

# הערכה מעצבת פנים רבות לה

פיתוח מבחן אינטראקטיבי שיכול לשמש

כהערכה בשירות הלמידה

מירב אריאלי-אטאלי

אוניברסיטת פורדהם

כנס אפ"י, ירושלים, פברואר 2017

\*העבודה המוצגת פותחה כעבודת דוקטוראט בהנחיית פרופ' דויד בודסקו

- אחד האתגרים של הערכה מעצבת - בעית מוטיבציה הנבחנים
  - איך להגביר מוטיבציה במבחנים שאינם high-stakes?
- אחת התשובות =< מבחנים "מעניינים" או "חדשניים"
  - לא שאלות בדידות, אלא מטלות שדורשות מעורבות (משימה אותנטית/שילוב אנימציה או אפקטים ויזואליים/ סימולציה / משחקים / מבחנים אינטראקטיביים)

# נסו לדמיין שאתם מתנסים במבחן הבא -

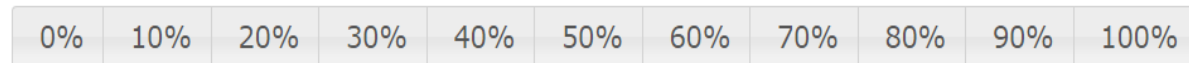
## General Knowledge Study

### Question 1 out of 40 (Level 1 - Very Easy)

What is the popular food used to carve jack-o-lanterns during Halloween?

Your answer:

My confidence is:



Submit Answer

## General Knowledge Study

### Question 1 out of 40 (Level 1 - Very Easy)

What is the popular food used to carve jack-o-lanterns during Halloween?

Your answer:

✓ Yes! Correct answer is Pumpkins.

**i** Points so far: 1

Select the difficulty level of your next question...

1. Very Easy 2. Easy 3. Easy-Medium 4. Medium 5. Medium-Hard 6. Hard 7. Very Hard

Next Question

שימו לב לשני מאפיינים של מבחן אינטראקטיבי:  
1. לנבחן אפשרות בחירה והשפעה על מהלך המבחן  
2. הנבחן מקבל משוב חוזר על נכונות תשובותיו, כמו גם ציון מתעדכן

# איך אנשים בוחרים רמת קושי?

ישנם לפחות שלושה גורמים שיכולים להשפיע:

1. ההסתברות לענות נכון על שאלות (שונה עבור רמות יכולות שונות)

2. רמת המוטיבציה שלך

• (כולל גם, צורך באתגר, ונכונות לקחת סיכון, תחושת הידע שלך, דימוי עצמי וכו')

3. שיטת הצייוןן במבחן (מערכת תגמול)

# מבחן אדפטיבי-עצמי (Self-Adapted Test; SAT)

- .1 לראשונה הוצע על ידי Rocklin and O'Donnell (1987)
- .2 מרבית המחקר סביב SAT נערך בין השנים 1987-2005 בעיקר על ידי שלוש קבוצות של חוקרים ואח"כ נעלם מהספרות המחקרית\*
- .3 מבחנים מרובי-ברירה (שאלות סגורות), במתימטיקה, ואוצר מילים
- .4 מבחנים הועברו במתכונת מדמה של "מבחן קובע" simulated high stakes

\*[e.g., Wise, Plake, Johnson, & Roos, 1992; Vispoel, Rocklin, & Wang, 1994; Hontangas, Olea, Ponsoda, Revuelta, & Wise, 2004 ]

# עניין מחודש במבחן אדפטיבי-עצמי

1. כמבחן פוטנציאלי מתאים למבחן שהוא low-stakes
2. כפלטפורמה לבחינת תהליך קבלת החלטות של נבחנים
3. שילוב טכנולוגיה בהערכה

