

Leveling up: Upskilling and its benefits to job satisfaction and retainment among Indonesian workers

Hamidah Alatas

In this project I tried to look at the relationship between training participation and several job outcomes which are job satisfaction and retention. This project is using the Indonesian Family Life Survey (IFLS) data, a longitudinal socio-economic survey in Indonesia. This project found that even though training participation does not significantly affect the job satisfaction, there are some evidences that together with age, participating in training might significantly predict workers' decision to stay in the same company

Introduction

The ever-expanding technological advancement is shifting the way we work. It changes not only the skills that industry demands but also the human capital investments of students and workers. The question of how to prepare the labor force with the skills they need has become more pressing now than ever. Moreover, the COVID-19 pandemic will most be expected to exacerbate these needs. Together with the fourth industrial revolution, the pandemic will accelerate a structural change of skills' needs that will harm the vulnerable workers if not properly addressed. Low-skill and low-educated workers will likely have more need to catch up with the upcoming changes by upskilling.

Upskilling, defined as an effort to help someone learning new professional skills to participate in the economy, is the key to helping people integrate into better and improved jobs ([WEF, 2020](#)). Moreover, job training program is proven to have a positive relationship with employee commitment and satisfaction ([Ocen, Francis and Angundaru, 2017](#)). However, the current uptake of any upskilling program in Indonesia is still low, as around 91 percent of unemployed people had not taken any skill training program in 2019 ([CMEA, 2020](#)).

This project aims to identify the relationship between employer-provided training participation and job outcome. There are two job outcomes this project are interested in, job satisfaction and retention. The first hypothesis is training participation, even after adjusting for other characteristics, has a positive relationship with job satisfaction. Participation in a job training will increase employee's skill thus they can do their job in more efficient way and make them more satisfied with their overall working environment. The second hypothesis is training participation, net other variables, has a negative relationship with job retention. This hypothesis is derived from an intuition that skill's improvement will increase employee's competitiveness and make it easier for them to find a new job. This project will explore the association (not necessarily causation) between job outcome and participation in skill training. Therefore, this project could not conclude whether upskilling participation is necessary and or sufficient for increase in job satisfaction and job retention.

Data and Variables

This project uses the Indonesian Family Life Survey (IFLS) data, a longitudinal survey in Indonesia. This survey covers a wealth of variables related to social indicators collected at the individual and household levels, including employment condition and demographic. Up until now there are 5 waves of IFLS; 1993, 1997, 2000, 2007, and 2014. The sample of IFLS is representative of about 83 percent of the Indonesian population with sample of 30,000 individuals living in 13 of the 27 at-the-time provinces in the country. The re-contact rate of IFLS is pretty high with 92 percent of households in IFLS 1 (1993) were re-contacted in IFLS 5 (2014). However, for this project, I will only be using 2007 and 2014 waves since they are the only waves with question about training. Moreover, I am filtering the data to only include people who are working with employer in the last 7 days.

Dependent Variables

The dependent variable of interest in this project is job outcome. Specifically, there are two variables that is hypothesized to have an association with job training which are job satisfaction and retention. Both 2007 and 2014 IFLS data have these variables. Below are the list of the variables:

1. **job_satif:** General job satisfaction is derived from IFLS question “How satisfied are you with your current job?” under tk16c1 variable. For this variable, I recode the original variable so it will be in increasing manner with 1 as “Very Unsatisfied” to 4 is “Very Satisfied”
2. **stay:** I will also look at whether the employee stay in the same company. IFLS asked the history of employment of a person in-between the waves. In 2014 data, the questionnaire asked the name of company that the person worked at in 2007 and 2014 (tk30), I will match these two information to get 1 if the person stay in the same company and 0 if the person change employment.

Independent Variables

There are three questions about employer-provided training in IFLS data: participation, duration, and type of training. These questions are only asked for people who are working with an employer (i.e. government and private worker). For this project, I am interested to look at participation and duration of training as the predictor of job satisfaction and retainment. Moreover, several variables that are likely to predict job satisfaction such as working hour, income, sector of employment, benefit, will also be regressed to the predictors. Below is the list of the main independent variables and control variables that I am interested to look at:

1. **training:** This is the key independent variables of this project, whether the respondent has ever participated in an employer-provided training in his/her current job. Both of 2007 and 2014 IFLS asked this question under tk25a8 variable “Have you ever received any training from your employer?”. This variable will take value of 1 if this respondent ever received a training from this employer and 0 if not. The hypothesis is participating in training, will increase someone’s job satisfaction and decrease the probability of them to stay in the same job.

