An Analysis of Some Variables Affecting the Internet Dependency Level of Turkish Adolescents by Using Decision Tree Methods

Murat KAYRİ*, Selim GÜNÜÇ**

Abstract
Internet dependency is going to expand into social life in wide area whereas it has been accepted as a pathological and psychological disease. Knowing the basic effects of internet dependency is an inevitable approach to use the internet technology healthy. In this study, internet dependency levels of 754 students were examined with the Internet Dependency Scale. Total score which will be obtained from the scale is the dependent variable and using daily hours of internet, how many years the users are acquainted with the internet, the purpose of using the internet, education level of users’ mother and father, income of users’ families and gender variables are included in the model as independent variables. Model was tested both Classification Tree and Chaid Analysis. In the Classification Tree method; purpose of using the internet, using daily hours of the internet, gender, income of users’ families and education level of users’ father which affected dependent variable were found meaningful. In Chaid Analysis, purpose of using the internet and using daily hours of the internet users’ independent variables were just found meaningful.

Key Words
Chaid Analysis, Classification Tree, Decision Trees, The internet Dependency.

* Correspondence: Assist. Prof., Murat Kayri, Yuzuncu Yıl University, Educational Faculty, Van/Turkey
E-mail: mkayri@yyu.edu.tr
** Slim Günüş, PhD Candidate, Yuzuncu Yıl University, Educational Faculty, Van/Turkey

© 2010 Eğitim Danışmanlığı ve Araştırma İletişim Hizmetleri Tic. Ltd. Şti.
Dependence is defined as being unable to give up substances or behaviors or control them (Egger, & Rauterberg, 1996). The notion is used mostly as “addiction” or “dependence” in the international literature. However, in DSM-IV (The Diagnostic and Statistical Manual, 4th edition) the terms “dependence” or “abuse” are used instead of “addiction” (Henderson, 2001).

Individuals might get addicted to many substances such as alcohol and drugs, but the term is not only used for cigarette smoking, alcohol or drugs, but also there are non-physical behavior based dependence such as eating dependence, game dependence, sex dependence, computer dependence, television dependence, shopping dependence and the internet dependence (DiMaggio, Hargittai, Neuman, & Robinson, 2001; Griffiths, 1999; Kim, & Kim, 2002). Both substance dependence and behavior based dependence are characterized by uncontrolled behavior or action and continuity of behavior or action despite bad results (Henderson, 2001).

Although the internet dependence might be seen in all age groups, it is mostly the problem of the male and the youth (Gonzalez, 2002). In Turkish, the term “Internet dependence” was first used in the literature through an e-mail sent as a joke by Dr. Ivan Goldberg in 1996 (Günüş, 2009). The term first appeared as “internet addiction” in the literature was later called differently by researchers and clinicians. These were “internet dependency” (Tvedt, 2007), “pathological internet use” (Davis, 2001; Young, 2004), “problematic internet use” (Caplan, 2002; Kaltiala-Heino, Lintonen, & Rimpela, 2004), “excessive internet use” (Yang, Choe, Bailey, Lee, & Cho, 2005), “internet abuse” (Young, & Case, 2004), and “internet addiction disorder” (Gonzalez, 2002; Kiralla, 2005). Moreover, in some studies the term “cyber addiction” is used to express online and offline dependence cases (Vaugeois, 2006). Young (2004) first used the term “internet addiction”, but in later studies the term “pathological internet use” was preferred since it was a clinical term. Other researchers also tended to use the terms “problematic” and “pathological” instead of “addiction”. Basically, all the terms are similarly used and the main point is excessive and problematic Internet use (Beard, & Wolf, 2001; Ceyhan, Ceyhan, & Gürcan, 2007). Briefly, all the terms suggest undesirable cases particularly caused by excessive Internet use.

