Session 3
What's in an application or a resume – characteristics that define a good candidate

October 2, 2018

Carlos Macuada and Grant Benson

Alternative Title

Using unstructured data sources to find the best interviewers and focusing on how to keep them

Agenda

- Session Goals
- Text analysis to identify strong applicants
  - Narrowing the field to match past applicant pool
- Characteristics of a strong applicant
  - Interviewer Satisfaction Survey 2018
- Suggestions? Feedback?

Recruitment Brownbag Series

- Session 1: Attrition measures, hypotheses, interventions, and new data
  - Sharon Parker, Grant Benson
- Session 2: The recruitment process – requirements, constraints, and goals
  - September 18th, 2018, Ken Szmigiel, Vivienne Outlaw
- Session 3: What's in an application or a resume – characteristics that define a good candidate
  - October 2nd, 2018, Carlos Macuada, Grant Benson
- Session 4: Screening questions, interviewer pay, job stressors, and interviewer attrition
  - November 13, 2018, Sharon Simonton, Grant Benson
- Possible Session 5: Follow-up on feedback, suggestions

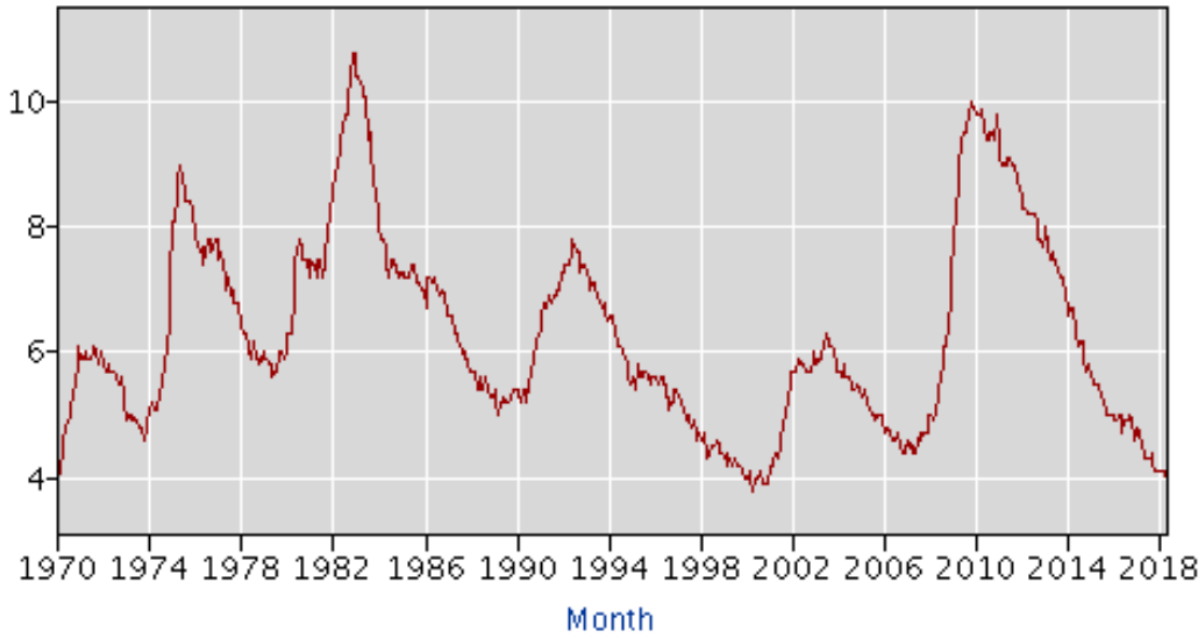
Session Goals

The Good Applicant	• Definitions and expectations
Sources and Methods	• Leveraging unstructured data for predictive skills • Proactively reaching out to potential applicants
Seeking input and suggestions	• What areas are we not looking at? • Are there other measures we should be using? • How can we improve?

Challenge of a Generation

Labor Force Statistics from the Current Population Survey

Series Id: LNS14000000
Seasonally Adjusted
Series title: (Seas) Unemployment Rate
Labor force status: Unemployment rate
Type of data: Percent or rate
Age: 16 years and over



Source:
https://data.bls.gov/pdq/SurveyOutputServlet

What Are We Looking For in an Interviewer?

- Intrinsic motivation – people who are problem solvers, self-guided, are compelled by the mission
- Unafraid of cold calling and speaking to strangers
- Flexible schedule
- Strong adherence to General Interviewing Techniques

Qualities of a Good Interviewer

- "people bright with ability and energy who are willing to work for rather drab wages. […] list of qualities desirable in an interviewer [included] … intelligence, perseverance, judgment, honesty personableness, courteousness, physical stamina, and so on."

-- Jean M. Converse. 1987. *Survey Research in the United States: Roots and Emergence 1890-1960*. pp. 96-97, referencing qualities market researchers were looking for in interviewing staff during the post-World War II era

Finding a Good Interviewer: Applicants Who Look Like Our Current Staff

"Passive" and Active Searching

- Session 2 outlined current hiring requirements, steps, and constraints
- "Passive" search process includes
  - Tailoring postings to projects, adapting based on historical results
  - Tailoring and adapting posting sources, posting length, even hourly rates
  - Review applicants in real time and adjust as necessary
- But what do you do when this doesn't yield a sufficient applicant pool?

Indeed Resume Search

- Indeed (and other sources) have millions of individuals who are actively searching for employment
- Resumes are posted and updated regularly
  - Can we proactively reach out to potentially interested and qualified applicants?
  - Can we identify potentially qualified applicants based on the words they use on their resumes?
  - Can we get a quick turn-around in some areas that we may be short in (in terms of qualified applicants)?

Sample Resume Availability

PSU area	50 mile Search NO Keywords	Filtered 1 month last updated resume	Filtered 1 week last updated resume
Greeley, CO	457,046	31,809	5,591
St Louis, MO	493,945	33,941	5,987
Kansas City, MO	246,186	28,180	4,994
Brooklyn/Queens, NY	555,284	227,516	39,857
Nassau Suffolk Counties, NY	3,324,849	223,014	39,160
Akron, OH	684,039	48,764	8,212
Columbus, OH	421,643	29,191	4,918
Portland, OR	374,952	25,558	4,551
Roanoke, VA	148,906	10,480	2,015
Seattle, WA	600,165	40,390	7,516
Oshkosh-Neenah, WI	156,962	11,103	1,832

Building an Interviewer Resume Profile

- Data Source
  - Applicants to one specific SRO project 2015-2017
  - Up to 4 job titles listed by applicants in their resume / online application
  - Up to 4 narratives explaining duties in online application

Hire Status (DCSR Codes)	Frequency (%)
Hired (101)	74 (3.28%)
Qualified Not Hired (608, 899, 900)	87 (3.86%)
FTF Iw, Not Qualified (607, 609, 610)	54 (2.39%)
No Iw, Qualified for FTF (402, 403, 404, 555)	567 (25.13%)
Didn't Pass Screener/Assessment (401, 601, 602, 603, 606)	1474 (65.34%)
Total	2256 (100%)

Building a Profile (2)

- Created “Stem” words
  - INTERVIEW for interviewer/interviewing/ interviewed, etc.
  - MANAG for management/manager/managed, etc.
  - Automated process using dictionary features

Keywords Extracted

	FREQUENCY	% SHOWN	% PROCESSED	% TOTAL	NO. CASES	% CASES	TF • IDF
CUSTOM	2043	2.44%	1.50%	1.05%	895	39.67%	820.3
ASSIST	1972	2.36%	1.45%	1.01%	1044	46.28%	659.9
MANAG	1900	2.27%	1.40%	0.98%	1007	44.64%	665.6
SERVIC	1651	1.98%	1.21%	0.85%	913	40.47%	648.6
INTERVIEW	1318	1.58%	0.97%	0.68%	554	24.56%	803.8
PROVID	1172	1.40%	0.86%	0.60%	699	30.98%	596.4
STUDENT	1086	1.30%	0.80%	0.56%	504	22.34%	706.9
CLIENT	914	1.09%	0.67%	0.47%	538	23.85%	569.0
SALE	910	1.09%	0.67%	0.47%	460	20.39%	628.4
DATA	898	1.07%	0.66%	0.46%	536	23.76%	560.5
COMMUN	785	0.94%	0.58%	0.40%	533	23.63%	491.9
PROGRAM	770	0.92%	0.57%	0.40%	469	20.79%	525.3
MAINTAIN	763	0.91%	0.56%	0.39%	541	23.98%	473.2
RESEARCH	757	0.91%	0.56%	0.39%	403	17.86%	566.3
FIELD	745	0.89%	0.55%	0.38%	400	17.73%	559.7
RESPONS	743	0.89%	0.55%	0.38%	431	19.10%	534.1
DEVELOP	725	0.87%	0.53%	0.37%	476	21.10%	489.9
CONDUCT	711	0.85%	0.52%	0.36%	466	20.66%	487.0
COORDIN	698	0.84%	0.51%	0.36%	482	21.37%	467.9
PLAN	661	0.79%	0.49%	0.34%	453	20.08%	460.9
INFORM	647	0.77%	0.48%	0.33%	459	20.35%	447.4
TRAIN	642	0.77%	0.47%	0.33%	448	19.86%	450.7
OFFIC	621	0.74%	0.46%	0.32%	425	18.84%	450.2
PATIENT	612	0.73%	0.45%	0.31%	257	11.39%	577.4
NEW	564	0.67%	0.41%	0.29%	399	17.69%	424.3
CALL	563	0.67%	0.41%	0.29%	388	17.20%	430.4
PRODUCT	559	0.67%	0.41%	0.29%	357	15.82%	447.6