2. **train_duration:** This is the second variable I am interested to look at, how long the training takes place. This question is under variable tk25a9 in IFLS. I will recode this variable to have value 0 if the respondent never participated in the training, 1 if the training is more than 12 months ago, 2 if the training is less than 12 months ago but for less than a week, and 3 if the training is less than 12 months ago and for one week or more. The hypothesis is the longer and recent a training is the higher someone's job satisfaction is and the less likely they will stay in the same job.
3. **hour:** In IFLS actually there are two questions about hour per week, one is asking about the total number of hours respondent worked during the past week (tk21a) and the other one is asking about in normal condition, what is the approximate total number of hours she/he work per week (tk22a). However, for this analysis I will use the working hour under normal condition. A study in Australia, found that long hours of work are associated with lower job satisfaction and more job search activity (Brown and Rotundo, 2005). Therefore, I expect working hour will have negative relationship with job satisfaction and probability of staying at the same job.
4. **log_income:** IFLS asked about the salary/wage during the last month (including the value of all benefits) in Indonesia Rupiah (IDR) under question tk25a1. I will log this variable because income variable has a very wide range. By transforming the variable to ln variable, we will have narrower range, and having the narrower range would make the estimates of the linear model less sensitive to outliers. I assume that income will have positive relationship with job satisfaction and probability staying in the same job.
5. **contract:** Contract is an important feature to have in a job. Without contract, employee will have less bargaining power thus it might be associated with lower satisfaction in job. IFLS asked the contract type that an employee has in question tk24a5. I will use this variable and recode it so it has value 0 for no contract, 1 for contract with no fixed time, and 2 for contract with fixed time.
6. **medical_benefit:** This variable has a value of 1 if a respondent receive any medical benefit from the employers (tk25a3e1, tk25a3e2, tk25a3e3) and 0 if not at all. A study found that benefits were significantly, and positively, associated to overall job satisfaction (Ellickson and Logsdon, 2002). Therefore, I expect that this variable also has positive relationship with job satisfaction and retention.
7. **physical:** In this variable, I combine three questions asking about whether someones' job require many physical activities. The three variables are whether the job requires lots of physical effort" (tk26a5), lifting heavy loads (tk26a6), and stooping, kneeling, crouching (tk26a7). I recode these variables to have increasing value; 1 "None/Almost none of the time", 2 "Some of the time", 3 "Most of the time", 4 "All/Almost all the time". I decided to combine these variables to **physical** without standardization (due to similar scale), which represents how much physical activities are needed in this job. The raw Cronbach's alpha coefficient is 0.75, which indicates that the new variable has a good reliability to measures the three variables at once. A study found that physical demands at work negatively but weakly correlated with job satisfaction (Andersen, 2017). I would like to test whether this relationship also applies in Indonesia case.

Table 1: Cronbach Alpha for physical variable

raw_alpha	std.alpha	G6(smc)	average_r	S/N
0.75	0.75	0.67	0.5	3.02

8. **stress**: I also add variable in which employees were asked whether their work burden them with stress. This question is in both IFLS 2007 and 2014 under question number tk26a12. I recode this variables to have increasing value from 1“None/Almost none of the time” to 4“all/Almost all the time”. The hypothesis is the more stressing the work environment is, the more likely employee will feel less satisfied with their job.
9. **age and female**: For demographic variable, I add age and gender as control variables. A study found that gender and age are associated with both job satisfaction and pride in work ([Magee, 2015](#)). In age, I dropped respondents with age less than 15 because employment module was supposed to be asked for respondents in working age (15+).

Descriptive Statistics

In 2007, there are 44,190 respondents who were working for employer that will be the sample for our analysis. The average age of the respondent is 33 years and on average they work for 45 hours a week. This working hour average is higher than the global average weekly working time which is 43 hours ([ILO, 2015](#)). Respondents who are working with employers in Indonesia, on average, have monthly income of IDR 1,042,381 or equal to USD 114 in 2007. Since income variable shows high variability in this data, I added the `log_income` variable.

Table 2: Summary statistics for control variables, 2007

variable	Mean	Median	Std.Dev	Min	Max
age	33.19	31.00	10.86	15.00	85.00
hour	44.74	45.00	16.99	0.00	140.00
income	1,042,381.00	800,000.00	1,147,609.00	7,500.00	20,000,000.00
log_income	13.47	13.59	0.93	8.92	16.81
physical	2.24	2.00	0.84	1.00	4.00
stress	1.60	1.00	0.71	1.00	4.00

In 2014, the number of respondents who were working with employer is 90,731 with average age increased a bit to 34 years old compared to 2007. Although this survey follow the same people for both wave, it is possible that the increase of average age of 1 years instead of 7 years is because there are new people that in 2007 were not working due to young age but in 2014 started to work. The working hour, however, decreased to 43 hours per week which could indicate improvement in working condition in general. Average of monthly income is also increased to IDR 2,016,246, but this is probably due to inflation since both

income is nominal and we do not have information on real income. The variability of income is also still very high with standard deviation of IDR 2,358,833. As above, I also transform the income variable to its logarithmic value.

Table 3: Summary statistics for control variables, 2014

variable	Mean	Median	Std.Dev	Min	Max
age	34.47	33.00	11.07	15.00	87.00
hour	43.15	42.00	17.61	1.00	140.00
income	2,016,246.00	1,500,000.00	2,358,833.00	2,500.00	90,000,000.00
log_income	14.11	14.22	0.97	7.82	18.32
physical	2.27	2.00	0.84	1.00	4.00
stress	1.69	2.00	0.82	1.00	4.00

For the main independent variables, I observe the increase of percentage of respondents who participated in employer-provided training program in 2014 compared to 2007 (28 percent vs 34 percent). Not only the proportion of workers who participated in training increase, when looking at duration and recency of training, we see an improvement of respondents who had more recent training and longer duration of training in 2014 compared to 2007. In 2014, 15 percent of workers spent more than 1 week in a training that happened less than a year ago while in 2007 only 11 percent belonged in this group of respondents.