# שאלות המחקר

**1. כיצד מתפקד המבחן האדפטיבי-העצמי (SAT) בתנאים שבהם הציון לא קובע, יחסית לתנאים שבהם הציון קובע?**

- השוואה בין low-stakes לבין high stakes
- אופרציונליזציה דרך "מטרת המבחן" - orientation goal (\*Dweck, 2006)
  - performance goal (do your best) או -
  - learning goal (learn from the test)

**2. מהי ההשפעה של שיטת הצינון?**

- מערכת תגמולים reward system
  - ציון מתגמל קושי או לא
  - ציון מונוטוני עולה או לא

		ציון מתגמל קושי	
		לא	כן
ציון מונוטוני	כן	Number Right (NR)	Reward per level (RW)
	לא	Percent Correct (PC)	Ability Estimate ( $\theta$ )

\*Also termed later as "Mindset"

# נתונים

- 583 נבדקים, מתנדבים דרך מערכת אונליין (AmazonTurk), תושבי ארה"ב וקנדה
- גילאים 18-74 (ממוצע 33)
- 45% נשים
- השכלה: 40% בעלי תואר ראשון, 6% בעלי תואר מתקדם
- תעסוקה: 56% עובדים משרה מלאה, 16% משרה חלקית, השאר סטודנטים/מובטלים/בפנסיה



# בנק פריטים

- 412 פריטי ידע כללי, שאלות פתוחות שנאספו ממקורות אונליין
- פרמטרים עבור הפריטים נאספו בפיילוט (N=198) למודל של 2PL (נאמדו מדדי קושי ואבחנה, לא מדד ניחוש)
  - mean difficulty 54%, median 55%; SD=26%; range 3%-99%
  - mean a-parameter 1.20, SD=0.61
  - mean b- parameter -0.40 and SD=1.8

# מבוק הפרויטים בניתי:

## pretest

- 24 items
- Targeting average a-parameter of 1 ; difficulty level of 70% (for motivational consideration)

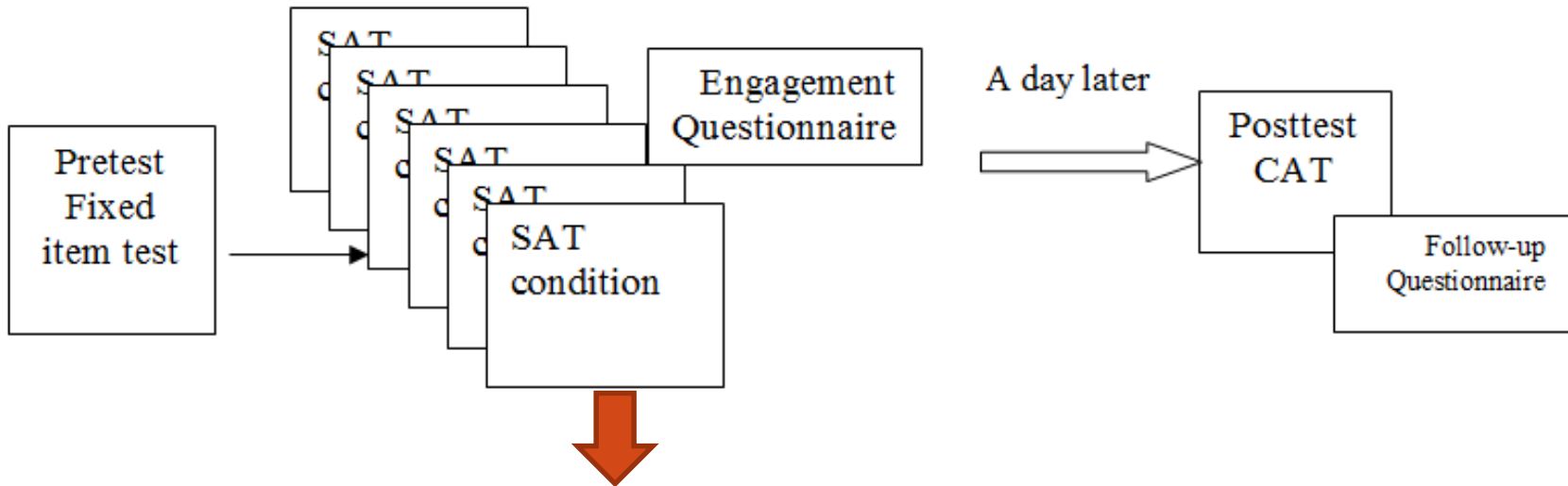
## SAT

- 40 items
- 7 difficulty levels (with 55 items in each )

