Working Students and Their Academic Performance – A Decision Tree Analysis

Guoying Zhang
Midwestern State University

Chris Y. Shao
Tarleton State University

Charles R. Johnston
Midwestern State University

In this study, a decision tree analysis is conducted to identify the effects from working. The result shows that student age is the first major indicator for better grade regardless of working status. Then, the factor of students' self-perception on the effect of working on academic performance matters a lot. A pessimistic student, who believes in the negative impact of working on studying, needs a balanced combination of course work and working load. However, for an optimistic student, academic standing is important. Senior students with a positive perception on working are more likely to validate this perception by good academic performance. While, for students in other academic standings, working for a job relevant to major can help. Otherwise, a moderate level of working load is still recommended. The analysis approach can be easily applied to any academic counselling: to identify when working intensity can matter, which group of students may be more vulnerable to a negative impact of working, and what working aspects may play a role in academic performance.

Keywords: Decision Tree Analysis, Working Students, Academic Performance

INTRODUCTION

The literature suggests that the majority of full-time university students in the U.S. have jobs. Factors contributing to this likely include increasing financial pressures from higher tuition and fees, the desire to maintain a certain level of lifestyle, the necessity to help support a family, attempting to gain social and work experiences, and academic requirements for practical internships. Regardless of the individual reasons, the fact is, it has become commonplace for many full-time, as well as part-time, students to work during some, if not all, of their undergraduate college careers. Gose's 1998 survey reported that 39% of freshman college students worked at least 16 hours per week, a 4% increase from 1993. Studies by Hawkins et al. (2005), Bradley (2006), Nonis and Hudson (2006), Bennett et al. (2007), and Miller et al. (2008) would support a conclusion that probably over 50% of full-time university students in the United

States now have jobs. Internationally, working is no less common for university students (Bradley, 2006; Holmes, 2008; Callender, 2008).

It is also commonly expressed by faculty and administrators that the academic performance of students suffers due to the increasing outside workloads they carry while pursuing their academic work. However, this perception has not been universally supported by past research efforts. These efforts include findings of both positive and negative relationships between working and academic performance. It was felt that further research utilizing traditional techniques would be unlikely to produce any significant new insights into these conflicting results. By combining decision tree modeling with a large, diverse sample from a total university population, this research seeks to take a new alternative approach to the complicated interaction between students' working and their academic performance. More importantly, the results of prediction rule can be followed up with more retention measures to reduce student at risks.

RELATED WORK

In The United States

One obvious approach to support a hypothesis of working producing a negative effect on academic performance incorporates the supposition that time spent working takes away from time available for studying. Corollaries to that supposition are a positive relationship existing between the time spent studying and academic performance, and a negative relationship between time spent working and academic performance. Even these simple intuitive propositions have not been consistently confirmed by past research. The confusion created by research in this area is perhaps best examined by dealing with past efforts in a chronological order. This examination is not intended to be totally inclusive, only comprehensive enough to set the context in which the present study approach was developed.

Pascarella and Terenzini (1991) and McFadden and Dart (1992) found a positive relationship between time spent studying and academic performance. No such relationship was found by Mouw and Khanna (1993). In looking at these and other past research, Furr and Elling (2000) investigated students working off-campus and postulated a benefit for a moderate amount of employment, with full-time employment producing a negative effect on GPA. While students perceived a negative impact of working, this was not supported by the quantifiable data collected. There was no significant relationship found between their cumulative GPAs and the number of hours they worked.

Conversely with respect to perceptions, Curtis and Lucas (2001) reported that few full-time undergraduate English students perceived any negative impact on their academic pursuits from working. In further support of "no negative relationship" between working and academic performance, Strauss and Volkwein (2002) found that there was a positive relationship and speculated that working students may actually bring that work ethic with them to the efforts thay make in school. Ackerman and Gross (2003) may have also contributed to this view by finding that students having less free time actually leads to higher GPAs, not lower.

Finding just the opposite, Hawkins et al. (2005) confirmed a statistically significant negative relationship between the perceived interference with academic performance (overall GPA) and the number of hours worked. Their sample from two universities consisted of 300 social work majors and it is perhaps noteworthy for this study that the data for the variables were gathered by self-reporting rather than objectively.

Nonis and Hudson (2006) considered all this past research and were also unable to establish any relationship between the time spent studying outside of class and academic performance represented by students' GPAs. Their research looked at the effect of the time spent working, as well as the time spent studying, on the academic performance of business students at an Association to Advance Collegiate Schools of Business (AACSB) accredited, public university in the United States. In addition to finding no significant relationship between the time spent studying and academic performance, they also could not find any relationship, either positive or negative, between the time spent working and academic performance as objectively measured by official semester GPAs.