Researchers used the DSM criteria to define Internet dependence and also benefited from sex dependence, pathological gambling and other
behavioral dependence within the framework of DSM (Thurlow, Lengel, & Tomic, 2004). Internet dependence was defined as a psychiatric disease by some researchers and clinicians and Internet dependent patients were provided with clinical treatment which gave positive results (Köroğlu et al., 2006; Murali, & George, 2007; Young, & Rodgers, 1998). Any substance use does not result in dependence as any behavior does not create dependence (Caplan, 2005). One of the main reasons why the Internet use appear as a kind of dependence or a factor with negative effects is “the amount of use” or the amount of time spent online (Chou, Condron, & Belland, 2005). Spending much time on the Internet might be considered as both a reason of Internet dependence (Irwansyah, 2005) and a symptom or result. The reason why the amount of time spent on the Internet is the most remarkable indicator of Internet dependence is that the weekly or daily amount of time dependent users included in studies spent on the Internet was considerably longer than non-dependent users (Cao, & Su, 2007; Everhard, 2000; Gonzalez, 2002; Hardie, & Tee, 2007; Kim, & Kim, 2002; Kubey, Lawin, & Barrows, 2001; Lin, & Tsai 2002; Nalwa, & Anand 2003; Simkova, & Cincera 2004; Young, 1996; Weitzman, 2000).

The amount of time spent on the Internet alone is not enough to define Internet dependent users. Also, the intended use of the Internet is a critical factor. Positive or negative results of the Internet use relate to the reasons of the Internet use and the intended use (Bayraktutan, 2005). Internet dependent users spend more time on online movies websites, music websites, game websites, chat rooms, pornographic websites, community websites and so on (Kim, & Kim, 2002; Murali, & George, 2007; Tvedt, 2007), while non-dependent users spend time on online news, online shopping and educational websites (Kim, & Kim 2002). In a study, Irwansyah (2005) concluded that the amount of the Internet use varied according to the intended use such as playing games, chatting, education and access to information. When compared to dependent users, non-dependent Internet users rather use the Internet for access to information (Chen, Chen, & Paul, 2001; Leung, 2004; Tekdal, 2005). In another study including young people, it was observed that the Internet users tended to use the Internet for access to information and receiving news and the amount of time spent online decreased when dependence creating factors such as chats and games were filtered or unallowable (Günüş, & Kayri, 2008).
In the present study, adolescents’ Internet dependence was defined and variables which potentially affected the Internet dependence were examined. These variables were examined using Classification Tree and Chaid analysis (Chi-squared Automatic Interaction Detection), two of data mining techniques. Another aim of the study was to compare the findings by Classification Tree and Chaid analysis. In this context, it was attempted to show cause-effect relationship in Internet dependence more unbiased by two different advanced statistical methods. Similarly, stability and unbiased estimating power of the two different statistical methods included in the research were tested (fit of model) and the parameters of the stronger-more stable method were taken as the model.

**Method**

**Research Model – Sample**

The research method was based on general survey model and it also included correlational survey model as it examined variables which might affect dependence.

The sample was national and one cosmopolitan province from each region and two schools from each province were chosen. The target group was secondary school students. The sample consisted of 754 randomly chosen students; 179 from Istanbul, 106 from Ankara, 112 from Izmir, 95 from Diyarbakir, 95 from Samsun, 88 from Mersin and 79 from Van. 39.9% of the sample (301) were female students, 58.8% (443) were male students. 1.3% (10) was lost data. The mean age of the sample was found as $15.82 \pm 1.18$.

**Research Instrument**

The Internet Dependence Scale (IDS), developed by Günüç (2009) was used for the study.

