Credibility Assessment

of the

Top Six Second-Tier Domain Extensions

among

High School Freshmen

Steve Cummings, Michelle Sousa, and John Rucker

OVERVIEW

A few months ago our research team asked how we could quantify websites' perceived credibility according to high school freshmen. We recognize that the internet's establishment in the modern western lifestyle ensures any data about how businesses can make internet-related decisions validates those decisions. We also recognize that technical aspects of websites could be measurable components affecting trust from the perspective of freshmen in high school, whose responses may indicate upcoming trends on online shopping and other online interactions.

The world wide web, which was once inaccessible to many, is now very much a part of everyday life. While our sample population is made up of young individuals, high school freshmen, they should not be disregarded based on their age. Based on their amount of online accessibility, as well as the hours spent browsing the web, we actually found that high school freshmen are great candidates and a relevant target group for this particular survey. Of the 121 students surveyed, 95% reported to have regular access to the internet at home, and 69% said they spend three or more hours using the internet on any given day. Within a given week, that's over 21 hours, at minimum, of internet usage.

This study is important because it can be used to forecast trends, as the worldwide consortium decides which extensions receive registration and set the cost standards, as the .com becomes less accessible, and other options must be considered. A current example is the internet powerhouse Google who created new domains ending in .xyz, and how those endings will be ranked among the young surveyors in comparison to other extensions currently being used. Ultimately, the purpose of our study is to learn if one or more domain extensions hold more merit than the others. While we know that the students use the internet regularly, we were curious if e extensions played a role in the sites that they visit; do the students feel that one URL extension is more 'trustworthy' than another, and are there trends in the sites that they choose? We were able to accomplish this task by focusing on freshmen from two high

schools varying in backgrounds, one being a traditional public rural school and the other an urban charter school with a focus on technology as the target groups for data collection.

We believe that these findings may have significant commercial impacts, as our target population will soon define the younger consumer bracket. We shared an online survey asking a variety of questions measuring a respondent's trust of popular website URLs and provide our findings herein.

• First, we explored the various URL extensions based on how they ranked via the World Wide Web online consortium eliminating all country extensions.

• Second, we attempted to understand how individuals ranked the various extensions based on the results from collection of survey questions administered to freshmen at two different high schools.

Impact of the Literature Review

In our concerted exploration, study, and analysis affecting website credibility as ascertained by high school freshmen and their responses to domain endings, a brief scan of existing studies helped hone our approaches. We knew we needed to investigate aspects of credibility of digital content found online, but as we learned about research already conducted, we were able to narrow our focus to domain endings. With the newer release and implementation of hosts of domain endings such as .biz and the Country-Code Top-Level Domains such as .us, .ca, .de, and .cn, questions of preference came to our minds. Although other studies have been used to rank domain endings based on total number registered, few studies allowed respondents to rank the endings or show trust levels associated with different endings. We also had access to school-aged students in high school and wanted to focus on up-and-coming shoppers to hopefully find research which could inform business decisions with lasting effects. We focused on high school freshmen from two cities.

In a very similar study, (Wogalter, 2008), the researchers examined components associated with Web sites that could affect credibility beliefs about Web site information: domain endings (e.g., .com, .edu), quality seals, and domain names. Their results showed that participants had difficulty discriminating between actual and fictitious domains. We did not use fictitious names in our study because we thought existing actual domain endings' credibility would be more helpful than to compare to false versions.

Trust and detecting deception: a recent meta-analysis showed that the largest factor in detecting deception in traditional direct communication is the perceived credibility of the sender, (George, 2016). Since credibility is so important to trust, our study aimed to assess whether the perceived credibility of computer communication can be linked to the URL extension. We found some likely linkages which correlate computer linkage and competency to at least noticing URL extensions.

A recent study showed the existence of domain preferences, even after factoring out issues such as position bias and relevance, (Ieong, 2012). Users have learned to trust some domains over others. Past work such as TrustRank measures user trust at a domain level, (Z. Gyongyi, 2004). A recent eye-tracking study also confirms that users pay attention to the displayed URL (E. Agichtein, 2006). Our study focused on the URL extension and our data analysis found some indicators of paying attention to URL endings based on demographic profiles.