## CAT

- 40 items
- 30 new items; 10 repeaters
- Maximizing information function

# מערך המחקר



	control	Equal weight scores		Rewarding for difficulty	
		NR	PC	RW	AE+
Learning goal	60	63	57	60	57
Performance goal	62	59	54	56	55

\*NR=number right;

PC=percent correct;

RW=reward points (1-7);

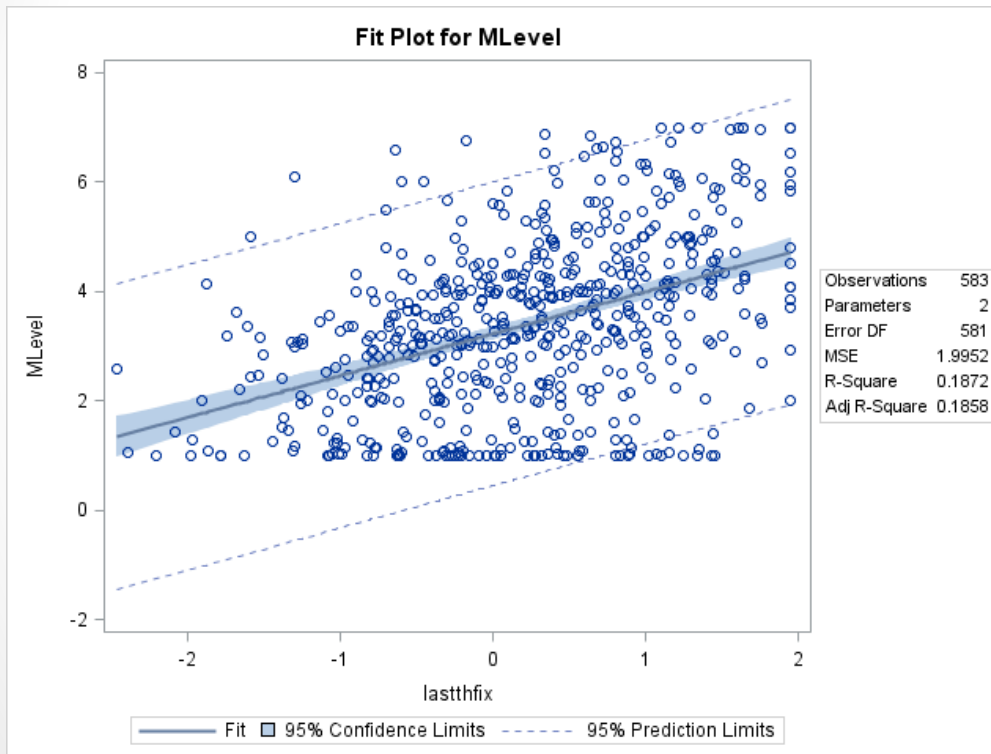
AE+=ability estimate; scale (100, 10)