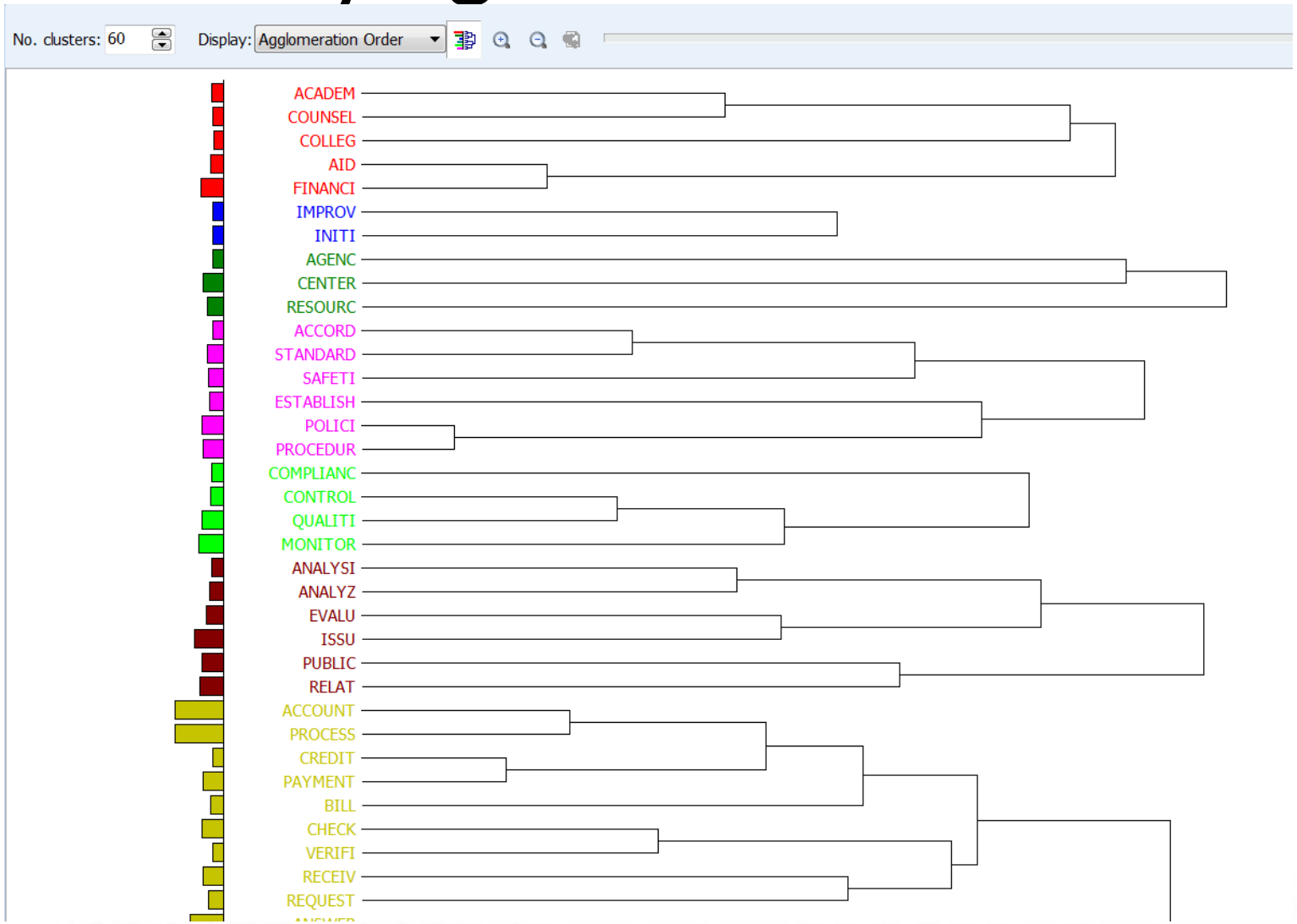
Relevance - Hired

	Hired	Qualified Not Hired	FTF Iw - Not Qualified	No Iw, Qualified for FTF	Didn't Pass Screener/Assessment	Chi2	P
TAUGHT	10.34%	6.90%	0.86%	19.83%	62.07%	23.056	0.000
HOUSEHOLD	10.10%	4.04%	6.06%	20.20%	59.60%	21.069	0.000
ENGLISH	9.84%	6.56%	1.09%	25.14%	57.38%	30.509	0.000
SELECT	9.38%	4.38%	3.75%	18.75%	63.75%	22.117	0.000
SAMPL	9.00%	3.00%	2.00%	20.00%	66.00%	11.284	0.024
HIRE	8.99%	2.81%	2.81%	26.97%	58.43%	19.858	0.001
MENTOR	8.65%	2.88%		14.42%	74.04%	17.851	0.001
EDUC	8.51%	4.68%	1.28%	23.62%	61.91%	43.751	0.000
INSTRUCTOR	8.16%	4.76%	1.36%	26.53%	59.18%	12.620	0.013
ADULT	7.84%	1.96%	2.94%	35.29%	51.96%	14.536	0.006
SENIOR	7.75%	3.88%	0.78%	27.13%	60.47%	9.950	0.041
FACE	7.41%	5.82%	4.76%	26.98%	55.03%	19.466	0.001
VOLUNT	7.35%	5.15%	3.68%	16.18%	67.65%	12.852	0.012
RESOURC	7.19%	4.79%	2.99%	23.35%	61.68%	8.948	0.062
COLLECT	7.19%	4.44%	3.59%	23.26%	61.52%	27.005	0.000
GRADE	7.14%	4.76%	1.79%	25.00%	61.31%	8.677	0.070
STAFF	7.04%	3.02%	1.76%	24.37%	63.82%	18.741	0.001
TECHNIC	7.02%	4.39%	4.39%	16.67%	67.54%	10.164	0.038
GROUP	7.02%	3.07%	2.19%	27.63%	60.09%	11.641	0.020
TREATMENT	6.73%	5.77%	0.96%	32.69%	53.85%	10.119	0.038
CLERK	6.36%	5.45%	4.55%	29.09%	54.55%	8.691	0.069
RESPOND	6.32%	10.88%	4.56%	20.00%	58.25%	55.212	0.000
GRADUAT	6.25%	8.93%	2.68%	22.32%	59.82%	11.395	0.022
DESIGN	6.25%	6.25%	1.79%	25.89%	59.82%	10.791	0.029
RECRUIT	6.25%	5.06%	2.08%	22.32%	64.29%	11.545	0.021
LEADER	6.25%	4.46%		34.82%	54.46%	12.009	0.017
IMPROV	6.25%	3.57%	0.89%	29.46%	59.82%	5.447	0.244
SUPERVIS	6.23%	2.86%	1.56%	26.75%	62.60%	13.203	0.010
LEVEL	6.16%	1.37%	2.05%	20.55%	69.86%	7.793	0.099
OWNER	6.14%	6.14%	6.14%	23.68%	57.89%	12.133	0.016
INTERVIEW	6.07%	6.83%	2.66%	23.67%	60.77%	67.157	0.000
PARENT	5.98%	2.72%	1.09%	26.63%	63.59%	6.265	0.180
ISSU	5.94%	3.50%	3.85%	25.87%	60.84%	9.753	0.045
TRAVEL	5.92%	4.73%	1.78%	27.81%	59.76%	5.476	0.242