Svanum and Bigatti (2006) continued this line of research and produced an article. Their conclusions again seemed to go in the opposite direction and further confuse the issue. They found a lower "GPA-indexed potential for course success" when working resulted in students being unable to devote as much effort to the course. They also postulated a greater effect on average students than above-average students over time. This result has some bearing on the approach of the current research effort as will be seen later in the paper.

In related research, Miller et al. (2008) sampled 903 students at a southeastern university for health risks associated with working and attending college. Self-reported GPA was divided into two categories, 3.0 or greater being "good" and less than 3.0, to measure academic performance. Students who worked at least 20 hours per week exhibited significantly lower performance in terms of GPA. In addition to the work based on the proposition of the negative impacts of working on academic performance, findings of students' perceptions of a positive impact also exist (Holmes, 2008).

Finally, Zhang and Johnston (2010) investigated the relationship between employment and the academic performance of business students during a specific semester and could find no significant relationship between the number of hours worked per week and semester grade point average.

Internationally

Studies outside the United States have also produced similarly confusing outcomes. Bradley's study (2006) indicated that about 85% of the Australian university students in his sample of 246 were employed during the semester in which the research was conducted. This was a surprisingly high percentage but supportive of the 72% found to be working in a national survey of university students by Devlin et al. (2007). For students that worked 20 hours or more per week in Bradley's research, no significant relationship existed between the number of hours worked and their GPA's. This was despite the fact that a perceived negative effect of working was present and increased with hours worked. Interestingly, an indication of job satisfation having a positive correlation with GPA may suggest why the highest GPA's were associated with students that didn't work and those that worked more than 20 hours per week, even though the finding was not significant.

In a 2008 study by Callender in the UK involving 1012 students from six different universities seemingly contradictory conclusions were reached. Engaging in paid employment while university students was found to produce negative effects on both grades in the short term and on their careers in the long term by relegating them to lower paid jobs. Increasing the number of hours worked by the students also increased these effects.

A Different Approach Utilizing Decision Tree Analysis

The previously discussed studies have served to confirm to many researchers that the relationship between working and academic performance must involve the interaction of more factors than simply the extent of student employment and study time. However, even while considering more factors, many of these studies were conducted with a "cause and effect" approach in attempting to establish relationships among the factors utilized.

In fact, studies seeking to imply cause and effect relationships between working and academic performance can be complex to compare due to their specific sample demographics, use of non-objective self-reported predictors, and a lack of linkage to past academic performance. In order to provide a more comprehensive study in this research stream, one objective of this paper is to utilize a larger student sample in a university domain without restriction to a particular college or major, establish a more reliable framework to capture student perceptions on the effect of working, while incorporating predictors of past academic performance and working experiences. Academic performance is also evaluated through both long-term GPA (cumulative GPA) and short-term GPA (semester GPA).

In order to perform a comprehensive analysis to reveal the role of employment in academic performance at a university level, student information on a comprehensive set of factors needs to be collected. The factors include demographics, current academic performance, past academic performance, current working information, past working information, social commitments other than working and study, and students' own perceptions of how working affects their academic performances. Hence, working characteristics investigated here encompass not only how much students worked during the semester of investigation and in the past, the nature of their employment and why they worked, but also their perceptions of working effects on academic performance, such as low attendance due to working schedule conflicts, no job relevance to academic majors, low physical preparedness for study due to full time working, low mental preparedness for study, few out-of-class communication with fellow students and instructors, commuting and parking distractions, and so on.

Due to the mixed results of attempting to establish a relationship between student employment and academic performance generated from tradtional approaches such as ANOVA, linear regression, and logistic regression, among others, it was decided that a different approach must be undertaken with this research to produce contribution to the literature. Decision tree techniques, as some of the most important data mining techniques, have been widely applied in the research field of education.

Chan (2007) used decision tree analysis to predict student performance in Web-based e-learning systems. Thomas and Galambos (2004) used decision tree techniques to predict university students' satisfaction. Kotsiantis et al. (2003) studied how to prevent student dropout in distance learning systems using decision trees. Superby et al. (2006) determined factors influencing the achievement of first-year university students using data mining, and one of the techniques used was decision tree analysis. Mihaescu and Burdescu (2006) also used decision trees for classifying students according to their accumulated knowledge in e-learning systems. Hsia et al. (2008) conducted decision tree analysis in course planning to analyze enrollees' course preferences and course completion rates. Ranjan and Khalil (2008) explored the potential effects of changes in recruitments, admissions and courses using deicison tree analysis.

Decision tree induction has also been widely used in various business fields. It is a common method used in data mining (Shmueli et al. 2007). The goal of buiding deicion trees is to create a model that predicts the value of a target variable based on input variables (sometimes refered to as feature variables). In this situation, a decision tree framework can help univeristy or college to identify working student at-risk at an earlier stage, and therefore, provide corresponding retention measures to help students.