Table 4: Training Participation, 2007 and 2014

Year	Count	Percent
2007	2,052	27.86
2014	3,150	34.07

In this analysis, I looked at two variables that indicate job outcome: satisfaction and retention. Overall, in 2014 employee had a slightly higher job satisfactory level as 83.7 percent of workers were either satisfied or very satisfied with their job compared to 80.7 percent of workers in 2007 who shared the same satisfactory level. Below we could also see that majority of Indonesian employee were satisfied with their job, with only a very little proportion of them were very unsatisfied with their job.

Moreover, among respondents who were working in 2014, 18 percent of them worked for the same employer as 7 years ago. From the table below we could also see that around 14 percent of working respondents in 2014 did not have a job in 2007.

When looking at the relationship between training participation and job satisfaction, we can see that employee who received training are more likely to feel very satisfied on their job compared to workers without training by 6 percentage points (Figure 1). The percentage of employee who are very unsatisfied

Table 5: Training duration and recency, 2007 and 2014

Training Duration & Recency	Count	Percent
2007		
No Training	5,313	72.14
More than a year ago	751	10.20
Less than a year ago for less than 1 week	485	6.59
Less than a year ago for more than 1 week	816	11.08
2014		
No Training	6,097	65.93
More than a year ago	909	9.83
Less than a year ago for less than 1 week	821	8.88
Less than a year ago for more than 1 week	1,420	15.36

Table 6: Job Satisfaction, 2007 and 2014

Job Satisfaction	Count	Percent
2007		
Very Unsatisfied	29	0.39
Unsatisfied	1,395	18.94
Satisfied	5,583	75.80
Very Satisfied	358	4.86
2014		
Very Unsatisfied	115	1.24
Unsatisfied	1,391	15.04
Satisfied	6,589	71.26
Very Satisfied	1,152	12.46

Table 7: Job Movement during 2007-2014 period

Movement	Count	Percent
Move	6,333	68.49
Prev Not Working	1,270	13.73
Stay	1,644	17.78

with the job however does not differ by training participation. However, other variables might mediate this relationship therefore, in the next chapter, we will be looking at the multivariate linear regression.

Same pattern also found when looking at the relationship between job movement and training participation. Among workers who held job in both waves, those who received training in 2007 were more likely to stay in their previous job in 2007. This actually contradicts the hypothesis that workers with training experience in previous job are more likely to move. There are two reasons why this might happen. The first

one is because this positive relationship is mediated by other variables, or the employers successfully held the equilibrium where quits are low and training high (Acemoglu and Pischke, 1998). The former reason will be checked using multiple linear regression in the next chapter.

