

# Finding a Jack-of-All-Trades:

An Examination of Transfer Learning in Text Comprehension

Kadlec, R., Bajgar, O., Hrinčár, P., Kleindienst, J.

IBM Watson, Prague lab

based on <https://openreview.net/pdf?id=rJM69B5xx>

Generalization is the key

# Cloze style questions

## Children's Book Test (Hill et al 2015)

"Well, Miss Maxwell, I think it only fair to tell you that you may have trouble with those boys when they do come. Forewarned is forearmed, you know. Mr. Cropper was opposed to our hiring you. Not, of course, that he had any personal objection to you, but he is set against female teachers, and when a Cropper is set there is nothing on earth can change him. He says female teachers can't keep order. He 's started in with a spite at you on general principles, and the boys know it. They know he'll back them up in secret, no matter what they do, just to prove his opinions. Cropper is sly and slippery, and it is hard to corner him."

"Are the boys big ?" queried Esther anxiously.

"Yes. Thirteen and fourteen and big for their age. You can't whip 'em -- that is the trouble. A man might, but they'd twist you around their fingers. You'll have your hands full, I'm afraid. But maybe they'll behave all right after all."

Mr. Baxter privately had no hope that they would, but Esther hoped for the best. She could not believe that Mr. Cropper would carry his prejudices into a personal application. This conviction was strengthened when he overtook her walking from school the next day and drove her home. He was a big, handsome man with a very suave, polite manner. He asked interestedly about her school and her work, hoped she was getting on well, and said he had two young rascals of his own to send soon. Esther felt relieved. She thought that Mr. Baxter had exaggerated matters a little.

S: 1 Mr. Cropper was opposed to our hiring you .  
2 Not , of course , that he had any personal objection to you , but he is set against female teachers , and when a Cropper is set there is nothing on earth can change him .  
3 He says female teachers ca n't keep order .  
4 He 's started in with a spite at you on general principles , and the boys know it .  
5 They know he 'll back them up in secret , no matter what they do , just to prove his opinions .  
6 Cropper is sly and slippery , and it is hard to corner him . ''  
7 `` Are the boys big ? ''  
8 queried Esther anxiously .  
9 `` Yes .  
10 Thirteen and fourteen and big for their age .  
11 You ca n't whip 'em -- that is the trouble .  
12 A man might , but they 'd twist you around their fingers .  
13 You 'll have your hands full , I 'm afraid .  
14 But maybe they 'll behave all right after all . ''  
15 Mr. Baxter privately had no hope that they would , but Esther hoped for the best.  
16 She could not believe that Mr. Cropper would carry his prejudices into a personal application .  
17 This conviction was strengthened when he overtook her walking from school the next day and drove her home .  
18 He was a big , handsome man with a very suave , polite manner .  
19 He asked interestedly about her school and her work , hoped she was getting on well , and said he had two young rascals of his own to send soon .  
20 Esther felt relieved .

Q: She thought that Mr. \_\_\_\_\_ had exaggerated matters a little .

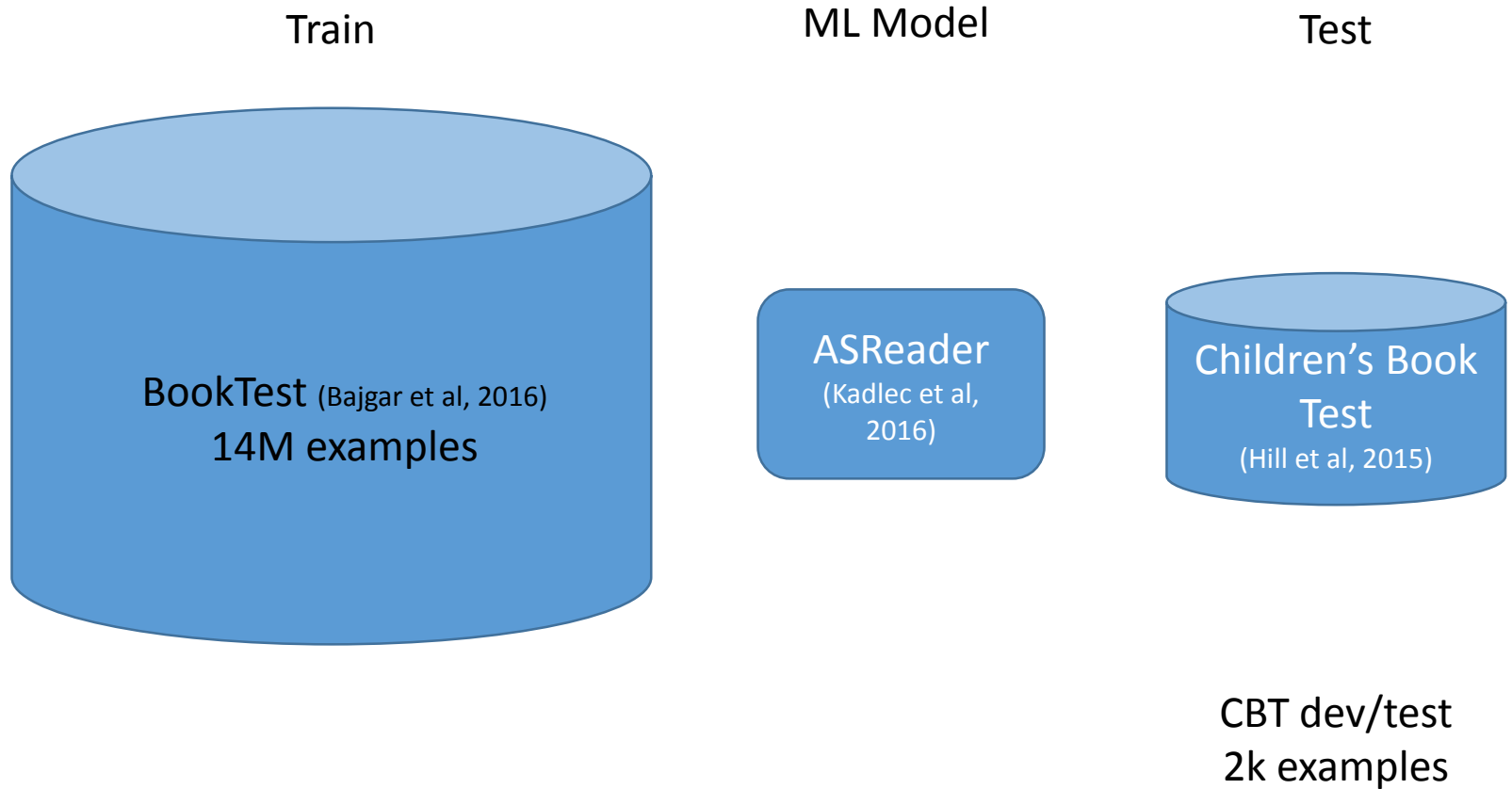
C: Baxter, Cropper, Esther, course, fingers, manner, objection, opinion, right, spite.