A decision tree can be generated by first splitting the whole data set into two subsets based on a specific value of a feautre variable. This process is repeated on each derived subset. The ideal split would divide a data set into two child sets so that each set contains nodes of the same value for the target variable. Unfortunately, perfect splits do not occur often, so it is necessary to evaluate and compare the quality of imperfect splits. Various splitting criteria have been proposed for evaluating splits, but they all have the same goal which is to favor homogeneity within each child set and heterogeneity between the child nodes (Safavian and Landgrebe 1991).

Another important determination in decsion tree induction is to establish when to end the splitting process. Ideally, the spliting process is completed when the resulting child set contains nodes of the same value of the target variable. However, in most cases, a tree can only reach such purity when the child set contains only one single node. Therefore, most decision tree induction adopts some other stopping criteria such as limiting the depth of the tree, limiting the minimum number of data nodes in a subset, or setting a tolerrable error rate (Safavian and Landgrebe 1991).

In this study, DTREG software was used to conduct the decision tree induction (www.dtreg.com). DTREG tries each feature variable to see how well it can divide a data set into two child sets. Once DTREG has evaluated each possible split for each possible feature variable, a dataset is split using the best split found. DTREG provides two methods for evaluating the quality of splits when building decision (classification) trees; Gini and Entropy.

The stopping criteria used by DTREG includes minimum size of the data set to split and maximum tree depth. However, DTREG does not use its stopping criteria as the primary means for deciding how large a tree should be. Instead, it builds an overly-large tree, and then analyzes the tree by using V-fold cross-validation and prunes it back to the optimal size. This is known as backward pruning.

With the decision tree incorporating best-splitting variables, it is possible to navigate the tree to identify when working intensity may matter, which group of students may be more vulnerable to a negative impact of working, and what working characteristics may play a role in academic performance. By looking at relevant variables branching the tree, it is anticipated that more meaningful insights can be generated than in past research with respect to the effects that student employment may have on their academic performance in colleges, and take follow up actions to help working students when the predication rule can be applied.

METHODOLOGY AND DATA COLLECTION

The target population consists of the approximately 6,000 students in all the colleges of the authors' university. Students' current and past employment information, social committmements status, and their perceptions of the effects of working on academic performance were collected through an online survey instrument utilizing the SurveyMonkey platform. With the results from piloting the survey instrument, a formal online questionnaire was set up to begin at approximately the 10th week of a fall semester. Students were guided to the survey via university postmaster emails, school website links, and follow-up emails. The online survey instrument was kept open until the end of the fall semester in order to maximize the number of respondents. To encourage participation throughout the process, several drawings of cash prizes for $75 and $25 were also conducted among the participants as incentives for student participation.

For the measurement construct of student perceptions of the effects of working on academic performance, 5-point Likert scale questions were used to measure items of the perceptions about working related factors. Reliability analysis were further conducted. Lastly, categorical variables were recoded accordingly.

Prior to the beginning of the spring semester, official information was retrieved from university records on the student participants' demographics, academic performances (both fall semester GPA and cumulative GPA), and total credit hours (for both the semester and overall) using student identification numbers. Therefore, this study avoids the inaccuracy of self-reporting problems in some of the past literature where the academic performance measurements were reported or estimated by the students themselves.

At the end of the fall semester, a total of 2103 unique and valid surveys were collected. Further, after the addition of the previously mentioned information from the institutional research department of the university, all individual student identifying data were eliminated from each record prior to statistical manipulation. Minor data cleaning and missing value handling were performed by deletion and imputation (Little, 1992).

RESULTS

Descriptive Statistics

The categorical variables for "All Students" are summarized with corresponding frequency in Appendix I. Inspection of the distribution of the categorical levels revealed that the demographic and academic information of our student participants are reasonable representations of the entire student body of the university. Appendix II further lists the descriptive statistics of continuous variables for "All Students." With regard to employment information, 70% of students were employed during the semester. 49% of students had commitments other than school and working with an average over 20 hours per week.

In order to preprocess the data, it is identified that graduate students in the data sample naturally had a relatively high GPA. Further, when considering working students, many of the graduate students work as "graduate assistants" which is not a typical job type for the majority of college students. Therefore, 276 graduate students data were removed from the sample set for analysis. Since we are studying the working effect on academic performance, the data set is further refined to include only undergraduate working students. Table 1 below illustrates the categorical information about our final data sample.