**Validity and Reliability**

Cronbach Alpha internal consistency coefficient of the scale was found as .94. Total explained variance of the four sub-dimensions was 47,463. The four sub-dimensions were respectively called: “Deprivation”, “Control Difficulty”, “Malfunctioning” and “Social Isolation” (Günüç, 2009).
Cronbach alpha (\(\alpha\)) reliability coefficients of the four sub-factors were calculated and internal consistency coefficient of the first factor which consisted of 11 items was found as .877; internal consistency coefficient of the second factor which consisted of 10 items was found as .855; internal consistency coefficient of the third factor which consisted of 7 items was found as .827 and internal consistency coefficient of the fourth factor which consisted of 7 items was found as .791. Item differentiation levels of the sub-dimensions ranged from .477 to .681 for the first factor, from .471 to .606 for the second factor, from .520 to .618 for the third factor and from .463 to .615 for the fourth factor. Attitude levels were rated with a five-point Likert scale. Likert type statements were as follows: “I totally agree”, “I agree”, “I am not sure”, “I disagree”, “I totally disagree”. Total scale score ranged from 35 to 175. A high scale score showed severity of Internet dependence.

**Procedure**

Total scale score was the dependent variable. The independent variables were as follows: Daily amount of time spent on the Internet, how long have students used the Internet, the most frequent intended use, educational background of mother and educational background of father, family income and gender. Findings of empirical studies in the literature formed a basis for independent variable choice. Regression model of the research was as follows: dependence score = Constant + \(b_1\) hour + \(b_2\) year + \(b_3\) intended use + \(b_4\) mother + \(b_5\) father + \(b_6\) income + \(b_7\) gender and covariances. The model was defined by two classification techniques: Classification Tree and Chaid analysis method. Total scale score was classified by two-stage cluster analysis as follows: low-moderate-risky-dependent. Therefore, the dependent variable was artificially made discontinuous in classification. In that case, only Classification Tree of SRA was used. The model findings were compared by Classification Tree and Chaid analysis.

SRA, defined as Classification Tree in the event that the predicted variable was obtained by classified scale, and as Regression Tree in case of continuous variable, is known as a classification technique which is not held responsible for parametric regression technique assumptions and defines correlations between dependent variable(s) and independent variables in its own population, without any inference with data set val-
UES (Chang, & Wang, 2006; Yamauchi, Ono, Baba, & Ikegami, 2001). SRA presents predictive variables which may affect the predicted variable in a tree shape structure according to significance level. Such a visual case makes it superior to other regression techniques (Hebert, Delphine, Daigneault, Nathalie, & Tremblay, 2006; Kayri, & Boysan, 2008). Algorithm of SRA is based on gathering similar variables on the same tree knot and all sub-branches are tied to the root knot, the predicted variable (Breimen, Freidman, Olshen, & Stone, 1998). In the literature, it is reported that SRA is considered as a data mining method and has been widely used particularly in medicine, industry, and engineering (Chang, & Wang, 2006). Decision trees, with estimation and definition making power, are the most frequently used classification models as they are cheaply established in data mining, they are easy to interpret, easily integrated with data base systems, and reliable (Pehlivan, 2006). One of the important statistical matters in scientific research is defining high effect levels as well as affecting factors (Doğan, & Özdamar, 2003). Chaid Analysis, which discusses factors that critically affect changes in predicted variables, attempts to determine interaction between variables in the model and moderate combinations (Doğan, & Özdamar, 2003). Chaid analysis is a statistical method which divides data cluster of categorical variables into detailed homogenous sub-groups to explain the dependent variable (Pehlivan, 2006). Furthermore, Chaid might be used in cases where the predicted variable is classified; categorical, ordinal categorical or continuous and predictive variables are continuous, categorical or classified categorical (Doğan, 2003). In SRA, the predicted and predictive variables are continuous, ordinal or classified. Although Chaid have functions of Classification Tree method, they are different in terms of algorithms they use. SRA obtains the most eligible tree structure according to the predicted variable type, by impurity criterion and comparison of sub-group impurity criteria, while Chaid creates the ideal tree structure (Pehlivan, 2006). Similarly, Chaid is defined as a method which continuously divides a population into different sub-groups or sections with a minimum intra-group predicted variable variance and a maximum intergroup predicted variable variance and determines interactions between variables or combinations (Doğan, 2003; Ratner, 1998).
Discussion

Median value, descriptive statistics of total score variable, was found as 68,500 and quartile deviation was found as 18,125. The lowest item-total score in the sample was 35, whereas the highest item-total score was 166. Median value of the sample was taken as threshold at the same time as the research sample was not normally distributed. When the Internet dependence level was taken as low-moderate-high, the following equation was considered as a defining criterion: “median-quartile deviation (low)”; “median (moderate-threshold)” and “median + quartile deviation (high)”. However, the equation was later given up thinking that it was not satisfactory for diagnosis and two-stage cluster analysis was used to examine all the extreme individuals in the sample.