a: Baxter

~ 200k examples (CN+NE)

Hill, F., Bordes, A., Chopra, S., & Weston, J. (2015). The Goldilocks Principle: Reading Children's Books with Explicit Memory Representations

# Starting point

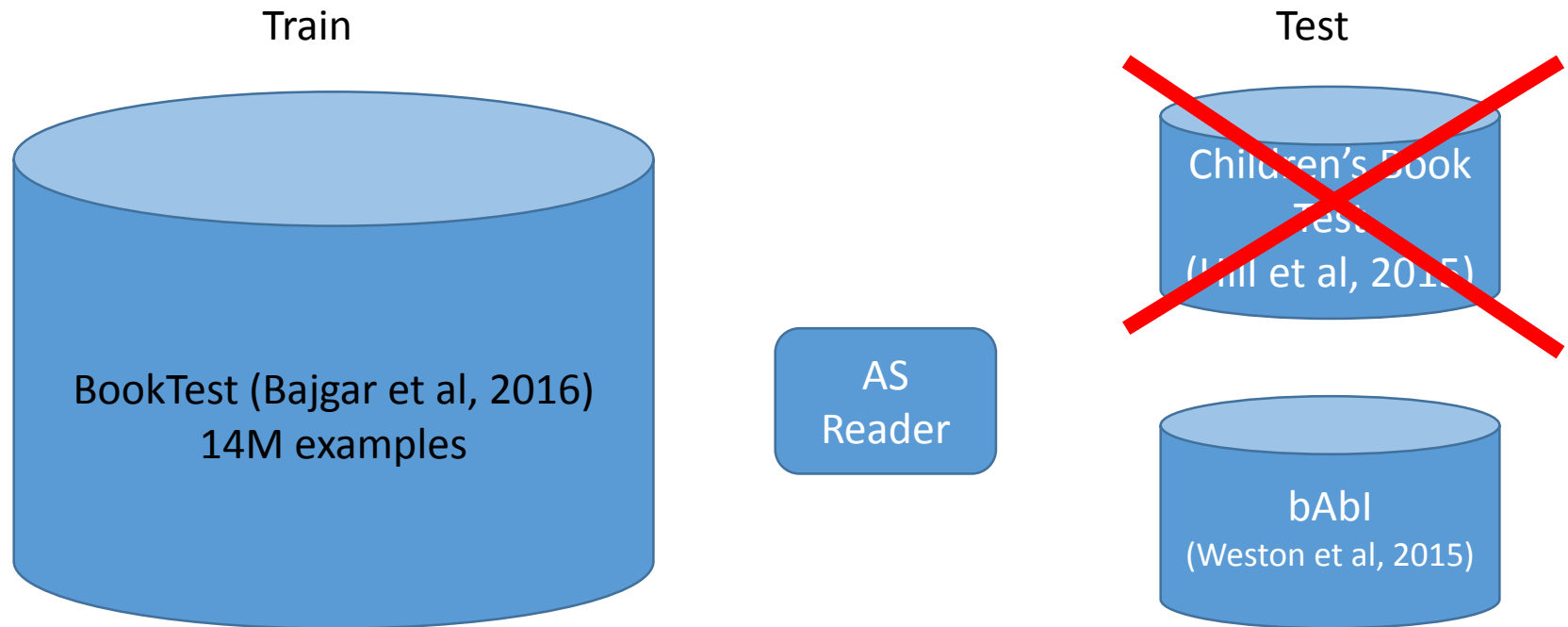


# BookTest

	Named entity		Common noun	
	valid	test	valid	test
Humans (context+query) (Hill et al., 2015)	NA	<b>81.6</b>	NA	<b>81.6</b>
AS Reader (ensemble) (Kadlec et al., 2016)	76.2	<b>71.0</b>	72.4	67.5
GA Reader (ensemble) (Dhingra et al., 2016)	75.5	71.9	72.1	69.4
EpiReader (ensemble) (Trischler et al., 2016b)	76.6	71.8	73.6	70.6
<b>IA Reader (ensemble)</b> (Sordoni et al., 2016)	76.9	<b>72.0</b>	74.1	<b>71.0</b>
<b>AoA Reader (single model)</b> (Cui et al., 2016a)	77.8	<b>72.0</b>	72.2	69.4
AS Reader (ensemble) (Kadlec et al., 2016)	82.3	<b>78.4</b>	85.7	<b>83.7</b>

Trained on more data (BookTest) than the previous models!

# Transfer learning?



Weston, J., Bordes, A., Chopra, S., Rush, A. M., van Merriënboer, B., Joulin, A., & Mikolov, T. (2015). Towards AI-Complete Question Answering: A Set of Prerequisite Toy Tasks.

# Simple testing tasks: bAbl tasks

## Task 1: Single Supporting Fact

Mary went to the bathroom.  
John moved to the hallway.  
Mary travelled to the office.  
Where is Mary? A: office

## Task 2: Two Supporting Facts

John is in the playground.  
John picked up the football.  
Bob went to the kitchen.  
Where is the football? A: playground

## Task 3: Three Supporting Facts

John picked up the apple.  
John went to the office.  
John went to the kitchen.  
John dropped the apple.  
Where was the apple before the kitchen? A: office

## Task 4: Two Argument Relations

The office is north of the bedroom.  
The bedroom is north of the bathroom.  
The kitchen is west of the garden.  
What is north of the bedroom? A: office  
What is the bedroom north of? A: bathroom

## Task 5: Three Argument Relations

Mary gave the cake to Fred.  
Fred gave the cake to Bill.  
Jeff was given the milk by Bill.  
Who gave the cake to Fred? A: Mary  
Who did Fred give the cake to? A: Bill

# Simple testing tasks: bAbl tasks

## Task 11: Basic Coreference

Daniel was in the kitchen.  
Then he went to the studio.  
Sandra was in the office.  
Where is Daniel? A: studio

## Task 12: Conjunction

Mary and Jeff went to the kitchen.  
Then Jeff went to the park.  
Where is Mary? A: kitchen  
Where is Jeff? A: park

## Task 13: Compound Coreference

Daniel and Sandra journeyed to the office.  
Then they went to the garden.  
Sandra and John travelled to the kitchen.  
After that they moved to the hallway.  
Where is Daniel? A: garden

