

**האומנם שינויים "קלים" פפריטים הם קלים?**

**ניתוח ההשפעה של שינויים בפריטי מבחני  
הישגים על מאפייניהם הפסיכומטריים**

נתנאל גולדשמיט  
עינת נוטע-קורן  
חגית גליקמן

# הקדמה

- מבחני המיצ"ב מתקיימים מידי שנה בארבעה מקצועות יסוד: שפת אם (עברית/ערבית), מתמטיקה, אנגלית, מדע וטכנולוגיה.
- המבחנים מועברים בכיתות ה' ו-ח' (שפת אם – כולל גם כיתה ב').
- כל מבחן כולל מספר נוסחי בחינה. כל נוסח מבחן מועבר לכ- 7000 תלמידים במגזר דוברי העברית ולכ- 3000 תלמידים במגזר דוברי הערבית.

✦ פריטים רבים משותפים למספר נוסחים, אולם לעתים

• קיימים הבדלי מיקום בין הפריטים או בין מקבצי פריטים בכל נוסח.

• בשאלות סגורות, ישנם הבדלים בסדר אפשרויות התשובה.

כמו כן:

◀ הליך פיתוח המבחנים כולל העברה ניסיונית במדגם של תלמידים.

✦ בעקבות הליך זה מתבצעים שינויים בפריטים. חלקם הכרחיים וחלקם "קוסמטיים".

## לאילו שינויים נתייחס?

שינויים אשר נתפסים כקלים למתבונן "הנאיבי". כלומר, נדמה שאינם משנים את מהות המדידה המתבצעת על ידי הפריט (או קבוצה של פריטים).

### שינויים בהם נדון:

א. שינויי "עריכה"

- שינויי ניסוח מזעריים, שינויי פיסוק, הדגשת מילים
- עיצוב מחודש של ציורים, תרשימים

ב. שינויי מיקום של מקבצי פריטים במבחן או של פרקים שלמים (למשל, טקסט והשאלות הנלוות אליו)

ג. ערבול סדר מסיחים בשאלות מסוג רב-ברירה

# המחקר הנוכחי

- מטרת המחקר היא ללמוד על ההשפעה של שינויים מסוגים אלה בפריטים על המאפיינים הפסיכומטריים שלהם.
- המחקר מתייחס למבחני הישגים שאינם עתירי סיכון ( Low Stakes) ושאין לחוצי זמן.
- אוכלוסיית הנבחרים: תלמידים דוברי עברית שנבחנו במיצ"ב התשס"ח.

## שאלות המחקר:

- האם קיימת השפעה לשינויים על ביצועי התלמידים?
  - האם ההשפעה זהה בדרגות כיתה שונות?
  - במקצועות שונים?
  - בקרב תלמידים בעלי רמות הישגים שונות?

מחקר אקספלורטורי (גישושי)