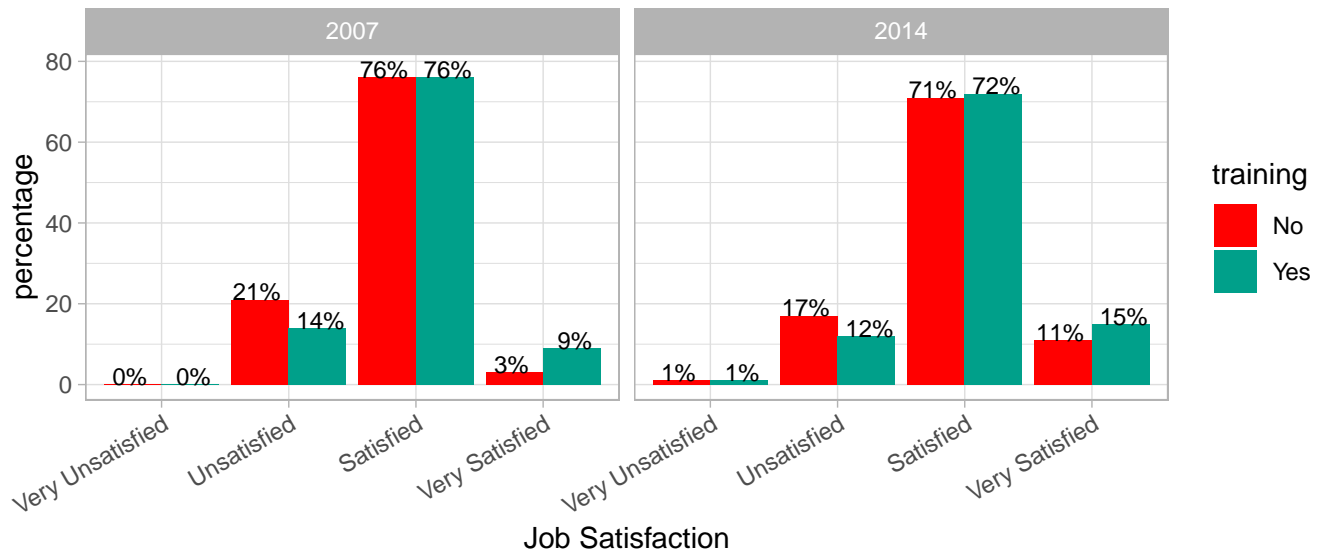


Figure 1: Job satisfaction by training participation

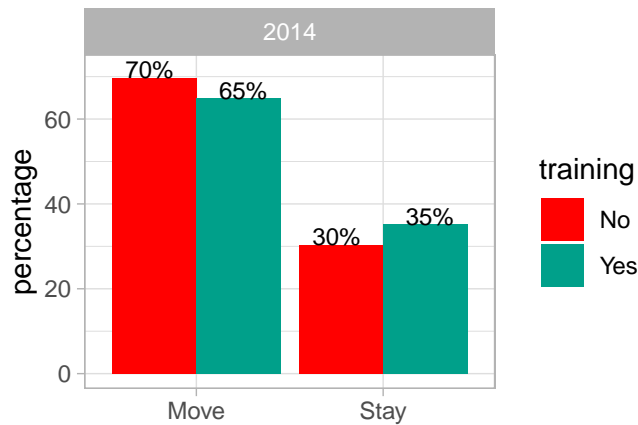


Figure 2: Job Movement by training participation

Initial Model

Training Participation and Job Satisfaction

In the first analysis, I ran naive (pooled) linear regression with and without demographic variables as control variables.

Model 1 (without demographic variables):

$$\begin{aligned} JobSatisf = & \alpha + \beta_1 Training + \beta_2 Hour + \beta_3 LogIncome \\ & + \beta_{41} Contract.1 + \beta_{42} Contract.2 + \beta_5 MedicalBenefit \\ & + \beta_6 Physical + \beta_7 Stress + \epsilon \end{aligned}$$

Model 2 (with demographic variables):

$$\begin{aligned} JobSatisf = & \alpha + \beta_1 Training + \beta_2 Hour + \beta_3 LogIncome \\ & + \beta_{41} Contract.1 + \beta_{42} Contract.2 + \beta_5 MedicalBenefit \\ & + \beta_6 Physical + \beta_7 Stress + \beta_8 Age + \beta_9 Gender + \epsilon \end{aligned}$$

In regression table below, clustered standard error by individual identification variable is used. Clustered standard error is a more suitable estimation for these models considering in panel data, error in the original model are more likely to be not independent. In this analysis, we can see that training has positive significant relationship to job satisfaction even after adjusting for working hour, log of income, contract ownership, medical benefit, physical activity level, and stress level. However, the effect is very small since on average, employee that received training has 0.07 higher job satisfaction score compared to employee without training net other variables. This relationship is significant with p-value <0.01 indicating that the probability to get t-value as large as 30 under the condition when there is no relationship between these two variables is less than 1 percent.

The positive and significant relationship holds true even after adding demographic variables, gender and age. Gender and age albeit positively associated with job satisfaction, did not change the coefficient of training by much. Net of other variables (including demographic variables), workers with training, on average, would still have higher job satisfaction score by 0.07 compared to non-trained workers. However, this model has a very low R-squared of 6 percent which indicates this set of independent variables can only explain 6 percent of variance in job satisfaction.

When training variable were broken down based on duration and recency, I found that net other variables, no matter the recency and duration group, workers with training on average has significantly higher job satisfaction compared to workers without training. For instance, workers who had a training more than a year ago, on average has higher job satisfaction by 0.07 point compared to workers without training, net other variables. However, the more interesting finding here is the association between job satisfaction and whether workers had no training or had a training for more than a week is actually not as strong as the relationship between job satisfaction and whether workers with no training or had a training for less than one week. This indicates that duration of training does not affect the job satisfaction.

Model 3 (without demographic variables):

$$\begin{aligned} JobSatisf = & \alpha + \beta_{11}TrainingDuration.1 + \beta_{12}TrainingDuration.2 \\ & + \beta_{13}TrainingDuration.3 + \beta_2Hour + \beta_3LogIncome \\ & + \beta_{41}Contract.1 + \beta_{42}Contract.2 + \beta_5MedicalBenefit \\ & + \beta_6Physical + \beta_7Stress + \varepsilon \end{aligned}$$

Model 4 (with demographic variables):

$$\begin{aligned} JobSatisf = & \alpha + \beta_{11}TrainingDuration.1 + \beta_{12}TrainingDuration.2 \\ & + \beta_{13}TrainingDuration.3 + \beta_2Hour + \beta_3LogIncome \\ & + \beta_{41}Contract.1 + \beta_{42}Contract.2 + \beta_5MedicalBenefit \\ & + \beta_6Physical + \beta_7Stress + \beta_8Age + \beta_9Gender + \varepsilon \end{aligned}$$

Table 8: Linear Regression Models Predicting Job Satisfaction

	Job Satisfaction (1-4 scale)	
	Model 1	Model 2
(Intercept)	0.068*** (0.010)	0.066*** (0.010)
Training = Yes	-0.001** (0.0003)	-0.0003 (0.0003)
Hour	0.076*** (0.005)	0.083*** (0.005)
Log Income	-0.073*** (0.013)	-0.062*** (0.013)
Contract = Fixed-Term	0.035*** (0.013)	0.034** (0.013)
Contract = Permanent	0.055*** (0.010)	0.051*** (0.010)
Medical Benefit = Yes	-0.018*** (0.005)	-0.005 (0.005)
Physical Level	-0.092*** (0.006)	-0.089*** (0.006)
Stress Level		0.002*** (0.0004)
Age		0.087*** (0.009)
Gender: Female	2.043*** (0.068)	1.790*** (0.073)
R ²	0.05	0.06