## Task 14: Time Reasoning

In the afternoon Julie went to the park.  
Yesterday Julie was at school.  
Julie went to the cinema this evening.  
Where did Julie go after the park? A: cinema  
Where was Julie before the park? A: school

## Task 15: Basic Deduction

Sheep are afraid of wolves.  
Cats are afraid of dogs.  
Mice are afraid of cats.  
Gertrude is a sheep.  
What is Gertrude afraid of? A: wolves

## Task 16: Basic Induction

Lily is a swan.  
Lily is white.  
Bernhard is green.  
Greg is a swan.  
What color is Greg? A: white



# Can it generalize what it learned?

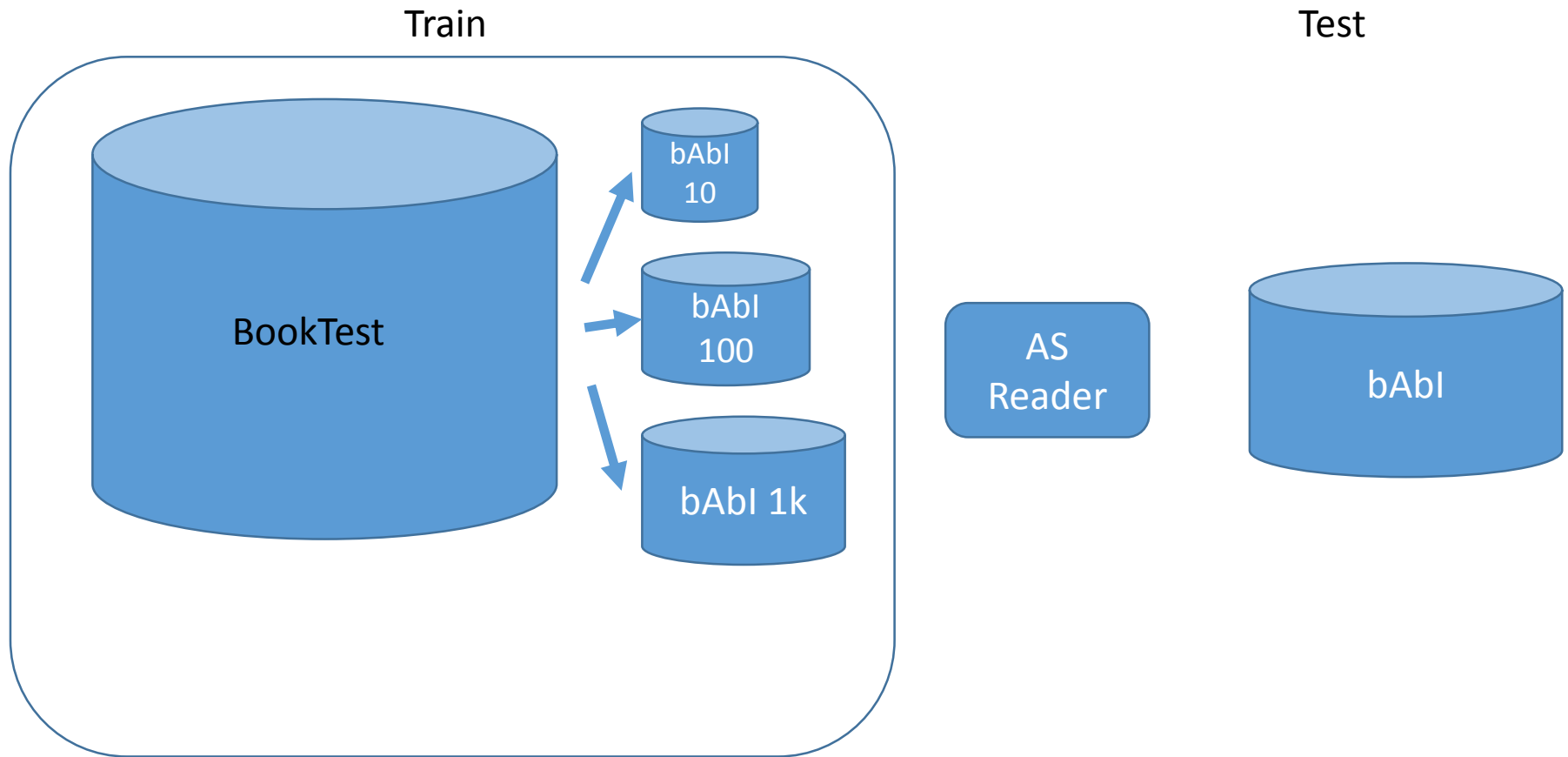
## Not really ...

Model:		Random	Rnd cand.	MemN2N (single) (PE LS RN)	MemN2N (single) (PE LS LW RN)	DMN+ (single)	ASReader	
	Train dataset	not trained	bAbI 10k	bAbI 1k	bAbI 10k	bAbI 10k	bAbI 10k	BookTest 14M
	Test dataset							
1	Single supporting fact	7.80	31.20	100.00	100.00	100.00	100.00	37.30
2	Two supporting facts	4.40	26.96	91.70	99.70	99.70	91.90	25.80
3	Three supporting facts	3.40	19.14	59.70	97.90	98.90	86.00	22.20
4	Two-argument relations	10.50	33.58	97.20	100.00	100.00	100.00	50.30
5	Three-argument relations	4.40	21.42	86.90	99.20	99.50	99.80	67.60
11	Basic coreference	6.20	30.42	99.10	99.90	100.00	100.00	33.00
12	Conjunction	6.70	27.25	99.80	100.00	100.00	100.00	30.40
13	Compound coreference	5.60	27.73	99.60	100.00	100.00	100.00	33.80
14	Time reasoning	5.00	27.82	98.30	99.90	99.80	95.00	27.60
15	Basic deduction	5.20	37.20	100.00	100.00	100.00	96.70	39.90
16	Basic induction	7.50	45.65	98.70	48.20	54.70	50.30	15.10
bAbI mean (11 tasks)		6.06	29.85	93.73	94.98	95.69	92.70	34.82