Cluster analysis was used to obtain a more detailed result for the Internet dependence and it was observed that the sample consisted of four sub-groups. Accordingly, the first cluster was “dependent group”, the second cluster was “dependence risk group”, the third cluster was “threshold group” (non-dependent or unapt), the fourth cluster was “non-dependent users.” The mean item-total score of the third group was 67,000 ± 5.981. The group was taken as threshold group because the value was very close to the threshold value which was found as 68,500. The threshold group consisted of 222 (29.4%) individuals. Therefore, those who had item total scores higher than the threshold value were dependent or potentially dependent (risky) and those who had item total scores lower than the threshold value were non-dependent users. Accordingly, as it is clear from Table 2, the second cluster consisted of potential dependent users (risk group) with a mean item total score of 89,306 ± 7.689. The group consisted of 199 (26.4%) individuals. The first group consisted of dependent users with a mean item total score of 118,236 ± 10.792. The group consisted of 76 (10.1%) individuals and dependence level of the group was high. The fourth group (below the threshold group) consisted of non-dependent users with a mean item total score of 44,786 ± 6.830. The group consisted of 257 (34.1%) individuals and the non-dependent user group had the highest number of individuals in the sample.

In Classification Tree; CT method, the intended use of the Internet was taken as the most critical factor which affected the predicted variable. SA examined the intended use of the Internet for the sample in two separate homogenous knots. It was seen that dependence levels of those
who reported that they used the Internet for online news and research were rather low, while dependence levels of those who said that they used the Internet for web surfing, chat, games, music, movies, pornography and online shopping were high. Similarly, there was a correlation between daily amount of time spent on the Internet and dependence levels. It was seen that the non-dependent user group spent less time on contrary to the dependent user group. Gender was significant for those who were not included in the dependent user group, but spent 1-3 hours online and income level was significant for those who were included in the dependent user group or the risk group and spent 1-3 hours online. Finally, CT findings showed that none of the variables was influential on those with low income, but educational background of father affected those with high income levels. As it is clear from Figure 1, CT findings showed that fathers of those with high dependence levels were high school graduates or university graduates, and fathers of those with low dependence levels were literate, primary school graduates or secondary school graduates. Nevertheless, it was noteworthy that individuals whose fathers were literate, primary school graduates or secondary school graduates were largely included in the risk group.

Like CT, Chaid analysis showed that “the intended use of the Internet” was the most influential factor ($\chi^2 = 101.423$, df = 6, $p<.01$). According to Figure 3, the daily amount of time spent online was influential on those who used the Internet for chat, games, pornography, and online shopping ($\chi^2 = 17.672$, df = 3, $p < .01$). Here, a high daily amount of time spent online was heavily influential on those with high dependence levels. Similarly, a high daily amount of time spent online was influential on the risk group. At the same time, a low daily amount of time spent online was heavily influential on the risk group. Moreover, Chaid analysis did not find gender, educational background of parents and income significant in the model. Yet, unlike Chaid analysis, CT found educational background of father, income and gender significant in the model and this case was shown in the tree shape structure. Although the dependent and independent variables were categorical, Chaid analysis, based on chi-square, could not give as many details as CT. As a result of analysis, it was seen that CT gave accurate classification (62.5%). However, the percentage was 61.5% for Chaid, and accurate classification estimation was slightly lower than Classification Tree method. In this respect, Chaid analysis obtained different findings from CT. Generally,
CT estimated the whole model with a higher percentage than Chaid analysis (CT accurate estimation level =62.5%, Chaid accurate estimation level =61.5%). However, the difference was not very high.