Note:

*p<0.1; **p<0.05; ***p<0.01

Standard error is clustered by individual ID

Table 9: Linear Regression Models Predicting Job Satisfaction

	Job Satisfaction (1-4 scale)	
	Model 3	Model 4
(Intercept)	0.075*** (0.014)	0.072*** (0.014)
Training: More than a year ago	0.076*** (0.016)	0.073*** (0.016)
Training: Less than a year ago for < 1 week	0.057*** (0.013)	0.058*** (0.013)
Training: Less than a year ago for > 1 week	−0.001** (0.0003)	−0.0003 (0.0003)
Hour	0.076*** (0.005)	0.083*** (0.005)
Log Income	−0.073*** (0.013)	−0.062*** (0.013)
Contract = Fixed-Term	0.035*** (0.013)	0.034** (0.013)
Contract = Permanent	0.055*** (0.010)	0.051*** (0.010)
Medical Benefit = Yes	−0.018*** (0.005)	−0.005 (0.005)
Physical Level	−0.092*** (0.006)	−0.089*** (0.006)
Stress Level		0.002*** (0.0004)
Age		0.087*** (0.009)
Gender: Female	2.040*** (0.068)	1.788*** (0.073)
R ²	0.05	0.06

Note:

*p<0.1; **p<0.05; ***p<0.01

Standard error is clustered by individual ID

Training Participation and Job Retention

For job retention model, I use the working conditions in 2007 as independent variables, and whether employee stay in 2014 or not as the dependent variables. Logistic regression is used since the dependent variable in this model (whether the employee stay or not) is binary variable. In the table below, we can see if we only include training participation variable (regardless of the duration and recency), there is no significant relationship between participation in the upskilling program to job retention. Net of other variables, on average, the odds of staying in the same job seven years later for workers who participated in training in 2007 over the odds of staying in the same job for workers who did not participate in training is 0.9. This indicates the negative but not significant relationship between training participation and job retention which align with the first hypothesis.

However, when we dissect the training participation based on recency and duration, we can see in the model 6 that workers who participated in a more than a week training in 2007 were significantly less likely to work in the same employer in 2014. Net of other variables, workers who did not participate in the training, on average, had a 1.41 (=1/0.707) higher odds of staying in the same company compared to workers who participated in a training for more than a week. This finding indicates that after controlling other variables, the more intense the training is, the more likely a worker to move to another job which probably due to improved skill.

Model 5 (Training):

$$\begin{aligned} \log\left(\frac{p_{stay}}{1-p_{stay}}\right) = & \alpha + \beta_1 Training + \beta_2 Hour + \beta_3 LogIncome \\ & + \beta_{41} Contract.1 + \beta_{42} Contract.2 + \beta_5 MedicalBenefit \\ & + \beta_6 Physical + \beta_7 Stress + \beta_8 Age + \beta_9 Gender + \epsilon \end{aligned}$$

Model 6 (Training Duration):

$$\begin{aligned} \log\left(\frac{p_{stay}}{1-p_{stay}}\right) = & \alpha + \beta_{11} TrainingDuration.1 + \beta_{12} TrainingDuration.2 \\ & + \beta_{13} TrainingDuration.3 + \beta_2 Hour + \beta_3 LogIncome \\ & + \beta_{41} Contract.1 + \beta_{42} Contract.2 + \beta_5 MedicalBenefit \\ & + \beta_6 Physical + \beta_7 Stress + \beta_8 Age + \beta_9 Gender + \epsilon \end{aligned}$$

Table 10: Logistic Regression Models Predicting Employee Retention

	Odds Ratio for Stay in the Job	
	Model 5	Model 6
Training = Yes	0.867 (0.077)	
Training: More than a year ago		1.097 (0.134)
Training: Less than a year ago for < 1 week		0.837 (0.124)
Training: Less than a year ago for > 1 week		0.707*** (0.087)
Hour	0.995* (0.003)	0.994** (0.003)
Log Income	1.346*** (0.081)	1.355*** (0.081)
Contract = Fixed-Term	0.627*** (0.087)	0.630*** (0.088)
Contract = Permanent	0.987 (0.121)	0.988 (0.121)
Medical Benefit = Yes	1.853*** (0.168)	1.844*** (0.167)
Physical Level	1.134** (0.058)	1.137** (0.058)
Stress Level	1.081 (0.061)	1.083 (0.061)
Age	1.010** (0.005)	1.009** (0.005)
Gender: Female	1.437*** (0.125)	1.443*** (0.125)
Constant	0.004*** (0.003)	0.003*** (0.003)
Corrected AIC	3919.1	3914.4

Note:

*p<0.1; **p<0.05; ***p<0.01

Remarks on Initial Models

In model 1-4, we used linear regression model to look at the relationship between training participation and job satisfaction. In these models, we assume equal distance between job satisfaction categories (i.e. the distance from very unsatisfied to unsatisfied is the same as the distance from satisfied to very satisfied). Therefore, we assume the relationship between job satisfaction and training participation is linear when it could be something else. Moreover, even though we found statistically significant results after controlling other variables and clustering standard error in these models, workers who participated in training may be fundamentally different from workers who did not participate in training in things that we could not observe (individual heterogeneity).