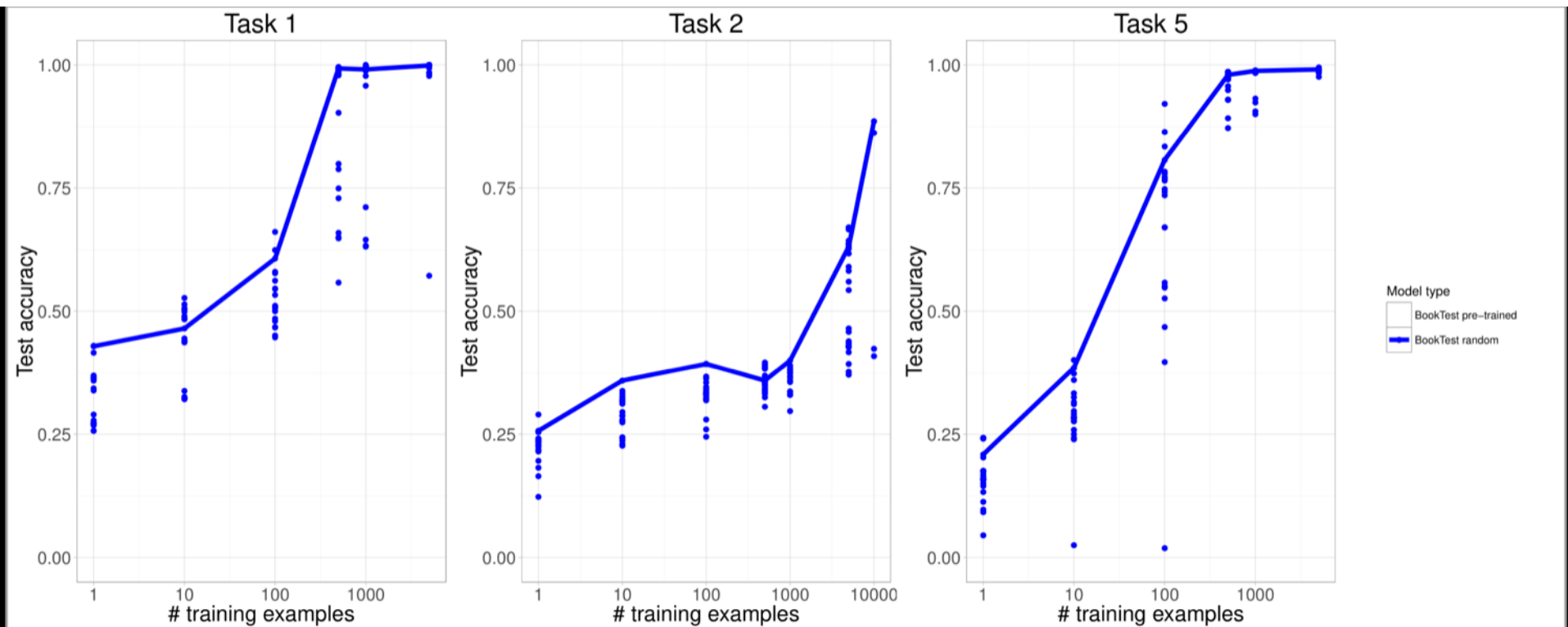


**BAD!**

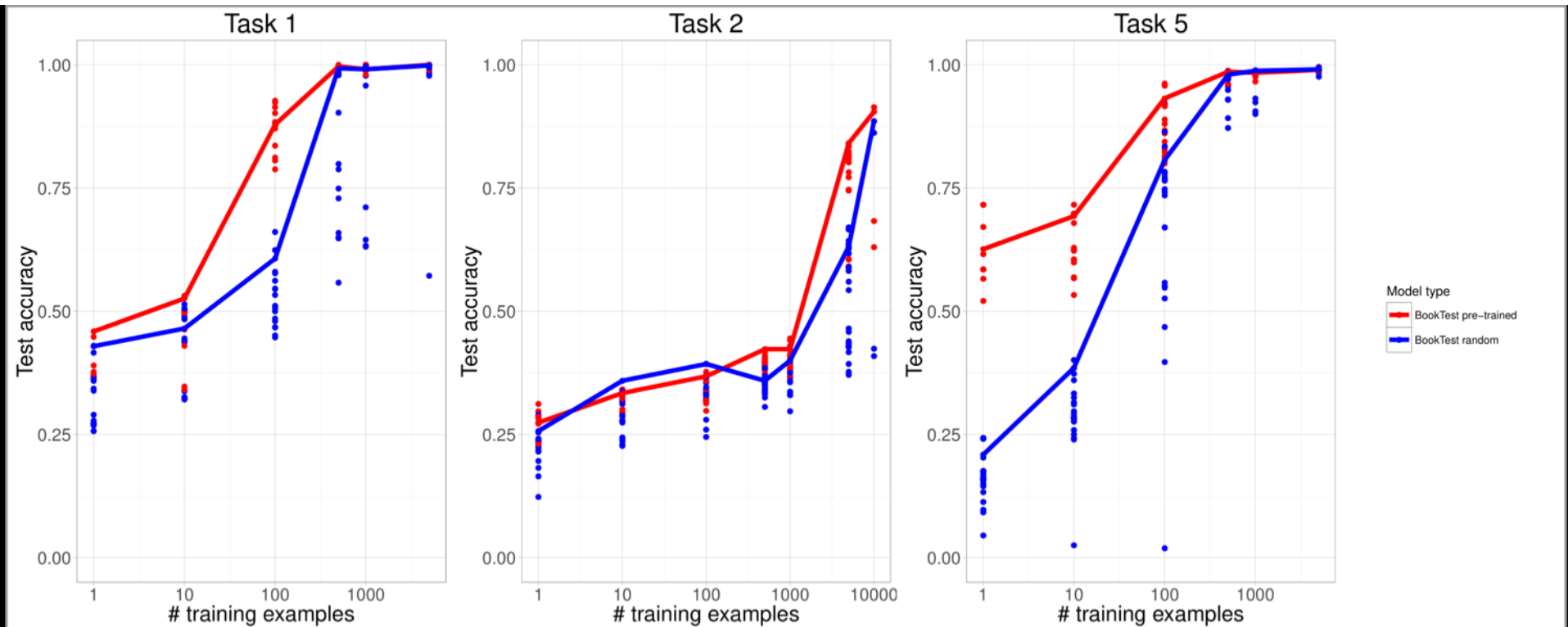
# Finetuning - bAbl

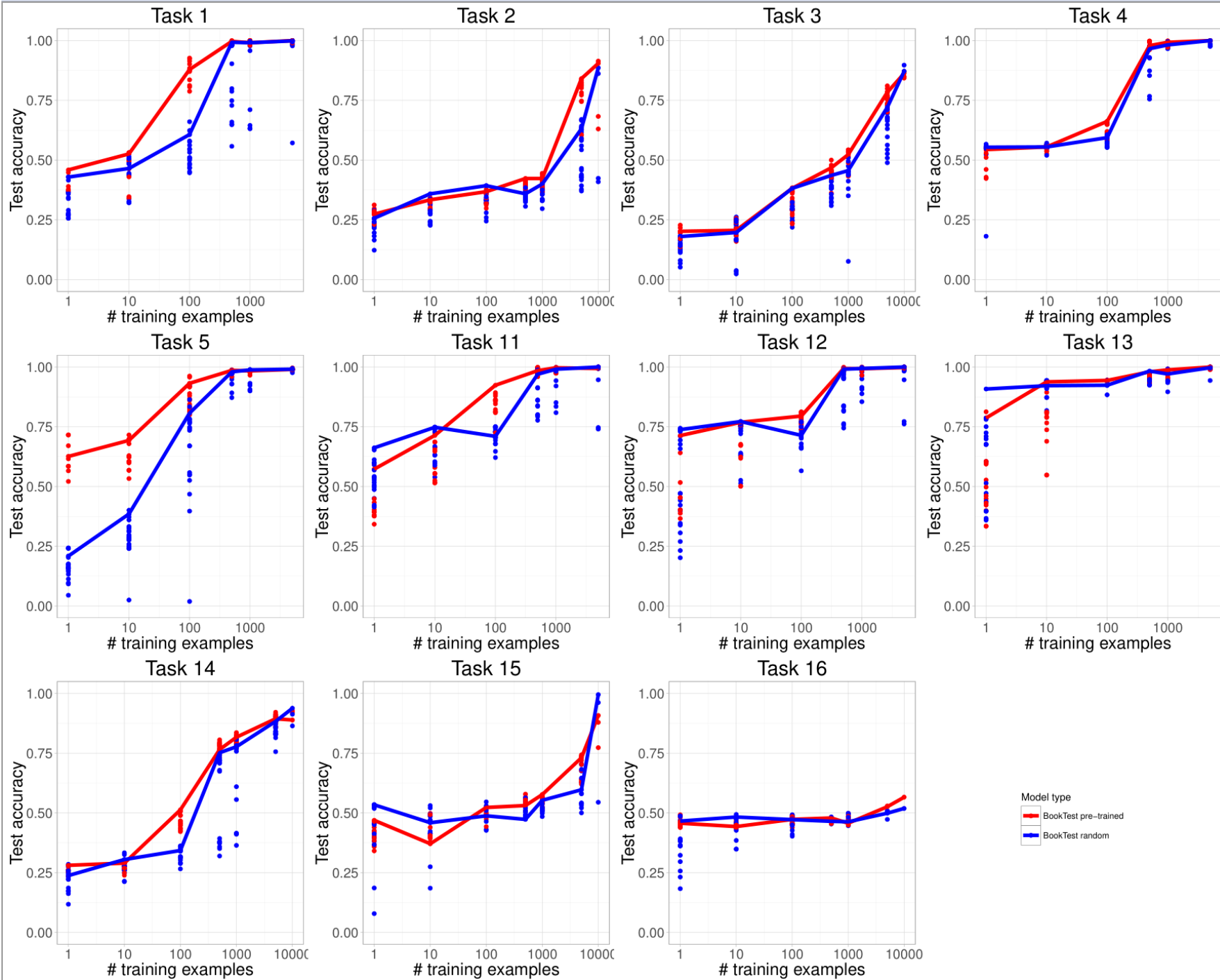