Identified Existing Bias

There are various opinions regarding the topic of website domain extension credibility. A series of secondary sources, however, formalized these stereotypes while propagating the bias to the masses. Kevin Schwartz's (2014) study, published online via https://moz.com claimed two assumptions to validity: "1) whether users even notice Country-Code Top-Level Domains (ccTLDs) and 2) if so, do they really prefer the Top-Level Domain (TLD) of their country." The results, however, apply to our study with the

exception that Schwartz targeted adults of the United States and Australia. We will target freshmen in high school on the California coast in various cities, but at public schools. Because new gTLDs (generic TLDs) are newer and the freshmen may not have been exposed to them, our freshmen population may have preferential biases towards more established domain extensions such as .com and .org. Also, the target populations may have more trust in .edu domain extensions due to the nature of our populations being in the education system during our study. Finally, most of our populations cannot legally own and use credit card accounts (18+ year old freshmen excluded), so they might have limited experience with more commercial online extensions (if such exist including and beyond ".com"). This limited current experience might have skewed results and their implications for the future of these internet consumers.

As part of identifying our target populations' demographics, we expanded our survey of domain extension credibility to include some digital literacy factors. According to Wogalter (2008), "participants who reported spending greater time on the Internet showed significantly higher trust ratings." We added one specific survey question to examine this in our study group (e.g. How many hours do you spend on the internet per day?)

Persuasive intent or commercial content is also a strong influence. Metzger (2010) says, "strong commercial motivations nearly uniformly produced negative responses and unsolicited and unwelcome information negatively affects users' perceptions." We definitely saw negative responses to terms such as "spam" and to URL extensions such as .biz and .tv, all highly commercialized constructs.

According to Peiper, (2003) in Guessing URL, a popular way people are finding web sites, some individuals are skipping the process of inputting the web URL and searching for the site by name through search databases. This may suggest that URL extensions may not be all that important, but the numbers of individuals who do this is not listed in the study. The study found that some respondents do not consider

URLs at least, which may imply they would not type them if visiting a site. More research is needed to confirm or reject those possibilities.

Study Design Prototypes

As creative and unique as we find ourselves, those who have gone before have paved with helpful stepping stones and offer insightful pathways to useful data regarding how the up-and-coming internet consumers view website domain extensions. Eli Schwartz's (2014) study comparing United States respondents to Australian respondents asked questions such as "Where is the establishment that owns the website www.joesplumbing.ca most likely to be located?" with four possible responses including "Canada, California, Cameroon, and I can't tell". This style of questioning allows respondents to share their ideas mostly regarding the domain extension ".ca" and responses allow comparison between people of the two continents.

Another question in the Schwartz study asked, "Of the two website addresses below, which one do you think is most likely to be the domain for a registered non-profit?" Three options followed: "www.stopbullying.org", "www.stopbullying.com", and "Don't know". The simplicity of this question and three distinct responses allowed for simple analysis to determine differences in the study's target populations.

The review of existing accessible literature showed us we had a niche question and target population. We molded our ideas from the above mentioned studies and publications. Ultimately, we tried to focus on bipolar trust/no trust and some scaled responses to six main domain extensions. We used those various responses to create a "Trust Index," which we had not seen in other literature.

Contextual Factors

The Time

In an effort to propose our study, research similar existing data, build tools as needed, conduct the survey to gather data, and analyze the data all within the span of about 100 days, we did not survey all high school freshmen in the world. Geographically our team is located in southern California and had access to two sample populations of high school freshmen, one residing in and around Paso Robles, California, and the other in the San Diego metropolitan area. We contacted teachers of freshmen students at Paso Robles High School and three campuses of High Tech High Schools in San Diego. One of our team is a teacher at Paso Robles High School. Our professor's son attends a High Tech High of San Diego. These two options were perfect given we could easily contact teachers to disseminate the survey to their students.

The Survey

Given the speed and prevalence of the internet, we knew that whatever measurement tools we used needed to be internet based for cheap, quick dispersal and response collection. We used online-based Qualtrics survey-creation and delivery tools. This platform proved extremely helpful because our team is spread across southern California yet Qualtrics allows multiple users to simultaneously build and edit online surveys as well as online delivery for responses. We found many tools in Qualtrics very helpful and realized the online survey software only lacked in data analysis after gathering. The data was exported and analyzed further in SPSS by IBM.

The Differentiation

Had we more time, been more organized, and had more connections, our total respondent number could have been much higher. As it is, we had no specific number of responses we were shooting for. We ended up analyzing 121 responses. The biggest deviation from our plans pre-survey activation turned out

to be throwing out the idea of comparing high school freshmen to seniors. There were only 9 total seniors who responded (and a handful of sophomores and juniors), but those were discounted. We did not want to compare over a hundred freshmen's perceptions of website domain trust to only 9 seniors and thought any conclusions we came to would be totally premature with such low senior responses.

Methodology

We studied the population of high school freshmen in southern California. We did this by using sample groups at two southern California high schools. This was a sample of convenience due to John Rucker's teaching position at Paso Robles and Dr. Rebecca Frazee's association with High Tech High (HTH) in San Diego.