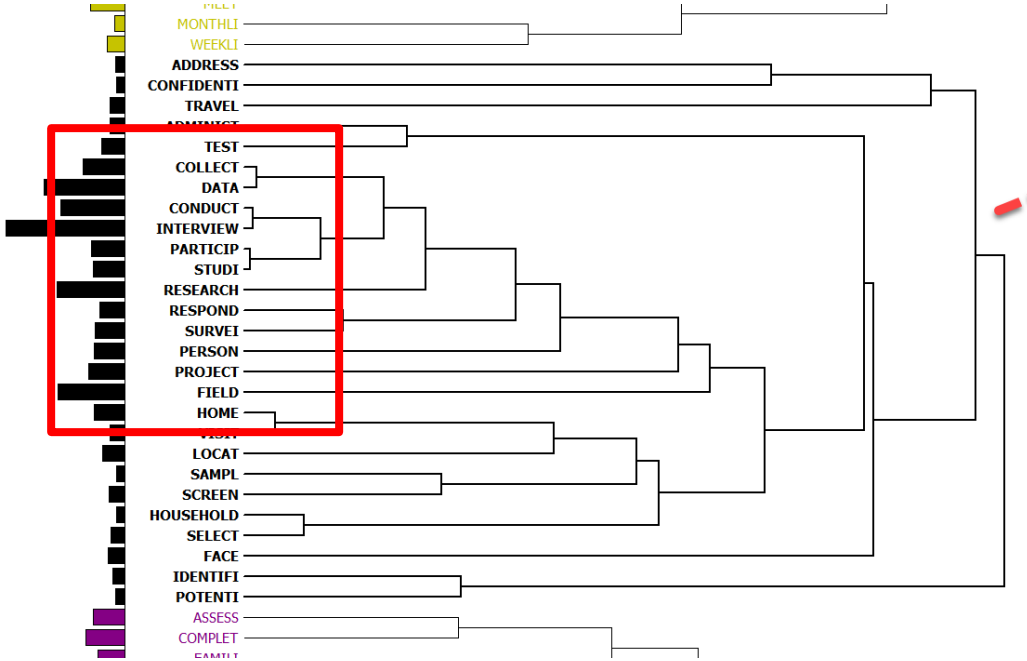
Relevance – Not Qualified

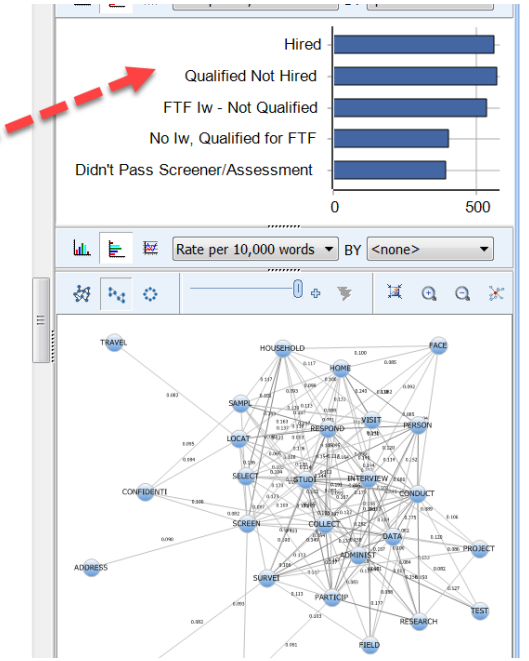
	Hired	Qualified Not Hired	FTF Iw - Not	No Iw, Qualified for FTF	Didn't Pass Screener/Assessment	Chi2	P
DOOR	1.65%	4.96%	8.26%	23.14%	61.98%	19.181	0.001
BEHAVIOR	4.97%	6.83%	8.07%	20.50%	59.63%	28.985	0.000
SELL	0.81%	4.03%	8.06%	19.35%	67.74%	20.740	0.000
ASSIGN	3.41%	3.75%	7.85%	21.16%	63.82%	38.408	0.000
CHECK	4.59%	5.96%	7.34%	19.72%	62.39%	28.751	0.000
POLICI	0.91%	4.11%	7.31%	18.72%	68.95%	29.874	0.000
APPOINT	1.26%	5.46%	6.72%	30.25%	56.30%	28.640	0.000
VERIFI	1.83%	5.50%	6.42%	25.69%	60.55%	9.247	0.055
BILL	2.33%	10.08%	6.20%	29.46%	51.94%	25.624	0.000
OWNER	6.14%	6.14%	6.14%	23.68%	57.89%	12.133	0.016
SCREEN	2.76%	5.52%	6.08%	24.86%	60.77%	12.298	0.015
HOUSEHOLD	10.10%	4.04%	6.06%	20.20%	59.60%	21.069	0.000
MONTHLI	3.31%	4.96%	5.79%	28.10%	57.85%	7.657	0.105
CHANG	2.75%	2.75%	5.50%	28.44%	60.55%	5.701	0.223
SERV	4.95%	2.97%	5.45%	24.26%	62.38%	10.322	0.035
ADDRESS	3.60%	4.50%	5.41%	29.73%	56.76%	6.547	0.162
RESID	2.44%	2.93%	5.37%	31.22%	58.05%	13.156	0.011
ONLIN	1.77%	4.42%	5.31%	30.97%	57.52%	7.485	0.112
ADMINISTR	3.18%	2.33%	5.30%	29.45%	59.75%	25.240	0.000
FRONT	5.22%	7.46%	5.22%	31.34%	50.75%	16.969	0.002
SYSTEM	3.32%	2.58%	5.17%	25.09%	63.84%	9.937	0.042
APPROPRI	3.23%	1.29%	5.16%	27.10%	63.23%	7.952	0.093
SURVEI	3.89%	6.89%	5.09%	29.04%	55.09%	25.874	0.000
PAYMENT	5.00%	6.50%	5.00%	30.00%	53.50%	17.278	0.002
EVENT	2.15%	1.54%	4.92%	26.15%	65.23%	14.607	0.006
TRANSPORT	1.63%	3.25%	4.88%	20.33%	69.92%	5.841	0.211
INDIVIDU	2.42%	9.00%	4.84%	28.03%	55.71%	32.763	0.000
GREET	3.45%	1.38%	4.83%	32.41%	57.93%	10.184	0.037
GUEST	3.37%	6.25%	4.81%	27.88%	57.69%	10.646	0.031
FACE	7.41%	5.82%	4.76%	26.98%	55.03%	19.466	0.001
ANALYZ	3.40%	6.80%	4.76%	25.17%	59.86%	7.434	0.115
LOCAT	5.10%	9.80%	4.71%	20.78%	59.61%	34.855	0.000
ANSWER	2.46%	2.77%	4.62%	30.77%	59.38%	14.233	0.007
HEALTH	5.17%	5.17%	4.57%	27.44%	57.65%	23.300	0.000

Identifying Clusters of Words

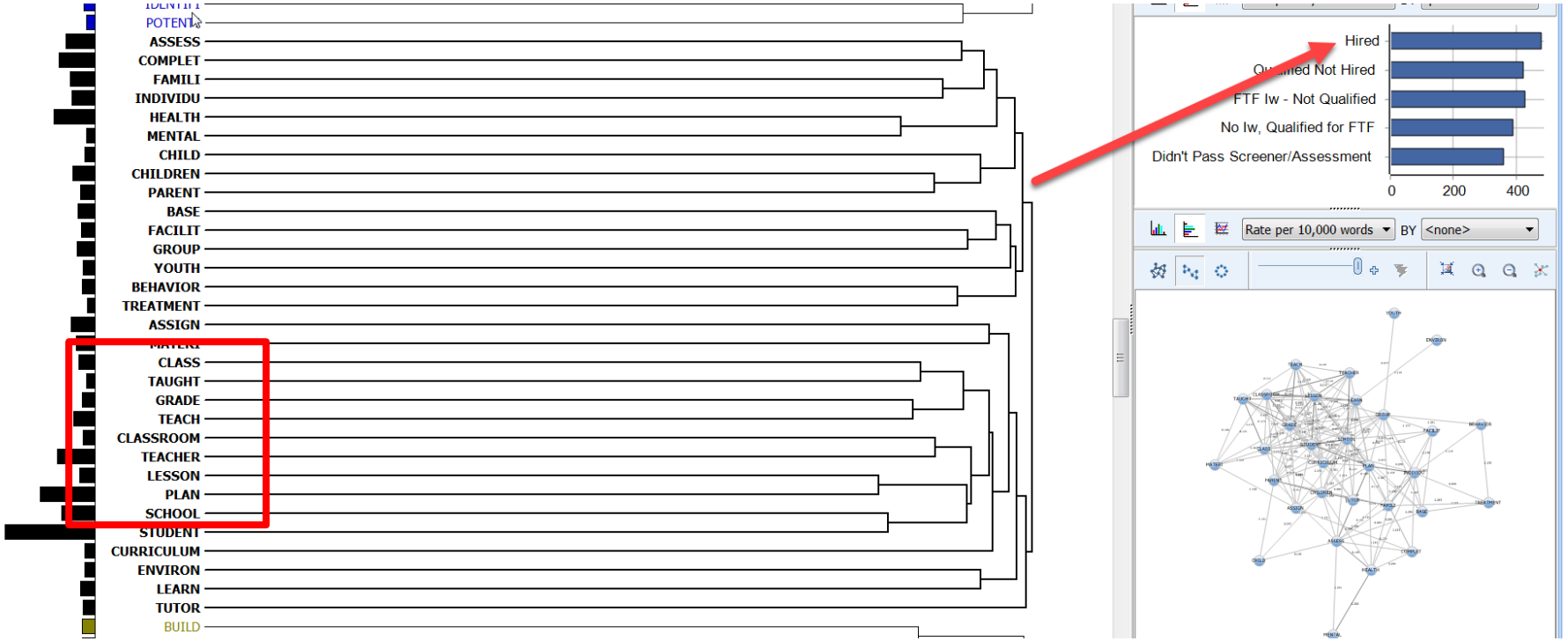


Clusters – Some Effect, but Surprisingly Small

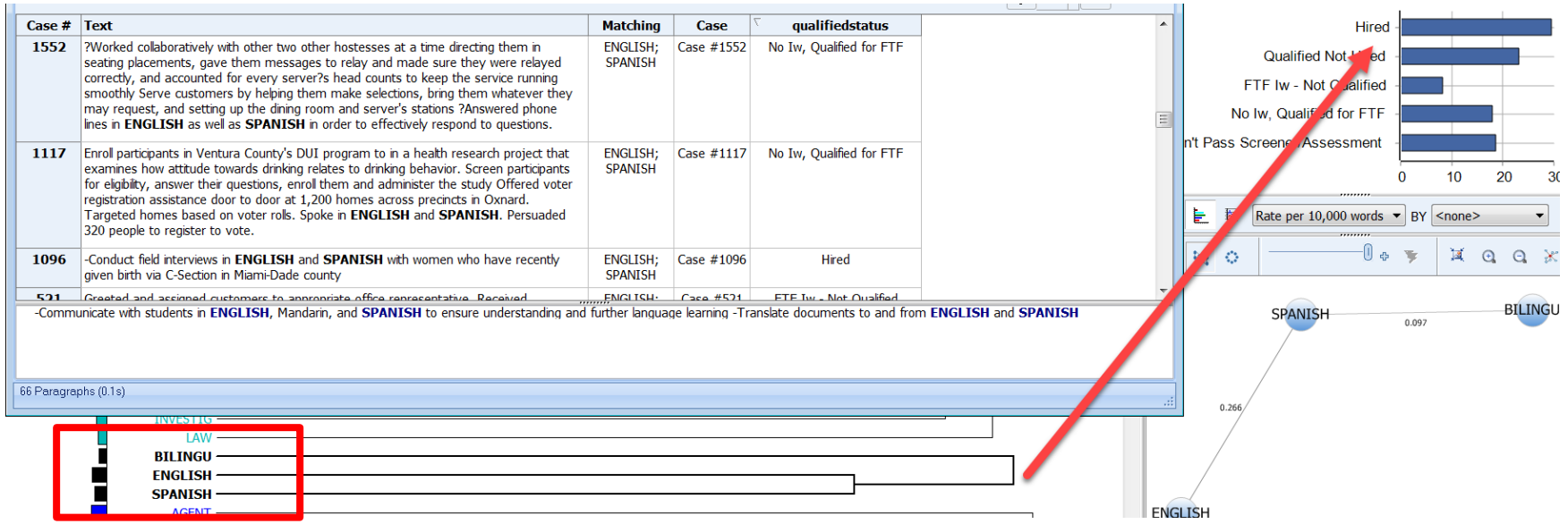




Clusters – Unexpected Predictor



Cluster – Unexpectedly Strong Predictor



Selecting Keywords

Include Words – words in parentheses removed as not statistically significant	Exclude Words – words in parentheses removed as not statistically significant
TAUGHT, ENGLISH, EDUC	(DOOR, SELL, PAYMENT)
SELECT, COLLECT	(BEHAVIOR)
FACE, RESPOND, INTERVIEW	(ADMINISTR, ANALYZ)
(STAFF, HIRE, TRAIN)	ANSWER, (PHONE), GREET

Include in Iw, Dummy Var	Exclude from Iw, Dummy Var 0	1	Total
0	**989**	**163**	**1,152**
1	**981**	**181**	**1,162**
Total	**1,970**	**344**	**2,314**

What If We Used the Keywords With Our Past Applicants…

Qualified Status from Hiring	Include Variables, Dummy Var No	Yes	Total
Hired	16 21.62	58 78.38	74 100.00
Qualified Not Hired	29 33.33	58 66.67	87 100.00
FTF Iw - Not Qualifie	22 40.74	32 59.26	54 100.00
No Iw, Qualified for	272 47.97	295 52.03	567 100.00
Didn't Pass Screener/	779 52.85	695 47.15	1,474 100.00
Total	1,118 49.56	1,138 50.44	2,256 100.00

Pearson chi2(4) = 40.9022 Pr = 0.000

Qualified Status from Hiring	Exclude from Iw, Dummy Var No	Yes	Total
Hired	66 89.19	8 10.81	74 100.00
Qualified Not Hired	78 89.66	9 10.34	87 100.00
FTF Iw - Not Qualifie	40 74.07	14 25.93	54 100.00
No Iw, Qualified for	474 83.60	93 16.40	567 100.00
Didn't Pass Screener/	1,262 85.62	212 14.38	1,474 100.00
Total	1,920 85.11	336 14.89	2,256 100.00

Pearson chi2(4) = 8.9001 Pr = 0.064

Multivariate approach

- DV= Qualified (binary 0,1)
  - Constructed using RecruitCode

Hire Status (DCSR Codes)	Qualified
Hired (101, 102)	1
Qualified Not Hired (608, 899, 900)	1
FTF Iw, Not Qualified (607, 609, 610)	0
No Iw, Qualified for FTF (402, 403, 404, 555)	1
Didn't Pass Screener/Assessment (401, 601, 602, 603, 606)	0

- Independent Variable = Selected words as indicator variables
- Logistic regression model was run to predict being qualified over different words.

