AI and technology: Applications in educational assessment, recruitment and last-mile healthcare

Shashank Srikant
shash@mit.edu

Result of work done at

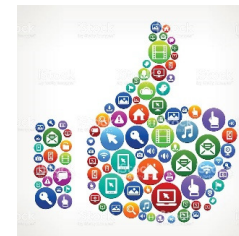Varun Aggarwal   research.aspiringminds.com

Nishant Ojha   sevasetu.org

Information Asymmetry

Information Asymmetry

Technology to the rescue

Many million of graduates

Millions of companies

I need a job.
Am I employable?

How do we sift through
million of candidates?

How do I signal my
employability to companies?

How do we hire those who
would succeed in the job?

Many million of graduates
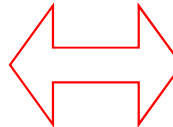
Millions of companies
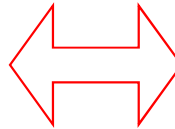
Can technology rescue this problem?
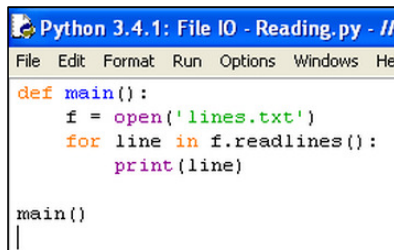
Many million of graduates

Millions of companies
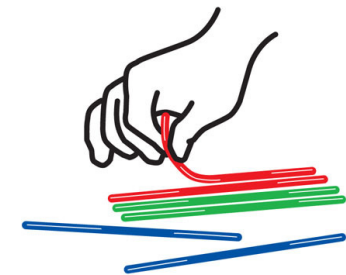

YES

Can technology rescue this problem?

New age assessments

Writing code

Spontaneous speech
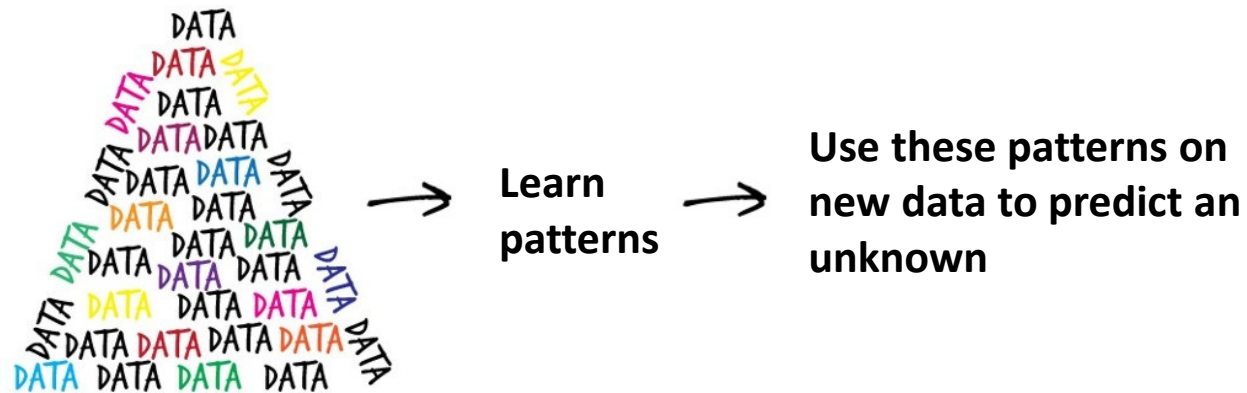
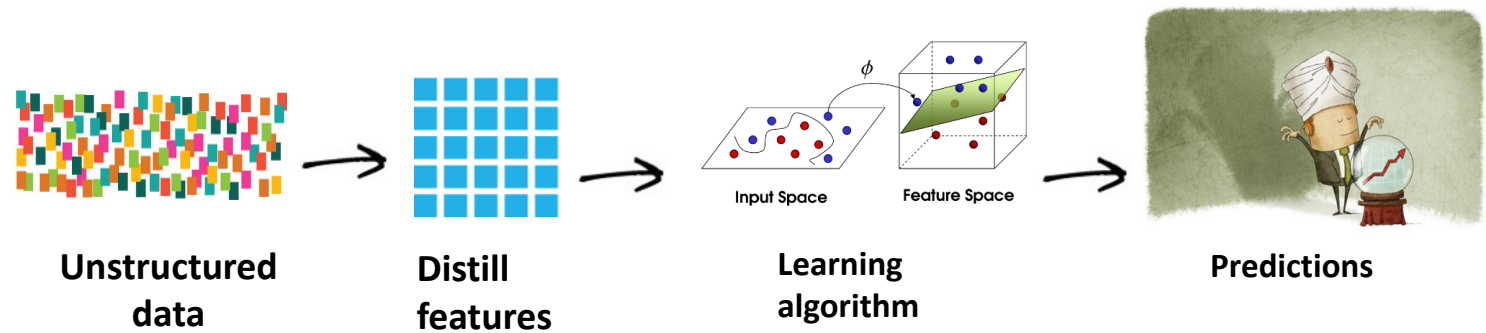Motor skills

Predicting hiring outcome

Labor market insights

Machine Learning
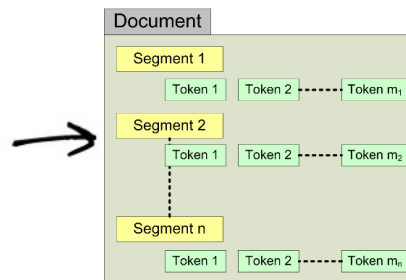
Key idea

Predict an unknown from information you know

Learn patterns → **Use these patterns on new data to predict an unknown**

Machine Learning

Unstructured data → **Distill features** → **Learning algorithm** (Input Space, Feature Space) → **Predictions**

Document — Segment 1, Token 1, Token 2 ... Token m_1; Segment 2, Token 1, Token 2 ... Token m_2; Segment n, Token 1, Token 2 ... Token m_n → **Automatically learn** →

SCARED

afraid, petrified, anxious, shaken, fearful, startled, panicked, terrified

New sentence - *I am afraid of dogs* → ***Scared*** [EMOTION DETECTED]

ML for skill identification – Basic setup
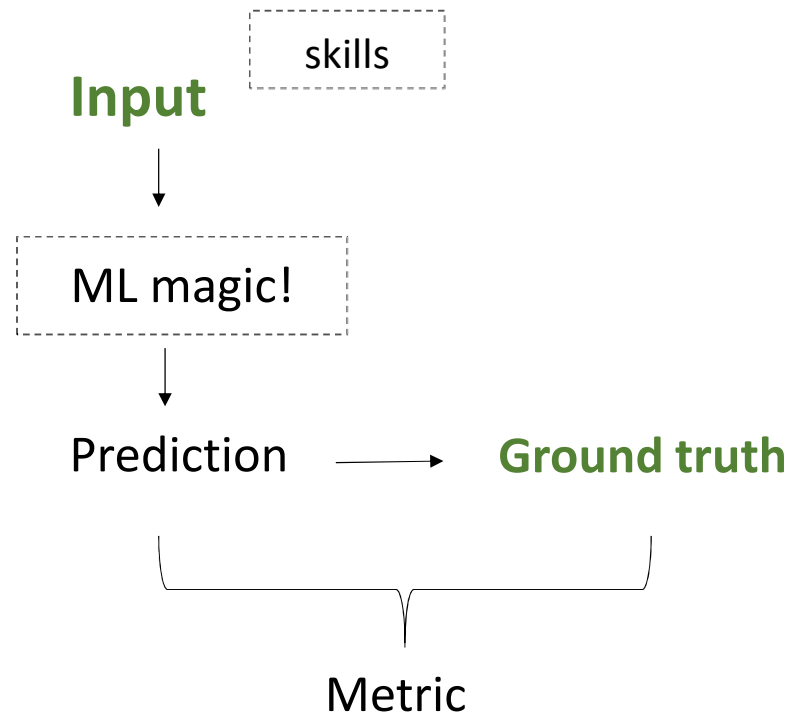
skills

Input

↓

ML magic!