# תוצאות

# באופן כללי,

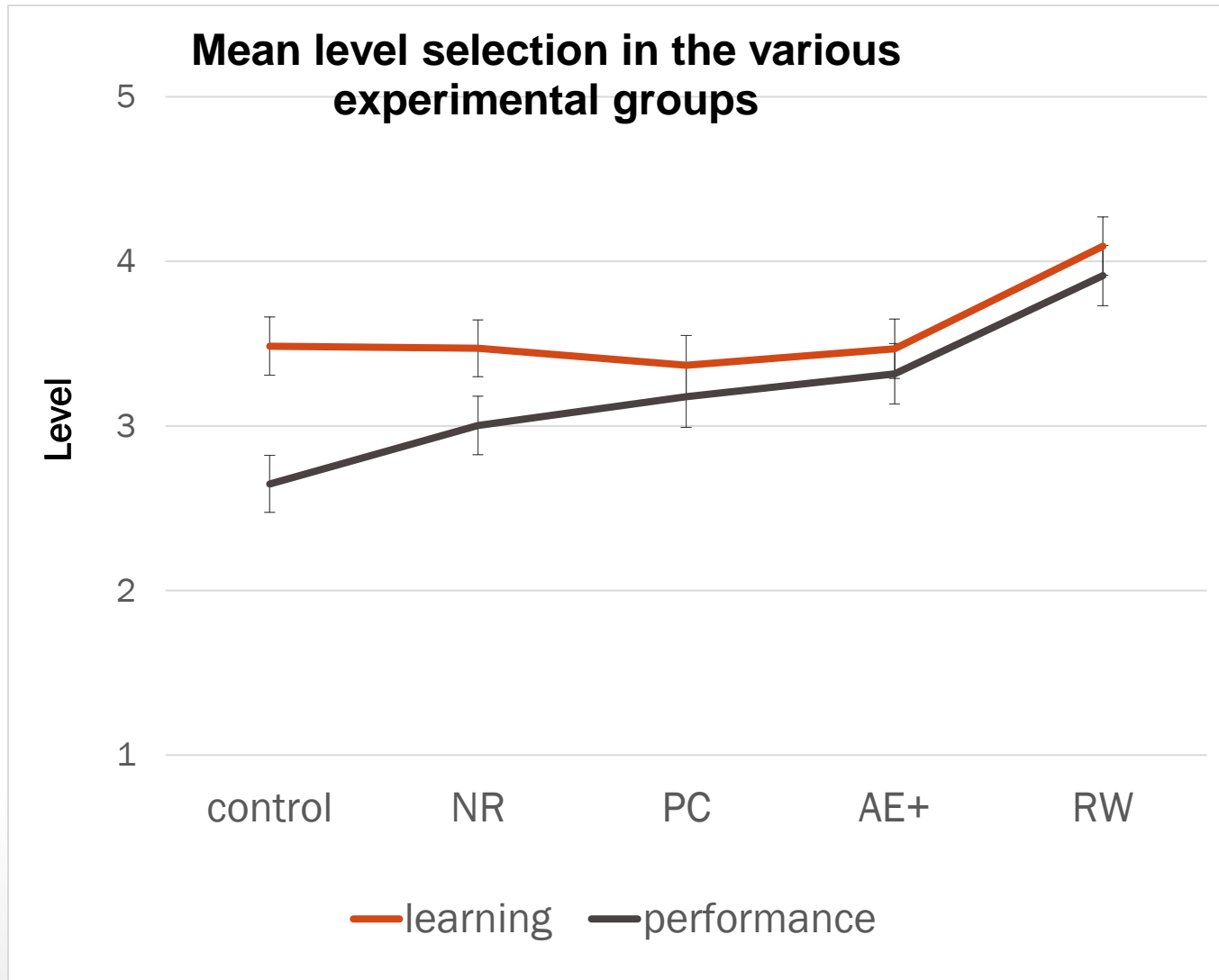
## נבחנים בוחרים בהתאם לרמת היכולת שלהם

Mean level selection by pretest ability score



- מתאם בין אומדן יכולת (ממבחן מקדים) לבין בחירת הקושי הממוצעת - 0.43
- שיפוע הרגרסיה - 0.77
- מודל מסביר בערך 20% מהשונות
- אחוז תשובות נכונות עמד על 67%