Multivariate approach

- Model 1
  - IV: Relevant words found in previous analysis (bivariate associations)
    - Overall, low model fit.
    - When evaluating all variables together, the only significant words were “Interview(er)” and “English”, both with a positive effect.
    - No significant words with negative impact (“exclude words”)

VARIABLES	(1) Odds Ratios
Qualified	
teach	1.143
english	1.541**
education	0.905
select	0.843
collection	1.068
face	1.168
respond	1.034
iwer	1.472***
answer	1.126
greet	1.086
Constant	0.438***
R2	0.009
Observations	2,314

*** p<0.01, ** p<0.05, * p<0.1

- Using the word “English” makes a candidate 1.54 times (or 54%) more likely to be qualified than those who don’t.
- Using the word “Interview(er)” makes a candidate 1.47 times (or 47%) more likely to be qualified than those who don’t use it.

Multivariate approach

- Model 2: General Model
  - Selection of words using descriptive statistics (more frequent words used by those qualified and not qualified).
  - Words used by both groups are not that different!
    - Table present bigger differences.

Word	Word % by Qualified	Word % by NotQ	Difference
data	0.76%	0.63%	0.13%
medical	0.44%	0.31%	0.13%
assistant	0.63%	0.51%	0.12%
office	0.47%	0.37%	0.10%
school	0.53%	0.43%	0.10%
health	0.72%	0.62%	0.10%
customers	0.40%	0.50%	-0.10%
service	0.93%	1.04%	-0.11%
sales	0.57%	0.72%	-0.16%
Work	0.01%	0.36%	-0.35%

Multivariate approach

- Model 2: General Model
  - 26 predictors (words) were tested simultaneously and those correlated or with low impact were excluded from the model.
  - Overall, low model fit (but a bit higher than previous)
  - Results from previous model maintain (Iwer and English), but there are other significant predictors:

- All predictors painted in blue are significant with a positive relation with being qualified.
- Using the word “English” makes a candidate 1.68 times (or 68%) more likely to be qualified than those who don’t.
- The word “education” is also significant but has a negative impact. Using this word makes a candidate 20% less likely to be qualified than those who don’t use it.

	(1) Full Model OR	(2) Reduced Model 1 - OR	(3) Reduced Model 2 - OR
iwer	1.417***	1.423***	1.362***
health	1.152	1.152	1.159
assistant	1.304**	1.298**	1.308***
english	1.685**	1.683**	1.722**
spanish	0.673	0.673	0.673
social	1.344**	1.328**	1.315**
university	1.176	1.175	1.15
manager	1.178	1.176	1.174
language	1.226	1.228	1.257
service	1.109	1.104	1.127
education	0.804*	0.805*	0.821*
community	1.283**	1.282**	1.300**
data	1.264**	1.254**	1.240**
teach	1.073	1.073	
select	0.841	0.844	
face	1.109	1.111	
provide	1.074	1.071	
answer	1.09	1.087	
greet	1.132	1.127	
research	0.922	0.925	
field	1.049		
customer	1.004		
professions	0.956		
collection	0.962		
sales	0.989		
respond	0.973		
Constant	0.323***	0.324***	0.333***
R2	0.022	0.022	0.021
Observatio	2,314	2,314	2,314

seEform in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Multivariate approach

- "Paradata" variables were tested
  - Amount of characters written
  - Number of previous job experience listed
  - Additional resume attached

- … But no effect over being qualified

How Do We Keep Our Interviewing Staff? Interviewer Satisfaction Survey

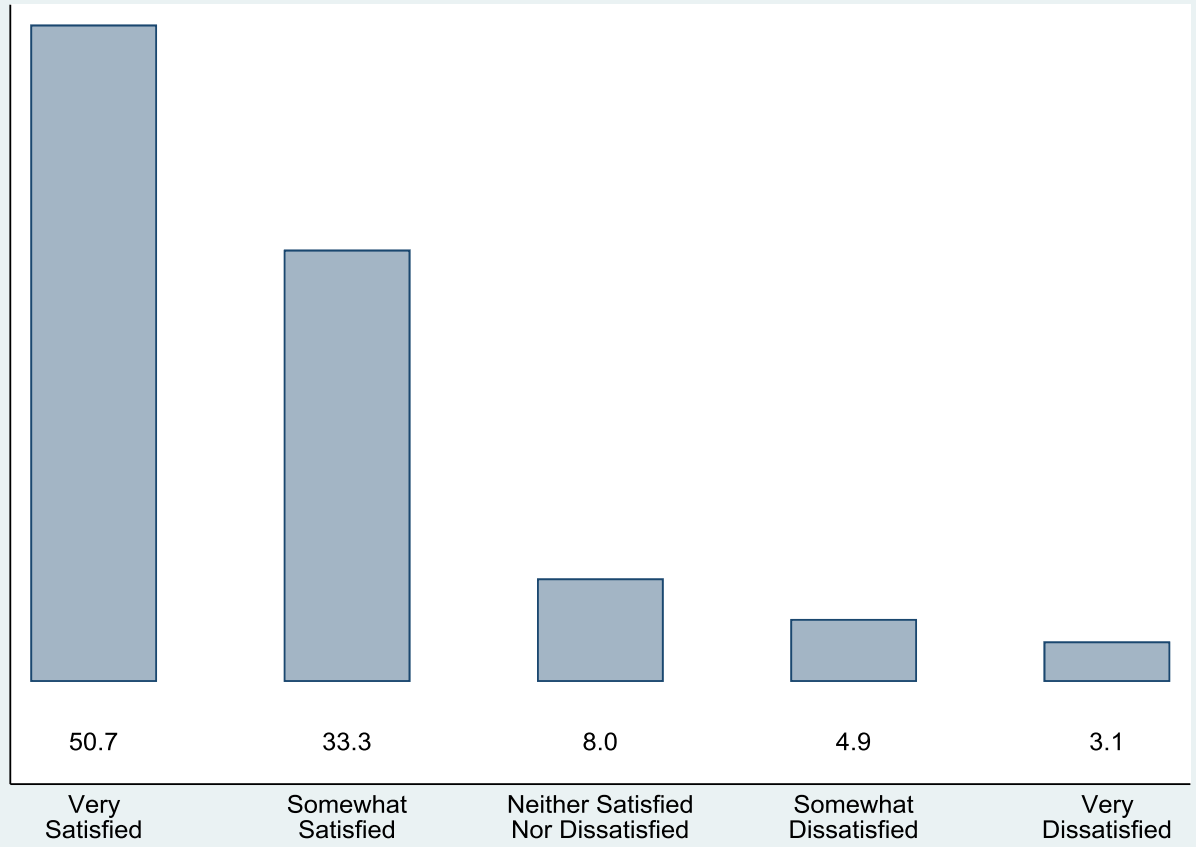
Interviewer Satisfaction Survey Responses

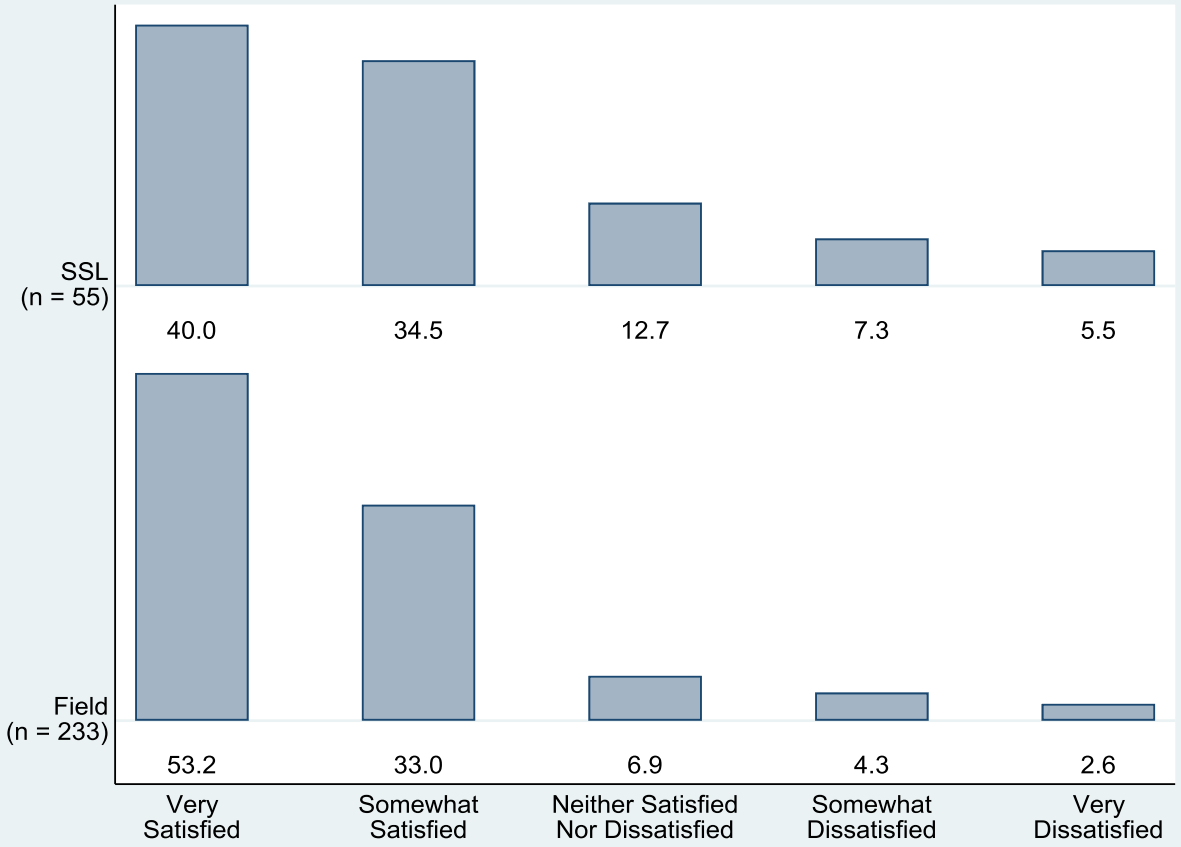
Frame	Population: 660 Active Iwers Sample: 303 RR = 46%
Dates	Start: June 11, 2018 Email Reminder: June 22, 2018 End: July 16, 2018
IW Length	Mean: 36 mins (sd = 27) Median: 27
Main DV	Overall Satisfaction Intent of Stay