↓

Prediction → **Ground truth**

Metric

Burger Rubric

Wow!

Good

Almost

Oops

New age assessments

Writing code
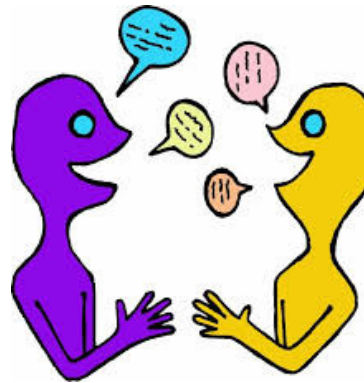
Spontaneous speech
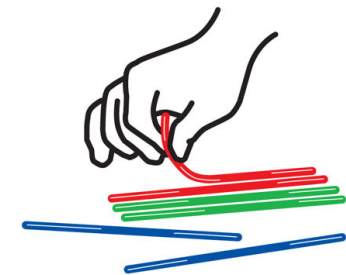
Motor skills

Predicting hiring outcome

Labor market insights

Grading programs

Every TA's nightmare

A 2-3 hour fling for test takers with no *objective* feedback to improve

Helps companies find talent at scale

Existing technology

- **Manual evaluation:** Doesn't scale; not standardized ❌

- **Test-case based evaluation**

  - High false-positives – hard code, not efficient
  - High false-negatives – inadvertent errors ❌

Rubric

An implementation

OBJECTIVE

To print the pattern of integers

1

2   3

3   4   5

4   5   6   7

```
void print_1(int N){
  for(i=1 ;i<=N; i++){
    print newline;
      count = i;
    for(j=0; j<i; j++)
      print count;
        count++;
  }
}
```

1. Are there loops? Are there print statements?

2. Is there a nested-loop structure?

3. Is the conditional in the inner loop dependent on

- a variable modified in the outer loop?

- a variable used in the conditional of the outer loop?

TARGET PROGRAM

```
void print(int N){
for(i =1 ; i<=N; i++){
    print newline;
        count = i;
        for(j=0; j<i; j++)
        print count; count++;
            }
}
```

CONTROL FLOW GRAPH

i = 1 — Parent scope

i <= N — Loop 1

count = i
j = 0

j < i — Loop 2

print(count)
count++
j++

i++

END

OBJECTIVE

To print the pattern of integers

```
1
2    3
3    4    5
4    5    6    7
```

count(**block1**:**loop(loop(++))**)   = 2
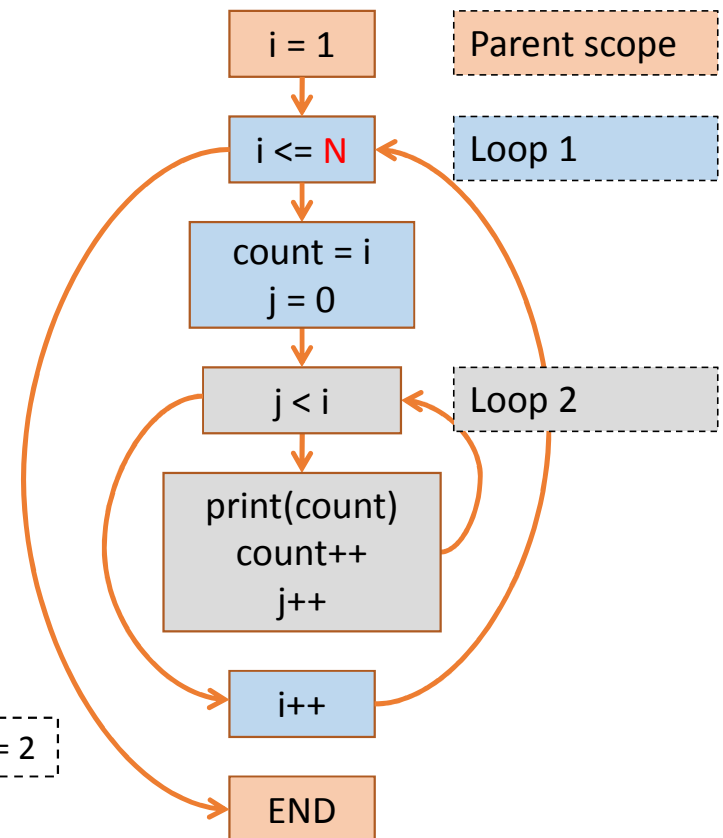
Automata - Our technology

Get an expert to grade these codes

$$A = \begin{bmatrix} 3 & 0 & 1 & 1 \\ 1 & 2 & 5 & 0 \\ 1 & 1 & 3 & 0 \\ 2 & 0 & 1 & 1 \end{bmatrix}$$

Use our grammar to extract features from each student's code

Input Space   Feature Space

Build a model for each question in our question bank
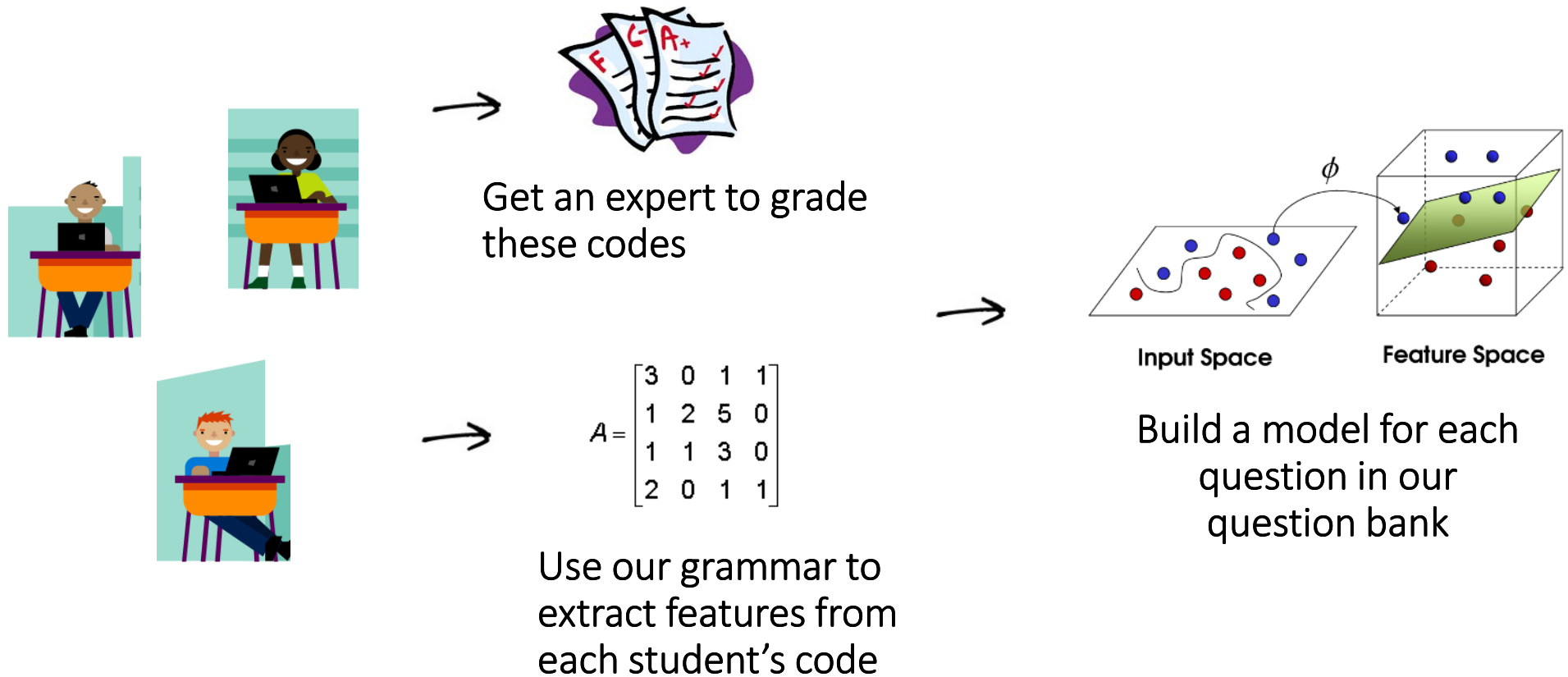
Results