In models 5 and 6, by using logistic regression we found that while general participation in training does not have a significant relationship with log-odds of staying in the same job, workers with more intense training are more likely to move to another job compared to workers without training. However, looking at the AIC, which is a measure that can be used for model selection, model 6 that includes the training intensity only improve model 5 by a little. Therefore, in the next iteration I will be looking at interaction to improve the models.

Model Improvement

First Difference Model

As mentioned above, by using linear regression, we fail to look after the individual heterogeneity in our models. However, by taking the difference between two time period in panel data we can look at the change in job satisfaction when training participation is changed. In this model, we can eliminate error due to unobserved, time-invariant factor that affect job satisfaction such as upbringing, exposure to job offer, their personality, etc.

The result of the first difference model in table 11 shows that the change in training participation did not have statistically significant relationship to change in job satisfaction anymore. On average, a change in worker's training participation from not participated to participated in a training translates to a 0.02 points increase in their job satisfaction net of other variables. However this relationship is not significant as the p-value is very high (0.42). The change of significance of a relationship between training and job satisfaction from significant relationship in linear regression to not significant in first difference shows that there is a spurious relationship in the first model. The job satisfaction is more likely to be driven by the type of workers participating in training not the training participation only.

First Difference Model:

$$\begin{aligned}\Delta JobSatisf = & \alpha + \beta_1 \Delta Training + \beta_2 \Delta Hour + \beta_3 \Delta LogIncome \\ & + \beta_{41} \Delta Contract.1 + \beta_{42} \Delta Contract.2 + \beta_5 \Delta MedicalBenefit \\ & + \beta_6 \Delta Physical + \beta_7 \Delta Stress + \beta_8 \Delta Age + \beta_9 \Delta Gender + \Delta \epsilon\end{aligned}$$

Table 11: First-Difference Models Predicting Job Satisfaction

	Job Satisfaction (1-4 scale)	
	Improved Model 1	Improved Model 2
Training = Yes	0.019 (0.024)	
Training: More than a year ago		−0.009 (0.029)
Training: Less than a year ago for < 1 week		0.051 (0.037)
Training: Less than a year ago for > 1 week		0.025 (0.034)
Hour	−0.001 (0.001)	−0.001 (0.001)
Log Income	0.103*** (0.017)	0.104*** (0.017)
Contract = Fixed-Term	0.015 (0.039)	0.013 (0.039)
Contract = Permanent	0.006 (0.027)	0.006 (0.027)
Medical Benefit = Yes	0.071*** (0.026)	0.071*** (0.026)
Physical Level	0.011 (0.015)	0.011 (0.015)
Stress Level	−0.052*** (0.014)	−0.052*** (0.014)
Age	−0.030** (0.012)	−0.030** (0.012)
Constant	0.234*** (0.084)	0.234*** (0.084)
R ²	0.02	0.02

Note:

*p<0.1; **p<0.05; ***p<0.01

Standard error is clustered by individual ID

This model however has a very strange pattern on age variable. In theory, age should be going up constantly for everyone as everyone was interviewed in the same time seven years later. However, as seen in table 12, while most of respondents' difference in age is seven years, we see that some respondents have -3 to 20 years difference in age between two waves. This might affect the previous first difference models. Therefore, in the next models I will only include observations with constant change in age.

Table 12: Summary statistics of age difference between two waves

Stats.	Value
Min.	-3.00
1st Qu.	7.00
Median	7.00
Mean	6.87
3rd Qu.	7.00
Max.	20.00
NA's	13,159.00

After filtering the data, the new improved first difference models show similar result as before. A change in training variables does not have a significant effect to change in job satisfaction. This finding confirms that the significant result we observed in linear regression might only be happened because a spurious relationship exist between these two variables.

Moreover, this new model in table 13 shows that change in log of income, stress level, and medical benefit significantly translates to a change in job satisfaction. On average, a respondent who had an increase of income by one percent also experienced increase in job satisfaction level by 0.085 points net of other variables. Similar pattern happened when someone's medical benefit status changed from non-receiver to receiver, on average, their job satisfaction increased by 0.108 points after controlling from other variables. While the previous two variables have positive relationship, a change in stress level negatively correlates with job satisfaction even after controlling for other variables. An increase in one point of work stress level, on average, translates to 0.05 decrease in job satisfaction. All of these variables has low p-value of <0.01 indicating the significance of the relationships.