TABLE 1
CATEGORICAL VARIABLES (UNDERGRADUATE WORKING STUDENTS)

Gender		
Female	786	67.0%
Male	388	33.0%
Other Commitment		
No	583	49.7%
Yes	591	50.3%
Academic Standing		
Freshmen	118	10.1%
Sophomore	201	17.1%
Junior	295	25.1%
Senior	560	47.7%
Job Status		
Both On and Off Campus	44	3.7%
Off Campus	952	81.1%
On Campus	178	15.2%
More than One Employer		
Not Available	26	2.2%
No	948	80.7%
Yes	200	17.0%
Job's Relevance to Major		
Not Available	26	2.2%
No	758	64.6%
Yes	390	33.2%
Working Schedule		
Not Available	26	2.2%
4:00 pm - midnight	494	42.1%
8:00 am - 4: 00 pm	601	51.2%
Midnight - 8:00 am	53	4.5%
Primary Reason for Working		
Not Available	26	2.2%
Internship Opportunity	4	0.3%
Financial Needs	1015	86.5%
Other	53	4.5%
Social Skills	14	1.2%
Gain Experiences	62	5.3%
Perceived Job Effect on Academic		
Not Available	31	2.6%
Negative	425	36.2%
Neither	376	32.0%
Positive	342	29.1%
Grand Total *N*	**1174**	**100.0%**

According to Table 1, 87% of working students worked because of financial needs. 65% of working students had a job with no relevance to their academic major. 42% of working students had to work mainly outside of regular business hours. Finally, 36% of working students believe that employment has a negative effect on their academic performance. These numbers shows that working is a common norm now for college students due to financial reasons. In addition, most of them working in a field not relevant to their study, and they do have to work out of regular business hours. Table 2 further illustrates the continous variables of only undergraduate working students sample.

TABLE 2
CONTINUOUS VARIABLES (UNDERGRADUATE WORKING STUDENTS)

Variable	N	Minimum	Maximum	Mean	Std. Deviation
Age	1174	16.00	67.00	26.20	8.58
TotalCreditHours	1174	0.00	245.00	88.21	42.75
FallCreditHours	1174	0.00	21.00	11.19	3.96
FallGPA	1174	0.00	4.00	2.99	0.82
OverallGPA	1174	0.00	4.00	3.01	0.64
PreFallGPA	1173	0.00	4.00	3.01	0.65
OtherCommittment	501	0.00	168.00	20.40	24.71
OverallWorkingHours	1120	0.00	168.00	29.69	15.17
FallWorkingHours	1129	0.00	100.00	29.14	13.40
FallCommutingHours	1075	0.00	15.00	2.35	2.75
NegativeOpinion	422	1.14	5.00	3.78	0.62
Scaletime	422	1.00	5.00	4.46	0.91
Scalephysical	422	1.00	5.00	4.14	0.92
Scalemental	422	1.00	5.00	4.21	0.90
Scalecommunication	422	1.00	5.00	3.49	1.10
Scaleschedule	422	1.00	5.00	2.38	1.25
Scaledistraction	422	1.00	5.00	3.48	1.18
Scalepressure	422	1.00	5.00	4.32	0.88

According to Table 2, the average age of these working students is 26. These working students in average took 11 semester credit hours. Average semester GPA is 3.0. Students have committements other than school and work with 20 hours per week. The average working hours per week is about 29. The average commuting hours for working is over 2 per week.

Students who answered that the perceived effect of working on academic performance was negative, were further requested to indicate their opinions as to why on seven 5-point Likert scale questions. The purpose of these questions is to validate the negative perception. The questions relate the negative opinion with the following items: less time, physical tiredness, mental tiredness, fewer communications, schedule conflicts, more distractions, and more pressure. *NegativeOpinion* is an average of the 7 item scores in the Likert scale questions, and the mean is 3.8 for our sample. The reliablity analysis of the variable shows an acceptable reliablity at a level of Cronbach's α=0.7. Therefore, 3.8 out of 5 indicates a strong level of agreement toward the negative perception of working toward academic performance.

Decision Tree Analysis

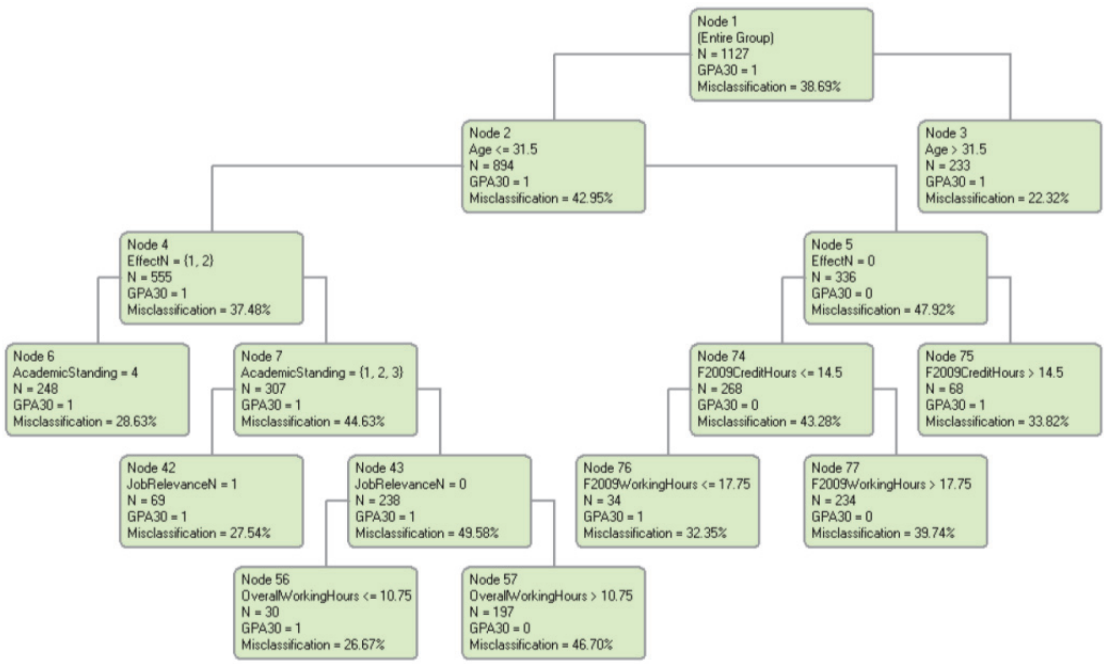
Decision Tree analysis being applied in a data mining context leads to it being necessary to determine how to handle the binary transformation of the target variable, fall semester GPA (representing academic performance). Many preliminary models were generated and examined utilizing the data from "Undergraduate Working Students." Eventually, several conclusions were reached with regard to the most useful of the models and it is those models that are discussed in the remainder of the paper.