Work published at KDD 2014 and KDD 2016

PROBLEM	# of features	Cross-val correl	Train correl	Validation correl
1	80	0.61	0.85	**0.79**
2	68	0.77	0.93	**0.91**
3	193	0.91	0.98	**0.90**
4	66	0.90	0.94	**0.90**
5	87	0.81	0.92	**0.84**

Validation correlation > 0.79

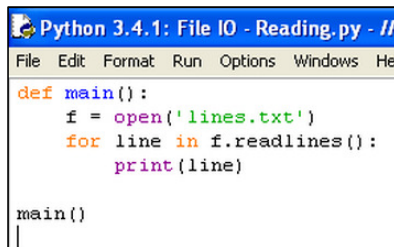**Matches inter-rater correlation between two human raters**

Impact
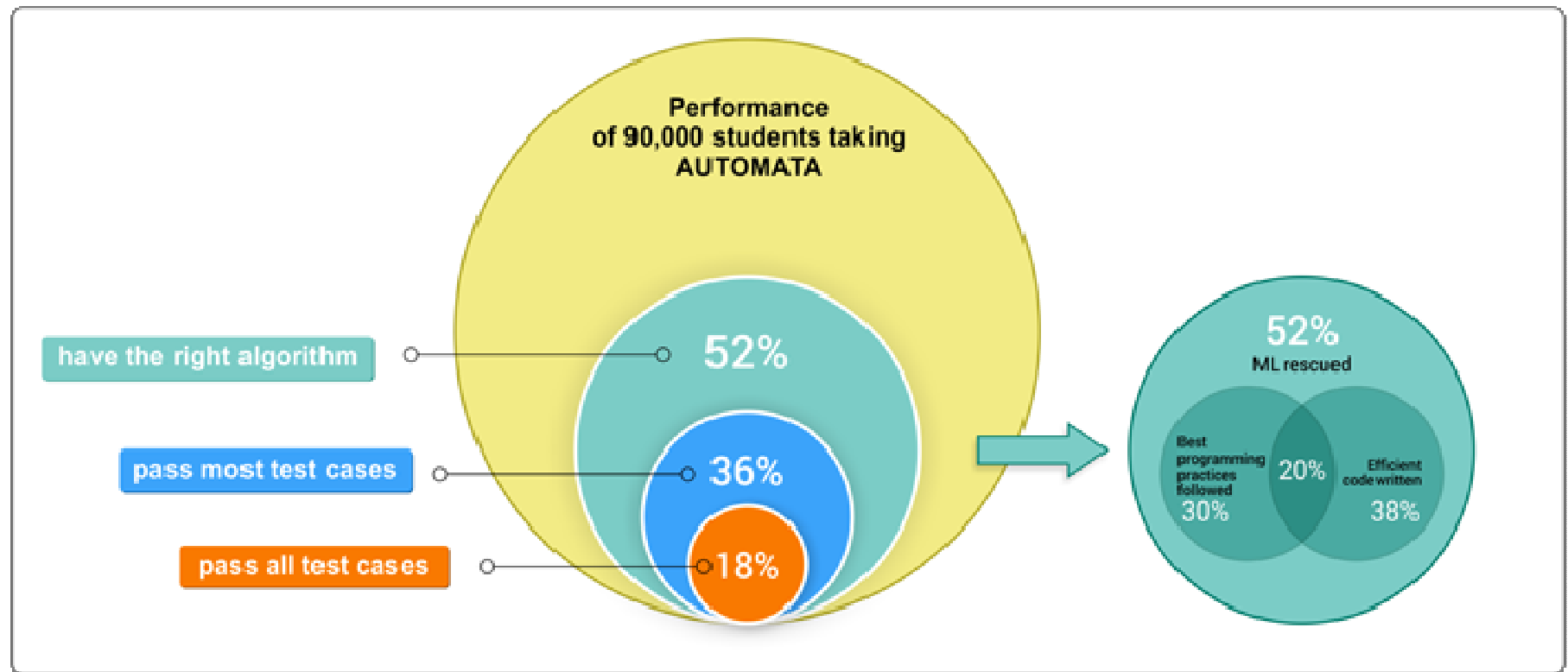
90K programs written by undergraduates in the US

New age assessments

Writing code
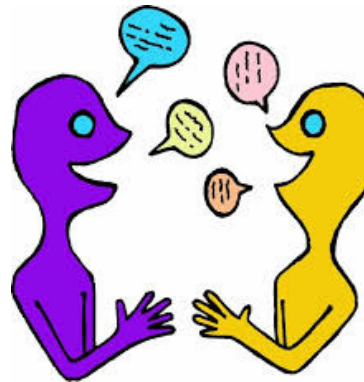
Spontaneous speech

Motor skills

Predicting hiring outcome

Labor market insights

SVAR

- **A first-of-its-kind IVR based Spoken English Assessment tool**
- **Built in-house**
- **Live since 2010**