Sample Characteristics (n=303)	n (303)	%
Gender		
Male	52	17.2
Female	250	82.5
Type		
SSL	63	20.8
Field	240	79.2
Race (3 Categories)		
White	213	70.3
Black, Non White	60	19.8
Other, Non White	30	9.9
Is Bilingual		
No	260	85.8
Yes	43	14.2
Years Working At SRO		
Less Than 1 Year	72	23.8
1 To 3 Years	75	24.8
4 To 9 Years	62	20.4
10 Or More Years	94	31.0

Two Main DV

- Overall satisfaction: "*Considering everything, how satisfied are you with SRO?*"

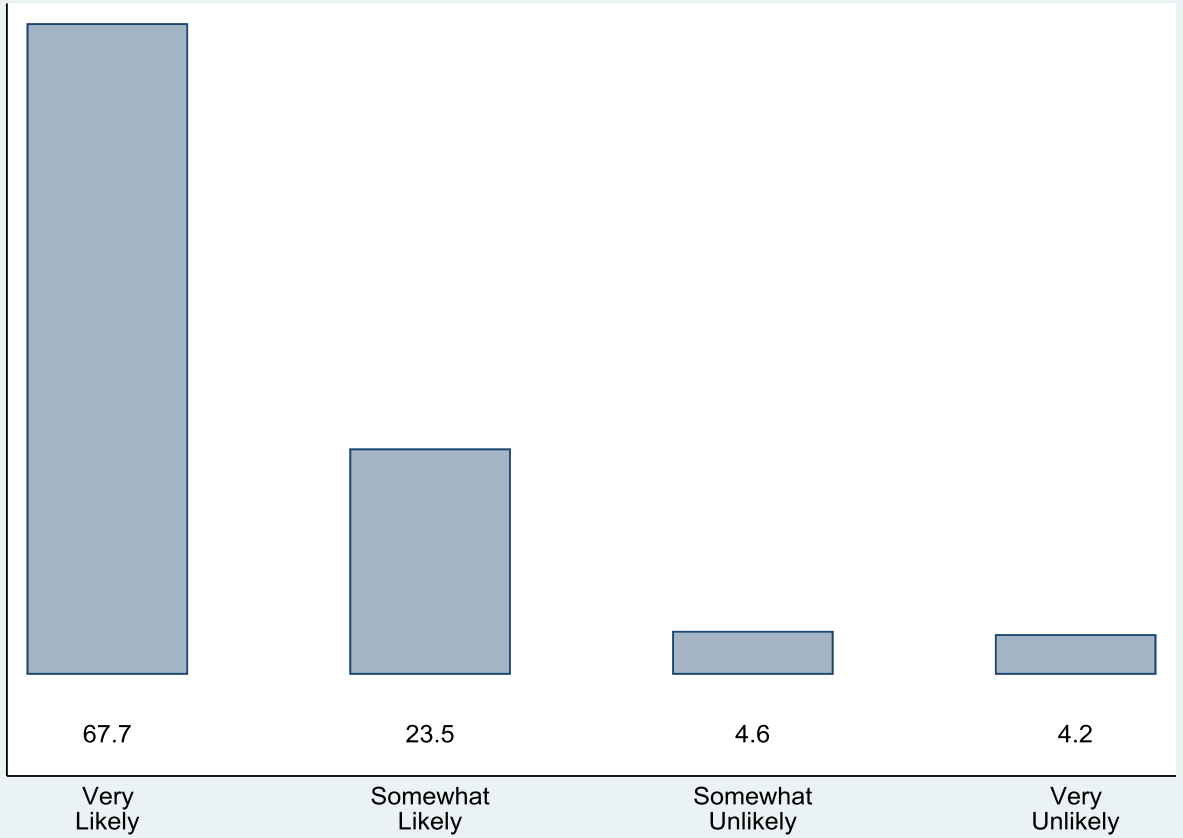


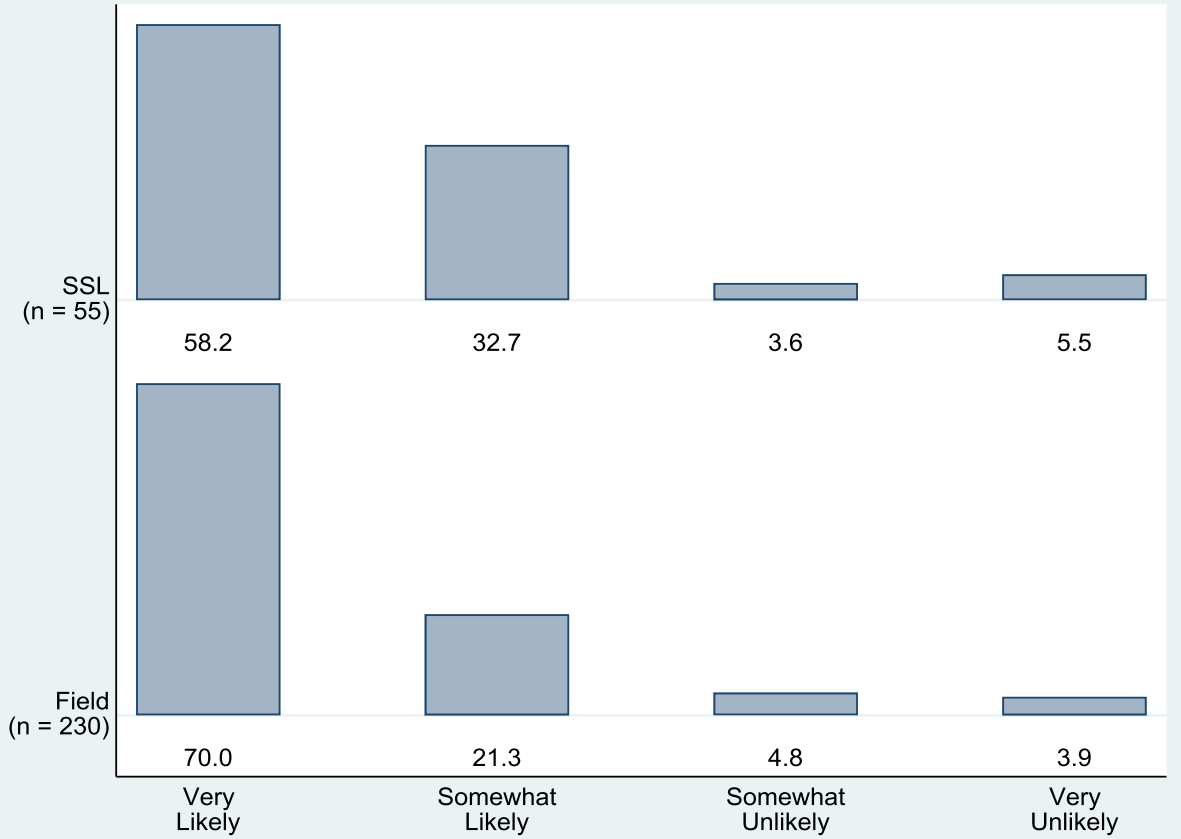


Two Main DV

- Intent of Stay:

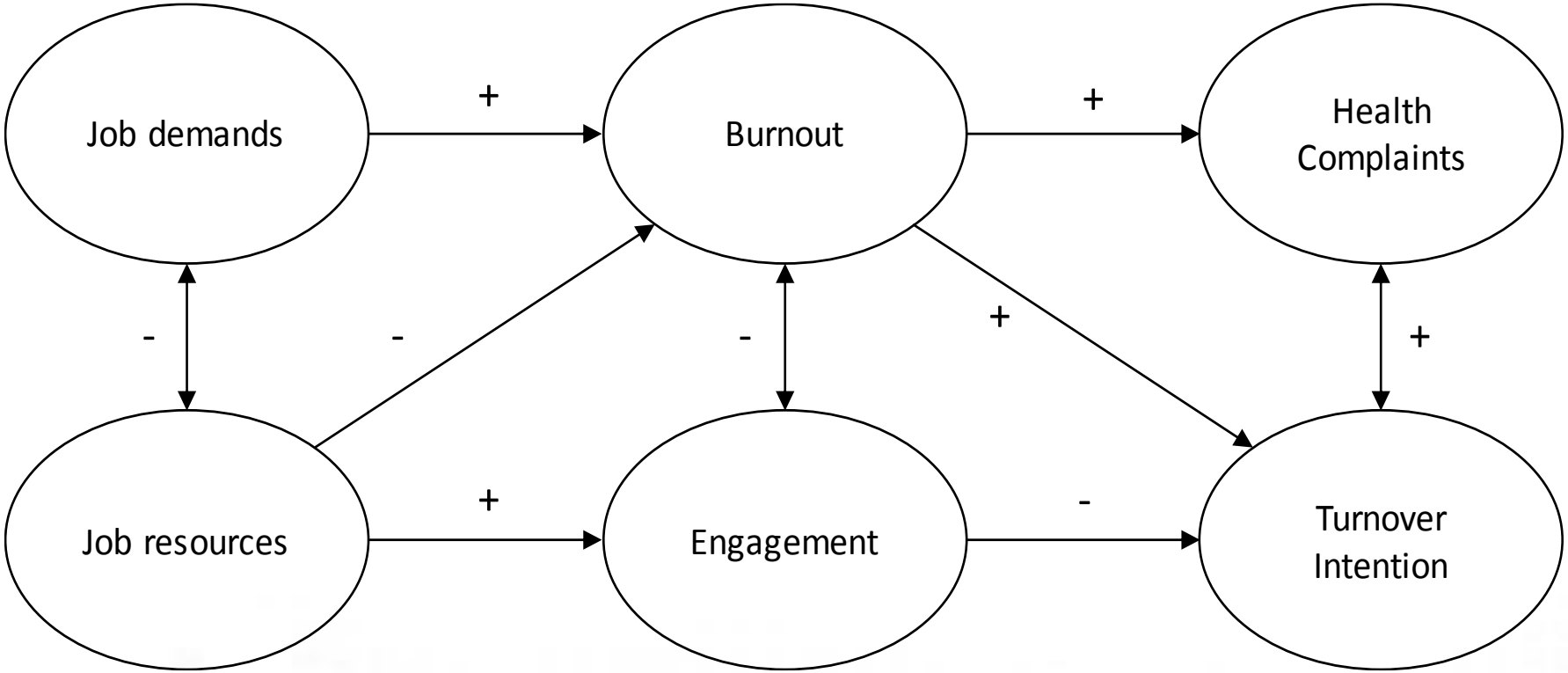
"As of today, how likely are you to continue to work for SRO next year, as long as there is work available"



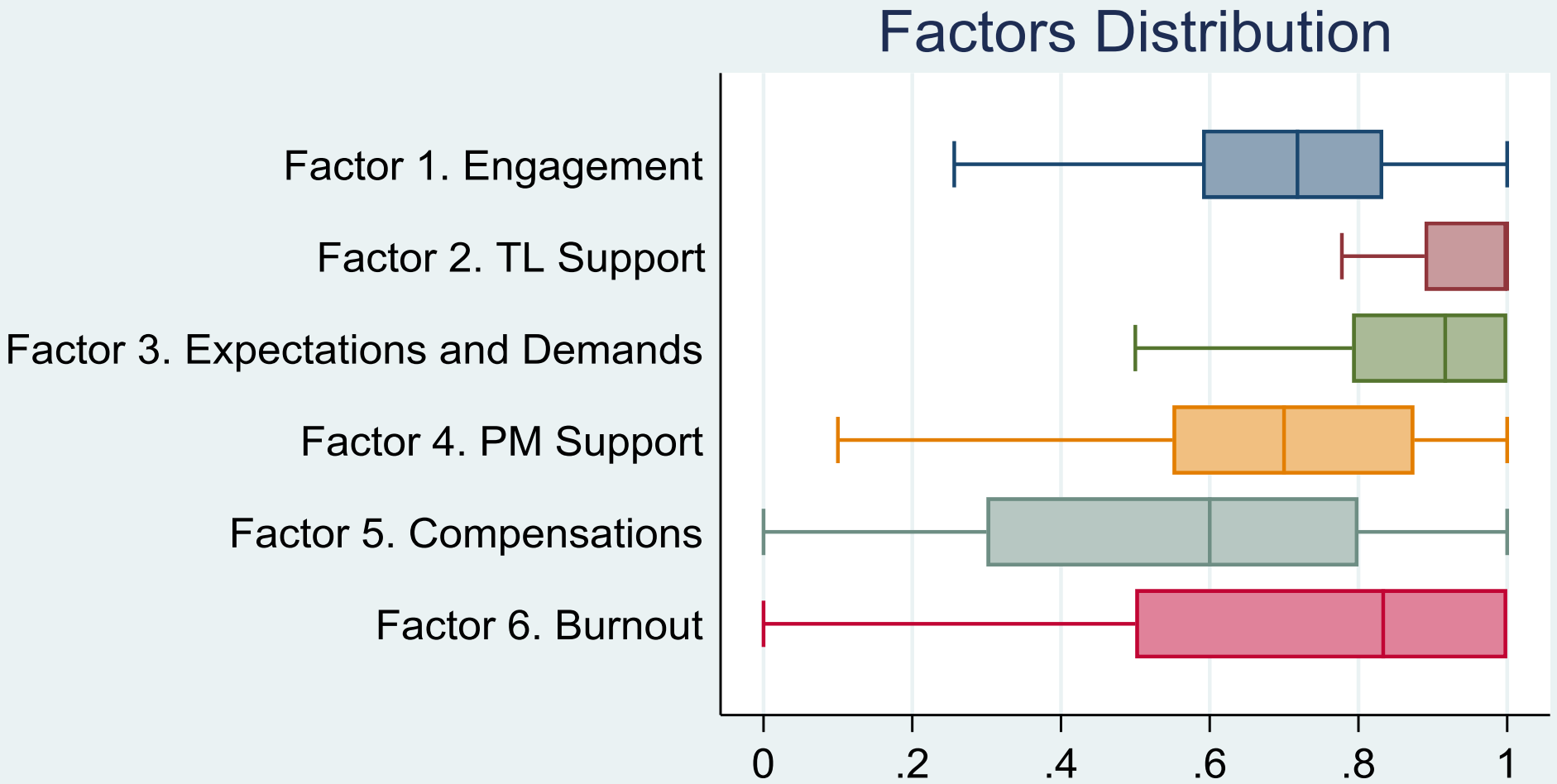


General Framework: JD-R

- Job Demands – Resources (Korunka et. al. (2009))