The instruments used in our survey consisted of a survey website called Qualtrics and another program called SPSS. Qualtrics is a website that allows you to create surveys in various formats. You can edit, collaborate, and share the survey in real-time. Once the survey goes "live", changes can no longer be made and you can then begin to collect and sort the data. Once we gathered and sorted the data, we ran it through the SPSS statistical program for data anlaysis.

Using Qualtrics, we created a survey and gathered background information, such as the student's school name, gender, hours of internet usage, as well as internet accessibility. We also used a self-rating system for students to rate their computer savviness. Lastly, and also very important, we asked the students if they consider the URL extensions.

In addition to the background questions, we also asked "scenario questions". For example, "Which of the following websites would you trust with your credit card information?" and "which of the following websites would you associate with a potential virus risk?". The options for the responses had the same website names, but varied in the URL endings.

One of the questions used word association for the URL extensions. Students were given a list of six URL extensions: .biz, .tv, .xyz, .edu, .pro, and .mobi and asked to match the URL extensions with the following words: honest, spam, credible, theft, phishing, ethical, cheats, and trustworthy. Once we collected the data, we clustered the descriptors into two categories: trusting and untrusting, allowing us to further breakdown the data in the search for overall trends.

Another of our questions used a ranking system allowing students to drag a URL name up and a down based on with the URL they find most credible at the top descending to the least credible. We also added one open-ended question, asking students which URL was most credible and why, but the results were fairly inconclusive due to a lack of responses.

The survey was piloted at the rural school and administered to 112 students. We soon realized, upon receiving the responses, that we were unable to see how some of the responses connected to the data because we had used images instead of text. We made changes accordingly.

In all, we received 191 surveyors and eliminated 70 surveyors due to lack of responses, such as skipping multiple questions, and we found that some students were not freshmen at all, but rather sophomores, juniors, and seniors. Since their data pool was too small, we eliminated them as well from the data.

Once we collected all of the data, we reviewed the information, discarded outliers, blank responses, and with the help of our professor, we ran the numbers through SPSS for trend collection.

Implementation

Our data was gathered using a Qualtrics online survey. A link to the survey was distributed by emails to teachers at High Tech High and by John Rucker at Paso Robles High. The survey link was then made available to freshmen (and a few seniors) and data acquired through the internet by Qualtrics. We acquired teacher's emails in two ways. For High Tech High, Dr. Frazee contacted a school administrator asking to conduct the survey and four ninth grade teachers from three different High Tech High campuses responded. John Rucker (group member) then emailed a description of the survey process as well as a link (see email below). For Paso Robles High School, John Rucker (teacher) emailed the whole teaching staff with a Qualtrics survey link and description of the survey process. See the email to HTH below. The Paso Robles High School email was similar with only differences in salutations.

Thank you for assisting our study! We look forward to sharing the results. The survey is open for one week (until Wednesday, April 20th). Please share the following link with your students:

https://sdsupsfa.co1.qualtrics.com/jfe/form/SV_6Qk1arggYfi2uMJ

If you have any questions or concerns, please email me (ruckerjohn3@gmail.com) or call 805-610-7652 for immediate assistance.

Here is a brief monologue to share with your students before they begin the survey:
We are about to take a brief survey regarding internet usage. This survey was created by some university students at San Diego State University as they study internet usage and trust in high school Freshmen and Seniors. The survey will not identify you, so please be completely honest. The university students will share the results and their conclusions with us at the end of their study. Please read the directions carefully. This will not be graded and is not linked to your identity. Thank you for your assistance in this study.

You have shown interest in supporting San Diego State University students in the Learning Design and Technology department. We are acquiring survey data from Freshmen and Seniors in high school in order to compare URL extension credibility opinions of these students. We expect this survey to take between 5 and 10 minutes for your students. It will not ask for specifically identifying responses. All of your students will remain anonymous. Tomorrow, Wednesday morning, we will email you the link to the survey. We do recommend you post the URL of the Qualtrics survey in an easy-to-access online space for the convenience of your students.

Again, we thank you for helping with this research and look forward to sharing the results with you!

Sincerely,

The Web Credibility Team, Michelle, Steve, John, and Dr. Frazee

SDSU Learning Design and Technology

This email was sent on Wednesday, April 13th. By Monday, April 18th, John sent another email to both schools' teachers as a reminder. It included the Qualtrics link and another brief description of the survey process and timeline (to close 1 week after starting). A final email was sent the morning of April 20th in a last effort to get more responses. The survey closed Wednesday afternoon after 4 p.m.