Models were initially run with the threshold of 3.0 for the binary transformation of the fall semester GPA target variable. This is the same as the mean fall semester GPA of the data sample (2.99 in Table 1) and on a 4.0 GPA scale (a 'B' average) is often considered a threshold of good academic performance. This target variable split in DTREG was set at GPA>=3.0 as 1 and GPA<3.0 as 0.

Additionally, since decision tree analysis using DTREG doesn't allow missing information in categorical variables, corresponding records were also eliminated. DTREG is used to build the tree for classifying data nodes based on values of other feature variables. Feature variables defined here include working information, general academic information, and demographic information. The decision tree model of classification doesn't assume any linear relationship of the fall semester GPA with other variables; instead it picks the best variable to split each tree node in a hierarchical fashion.

Decision trees were generated using all the working student records from the data set refined to include only undergraduates. The total number of this sample is 1174 students. For all of these students in the sample, 61% had a fall semester GPA >= 3.0 and 39% had a fall semester GPA < 3.0. Therefore, when assuming they all have a semester GPA >= 3.0, the "misclassification rate" is 39%. The resulting classification tree is presented in Figure 1.

FIGURE 1
DECISION TREE FOR UNDERGRADUATE WORKING STUDENTS



The derived decision tree includes factors of *Age*, *FallCreditHours*, and *AcademicStanding*. On top of those demographic and academic related variables, several working related variables are also presented in the tree to help with the classification: *JobRelevance, PerceivedJobEffect*, *FallWorkingHours*, and *OverallWorkingHours.* Several of interesting decision tree rules are derived. Please note that only when the tree leaf node has a misclassification rate lower than 39% (the root node misclassfication rate), can that leaf node be used as a rule.

Rule 1: If a working student is older than 31.5 years old, then semester GPA will be greater than or equal to 3.0 with a misclassification rate of 22.3%. Therefore, 31.5 is an important age threshold for working students. Among older students, working doesn't explain the difference in academic performance. In other words, older working students can balance better between working and studying, and working is not a factor in academic performance.

However, if a working student is *not* older than 31.5 years old, several working related aspects can definite influence the academic performance. We elaborate these rules below.

Rule 2: If a younger working student has a *negative* perceived job effect on academic performance, but with semester credit hours more than 14.5, then semester GPA will be greater than or equal to 3.0 with a misclassification rate of 33.8%. These students selfselect to enroll in more classes to dedicate more time to coursework, and consequently, achieve a better academic performance despite their negative perception of working.

Rule 3: If a younger working student has a *negative* perceived job effect on academic performance, with semester credit hours *no* more than 14.5, and with weekly working hours *no* more than 17.8, then semester GPA will be greater than or equal to 3.0 with a misclassification rate of 32.4%. These students have a less loaded course schedule and a less intensive working load. In other words, they might selfselect to enroll in moderate amount of coursework and choose low working intensity (note that average working hours per week is 29.1) to achieve a better academic performance.

Rule 4: If a younger working student has an indifferent or positive perceived job effect on academic performance, and in a senior standing, then semester GPA will be greater than or equal to 3.0 with a misclassification rate of 28.6%. These positive senior students are capable of manaing working and studying together, and likely to achieve a better academic performance.

Rule 5: If a younger working student has an indifferent or positive perceived job effect on academic performance, but *not* in a senior standing, while having a job relevant to major, then semester GPA will be greater than or equal to 3.0 with a misclassification rate of 27.5%. These positive non-senior students are fortunate to have a job relevant to major. This probably explains that working is facilitating studying and these students intend to achieve a better academic performance.