SVAR

SVAR
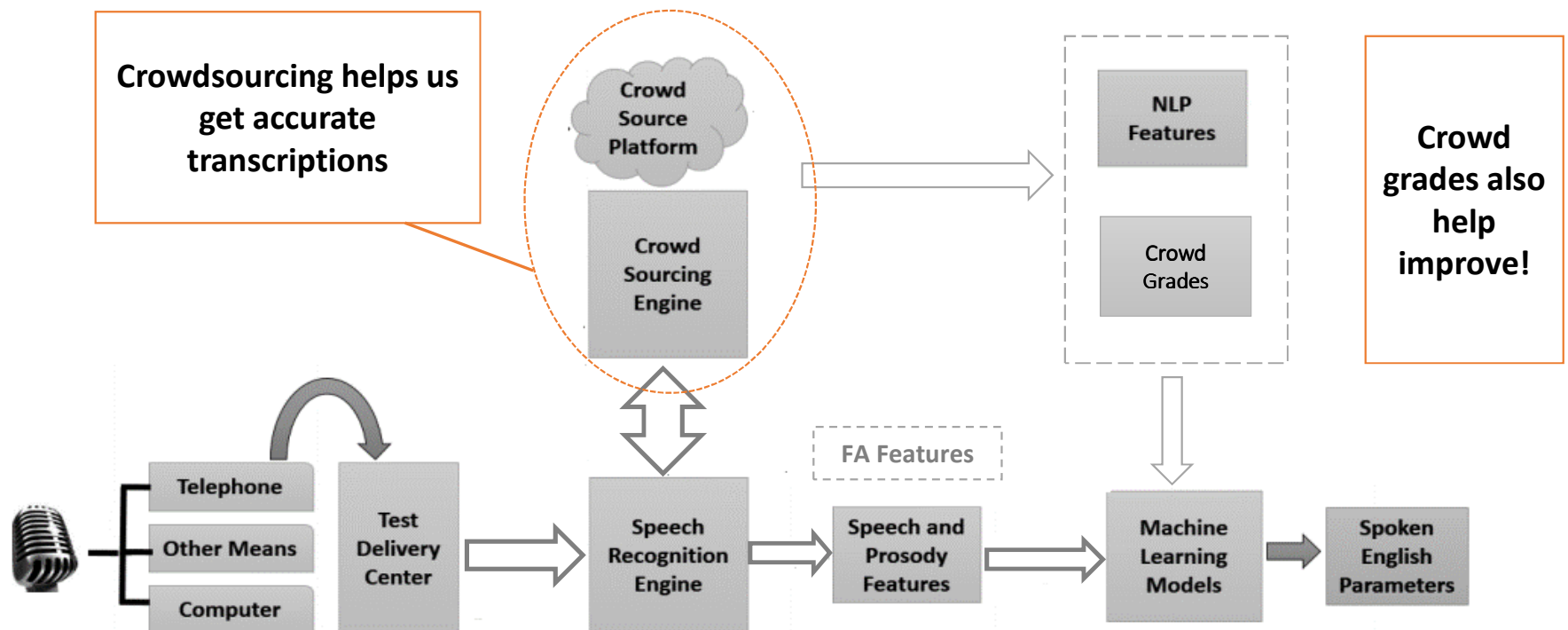
Results

Work published at KDD 2015 and ACL 2015
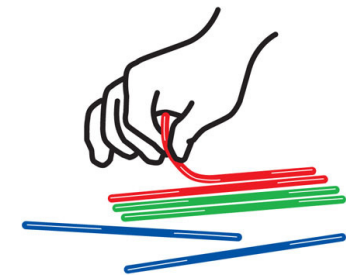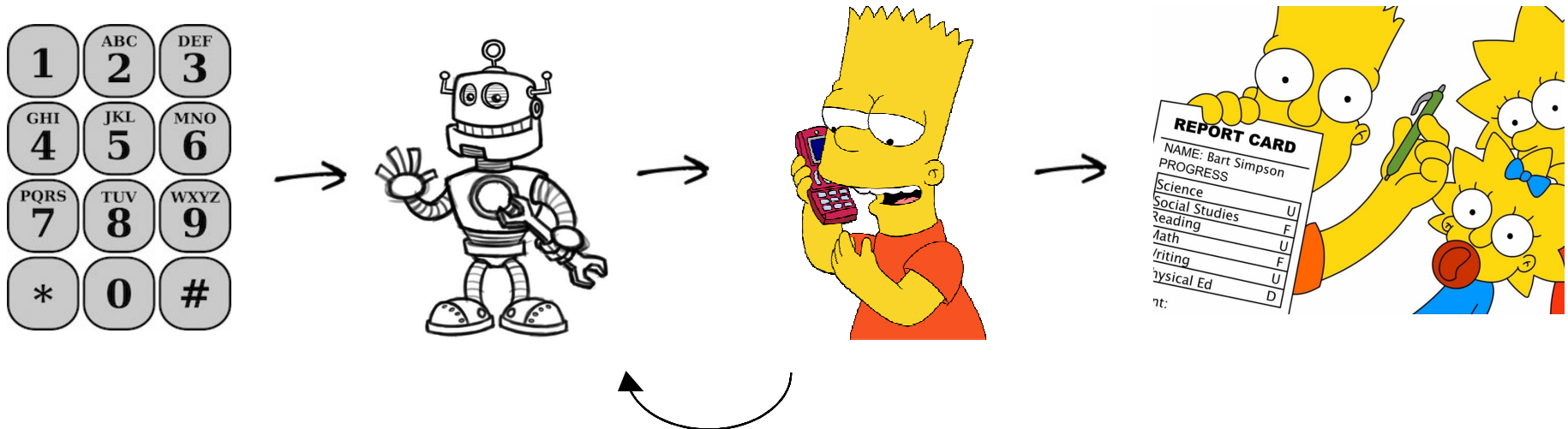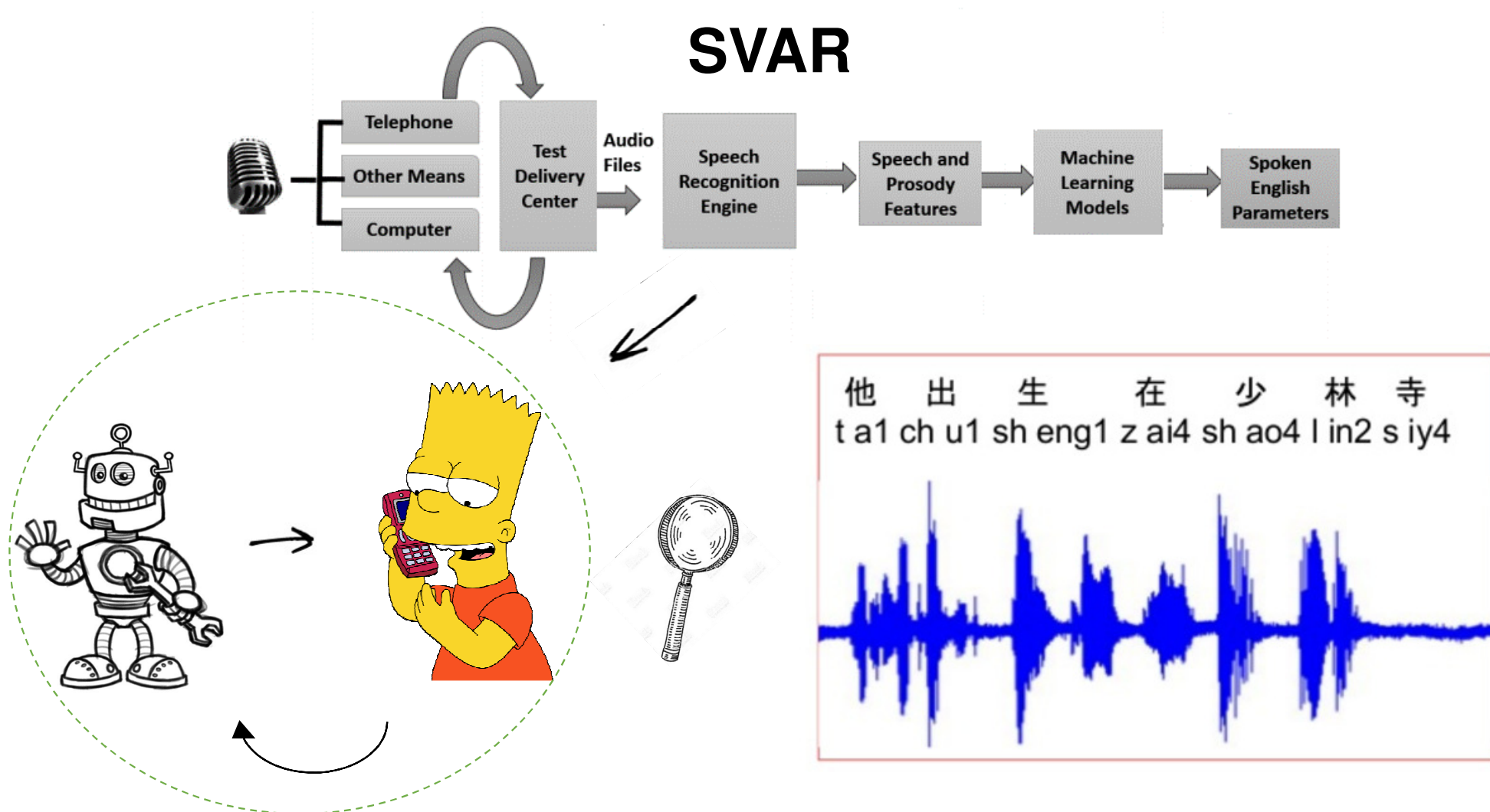
Technique	Model Code	Feature Type	Train r	IN Set Validation r	PH Set Train r	Validation r
Ridge Regression	RR-1	RS/LR	0.42	0.51	0.47	0.44
	RR-2	Pure ML	0.46	0.48	0.60	0.54
	RR-3	Crowd Grades	0.61	0.67	0.61	0.71
	RR-4	ML-CS	0.64	0.70	0.77	0.60
	RR-5	All	0.74	0.74	0.76	0.79

Crowdsourced transcriptions + Crowd grades outperforms all other methods

Accuracy nears inter expert agreement (~0.8).