Findings

Survey Analysis

A measure of Trust. We created a trust index (TI) by combining results of 13 "trust" questions. The original data were re-configured so that higher trust is specified by higher scores in each of the survey's "trust" questions. A higher TI indicates a higher tendency to trust the Domain Extension. The lowest possible trust index value is three and the highest is twenty-eight. We started with 191 survey responses. After removing the incomplete surveys, we had 121 responses; our n=121. The overall results are shown in Figure 1 and Table 1.
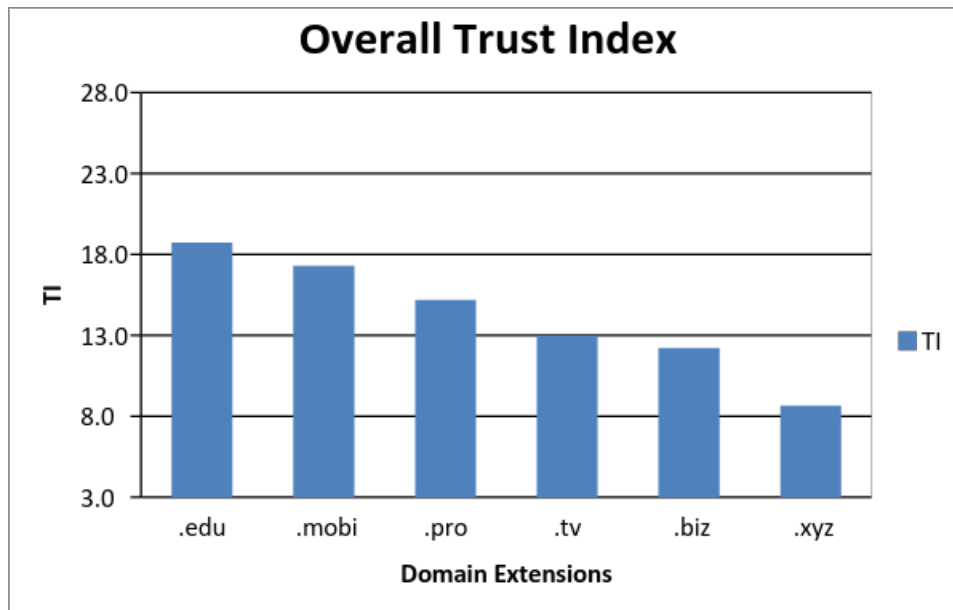
Figure 1. Overall Trust Index (TI)

Table 1. Overall Trust Index

	Mean TI					
	.edu	**.mobi**	**.pro**	**.tv**	**.biz**	**.xyz**
Overall Average	18.7	17.3	15.2	13.0	12.2	8.7
Standard Dev.	2.9	2.8	3.1	4.6	4.2	4.5

Our results indicated that that the .edu extension was the most trusted, with a mean trust index of 18.7, and the .xyz was the least trusted with a mean of 8.7. Relative trust of the other domain extensions is shown in Table 1. The higher-rated domain extensions showed less variability than the lower rated extensions. (Frequency distributions for the overall results are shown in the Appendix C.) This may indicate that there is more understanding of the high-trust sites.)

Demographic data. Data from the demographic questions were re-configured so that the responses are binary. The "geek rating" assessment of digital literacy, originally a 5 point scale, became "high literacy" for the 4 or 5 scores, and "low or average" literacy for scores of 1-3. Similarly, the "consider the URL" scores were converted to yes and no.

Discounted data. We found that only 5% responded that they did not have internet or wifi at home (n=X). These were not differentiated enough, so the categories were dropped from further consideration.

The index data was then considered in relation to each of four independent characteristic pairs; urban/rural, m/f, whether they considered the URL extension/or not and digital literacy high/low. The trust index averages are shown in Table 2 and Figure 2, below. In Figure 2, significant differences are labeled A, B, C, D, E, and F, and described below.

Table 2, Overall Mean Trust Index compared by respondent groupings/characteristics.

Responses			Mean TI					
Frequency	Percent	Category	.edu	.mobi	.pro	.tv	.biz	.xyz
43.0	35.5	**Rural**	17.9	17.5	15.6	13.5	11.2	10.1
78.0	64.5	**Urban**	19.2	17.2	15.0	12.7	12.8	7.9
66.0	54.5	**Male**	18.2	17.2	14.7	13.6	12.4	9.0
55.0	45.5	**Female**	19.4	17.4	15.7	12.2	11.9	8.3
61.0	50.4	**Yes Consider URL**	19.0	17.7	15.0	13.4	12.2	8.4
60.0	49.6	**No Consider URL**	18.4	16.9	15.4	12.6	12.3	8.9
64.0	16.5	**High Literacy**	19.1	17.5	15.1	14.0	12.2	7.9
57.0	33.9	**Low or Avg Literacy**	18.3	17.0	15.3	11.8	12.3	9.5
		Overall Average	18.7	17.3	15.2	13.0	12.2	8.7

Note: Significant differences are highlighted.