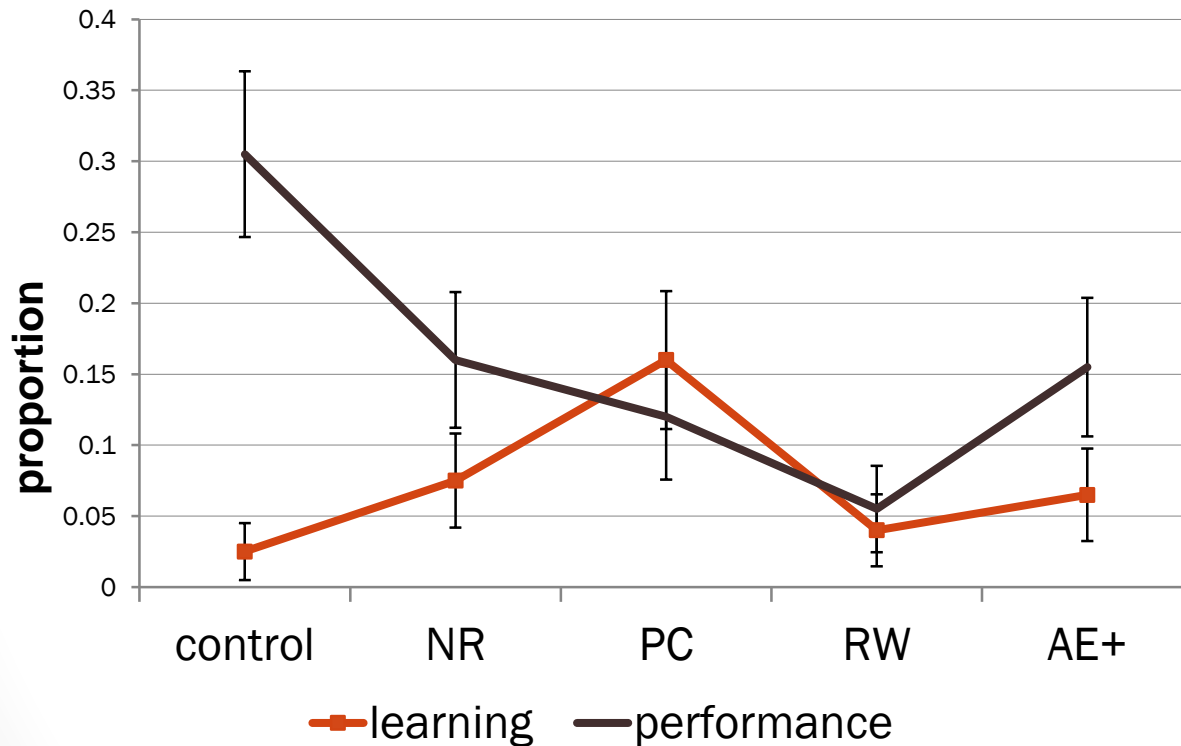
# בחירת קושי בקבוצות השונות



# זומה בעניין מוטיבציה

## Uniform Level 1 Selection

Percent of uniform level 1 selection in all experimental groups

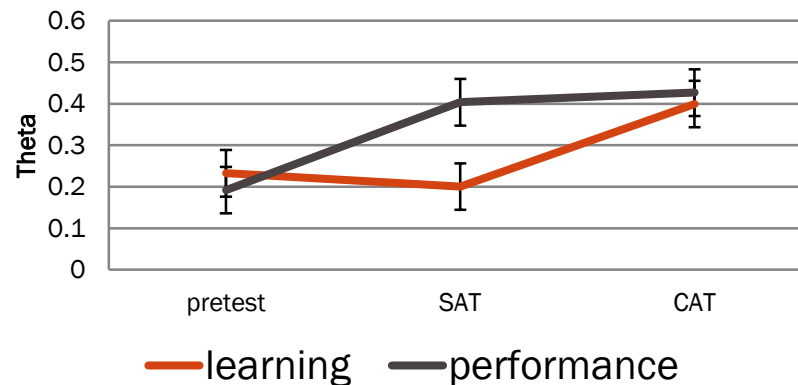


ממצא תומך נמצא בדיווח העצמי, שבו נבחנים בתנאי הביצוע בקבוצת הביקורת ובקבוצת NR דיווחו רמות מעורבות נמוכות יותר