# 2nd Experiment: It does better with target-adjustment!

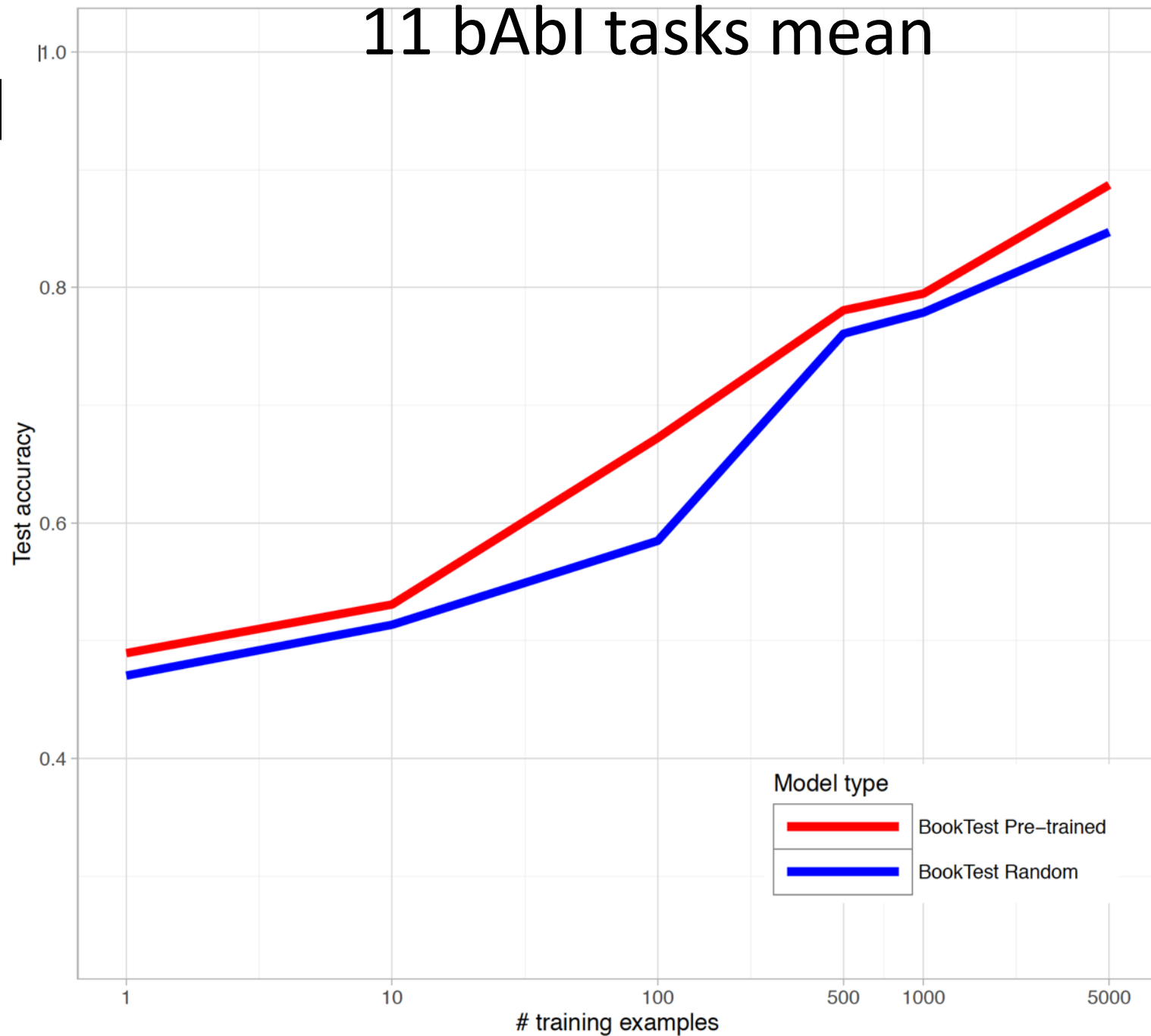


# 2nd Experiment: It does better with target-adjustment!