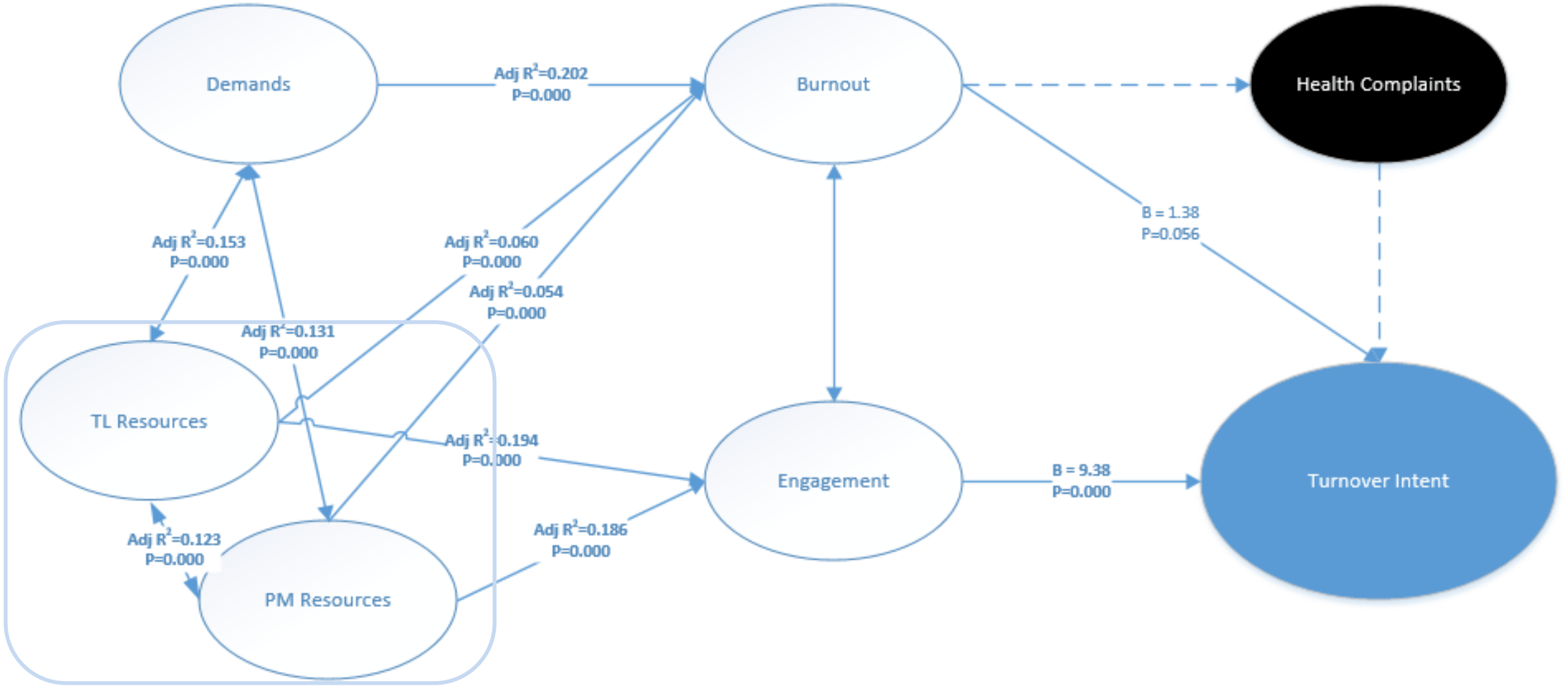


Factor Analysis Results

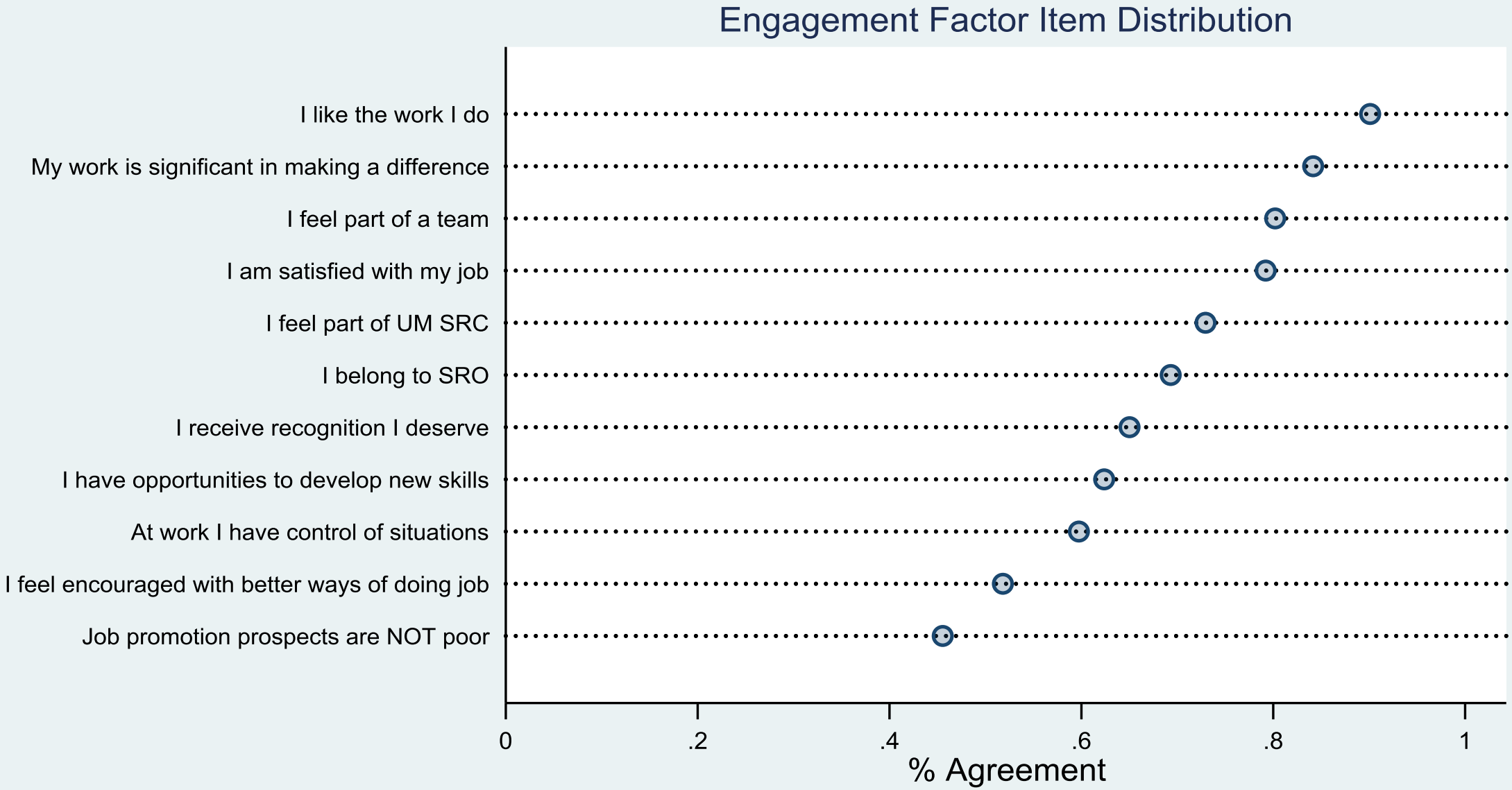


Note: Higher values indicate a positive evaluation of the topic

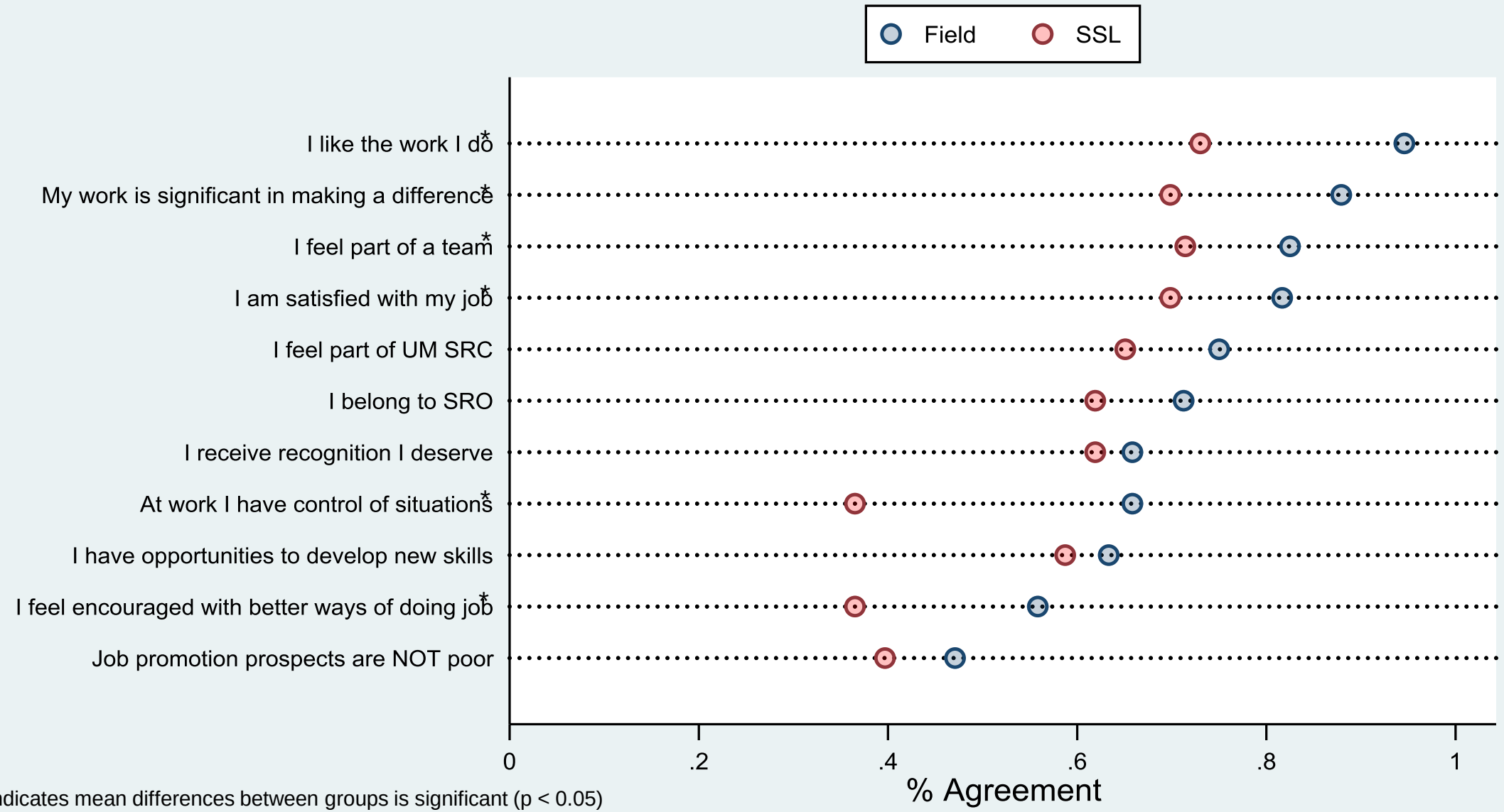
Tentative Survey Results (9-24-2018)



Correlations EXCEPT to turnover intent using bivariate regression results of unstandardized factor scores
Turnover intent correlations are bivariate logistic regression results

