the codebook. We continue to routinely communicate with our IRB and research ethics colleagues to maintain the highest standards of integrity in our data collection and reporting approaches.

CODER TRAINING

The training of coders who will use the codebook follows a similar process that was used in codebook development. Coders are first trained in the diagnostic criteria that underlie the variables of interest, as well as in clinical information relevant to the variable of interest. Coders are also trained in the theoretical constructs that guide our assessments of profiles. Coders are then taught about the codebook in several hands-on sessions in which active participation in discussion is encouraged. After this session, coders begin applying the

codebook to practice sets of profiles, and are encouraged to take their time to fully understand the process of profile evaluation and the theoretical basis underlying the codebook. During this pilot coding, assessments of the codebook are ongoing through the iterative process of meeting with the research team.

The codebook is also evaluated through frequent assessments of interrater reliability during the pilot coding phase. Both the categorization and the counts of references are tracked and evaluated frequently using Cohen's kappa [26, 27].

DISCUSSION

Codebook successes

Our codebooks have undergone considerable adaptation in the past five years. We have successfully applied our codebooks to both MySpace and Facebook to evaluate health risk behaviors. In one study we found associations among health risk behaviors displayed on profiles that were similar to those reported in previous work involving offline self-report [9]. These findings support the validity of displayed behaviors on SNSs. Our interrater assessments have been consistently high, with Cohen's kappa values ranging from 0.67 to 0.84 for evaluation of various health risk behaviors, which suggests high reliability of our coding scheme [9, 28]. Our iterative process has helped us to adapt our codebooks to account for new variables of interest as well as new trends and structural changes among SNSs.

Codebook challenges

This ongoing process of codebook development has required time, patience and team commitment. Development and maintenance of a codebook leads to challenges inherent in team research: the need for coordination of schedules and tasks to accommodate frequent meetings, the time required to reach group consensus. We have also learned that the time required in training a new team member as a coder is substantial, when a team member needs to periodically or permanently relinquish their involvement in projects this can impact the overall productivity of the team.

Another challenge has been adapting to changes within SNSs. A recent example includes changes in Facebook privacy settings in December 2009. During an ongoing study, we found that the data available to us within certain Facebook profile sections had changed, forcing us to determine how best to consider such changes in our analyses. Further, upgrades to SNSs can lead to server outages and changes in how search engines function within the website.

Future research using SNSs

Despite these challenges, our codebook development process has been a rewarding one. We will continue to revise and adapt our codebook as new research questions arise and new trends in social media use emerge. Our next steps include establishing the validity of displayed health risk behavior information on SNSs by linking it to self-report. Once this validity is better understood, efforts to incorporate displayed SNS information into screening or intervention efforts can move forward.

It is our hope that this paper will provide detailed information about one systematic approach to codebook development so that other researchers may use this structure to develop and test their own codebooks for use in SNS research. It is not our intent to present this paper as the only way to develop or use a codebook; rather, we hope to assist other researchers in adapting this process to fit their own areas of interest.

Acknowledgments

The project described was supported by Award Number K12HD055894 from the Eunice Kennedy Shriver National Institute Of Child Health & Human Development.

References

1. Lenhart, A.; Purcell, K.; Smith, A., et al. Social media and young adults. Washington, DC: Pew Internet and American Life Project; 2010.
2. Lewis K, Kaufman J, Christakis N. The Taste for Privacy: An Analysis of College Student Privacy Settings in an Online Social Network. *Journal of Computer-Mediated Communication*. 2008; 14(1): 79–100.
3. Pempek TA, Yermolayeva YA, Calvert SL. College students' social networking experiences on Facebook. *Journal of Applied Developmental Psychology*. 2009; 20(3):227–238.
4. Google. Google Ad Planner. 2010. Available from: https://www.google.com/adplanner/planning/site_details#siteDetails?identifier=facebook.com&geo=US&trait_type=1&lp=false
5. Childs, M. Facebook surpasses Google in weekly US hits for first time. *Business week*; 2010.
6. Christofides E, Muise A, Desmarais S. Information Disclosure and Control on Facebook: Are They Two Sides of the Same Coin or Two Different Processes? *Cyberpsychol Behav*. 2009 Mar 1.
7. Hinduja S, Patchin JW. Personal information of adolescents on the Internet: A quantitative content analysis of MySpace. *J Adolesc*. 2008 Feb; 31(1):125–146. [PubMed: 17604833]
8. Moreno MA, Parks M, Richardson LP. What are adolescents showing the world about their health risk behaviors on MySpace? *Med Gen Med*. 2007; 9(4):9.
9. Moreno MA, Parks MR, Zimmerman FJ, et al. Display of health risk behaviors on MySpace by adolescents: Prevalence and Associations. *Archives of Pediatrics and Adolescent Medicine*. 2009; 163(1):35–41. [PubMed: 19124701]
10. Cooper ML. Alcohol use and risky sexual behavior among college students and youth: evaluating the evidence. *J Stud Alcohol Suppl*. 2002 Mar.(14):101–117. [PubMed: 12022716]
11. Hingson R, Heeren T, Winter M, et al. Magnitude of alcohol-related mortality and morbidity among U.S. college students ages 18–24: changes from 1998 to 2001. *Annu Rev Public Health*. 2005; 26:259–279. [PubMed: 15760289]
12. Mulye TP, Park MJ, Nelson CD, et al. Trends in adolescent and young adult health in the United States. *J Adolesc Health*. 2009 Jul; 45(1):8–24. [PubMed: 19541245]
13. Halpern-Felsher BL, Ozer EM, Millstein SG, et al. Preventive services in a health maintenance organization: how well do pediatricians screen and educate adolescent patients? *Arch Pediatr Adolesc Med*. 2000 Feb; 154(2):173–179. [PubMed: 10665605]
14. Lehrer JA, Pantell R, Tebb K, et al. Forgone health care among U.S. adolescents: associations between risk characteristics and confidentiality concern. *J Adolesc Health*. 2007 Mar; 40(3):218–226. [PubMed: 17321421]
15. Brown JD. Adolescents' sexual media diets. *J Adolesc Health*. 2000 Aug; 27(2 Suppl):35–40. [PubMed: 10904204]
16. Bandura, A. *Social Foundations of Thought and Action: A Social Cognitive Theory*. Englewood Cliffs, NJ: Prentice Hall; 1986.
17. Bandura, A. *Social Learning Theory*. New York: General Learning Press; 1977.
18. Glanz, K.; Rimer, B.; Lewis, F. *Health Behavior and Health Education: Theory, Research and Practice*. San Francisco, CA: Wiley and Sons; 2002.
19. Bandura A. Social cognitive theory: an agentic perspective. *Annu Rev Psychol*. 2001; 52:1–26. [PubMed: 11148297]
20. Creswell JW, Miller D. Determining validity in qualitative inquiry. *Theory into Practice*. 2000; 39(3):124–130.
21. Palinkas LA, Prussing E, Landsverk J, et al. Youth-violence prevention in the aftermath of the San Diego East county school shootings: a qualitative assessment of community explanatory models. *Ambul Pediatr*. 2003 Sep&Oct; 3(5):246–252. [PubMed: 12974662]