New age assessments
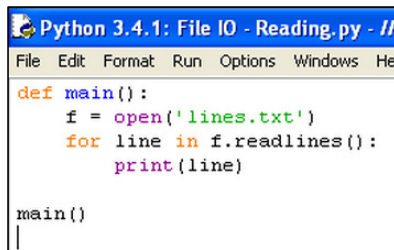
Writing code

Spontaneous speech
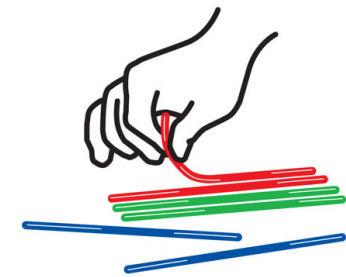
Motor skills

Predicting hiring outcome

Labor market insights

Many standardized tests predict performance of **knowledge workers**.

No automated tests for **vocational job workers**
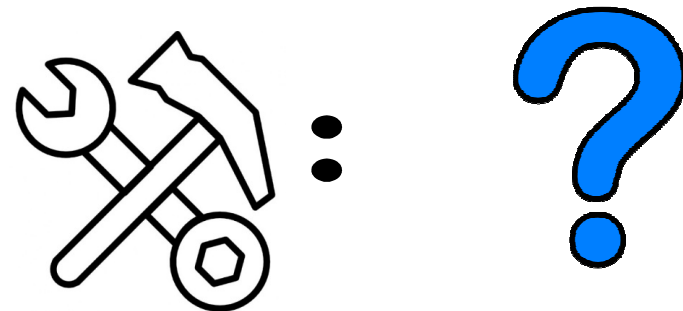
350/1650 jobs

require motor-skills

O'Connor Tweezer Dexterity Test

GATB Manual Dexterity Scale

350/1650 jobs

require motor-skills

o*net™

ML for skill identification – Basic setup

skills

Input

↓

ML magic!

↓

Prediction → **Ground truth**

Metric

Burger Rubric

Wow!
Good
Almost
Oops
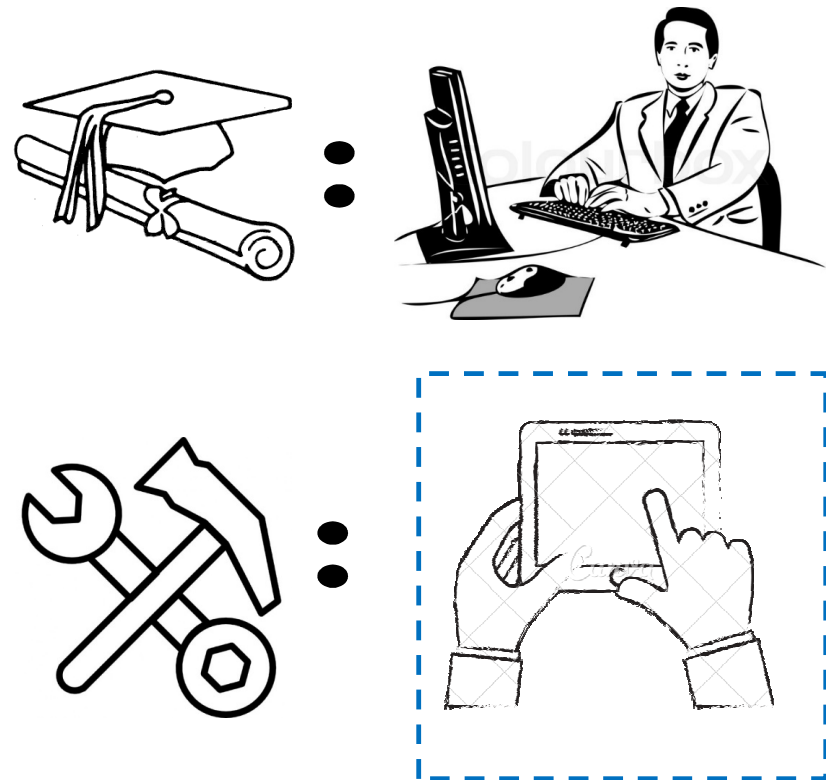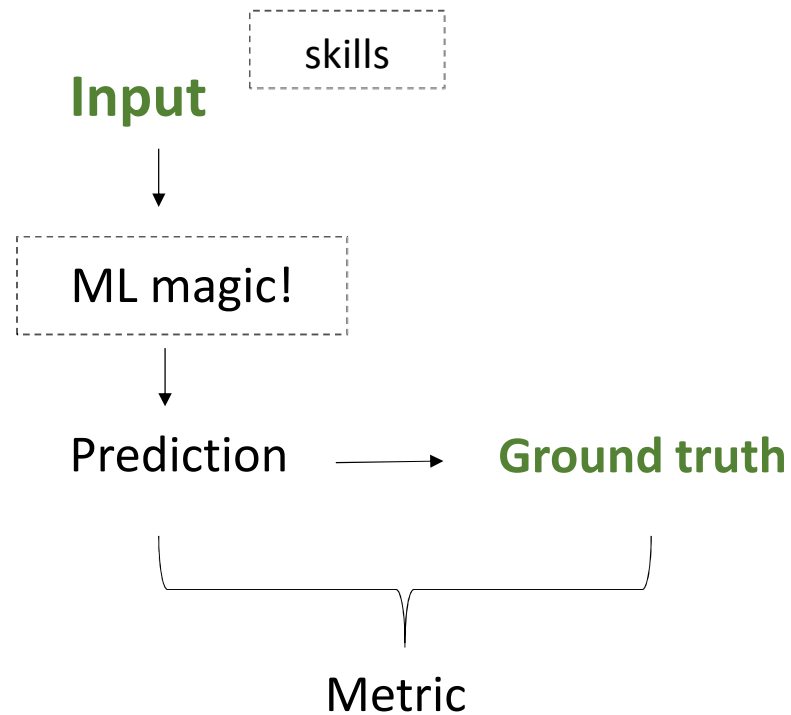
Which motor skills to measure?

```
┌─────────────────────┐      ┌─────────────────┐
│ Fleishman's Taxonomy │──────│  Motor Skills   │
└─────────────────────┘      └─────────────────┘
```
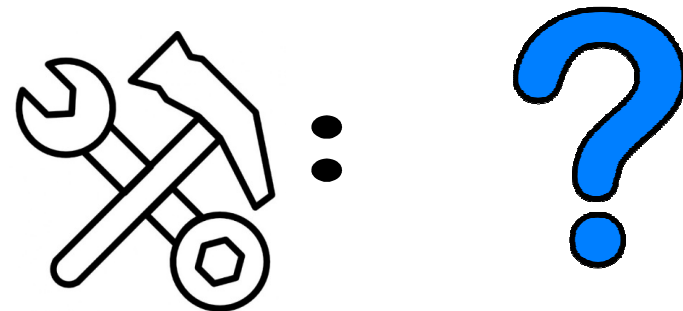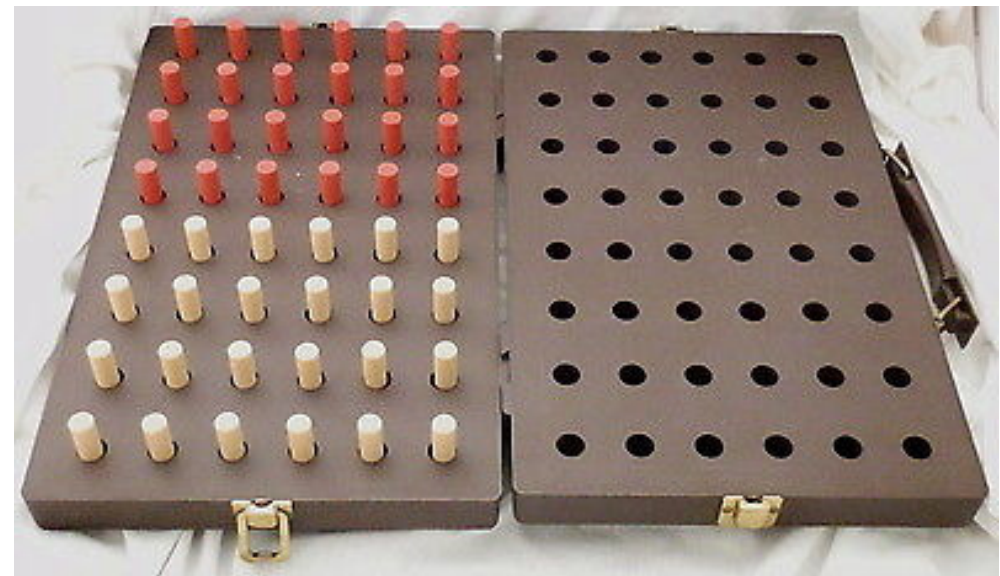
Finger Dexterity	Reaction Time	Multi-limb Coordination