Figure 2. Overall Mean Trust Index for All Domain Extensions by Respondent Characteristics
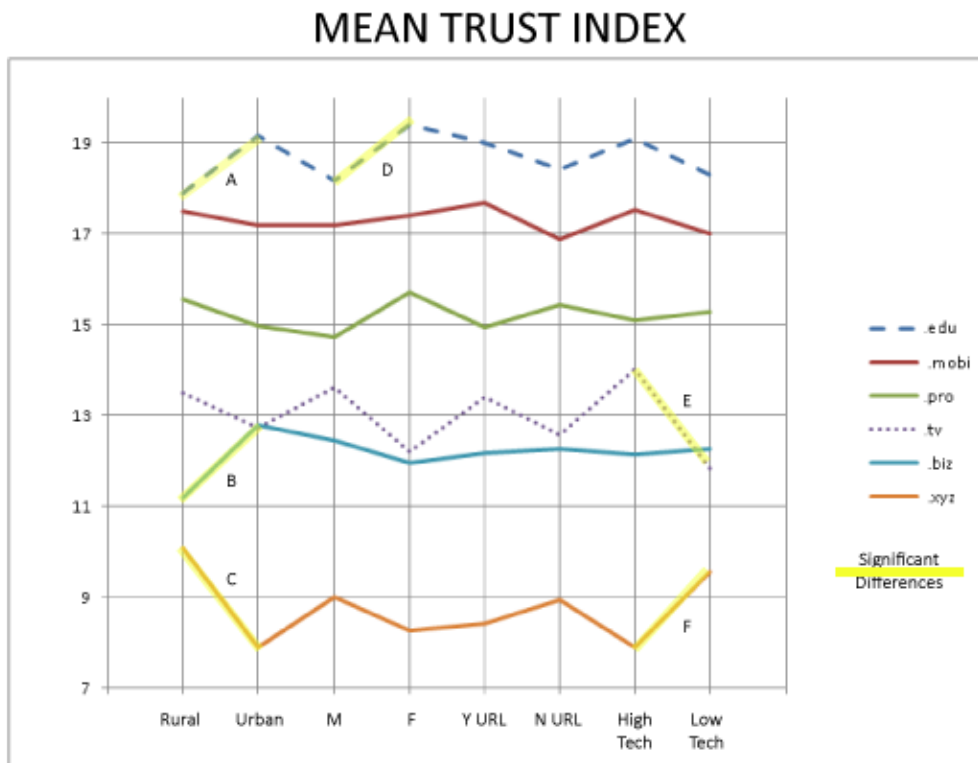
Significant differences were noted between the urban and rural sample for three domain

extensions:

    A) Urban participants found .edu to be more trustworthy. (F = 5.4, p = 0.02)

    B) Urban participants found .biz to be more trustworthy. (F = 4.152, p = 0.04)

    C) Urban participants found .xyz to be less trustworthy. (F = 6.863, p = 0.01)

Significant differences were noted between the male and female sample for one domain extensions:

D) Female participants found .edu to be more trustworthy. (F = 5.49, p = 0.021)

Significant differences were noted between the high digital literacy and low digital literacy sample (labelled "High Tech" and "Low Tech" in the graph for space reasons.) for two domain extensions:

E) High digital literacy participants found .tv to be more trustworthy. (F = 6.97, p = 0.009) Correlation analysis also found a significant relationship (at the 0.05 level) between High Digital Literacy and the .tv extension.

F) High digital literacy participants found .xyz to be less trustworthy. (F = 4.233, p = 0.042)

Data Interpretation

It was not surprising that the .edu extension ranked highly in our survey, since the extension is reserved for educational institutions. This served to validate our data because we expected that result. The interesting results are for the other five extensions. Figure 1 shows that overall, and Figure 2 shows that across all groups the rankings are consistent with the most trusted at the top:

.mobi

.pro

.tv

.biz

.xyz

This is in contrast to the relative number of registrations for these domain extensions (from the w3techs.com site):

.biz   .5%

.tv   .3%

.xyz   <.1%

.pro   <.1%

.mobi <.1%

This inconsistency is possibly due to the "familiarity breeds contempt" principle. Cognitive heuristics such a reputation, endorsement and consistency and, in this case probably expectancy violation have been noted as factors in detecting deception and trust (Metzger, 2010).

Attitudes for each of the demographic categories are generally similar for each of the domain extensions.

The significant differences within groups were noted above in Figure 2.

The .edu extension was seen as significantly more trustworthy by the urban group and females, however all other groups rated .edu higher than any other extension. By the same token, the .xyz extension was significantly less trustworthy for the urban group and the high digital literacy group, but also rated below all other extensions by all groups. These differences while statistically significant do not appear to be immediately important or practically significant.