Table 13: First-Difference Models Predicting Job Satisfaction

	Job Satisfaction (1-4 scale)	
	Improved Model 3	Improved Model 4
Training = Yes	0.013 (0.024)	
Training: More than a year ago		0.005 (0.032)
Training: Less than a year ago for < 1 week		0.027 (0.037)
Training: Less than a year ago for > 1 week		0.012 (0.031)
Hour	-0.001 (0.001)	-0.001 (0.001)
Log Income	0.085*** (0.020)	0.085*** (0.020)
Contract = Fixed-Term	0.007 (0.038)	0.006 (0.038)
Contract = Permanent	0.026 (0.029)	0.027 (0.029)
Medical Benefit = Yes	0.108*** (0.027)	0.108*** (0.027)
Physical Level	0.008 (0.015)	0.008 (0.015)
Stress Level	-0.048*** (0.014)	-0.048*** (0.014)
Age	0.037* (0.021)	0.036* (0.021)
R ²	0.023	0.023

Note:

*p<0.1; **p<0.05; ***p<0.01

Data is filtered to include only obs with constant change of age

Improved Logistic Regression

Many studies have found that age correlates negatively with training participation. The decrease in motivation for training of workers from older age group compared to their younger counterparts has been at the center of narratives around why the age of workers is negatively associated with workers' training participation (Kanfer and Ackerman, 2004). This relationship has been explained as both moderators or is mediated by other variables by many studies. These studies lead me to believe that some level of change in age might increase the effect of training on the odds that someone will stay in the same company. Therefore, I will consider the interaction effect of training and age for the model improvement. For younger people, training might matter more to their job retention decision because they more likely want to explore job opportunity thus their improved skills is translated to higher probability to move to another job. For older people, they are less likely to making decision of staying or moving from a job on training opportunity.

Model with Interaction:

$$\begin{aligned} \log\left(\frac{P_{stay}}{1 - P_{stay}}\right) = & \alpha + \beta_1 Training + \beta_2 Hour + \beta_3 LogIncome \\ & + \beta_{41} Contract.1 + \beta_{42} Contract.2 + \beta_5 MedicalBenefit \\ & + \beta_6 Physical + \beta_7 Stress + \beta_8 Age + \beta_9 Gender + \beta_{10} Age * Training + \epsilon \end{aligned}$$

The result in table 14 shows that after adding the interaction, the training variables and their respective age's interaction have significant effect in increasing the odds of someone staying at the same job after seven years. The result show a very interesting result. People with training, on average, has higher odd to stay in the same job by 2.88 net of other variables. The interaction, is also statistically significant, indicates that holding other variables constant, for respondent who participated in a training, each additional years of age, on average, decreases someone's odds to stay in the same job by 4 percent. This increase is significant as the probability to get z-value as large as 3.89 by chance is less than 0.1% under the condition the interaction is not significant. It means that odds to stay decrease as the age increasing for people with training compared to workers without training which align with our previous hypothesis. Similar patterns were observed when dissecting training variables by recency and duration.

Table 14: Logistic Regression Models Predicting Employee Retention

	Odds Ratio for Stay in the Job	
	Improved Model 5	Improved Model 6
Training = Yes	2.876*** (0.917)	
Training: More than a year ago		3.639*** (1.789)
Training: Less than a year ago for < 1 week		2.803* (1.543)
Training: Less than a year ago for > 1 week		2.829** (1.288)
Hour	0.994** (0.003)	0.994** (0.003)
Log Income	1.367*** (0.082)	1.378*** (0.083)
Age	1.021*** (0.005)	1.021*** (0.005)
Gender	1.430*** (0.124)	1.437*** (0.125)
Training * Age	0.965*** (0.009)	
Training: More than a year ago * Age		0.966** (0.013)
Training: Less than a year ago for < 1 week * Age		0.964** (0.015)
Training: Less than a year ago for > 1 week * Age		0.959*** (0.013)
Constant	0.002*** (0.002)	0.002*** (0.002)
Corrected AIC	3905.9	3903.3

Note:

*p<0.1; **p<0.05; ***p<0.01

Controlled variables include contract ownership, medical benefit, physical, and stress level

Figure 3 below confirms the relationship between age and training to log-odds of staying in the same job. For respondents with no training participation, the older they are, on average, increased their log-odds to stay in the same job, otherwise, age negatively correlates with log-odds to stay in the same job for respondents with training.