Decent job success correlation*

Finger Dexterity	Manual Dexterity	Wrist-Finger Speed	Multi-limb Coordination
0.19	**0.22**	**0.18**	**0.14**
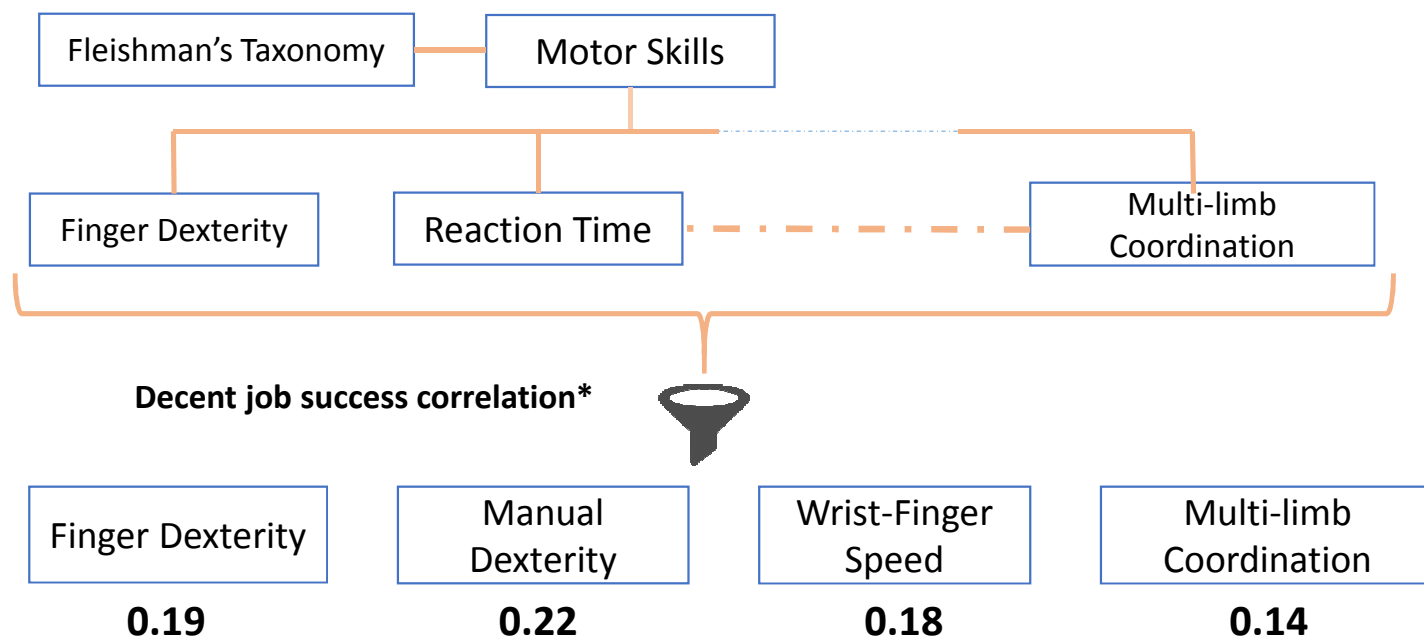
*J. J. McHenry and S. R. Rose. Literature review: Validity and potential usefulness of psychomotor ability tests for personnel selection and classication. Technical report, DTIC Document, 1988

Trace Path A & B
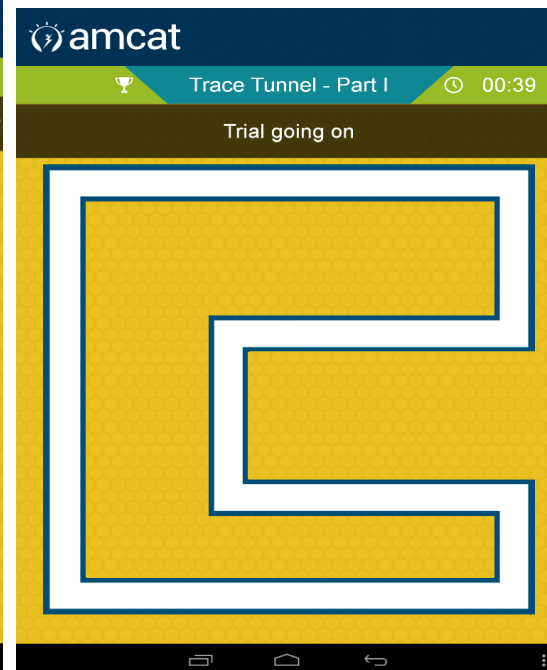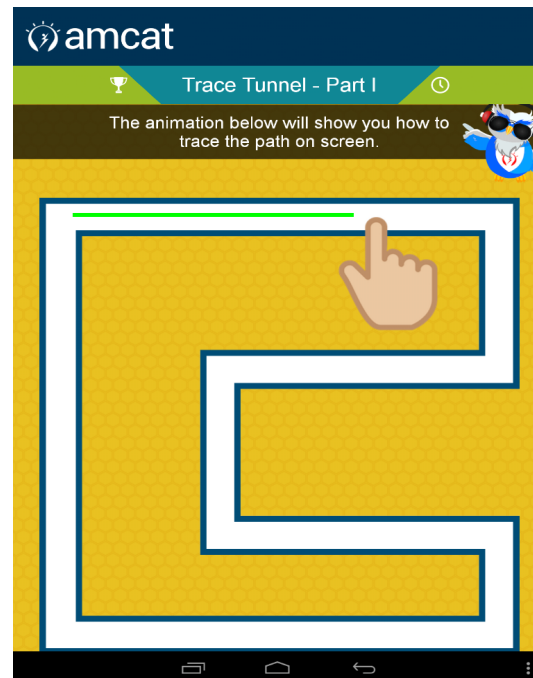
Requires **arm/hand, elbow and shoulder** movements.

