Empirical research focused on the effect of training on employment outcomes has received a great deal of attention in the labor economics literature (for example, Mincer, 1962; Card & Sullivan, 1988; Gritz, 1993). However, the professional sports landscape also allows for an opportunity to examine the relationship between training and employment outcomes. In Major League Baseball (MLB) the rights to amateur players are typically acquired by clubs via the June First-Year Player Draft. Here, players are traditionally selected from one of three sources – high schools, four-year institutions, and junior or community colleges. Since players are drafted by clubs at various stages in their development, there is variation in both the type and quality of training possessed by players at the time of the Draft. With the top clubs spending over $10 million annually on draftee signing bonuses (Castrovince, 2010), it is of practical significance to examine whether relationships exist between training and employment outcomes. Accordingly, this research will use 1966-2005 MLB First-Year Player Draft data to investigate.

Sport-specific research on labor market outcomes for players is robust. Among others, Frick, Pietzner & Prinz (2007) have examined employment outcomes as measured by employment spell and career duration in the German Bundesliga. In North America, Hoang and Rascher (1999) and Groothius and Hill (2004) investigated labor market exit discrimination for NBA players. Specific to baseball, Spurr and Barber (1994) examined pitcher demotion, promotion and turnover and Spurr (2000) studied the ability of clubs to find talent. Despite these contributions, there has been little mention of the link between training and employment outcomes. This study addresses this missing element by investigating this relationship in the context of pro baseball.

Theory regarding general on-the-job training (OTJ) specifies that the value of marginal product (VMP) for a worker is dependent on the amount of training acquired by that worker when holding other variables such as innate ability constant (Borjas, 2005). Therefore, in the two worker case where worker A possesses a larger amount of general OTJ than worker B, represented by TA > TB, all else equal, VMPA > VMPB. Here, three different treatments of players have been identified. Group A will represent players drafted directly out of high school, group B will signify players drafted out of four-year institutions, and group C will represent players drafted out of a junior or community college (JUCO). At the time of the draft, players in group A have accumulated four years of general OTJ, players in group B have accumulated either seven or eight years of OTJ, and players in group C have accumulated either five or six years of OTJ. Clearly, variation exists in both the quantity and quality of OTJ accrued at the time of the Draft. But it is not clear that TB > TC > TA = VMPB > VMPC > VMPA. Therefore, it is of interest to investigate this relationship at the time of selection into the labor market.

There are two primary employment outcomes of interest in this study. The first is to determine the rate at which individuals selected in the MLB Draft reach MLB and the second is to determine the Major League career duration of players drafted from each of the three training groups. In order to properly address the two labor market objectives identified, two separate empirical approaches must be utilized. In order to estimate the differences in the probability of entry into the major leagues, logistic regression is appropriate. To model the MLB career duration of drafted players, the Cox (1972) proportional hazard model will be utilized.

\[ P(Y = 1| \text{tr}, \text{dr}, \text{de}, \text{fr}) = FL(\beta_0 + \beta_1 \text{tr} + \beta_2 \text{dr} + \beta_3 \text{de} + \beta_4 \text{fr}) , \]

The functional form of the logistic regression equation is shown above, where Y = 1 if the drafted player played in MLB during his career, FL is the logistic distribution function, tr is a vector of training background variables, dr is a vector of draft specific variables, de is a vector of demographic variables and fr is a vector of franchise specific variables.

\[ \lambda(t|b,d,p,f) = \lambda_0(t) \exp(\beta_1 X_b + \beta_2 X_d + \beta_3 X_p + \beta_4 X_f) , \]

The conditional hazard function in the Cox model is shown above and provides the conditional probability of a player exiting MLB at time t based on the values of covariates included in the model. \( \lambda \) is the hazard rate, t is the time
variable, $\lambda_0(t)$ is the baseline hazard function, $X_b$ is a vector of training background variables. $X_d$ is a vector of draft specific variables, $X_p$ is a vector of demographic variables, and $X_f$ is a vector of franchise specific variables.

The logistic regression model produces a likelihood ratio chi-squared $(31) = 7595.21, p < .0001$, which suggests that the model fits significantly better than the null model. Initial results show that when holding all else equal, players drafted out of four-year institutions are 20.81% more likely to reach MLB as compared to high school players. Additionally, JUCO draftees are 16.61% more likely to reach MLB as compared to high school players. Both results are significant at the 0.01 level. Together, these results illustrate that when holding all else equal, players with larger amounts of accumulated training prior to selection in the MLB Draft have enhanced labor market outcomes as measured by the probability of reaching the major leagues.

The Cox proportional hazard model also fits significantly better than the null model ($X_2 = 184.81, 31$ d.f., $p < .0001$). Alternatively, these results do not support enhanced employment outcomes for players with larger amounts of training. All else equal, four-year college players have a 30% increase in the hazard rate of exiting MLB as compared to high school draftees. There is also no significant difference between JUCO and high school players in MLB exit probability. These results suggest that once draftees enter MLB, those with amplified levels of training do not show enhanced labor market outcomes as measured by career duration.

Overall, the results suggest that the relationship between training and the value of the worker to the firm holds in one of the two labor market outcomes identified. In terms of practical application, it appears that franchises are more easily able to identify players with enhanced levels of training in terms of projecting whether a not a player will reach MLB. Alternatively, these higher levels of accumulated training are not positively associated with the length of a player’s duration in the MLB labor market.