10.1% of the participants included in the study were found as Internet dependent users. The percentage was found as 1.1% (Bayraktar, 2001), 3.1% (Kaltiala-Heino et al., 2004), 1.98% (Johansson, & Götestam, 2004), 20.7% (Yen, Yen, Chen, Chen, & Ko, 2007), 2.4% (Cao, & Su, 2007), 8% (Hardie, & Tee, 2007) and 4.3% (Jang, Hwang, & Choi, 2008) in similar studies. Internet dependence percentage varied depending on generalibility of populations in studies, information technology opportunities in various countries and social structure.

In this study, the key to Internet dependence appeared as “the intended use of the Internet”. The findings showed that the most frequent use of the Internet was “research”, but “chat” was considerably high, as well. Still, “pornography”, “chat” and “games” were the most frequent intended uses for dependent users. In the literature, these results were supported and it was concluded that those intended uses caused Internet dependence (Bayraktutan, 2005; Chen, Chen, & Paul, 2001; Dinicola, 2004; Everhard, 2000; Görül, 2002; Kim, & Kim, 2002; Murali, & George, 2007; Thatcher, & Goolam, 2005; Tvedt, 2007). Besides, as in other studies (Cao, & Su, 2007; Everhard, 2000; Gonzalez, 2002; Hardie, & Tee, 2007; Kim, & Kim, 2002; Kubey et al., 2001; Lin, & Tsai, 2002; Nalwa, & Anand 2003; Simkova, & Cincera 2004; Young, 1996; Weitzman, 2000), the present study showed that the amount of time spent on the Internet was an important indicator of Internet dependence, as well as the intended use of the Internet.

Particular variables which might affect Internet dependence levels of the participants were modeled. Total score form Internet Dependence Scale was taken as the predicted variable and the predictive variables were as follows: daily amount of time spent on the Internet, how long have students used the Internet, the most frequent intended use, educational background of mother and educational background of father, economic family income levels and gender variables. Classification Tree and Chaid analysis, two of decision tree methods, were used and the findings from the two methods were compared. Classification Tree and Chaid analysis, which use different algorithms for tree construction, are known as recommended decision trees in the literature. In the research, it was attempted to examine which method was able to show the
stronger tree in order to reveal Internet dependence creating factors in a stable and strong way.

It is well-known that Classification Tree and Chaid analysis are two useful methods to determine predictive variables which might affect the predicted variable in cases where the predicted variable is ordinal or classified but in this study, Classification Tree method was considered more stable during involving all the possible combinations in the model in terms of correlations between the variables.

In the light of the research model obtained by Classification Tree method, the intended use of the Internet might be considered as the main triggering factor for Internet dependence for Turkish secondary school students. Use of the Internet for “Web surfing”, “music-movies”, “chat”, “games”, “pornography” and “online shopping” increased severity of dependence based on “use time”. Use of the Internet for “online news” and “research” by these students did not result in a significant dependence. Similar studies supported the view (Chen, Chen, & Paul, 2001; Leung, 2004; Tekdal, 2005). It was striking that when income levels were relatively high, Internet dependence levels of those who use the Internet for “Web surfing”, “music-movies”, “chat”, “games”, “pornography” and “online shopping” increased. It was seen in the research that students with low income levels were mostly included in the risk group. Although educational background of father was influential on Internet dependence, it was interesting that the students whose fathers were high school graduates or university graduates were more dependent users. It was also observed that the students whose fathers were literate, primary school graduates or secondary school graduates were intensively included in the risk group. In this context, educational background of father played an important role in shaping Internet dependence. As education level increased, Internet dependence level increased and as education level decreased, Internet dependence decreased, as well. However, the fact that the students whose fathers were literate, primary school graduates or secondary school graduates were intensively included in the risk group supported the assumption that it may increase the severity of Internet dependence in time. Based on the fact that social and individual dynamics may change over time, the findings of the study might differ in the process and in this context, further studies are recommended to reexamine Internet dependence levels with different measurement tools by taking possible changes into account.