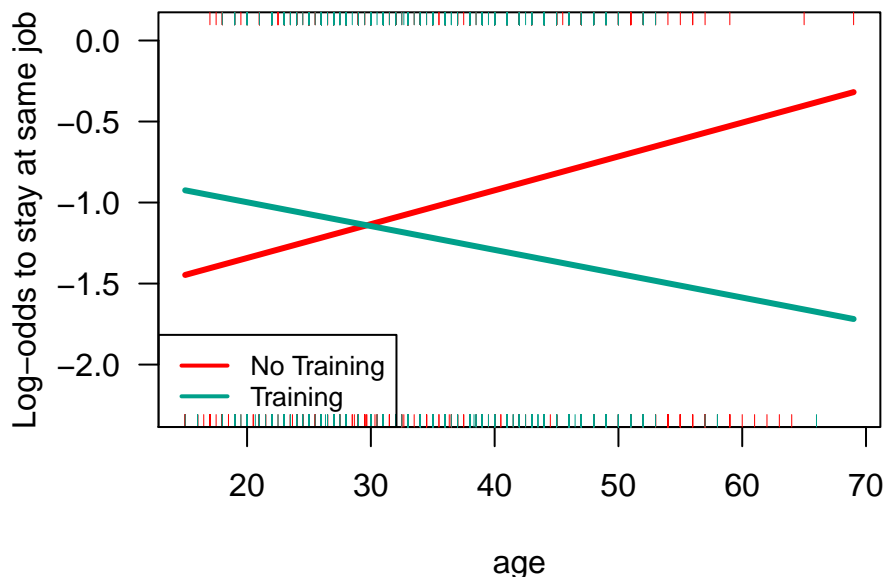


Figure 3: Log-Odds of Staying by Age and Training Participation

Conclusion

This project found that even though training participation does not significantly affect the job satisfaction, there are some evidences that together with age, participating in training might significantly predict workers' decision to stay in the same company. At first, we see that job satisfaction statistics seems to differ by whether the workers undertake training or not, moreover the relationship holds to be significant even after controlling for other variables under linear regression models. However, after controlling for individual heterogeneity, by using first differential models, the significant relationship disappear. This indicates that the linear regression significant coefficient is driven by the differences in workers' non-observable and time-invariant characteristics between those who participated in training and those who did not. Therefore, we fail to conclude there is a relationship between these two variables.

By using logistic regression, we also looked at whether training participation could predict job retention. The first model showed that training in general does not increase or decrease odds to stay in the same job. However, workers who received more intense training (more than one week) has a significantly higher

odds ratio of dropping out the company compared to workers without training. Moreover, in the improved model, we consider the interaction between age and training. This is to recognize that age might affect the relationship between training and job retention. This improved models show that odds to stay in the same job decrease as the age increase for trained workers compared to non-trained workers. Therefore we could conclude that controlling for other variables, training and age have a significant relationship with odds to stay in the same job.

However, all of the conclusions above were made with some things to keep in mind. Firstly, there is an issue of small sample size. The IFLS data only representative to 80 percent of Indonesian population. Moreover, by filtering the data to only include employee, those who answered the question about training, and those with both jobs information in both waves we exclude a lot of respondents and lose the representativeness even more. Therefore, the conclusion might only holds for IFLS respondents only. Secondly, the duration between wave is very long, seven years. Many might happened during those seven years that affect the change of training participation and change in job satisfaction. For example, in 2012 the Ministry of Manpower issue a Minister Decree on skills training funding ([MOM, 2012](#)). This new decree might push some employers to provide training to their employee. Therefore, the first difference model might not catch this time-varying error. Thirdly, the stay variable in the logistic regression capture only workers who works consistently in the same employers for seven years. IFLS data cannot differentiate if the employee come back to previous employers after transfer or more than one year of not-working period. These employees are categorized as “Move” in the retention variable. This failure to identify these workers might undermine our models of training participation and job retention. Finally, larger panel study which sampling design focus more to get the representative employees in Indonesia should be done to get better inference on the effect of training to some job outcomes.

References

- Acemoglu, Daron and Jorn-Steffen Pischke. 1998. "Why Do Firms Train? Theory and Evidence." *The Quarterly Journal of Economics* 113(1):79–119.
- Andersen, Lars. 2017. "Job Satisfaction is More Than A Fruit Basket, Health Checks and Free Exercise: Cross-sectional Study Among 10,000 Wage Earners." *Scandinavian Journal of Public Health* 45:476–484.
- Brown, Michelle and Maria Rotundo. 2005. "Working Long Hours, Job Satisfaction and Job Search Activity: A Longitudinal Investigation." *Convergence and divergence of management theory and practice. Proceedings of the 19th ANZAM Conference* pp. 1–13.
- CMEA. 2020. *Evaluation Report for the Implementation of 2020 Pre-Employment Card Program*. Jakarta: Coordinating Ministry for Economics Affair.
- Ellickson, Mark and Kay Logsdon. 2002. "Determinants of Job Satisfaction of Municipal Government Employees." *Public Personnel Management* 31(3):343–358.
- Hlavac, Marek. 2018. *stargazer: Well-Formatted Regression and Summary Statistics Tables*. Bratislava, Slovakia: Central European Labour Studies Institute (CELSI). R package version 5.2.2.
URL: <https://CRAN.R-project.org/package=stargazer>
- ILO. 2015. "Ensuring Decent Working Time for the Future."
- Kanfer, Ruth and Phillip Ackerman. 2004. "Aging, Adult Development, and Work Motivation." *The Academy of Management Review (Academy of Management)* 29(3):440–458.
- Magee, William. 2015. "Effects of Gender and Age on Pride in Work, and Job Satisfaction." *Journal of Happiness Studies: An Interdisciplinary Forum on Subjective Well-Being* 16(5):1091–1115.
- MOM. 2012. *Minister of Manpower's Decree Number 6/2012*. Jakarta: Ministry of Manpower.
- Ocen, Edward, Kasekende Francis and Gladies Angundaru. 2017. "The Role of Training in Building Employee Commitment: The Mediating Effect of Job Satisfaction." *European Journal of Training and Development* 41(9):742–757.
- WEF. 2020. "The Future of Jobs."