# מתאמים בין שלושת אומדני היכולת

	$\theta$ -pretest	$\theta$ -SAT	$\theta$ -CAT
$\theta$ -pretest	--		
$\theta$ -SAT	.69	--	
$\theta$ -CAT	.66	.74	--
Mean	0.21	0.30	0.41
SD	0.88	0.97	1.13

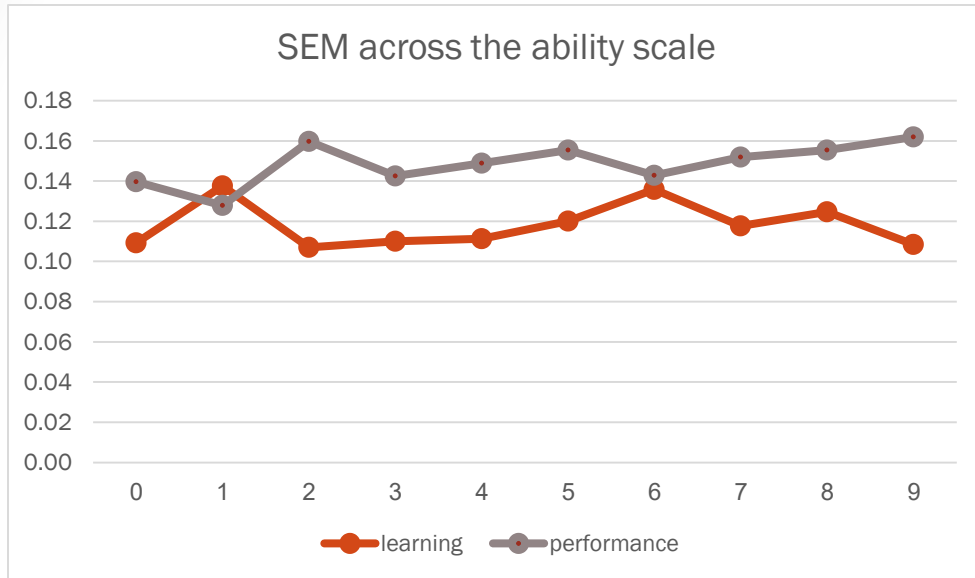
**Ability score in the three tests by orientation goal**



→ Stakes effect



# טעות התקן של המדידה



טעות התקן של המדידה  
 נמוכה יותר:  
 1. בקבוצות הלמידה  
 2. בקבוצת RW

	control	Equal weight scores		Rewarding for difficulty	
		NR	PC	RW	AE+
Learning goal	0.11	0.12	0.14	0.10	0.13
Performance goal	0.18	0.16	0.16	0.11	0.14

# סיכום ודיון

- באופן כללי, נבחנים משתמשים במבחן העצמי האדפטיבי כמצופה, בוחרים מכל רמות הקושי, ובעיקר בוחרים את הרמה שמתאימה להם
- המבחן העצמי-אדפטיבי מתפקד טוב יותר בתנאי למידה, דווקא כשהציון לא קובע (low-stakes) –
  - נבחנים מראים יותר מוטיבציה (בוחרים רמות קושי יותר גבוהות, מאתגרים את עצמם, מדווחים יותר מעורבות)
  - מדידה יותר מהימנה (טעות תקן קטנה יותר)
  - פחות מושפע מסוג הציון, ואפשר גם לא לתת ציון בכלל
  - אם כבר נותנים ציון אז –
    - ציון מתגמל על מאמץ, וקל להבנה
    - ציון מונוטוני עולה

**Thank you**

**[mattali@fordham.edu](mailto:mattali@fordham.edu)**