22. Knight JR, Shrier LA, Bravender TD, et al. A new brief screen for adolescent substance abuse. *Arch Pediatr Adolesc Med*. 1999 Jun; 153(6):591–596. [PubMed: 10357299]
23. Moreno MA, Fost NC, Christakis DA. Research ethics in the MySpace era. *Pediatrics*. 2008 Jan; 121(1):157–161. [PubMed: 18166570]
24. Parks MR, Roberts LD. 'Making MOOsic': The development of personal relationships on line and a comparison to their off-line counterparts. *Journal of Social and Personal Relationships*. 1998 Aug; 15(4):517–537.
25. Walther, JB.; Parks, MR. Cues filtered out, cues filtered in: Computer mediated communication and relationships. In: Miller, GR., editor. *The handbook of interpersonal communication*. Thousand Oaks: Sage; 2002. p. 529-563.
26. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977; 33:159–174. [PubMed: 843571]
27. Kimberlin CL, Winterstein AG. Validity and reliability of measurement instruments used in research. *Am J Health Syst Pharm*. 2008 Dec 1; 65(23):2276–2284. [PubMed: 19020196]
28. Moreno MA, Briner LR, Williams A, et al. A content analysis of displayed alcohol references on a social networking web site. *J Adolesc Health* 2010. 2010; 47(2):168–175.

Figure 1.
Example of study participant flow chart to demonstrate application of exclusion criteria.

Figure 2.
Example Facebook profile

Table 1

Example data collection sheet using codebook

Age (years)	Gender (0=m, 1=f)	Days since last profile activity	# of FB friends	First alcohol reference noted on profile: Image description or verbatim text	Second alcohol reference noted on profile: Image description or verbatim text	Third alcohol reference noted on profile: Image description or verbatim text	Fourth alcohol reference noted on profile: Image description or verbatim text	Total # alcohol refs in past year
20	0	1	451	Status up date: "Great friends, coronas, sunburns and a football win. Solid day."	Status update: "Can I have more red wine?"	Tagged photo: picture drinking beer bottle at party		3
18	0	3	352					0
20	0	0	570	Information section quote: "Alcoholism is a disease, but it's the only one you can get yelled at for having. 'Damn it you're an alcoholic.' 'Damn it you have cancer' just doesn't sound right."	Tagged photo: picture drinking beer bottle at party	Tagged photo: picture drinking beer bottle at bar		3
19	1	1	386	Status update: "Is it ok to take my birth control pill with beer?"	Status update: "I'm so hung over I forgot how to text"	Profile picture of profile owner with a beer	Tagged picture of profile owner with a bottle of alcohol	10
19	0	2	519	Tagged picture of profile owner playing beer pong, beer in hand				1
19	0	3	543	Status update: "My phone can survive water, now I know my phone can survive whisky."	Tagged picture of profile owner playing beer pong, beer in hand	Tagged picture of profile owner playing quarters, beer in hand	Tagged picture of profile owner playing beer pong, beer in hand	6
	0	13	214					0
20	0	0	619	Status update: "Face book, you are a dynasty. Someday my kids will add me as a friend and they'll see all the drunk college photos of their father."	Status update: "Alcohol levels low ... Needs to party!"	Bumper sticker with quote: "Enjoy it now because after college it's called alcoholism"	Tagged picture: Profile owner playing beer pong, beer in hand	5
	1	0	831	Status update: "Football, friends, too many beers, GREAT"	Status update: "Done with work and ready for Margarita weekend!"	Tagged picture of profile owner with Tequila shot	Tagged picture of profile owner drinking a glass of wine	10
20	0	3	496					0
0	0	437	Status update: "Needs a drink after this week, and its only Monday"	Status update: "First thirsty Thursday of the year!"	Status update: "Cold beer, hot fish, and good company ... Livin the good life"	Status update: "Shot a tequila, beer on tap, a good looking woman, to sit on my lap."		6

All examples have been altered to protect confidentiality