Fit and Resize Circle
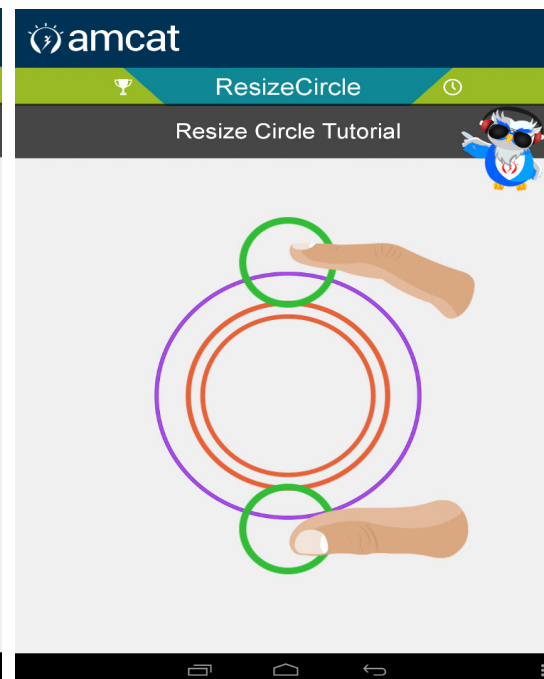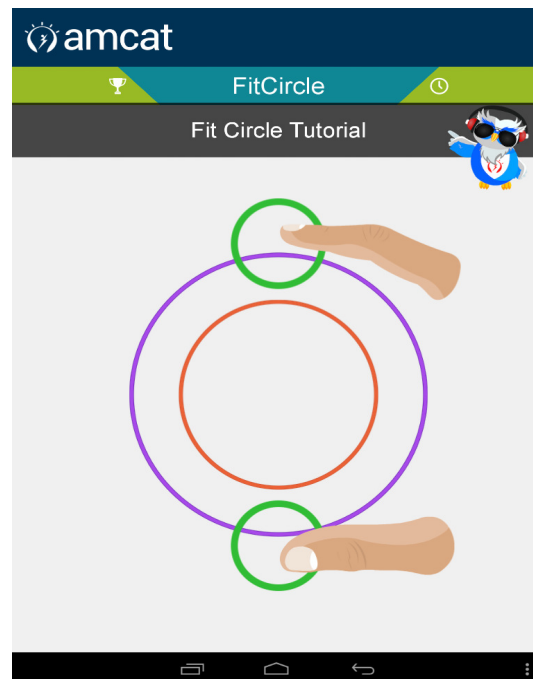
Requires **finger and wrist** movements.

Roll Ball A & B
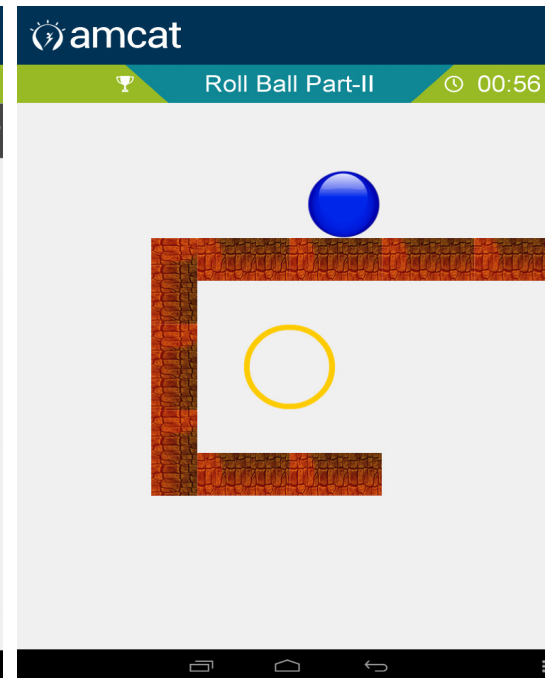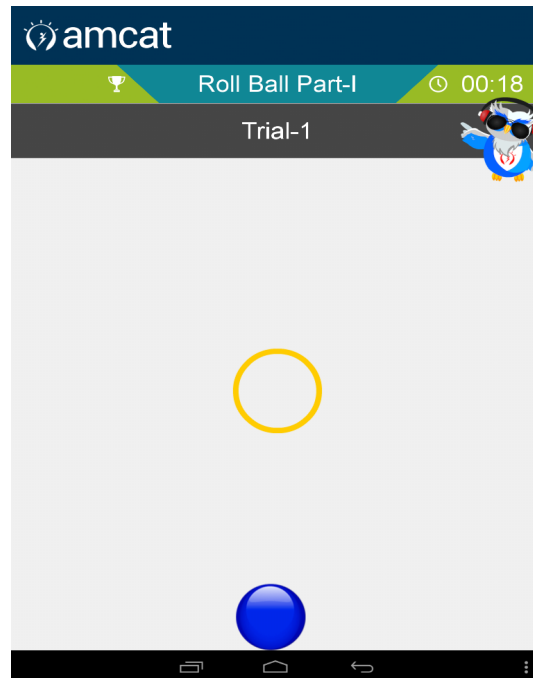
Requires **coordinated arm and shoulder** movements.

Draw an Arc A & B
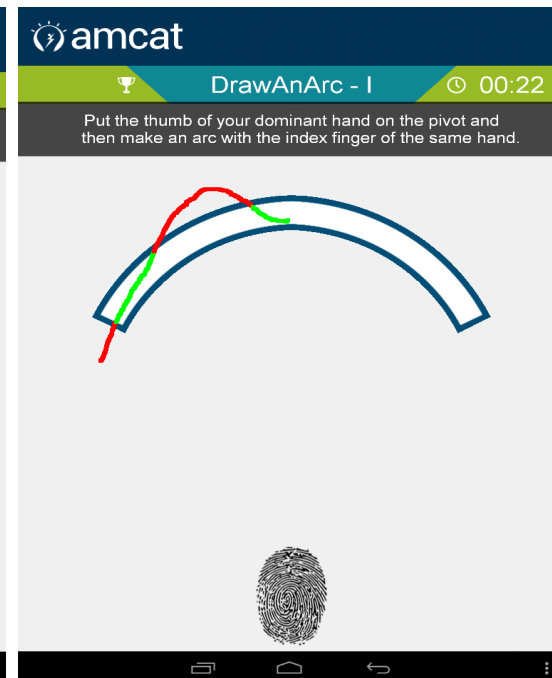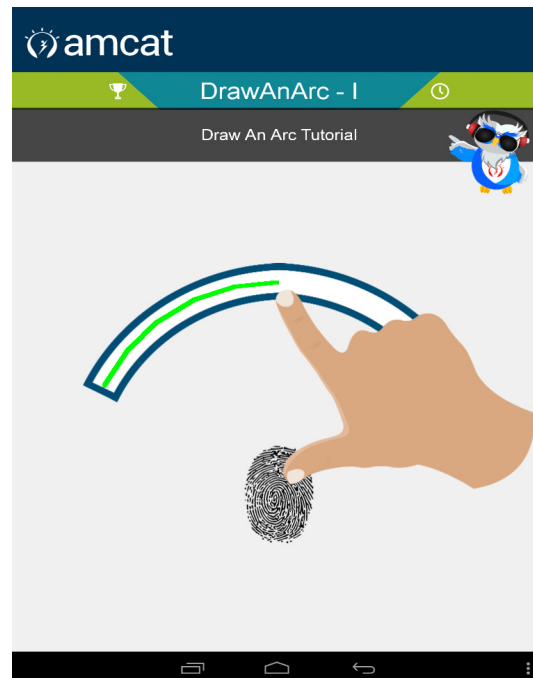
Requires **wrist and finger** movements.