Urban participants rated the .biz extension significantly higher. This could possibly be due to the urban environment having more businesses or more parental involvement in business. The urban group may have more positive experiences with business, but this would need further study to be certain.

The high digital literacy group seems to be exceptional in their trust of the .tv extension. This may be because this group possesses special knowledge or insight regarding the internet. This top level domain name has been marketed as a preferred web address for rich media content because of its' similarity to television. Perhaps the high digital literacy group is aware of this. This is supported by comments such as, "I know a couple sites that use .tv that are reputable." This group is also exceptional in their mistrust of the .xyz extension. Google uses abc.xyz for its corporate (Alphabet Inc.) website. Could there be a "big-brother" aversion to the extension because of this. Further study could clarify this question.

Conclusions

For High School freshmen in at two schools in Southern California, the most trusted generic top-level domain in our study was .edu. Next most trusted was .mobi and then .pro. Next, the .tv and .biz domains

shared about the same level of trust after .pro, and the .xyz domain was the least trusted. If I were setting up a new site targeted at High School Freshmen, I would seriously consider using the .mobi domain because of this study. And by contrast, if I wanted to appeal to an edgy-dangerous crowd I might consider using the .xyz domain.

References

Agichtein, E., Brill, E., Dumais, S., & Ragno, R. (2006). Learning user interaction models for predicting web search result preferences. *Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval - SIGIR '06,* 3-6. doi:10.1145/1148170.1148175

Bremer, K. L., & Salkind, N. J. (2015). *Study guide to accompany Neil J. Salkind's Statistics for People Who Think They Hate Statistics, 5th edition*. Los Angeles: SAGE.

Doyle, T., & Hammond, J. L. (2006). Net cred: Evaluating the internet as a research source. *Reference Services Review, 34*(1), 56-70. doi:10.1108/00907320610648761

Fraenkel, J. R., & Wallen, N. E. (2015). *How to Design and Evaluate Research in Education* (9th ed.). New York: McGraw-Hill Publishing Company.

Gardere, S. (2008, September 4). ICANN Announces New Expansion Of The Internet Domain Name System. *Mondaq Business Briefing*. Retrieved February 20, 2016, from http://www.highbeam.com/doc/1G1-184447054.html?refid=easy_hf

Grimes, D. J., & Boening, C. H. (2001). Worries with the Web: A Look at Student Use of Web Resources. *College & Research Libraries, 62*(1), 11-22. doi:10.5860/crl.62.1.11

Gyöngyi, Z., Garcia-Molina, H., & Pedersen, J. (2004). Combating Web Spam with TrustRank. *Proceedings 2004 VLDB Conference,* 576-587. doi:10.1016/b978-012088469-8.50052-8

Ieong, S., Mishra, N., Sadikov, E., & Zhang, L. (2012). Domain bias in web search. *Proceedings of the Fifth ACM International Conference on Web Search and Data Mining - WSDM '12,* 413-413. doi:10.1145/2124295.2124345

Lindenthal, T. (2014). Valuable words: The price dynamics of internet domain names. *Journal of the*

    Association for Information Science and Technology J Assn Inf Sci Tec, 65(5), 869-881.

    doi:10.1002/asi.23012

 Metzger, M. J., Flanagin, A. J., & Medders, R. B. (2010). Social and Heuristic Approaches to Credibility

    Evaluation Online. *Journal of Communication, 60*(3), 413-439. doi:10.1111/j.1460-

    2466.2010.01488.x

Mowshowitz, A., & Kawaguchi, A. (2005). Measuring search engine bias. *Information Processing &*

    Management, 41(5), 1193-1205. doi:10.1016/j.ipm.2004.05.005

Swartz, A. (2013). Aaron Swartz's A Programmable Web: An Unfinished Work.*Synthesis Lectures on the*

    Semantic Web: Theory and Technology, 3(2), 1-64. doi:10.2200/s00481ed1v01y201302wbe005

W3Techs - World Wide Web Technology Surveys. (2016, February). Retrieved February 20, 2016, from

http://w3techs.com/

Wogalter, M. S., & Mayhorn, C. B. (2008). Trusting the Internet. *International Journal of Technology and*