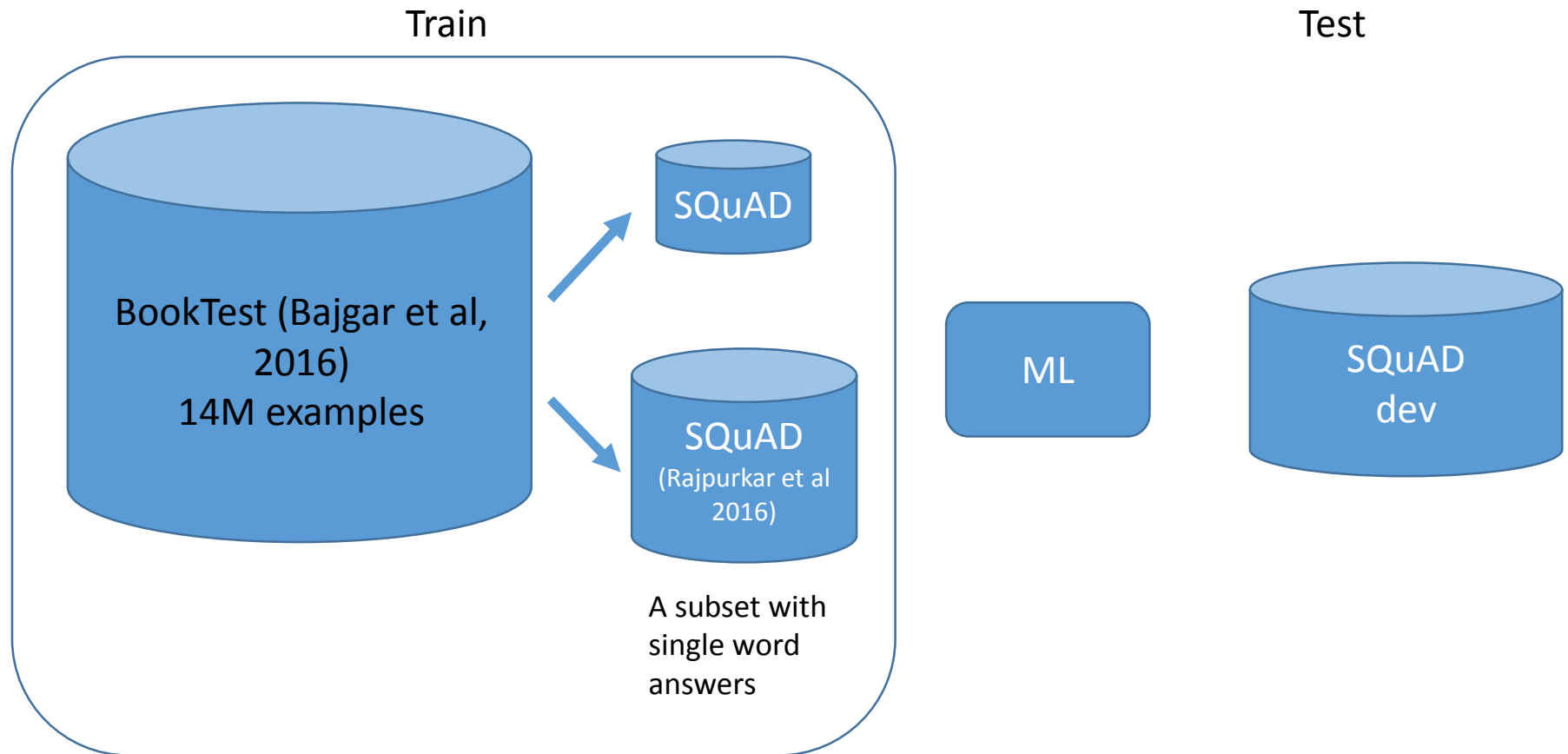




# 11 bAbI tasks mean

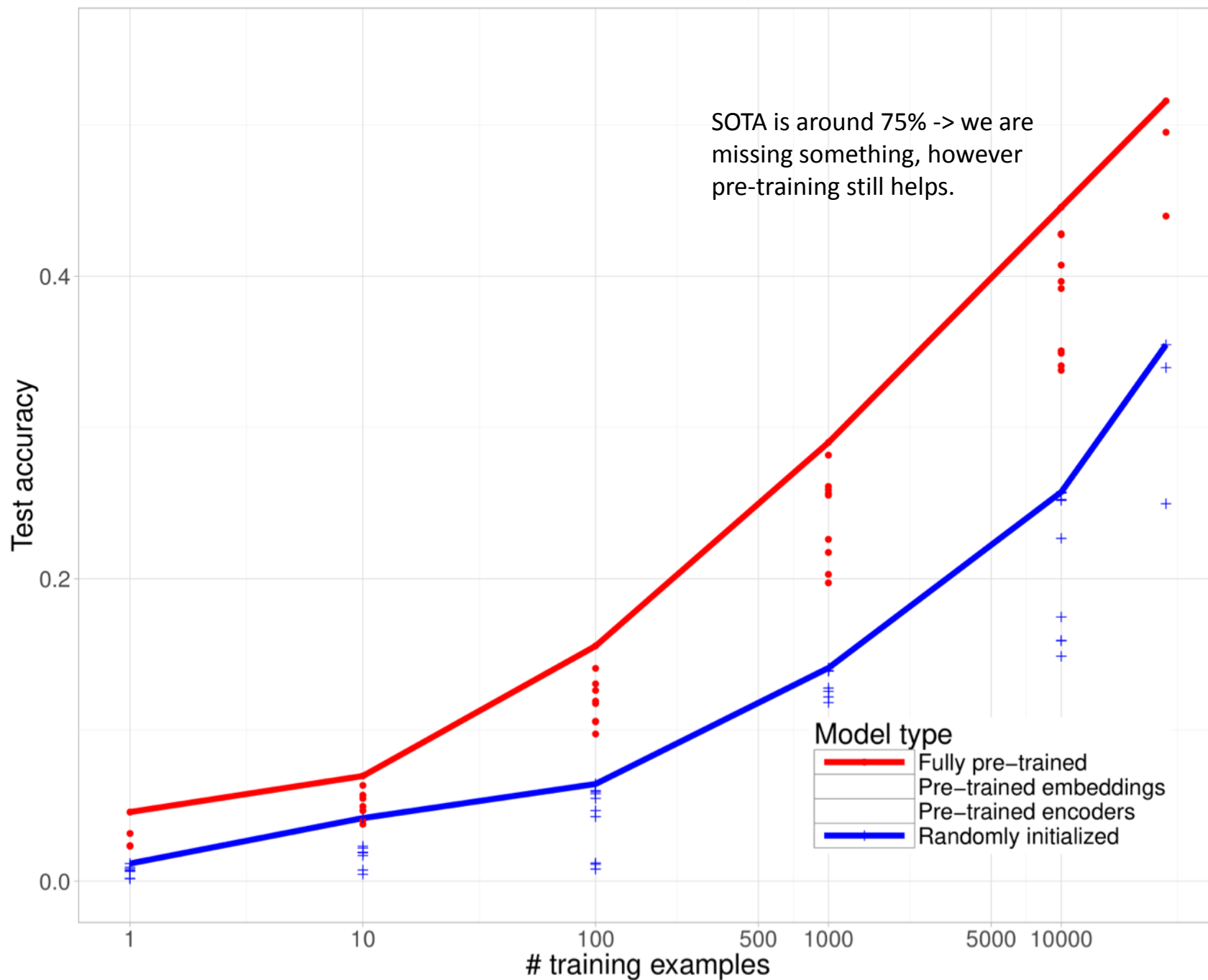


# Finetuning - SQuAD