Rule 6: If a younger working student has an indifferent or positive perceived job effect on academic performance, but *not* in a senior standing, and has a job *not* relevant to major, while having an overall average working intensity in college *no* more than 10.8 hours per week, then semester GPA will be greater than or equal to 3.0 with a misclassification rate of 26.7%. These positive non-senior students do not work intensively in college (average is 29.7), although with a job *not* relevant to major shows that these students can still achieve a better academic performance with light working load.

To better summarize these rules, we provide Table 3 for a clear visual resprenetation of these rules. The six rules incorporate different factors determining good academic performance.

TABLE 3
SUMMARY OF DECISION TREE RULES

<table>
<tr><td>>31.5 years old</td><td colspan="5"><= 31.5 years old</td></tr>
<tr><td rowspan="5">GPA>= 3.0</td><td colspan="2">Negative Perceived Job Effect</td><td colspan="3">Positive or No Perceived Job Effect</td></tr>
<tr><td rowspan="3">>14.5 CreditHours</td><td rowspan="3"><=14.5 CreditHours; and <=17.8 WeeklyWorkingHours (Semester)</td><td rowspan="3">Senior</td><td colspan="2">Freshman, Sophomore, Junior</td></tr>
<tr><td rowspan="2">Job Relevant</td><td>Job Not Relevant</td></tr>
<tr><td><= 10.8 WeeklyWorkingHours (Overall)</td></tr>
<tr><td colspan="2">GPA >= 3.0</td><td>GPA >= 3.0</td><td>GPA >= 3.0</td><td>GPA >= 3.0</td></tr>
</table>

DISCUSSION

The decision tree analysis shows that for working students, student age is the top prominient factor to determine good academic performance. If students are older than 31.5 years, they are very likely to have high semester GPA (3.0) regardless of working related factors. This might be explained by the age and maturity of these students. They usually have clear motivation toward college study, established study pattern, and relatively strong time management skills. In other words, college retention focus should not be placed toward these students.

The study further indicates that only when students are *no* more older than 31.5 years, some aspects of working come into play and additional rention intervention might help these younger students.

We start with students' perception toward job effect on academic performance. Based on their past experience of working, students might believe that working has *negative* effects on academic peformance. With this group of students, a simple way for intervention is to encourage more course work to occupy more time from students. Students taking more than 14.5 credit hours, which is equivalent to 5 courses per semester, might have to dedicate more time studying. This will prioritize studying above working for students, and consequently, leads to a desirable academic performance. However, due to various reasons (such as financial reason), if a student cannot commit to that many credit hours of study per semester, the proper suggestion to him or her is to work relatively less (with no more than 17.8 hours per week). This suggests a part time job (less than 20 hours per week) should be a better choice over a full time job.

On the other hands, some students might believe that working has *positive or no* effects on academic peformance. If they are senior students, no further actions are needed. They are positive toward working, and their senior status indicates they are already capable of balancing working and studying, and they are very likely to have good academic performane. For those who are not in their senior standing, it is highly suggested that they can benefit from working in major related field, either through an internship or recommendation from major advisors. In other words, if students are positive toward working and able to work in the major related field, the working experience can enhance their academic performance with hands on knowledge. Last but not the least, if a positive student in this group is not able to locate such a major related job, the suggestion is still simple – work less in general, or no more than 11 hours per week throughout the entire college period.

CONCLUSION AND FUTURE DIRECTION

There seems to be an overriding "maturity" theme associated with higher levels of academic performance. These students being in their early thirties or older is a solid indicator of superior academic performance. If students are younger, the following insights can be derived.

Students perception of working is one the most important factors. If they believe working negatively impacts their acadmic performance, then they need to pay close attention to the amount of both course load and work load. A moderate and balanced combination of course load and working load will be helpful for students to achieve a better academic performance.

However, if students are confident in their capabilities of balancing working and student, that is, having a either positive or indifferent perception on the relationship of working and academic performance, then the actual load might not matter a lot. In this scenario, it is found that Senior standing students usually carry out good acadmic performance regardless of any other working related factors. Freshman, Sophomore, and Junior standing students, on the other hand, need to find a job relevant to their major to increase the chance of a better grade. If this cannot be implemented, then they had better to committ a moderate level of working load though they don't feel working negatively impact academic performance.

Instead of drawing a conclusion of whether working is good or bad for academic performance, our study utilizing a decision tree technique to classify students based on their academic and work related information. Several rules are derived to understand the complexity of working effect on students. It is interesting to show that age, academic standing, working hours per week, credit hours per semester, job relevance are all determining acadmeic performance. In addition, students' self perception matters a lot. Different self percetions lead to different deciding factors.These rules help collge to embrace the fact of student working, and design relevant rention strategies.