Results

Work published at UbiComp 2016

**Tablet
vs
Job performance**

0.19 to 0.38

**Manual
vs
Job performance**

0.19 to 0.33

- For each performance metric, **atleast one tablet** score shows a significant correlation.

- There is maximum support for **Trace app** which shows the highest correlation with job performance in **five out of the seven** metrics.

Outcome predictions using test scores

Client: What's the right cut-off to have on our test scores?

Conditions:

- I get to choose roughly how many candidates I want to talk to

- Your recommendations give the least type-1 and type-2 errors

- Your recommendations make sense

Outcome predictions using test scores

Not allowed

L > 400
E + 0.4*L > 870
E + 0.7*Q > 980

Class 1

Class 0

Logical

English

Outcome predictions using test scores

1	• 0.42* English + Logical > 722
2	• Logical > 464

OR

1	• Computer > 360
2	• 0.47*English + 0.98*Computer + Logical > 1196

• Logical ability needs to above the mean.
• Above a minimum benchmark, lack of logical Ability can be offset by high English comprehension skills

• The first filter eliminates bottom 28% candidates in Programming.
•Above a minimum benchmark, lack of Programming knowledge can be offset by high Logical Ability and English skills.

Insights into the labor market

Who gets a job?

Variable	coefficient	p-value	Unit of change*	Odds (e^(coefficient*unit))
English score	**0.0026**	**0.00**	**100**	**1.29**
Quantitative Ability score	0.0003	0.38	100	1.03
Logical Ability score	**0.0014**	**0.01**	**100**	**1.15**
Domain Percentile	**0.0037**	**0.04**	**10**	**1.04**
10th class percentage	0.0083	0.16	10	1.09
12th class percentage	-0.0086	0.08	10	0.92
College Percentage	**0.0151**	**0.01**	**10**	**1.16**
Gender	-0.0442	0.60	1	0.96
Tier of college	**-0.1270**	**0.03**	**1**	**0.88**
Branch of study	**0.1515**	**0.05**	**1**	**1.16**
Tier of city	-0.0026	0.96	1	1.00
Openness to Experience score	-0.0253	0.58	1	0.98
Extraversion + Agreeableness Score	0.0001	1.00	1	1.00
Polychronicity score	0.0175	0.66	1	1.02
Constant	-4.1389	0.00		

Merit

Bias

A candidate with an AMCAT English & Logical score higher by 100 points each and domain percentile up by 10 points has **54%** higher odds to get a job.

A candidate from a tier 2 campus has 12% (25%) lower odds and tier 3 campus has 24% (33%) lower odds to get a job even if he/she has the equal merit.

Insights into the labor market

What salary does one get?

Variable	Coefficient	p-value	Unit of change*	Change in salary (in INR thousands)
English score	**0.0013**	**0.01**	**100**	**13325**
Quant Ability score	-0.0002	0.64	100	-1883
Logical Ability score	**0.0014**	**0.04**	**100**	**14164**
Domain Percentile	0.0013	0.57	10	1322
10th class percentage	0.0067	0.39	10	6709
12th class percentage	0.0007	0.91	10	732
College Percentage	**0.0220**	**0.00**	**10**	**21979**
Gender	0.1563	0.12	1	15631
Tier of college	**-0.3326**	**0.00**	**1**	**-33257**
Branch of study	-0.0376	0.69	1	-3762
Tier of city	0.0341	0.63	1	3407
Openness to Experience score	0.0157	0.78	1	1570
Extraversion + Agreeableness Score	-0.0393	0.28	1	-3934
Polychronicity score	0.0105	0.83	1	1049
Constant	-0.3968	0.52	1	

Merit

Higher English & Logical scores by 100 points each and higher college percentage by 10%, translate into a higher annual salary by Rs. 49500 (15% of standard package)

Bias

A candidate from a tier 2 campus has **12% (25%) lower odds** and tier 3 campus has **24% (33%) lower odds** to get a job even if he/she has the **equal merit.**

Information Asymmetry

Seva Setu

Mother care

Each one, reach one

Auto Aid

Monitor & improve Village Health and Nutrition Days (VHNDs)

Child care

NRC rehabilitation of SAM children

Citizen care

Pension schemes

Aam Sabha

Audits of existing state machinery
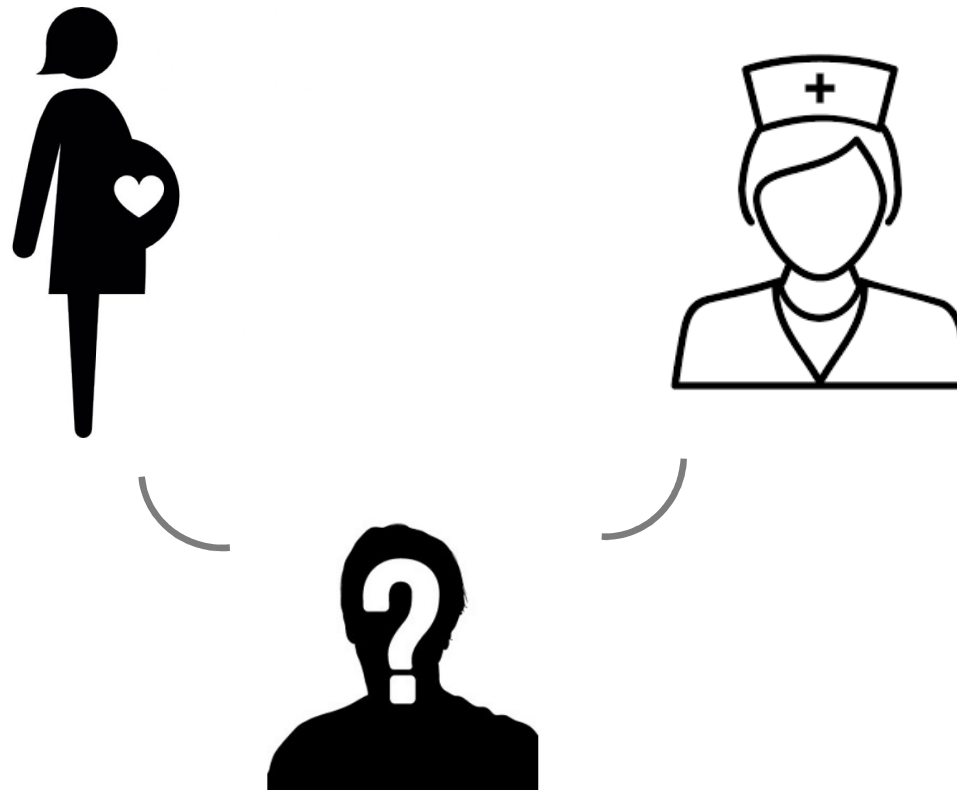RTIs, Publications

Skills training

Training and employing

SevaSetu

The problem

Rapid survey on Children tells us 10 million kids in Bihar are malnourished

Once a month government program where a nurse visits each village. Not personalized enough

MMR in Bihar is 219 per 100,000 births

IMR in Bihar is 43 per 1000 births

The problem

Each One, Reach One

sevasetu.org/mother_care

sevasetu.org/mother_care

Takeaway

Bunch of hard problems out there for us to deal with

Key is to identify the right problem and articulate it right

Hope that the articulation can leverage technology to make it sustainable, easily implementable