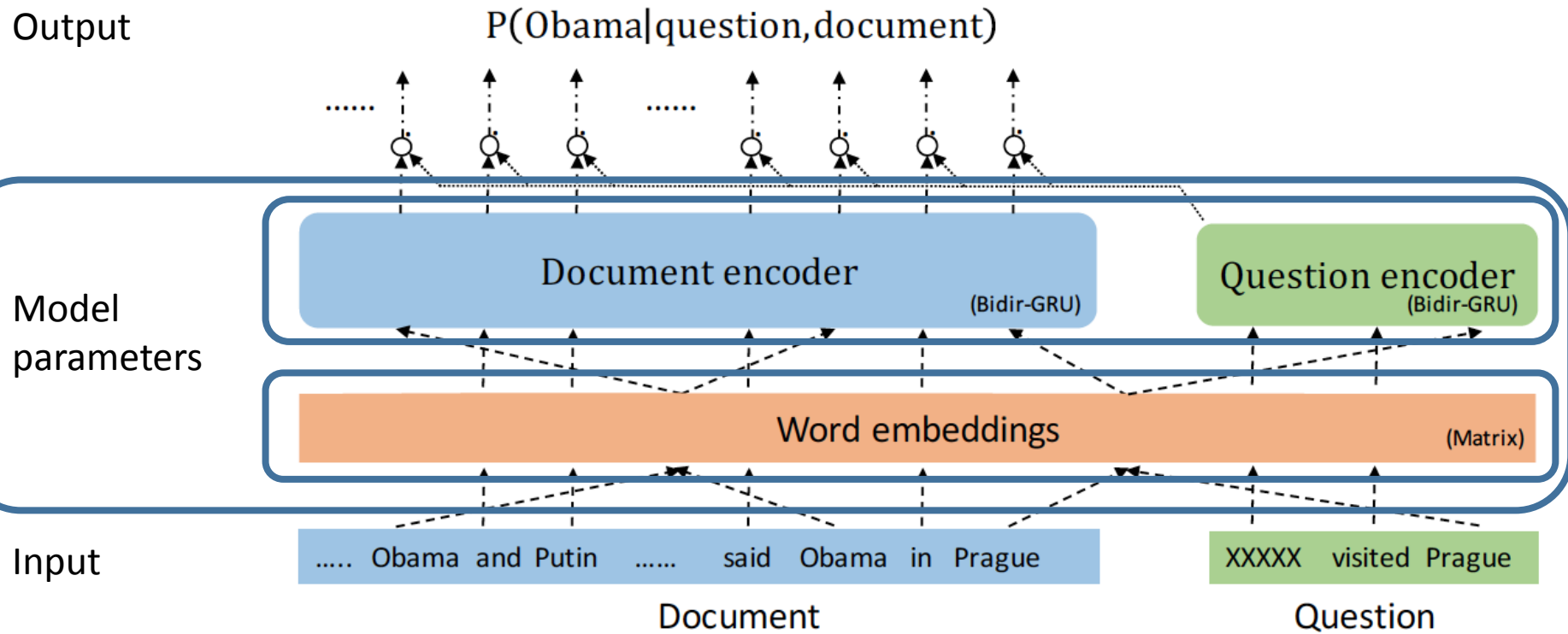
Rajpurkar, P., Zhang, J., Lopyrev, K., & Liang, P. (2016). SQuAD: 100,000+ Questions for Machine Comprehension of Text