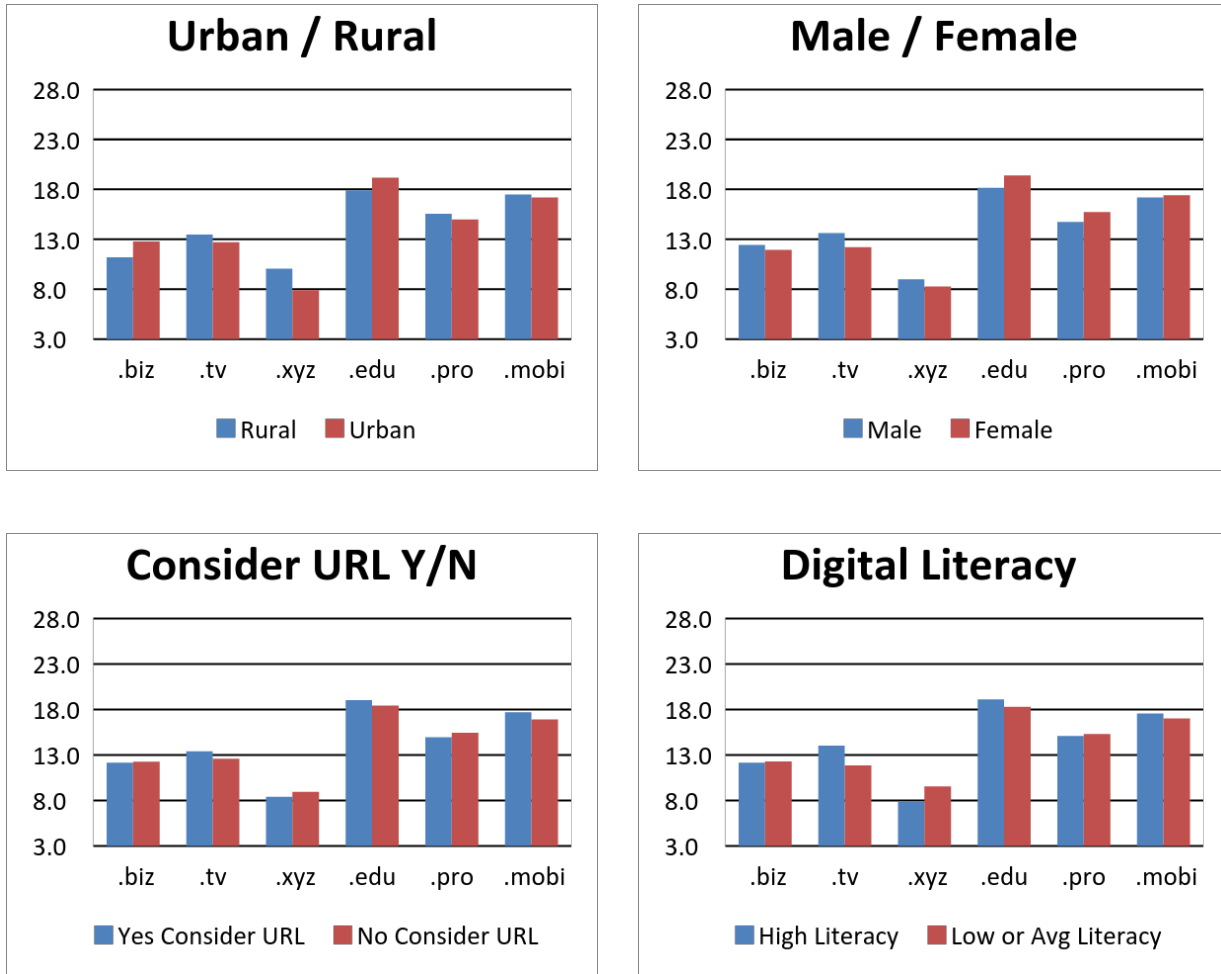   *Human Interaction, 4*(1), 75-93. doi:10.4018/jthi.2008010105

Appendices

Figure A.