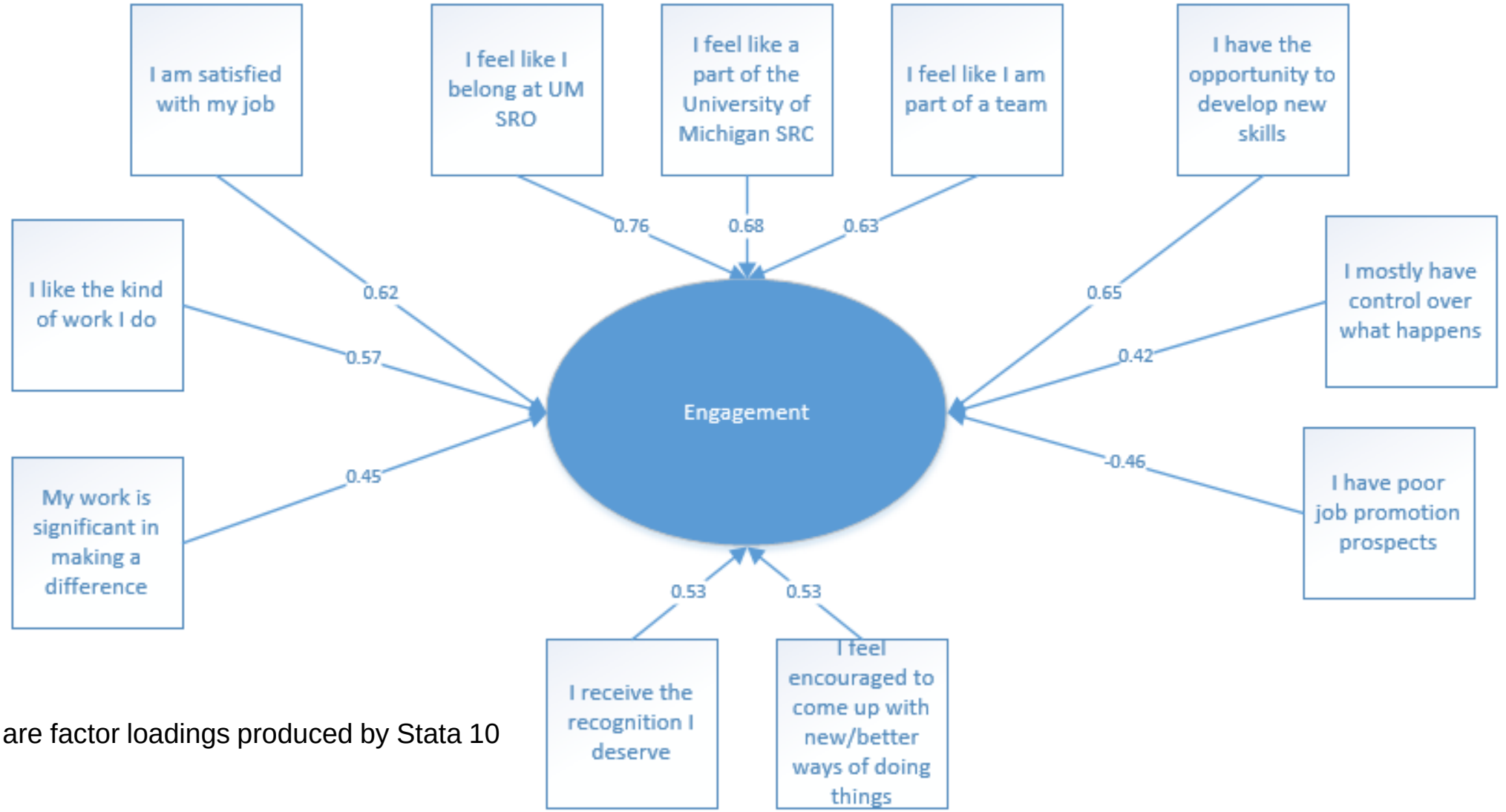


Engagement Factor Item Distribution



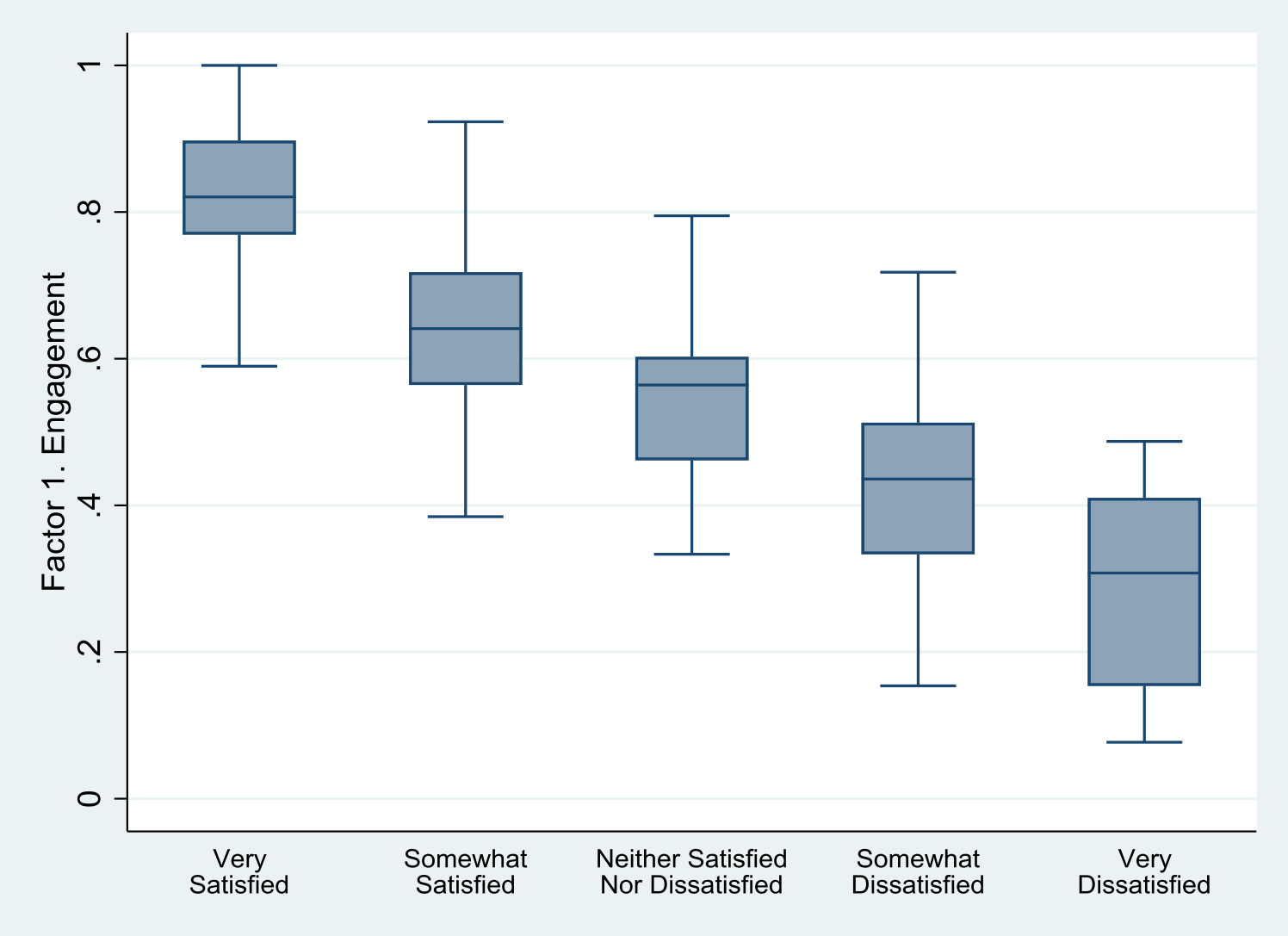
* Indicates mean differences between groups is significant ($p < 0.05$)

Intrinsic Motivation – Interviewer Engagement

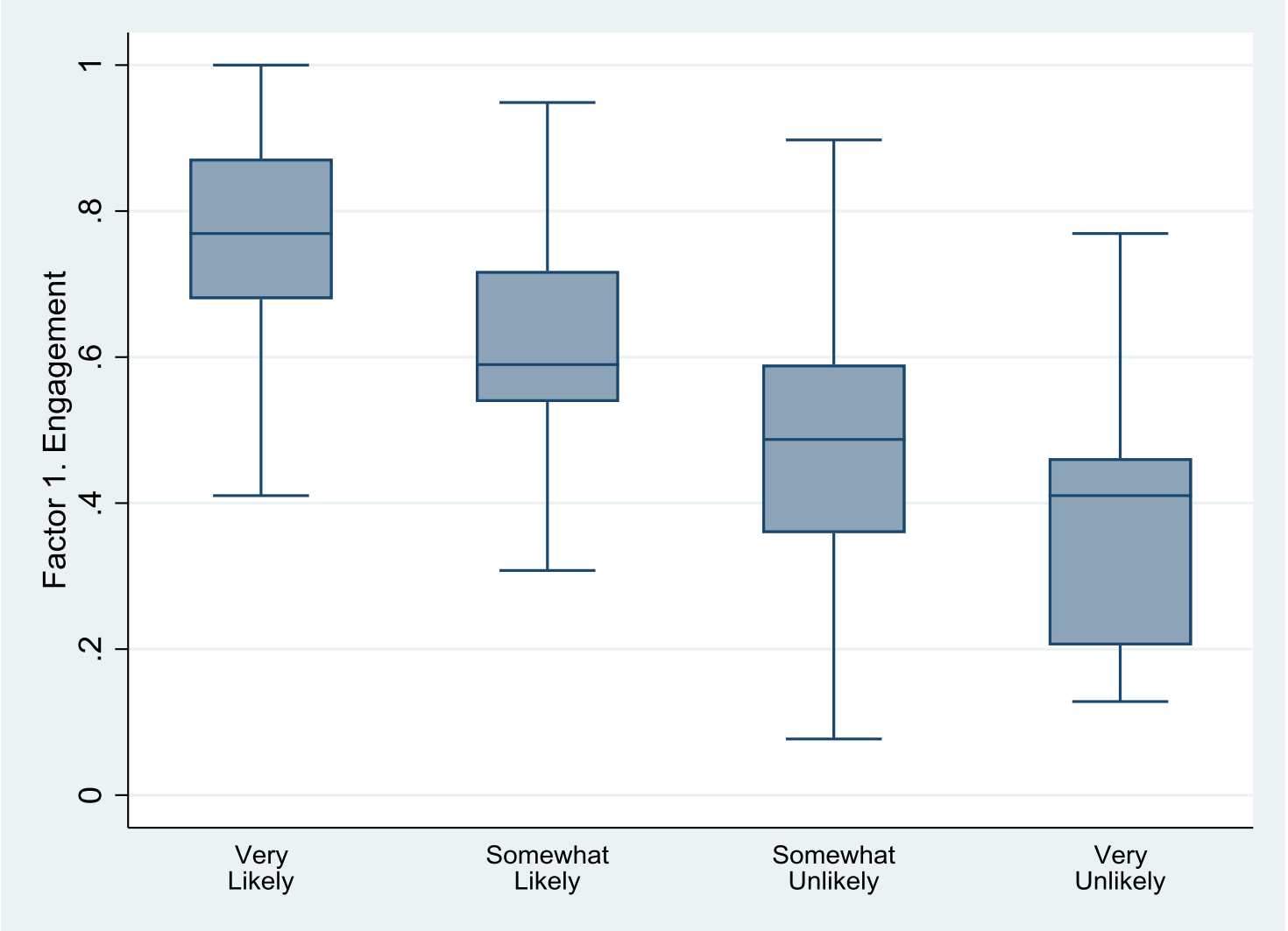


* Values are factor loadings produced by Stata 10

Engagement and Job Satisfaction



Engagement and Intent to Stay



How Do We Increase Engagement?

- Branding and UM / SRO Awareness
- Training changes
  - Orientation, campus tour
- Communications improvement
  - Better use of interviewer website
  - Work in the world; Sharing project results back to iwers
  - “Stay” interviews / testimonials
- On-going Feedback Process (repeat surveys)

Next Steps

- Resume Review
  - Overall, model fit was quite low
    - Test with other variables rather than just job duties descriptors (read CV in pdf?)
    - Volatility of word usage without context, review ngrams (previous and posterior word)
    - Since it seems to be more of a prediction problem (rather than estimating coefficients), try Random Forest analysis and Lasso models instead of the logit model.
  - Do these results vary study to study?
    - Are the words of qualified resumes for one project different from words use by those qualified in other studies?
    - Are there variables that apply across several studies?
- Interviewer Engagement and Retention
  - Other suggestions?

Suggestions / Feedback

- Resume review
  - Create factors / indices
  - Compare to high performers instead of hires
  - Do search process early, not last ditch effort
  - Beware of exclusionary variables
  - Create a dictionary rather than just use stem words
  - Map to LinkedIn profile
- Engagement
  - Add iwer profiles
  - Do more to include iwers in activities
  - Advise iwers where they can be creative
  - TLs to Iwers – smaller teams; have TLs run several smaller groups
  - Maybe an alumni engagement effort / newspaper