There is no single rule fitting to all working students. We believe college rention intervention should focus on mainly younger students. A self perception survey can be distributed first to seek for students' perception of job effect on academic performance. The result (negative, or otherwise) can be associated with student records, and communicated to academic advisors of students. Based on the result and student acadmic standing, advisors can help students to determine an appropriate amount of course load each semester, make suggestions of finding major relevant jobs, caution toward excessive work hours per week, and propose moderate working intensity accordingly. After all, the academic advisors can no longer just advise on "acadmic" related content, they have to take into consideration of how student can balance working with studying.

Overall, this study provides a useful framework to conduct decision tree analysis on the relationship of working and studying. The derived rules can be easily adoped in advising, counselling, policy making, among others.

There are several future directions for this study. It is expected that a regular survey will be conducted and the decision tree model will be re-evaluated with new supply of data for its robustness. Further, the scope of the data is limited to the university under investigation. In the future, it can be expanded to more than one higher education institution with the support of a regional interest. Last but not the least, the attempt of using decsion tree analysis for understanding students is simply a starting point, there are lots of other data mining techniques, which can be used to explore this field to generate up-to-data insights about student working and academic performance.

REFERENCES

Ackerman, D.S., & Gross, B.L. (2003). Is time pressure all bad? Measuring between free time availability and student performance perceptions. *Marketing Education Review*, 12, 21-32.

Bennett, S., Evans, T., & Riedle, J. (2007). Comparing Academic Motivation and Accomplishments Among Traditional, Nontraditional, and Distance Education College Students. *Psi Chi Journal of Undergraduate Research*, 12(4), 154-161.

Bradley, G. (2006). Work participation and academic performance: a test of alternative propositions. *Journal of Education and Work*, 19(5), 481-501.

Callender, C. (2008). The impact of term-time employment on higher education students' academic attainment and achievement. *Journal of Education Policy*, 23(4), 359-377.

Chan, C.C. (2007). A Framework for Assessing Usage of Web-Based e-Learning Systems. *Proceedings of International Conference on innovative Computing, Information and Control*, Washington, DC, 147- 151.

Chen, G., Liu, C., Ou, K., & Liu, B. (2000). Discovering decision knowledge from web log portfolio for managing classroom processes by applying decision tree and data cube technology. *Journal of Educational Computing Research*, 23(3), 305–332.

Cubeta, J.F., Travers, N.L., & Sheckey, B.G. (2001). Predicting the academic success of adults from diverse populations. *Journal of College Student Retention*, 2(4), 295-311.

Furr, S.R., & Elling, T.W. (2000). The Influence of Work on College Student Development. *NASPA Journal*, 37(2), 454-470.

Gose, B. (1998). More freshman than ever appear disengaged from their studies, survey finds. *The Chronicle of Higher Education*, A37-A39.

Hawkins, C.A., Hawkins, R.C., Smith M.L., & Grant, D. (2005). The Relationship among Hours Employed Perceived Work Interference, and Grades as Reported by Undergraduate Social Work Students. *Journal of Social Work Education*, 41(1), 13-27.

Hayes, A.F. (2005). *Statistical Methods for Communication Science*, 1st edition. Lawrence Erlbaum Associates, pp. 376.

Hocking, R.R. (1976, March). The Analysis and Selection of Variables in Linear Regression. *Biometrics*, 32, 1-49.

Holmes, V. (2008). Working to live: Why university students balance full-time study and employment. *Education + Training*, 50(4), 305-314.

Hsia, T., Shie, A., & Chen, L. (2008). Course planning of extension education to meet market demand by using data mining techniques – an example of Chinkuo Technology University in Taiwan. *Expert System Application Journal*, 34(1), 596-602.