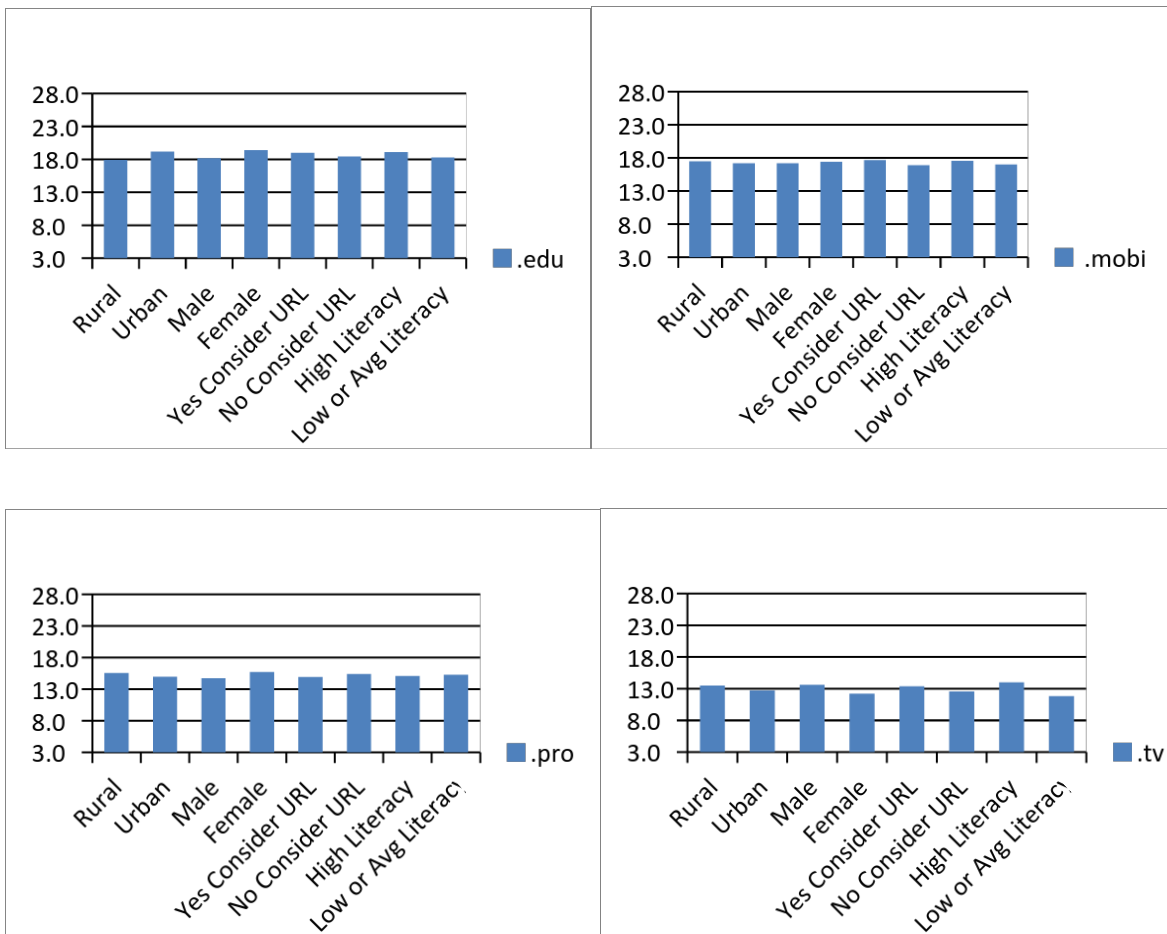Relative Trust Ratings by Demographic Category

Figure B.

Relative Trust Ratings by Demographic Category

28.0	28.0
23.0	23.0
18.0	18.0
13.0	13.0
8.0	8.0
3.0	3.0

Rural, Urban, Male, Female, Yes Consider URL, No Consider URL, High Literacy, Low or Avg Literacy

■ .biz

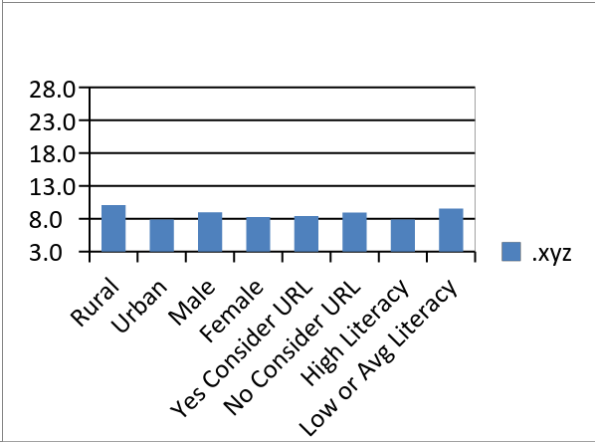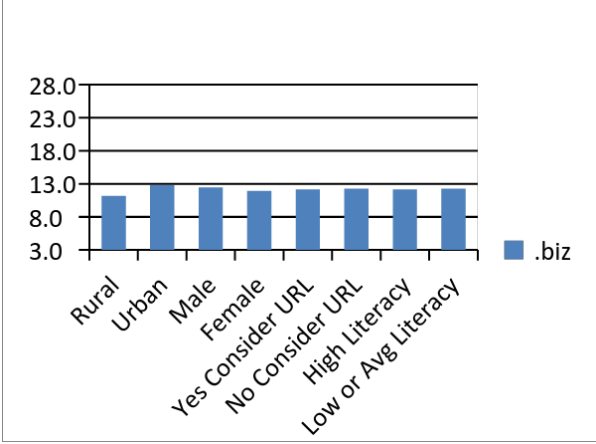Rural, Urban, Male, Female, Yes Consider URL, No Consider URL, High Literacy, Low or Avg Literacy

■ .xyz

Figure C.

Frequency Distributions for Trust Index