# Accuracies on SQuAD subset

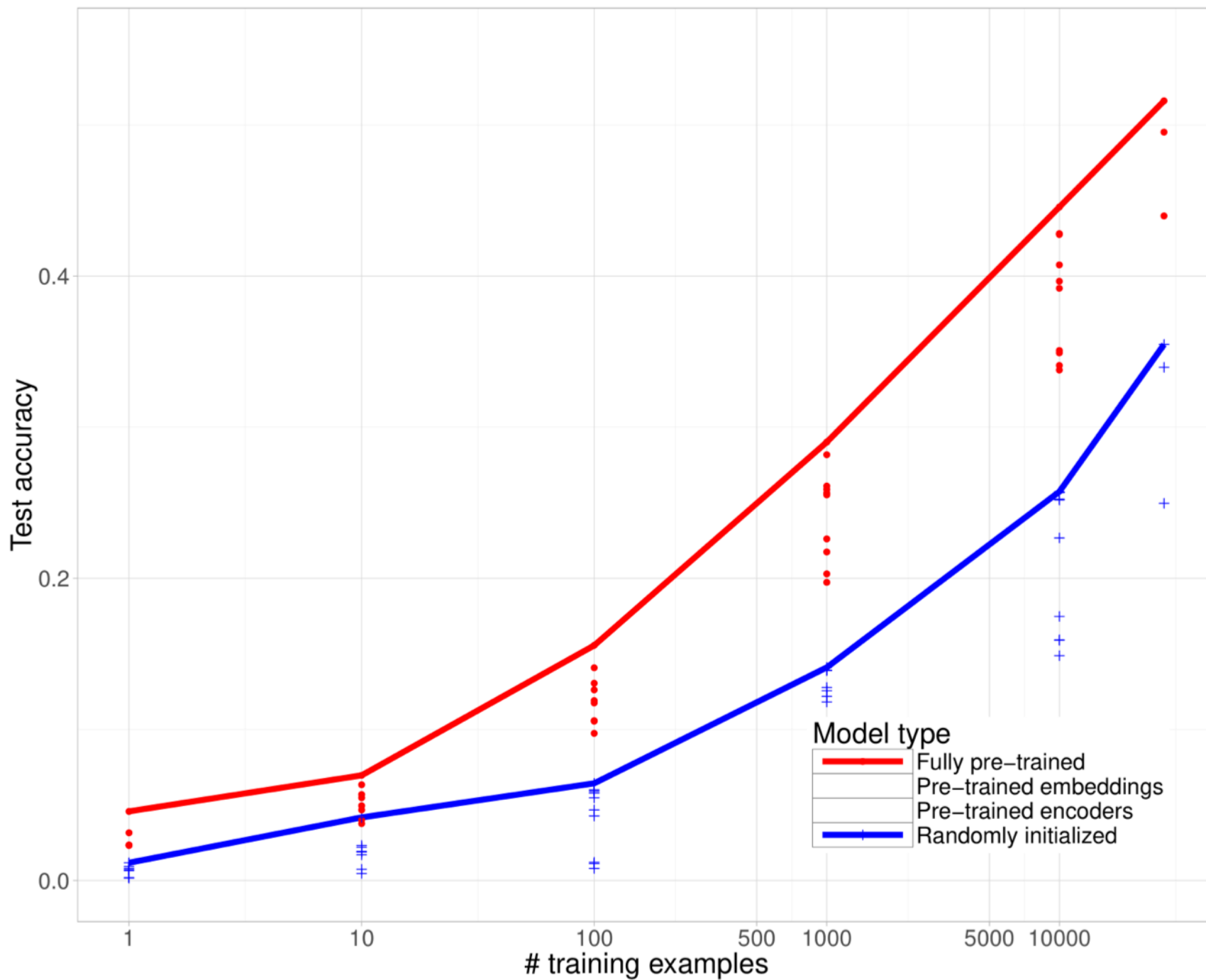




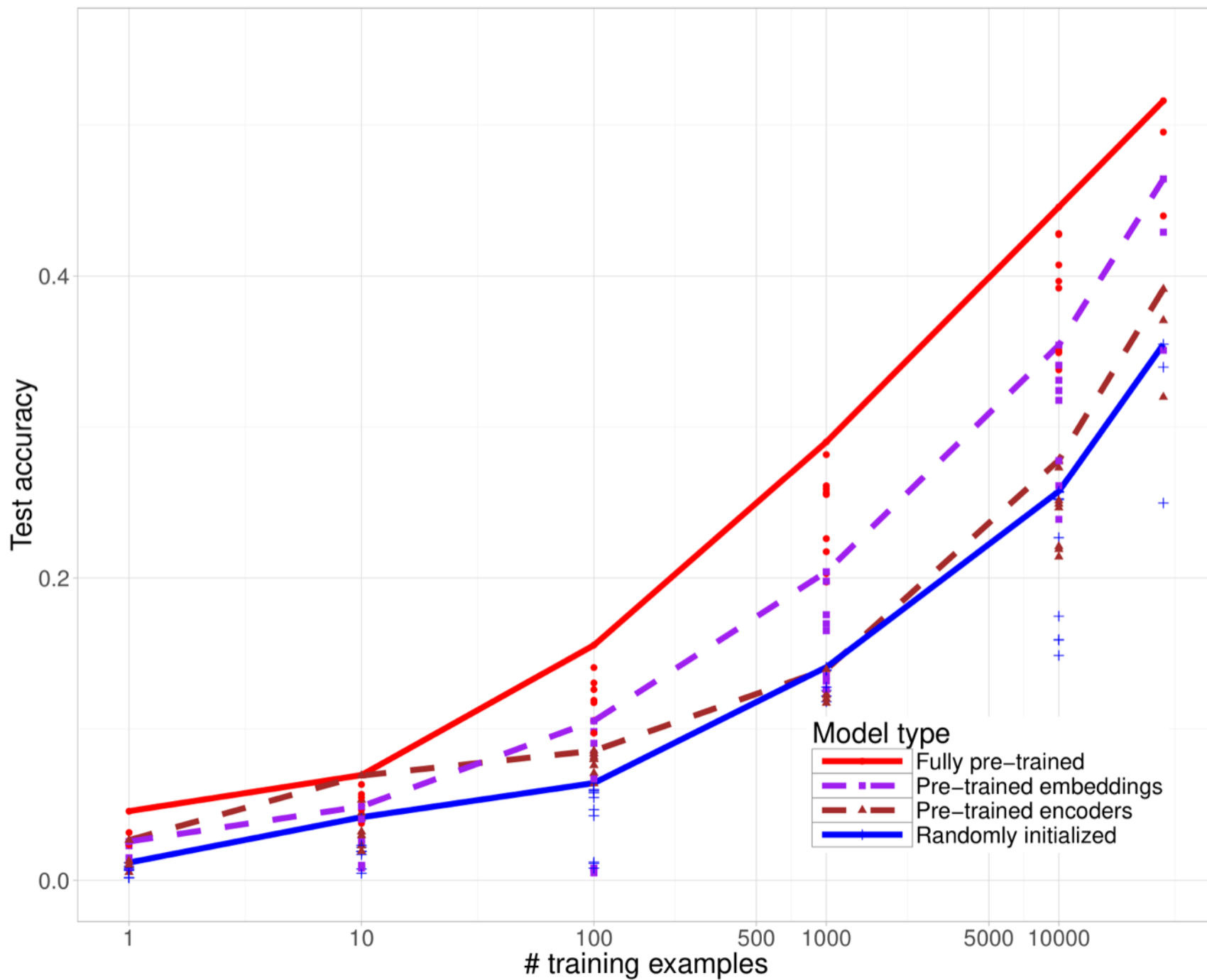
# 3rd Experiment: Where is the useful knowledge? Partial pretraining



# Accuracies on SQuAD



# Accuracies on SQuAD



# Conclusions:

- Pre-school helps
- But it's not enough!
- More work to be done!

Questions?