Kotsiantis, S., Pierrakeas, C., & Pintelas, P. (2003). Preventing student dropout in distance learning systems using machine learning techniques. *Proceedings of International Conference on Knowledge-Based Intelligent Information & Engineering Systems*, Oxford, 3-5.
Light, R.J. (2001). *Making the most of college*. Cambridge, MA: Harvard University Press, pp. 27.
Little, R.J.A. (1992). Regression with Missing X's: A Review. *Journal of the American Statistical Association*, 87, 1227-1237.
McFadden, K., & Dart, J. (1992). Time Management skills of undergraduate business students. *Journal of Education for Business*, 68(2), 85-88.
Mihaescu, C., & Burdescu, D. (2006). Testing attribute selection algorithms for classification performance on real data. *Proceedings of International IEEE conference Intelligent Systems*, Varna, Bulgaria, 581-586.
Miller K., Danner, F., & Staten, R. (2008). Relationship of Work Hours with Selected Health Behaviors and Academic Progress among a College Student Cohort. *Journal of American College Health*, 56(6), 675-679.
Mouw, J., & Khanna, R. (1993). Prediction of academic Success: A review of the literature and some recommendations. *College Student Journal*, 27(3), 328-336.
Nonis, S.A., & Hudson, G.I. (2006). Academic Performance of College Students: Influence of Time Spent Studying and Working. *Journal of Education for Business*, 81(3), 151-159.
Ranjan, J., & Khalil, S. (2008). Conceptual Framework of Data Mining Process in Management Education in India: An Institutional Perspective. *Information Technology Journal*, 7(1), 16-23.
Safavian, S.R., & Landgrebe, D. (1991). A Survey of Decision Tree Classifier Methodology. *IEEE Transactions on Systems, Man, and Cybernetics,* 21(3), 660-674.
Shmueli, G., Patel, N.R., & Bruce, P.C. (2007). *Data Mining for Business Intelligence*. Hoboken, NJ: John Wiley & Sons, Inc.
Stinebrickner, R., & Stinebrickner, T.R. (2003). Working during School and Academic Performance. *Journal of Labor Economics*, 21(2), 473-491.
Strauss, L.C., & Volkwein, J.F. (2002). Comparing Student Performance and Growth in 2- and 4-Year Institutions. *Research in Higher Education*, 43(2), 133-161.
Superby, J.F., Vandamme, J.P., & Meskens, N. (2006). Determination of factors influencing the achievement of the first-year university students using data mining methods. *Proceedings of International conference on intelligent tutoring systems*, Educational Data Mining Workshop, Taiwan, 1-8.
Svanum, S., & Bigatti, S.M. (2006). The Influences of Course Effort and Outside Activities on Grades in a College Course. *Journal of College Student Development*, 47(5), 564-576.
Thomas, E.H., & Galambos, N. (2004). What satisfies students? Mining student-opinion data with egression and decision tree analysis. *Research in Higher Education Journal*, 45(3), 251-269.
Velleman, P.F., & Wilkinson, L. (1993). Nominal, ordinal, interval, and ratio typologies are misleading. *The American Statistician*, 47(1), 65-72.
Zhang, G., &Johnston, C. (2010). Employment and the Academic Performance of Undergraduate Business Students. *Southwestern Business Administration Journal*, 10(2), 54-83.

APPENDIX 1
CATEGORICAL VARIABLES (ALL STUDENTS)

Demographic Information	***n***	**%**
Gender		
Not Available	56	2.7%
Female	1403	66.7%
Male	644	30.6%
Race/Ethnicity		
Not Available	56	2.7%
Hispanic	205	9.7%
Indian or Alaskan Native	18	0.9%
Asian or Pacific Islander	65	3.1%
Black or African American	159	7.6%
White	1362	64.8%
Non-Resident Alien	231	11.0%
Unknown	7	0.3%
Marital Status		
Not available	11	0.5%
Divorced	136	6.5%
Married	544	25.9%
Single, never married	1403	66.7%
Single, spouse deceased	9	0.4%
Other Commitment		
No	1073	51.0%
Yes	1030	49.0%
Academic Information		
Academic Standing		
Not Available	56	2.7%
Freshmen	257	12.2%
Junior	314	14.9%
Sophomore	420	20.0%
Senior	780	37.1%
Graduate Student	276	13.1%
Employment Information		
Job Status		
Not Employed	640	30.4%
Both On and Off Campus	64	3.0%
Off Campus	1149	54.6%
On Campus	250	11.9%
More than One Employer		
Not Available	670	31.9%
No	1178	56.0%
Yes	255	12.1%
Job's Relevance to Major		
Not Available	670	31.9%
No	836	39.8%
Yes	597	28.4%
Working Schedule		
Not Available	671	31.9%
4:00 pm - midnight	524	24.9%
8:00 am - 4: 00 pm	847	40.3%
Midnight - 8:00 am	61	2.9%
Primary Reason for Working		
Not Available	670	31.9%
Internship Opportunity	8	0.4%
Financial Needs	1249	59.4%
Other	75	3.6%
Social Skills	16	0.8%
Gain Experiences	85	4.0%
Perceived Job Effect on Academic		
Not Available	676	32.1%
Negative	497	23.6%
Neither	468	22.3%
Positive	462	22.0%
Grand Total *N*	**2103**	**100.0%**

APPENDIX 2
CONTINUOUS VARIABLES (ALL STUDENTS)

Variable	N	Minimum	Maximum	Mean	Std. Deviation
Age	2047	16.00	67.00	26.57	8.98
TotalCreditHours	2037	0.00	245.00	75.11	46.26
FallCreditHours	1730	0.00	21.00	10.81	4.53
FallGPA	2027	0.00	4.00	3.06	0.86
OverallGPA	2037	0.00	4.00	3.10	0.66
PreFallGPA	1486	0.00	4.00	3.09	0.61
OtherCommittment	890	0.00	168.00	22.10	27.30
OverallWorkingHours	2016	0.00	168.00	24.21	18.83
FallWorkingHours	1413	0.00	100.00	30.49	13.86
FallCommutingHours	1342	0.00	15.00	2.45	2.80
NegativeOpinion	493	2.37	4.46	3.79	0.73