# **שינויי עריכה על קצה המזלג**

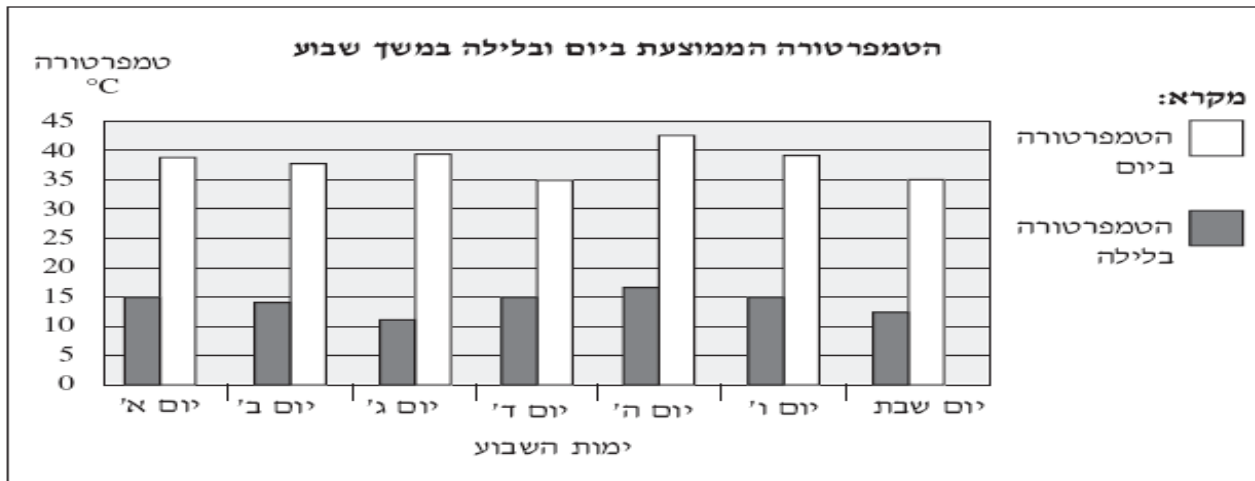
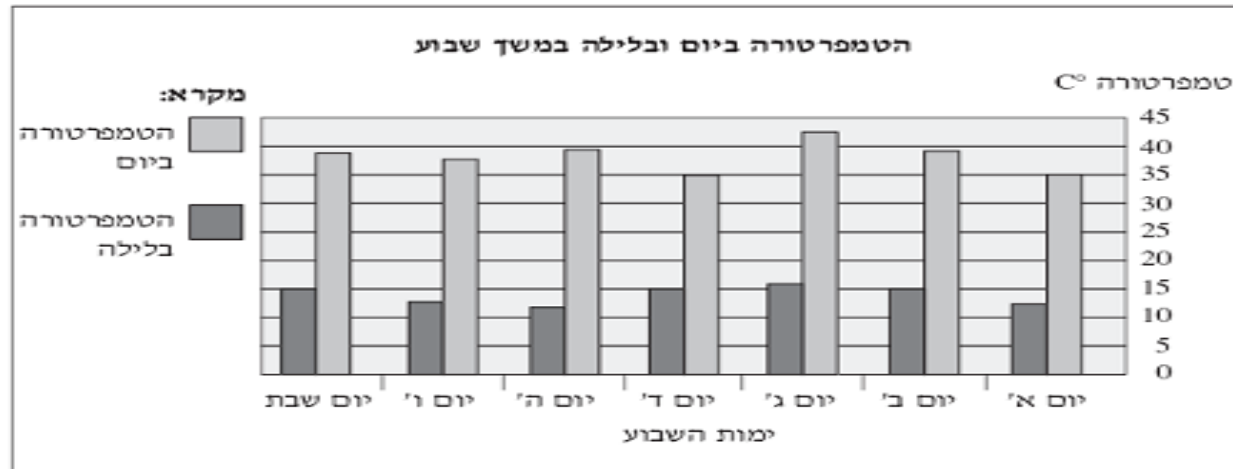
# דוגמא מספר 1 (מדעים ח') - שינויים קלים בניסוח

בהנחיות של שירותי הכבאות וההצלה **נכתבה** ההנחיה הבאה: "אם התלקחה אש בִּמְחֶבֶת-טיגון שבה שמן רותח, אל תשפוך מים! סגור את הגז וכסה את המחבת במטלית עבה. היזהר שהמחבת לא תתהפך ותגרום לכוויות."  
כיצד השימוש במטלית עבה גורם לכיבוי האש?

## לעומת:

בהנחיות של שירותי הכבאות וההצלה **כתובה** ההנחיה הבאה: "אם התלקחה אש בִּמְחֶבֶת-טיגון שבה שמן רותח, אל תשפוך מים! סגור את הגז וכסה את המחבת במטלית עבה. היזהר שהמחבת לא תתהפך ותגרום לכוויות."  
כיצד השימוש במטלית עבה (**סמרטוט**) גורם לכיבוי האש?

# דוגמא מספר 2 (מדעים ה')





# דוגמא מספר 3 - שינויי ערכים (מתמטיקה ח')

בסרטוט שלפניכם הנקודה D נמצאת על BC.  
נתון: משולש ADC הוא שווה-שוקיים ( $AD = DC$ ).

$$\angle DCA = 30^\circ$$

$$\angle CBA = 50^\circ$$

מהו הגודל של זווית  $\alpha$ ?

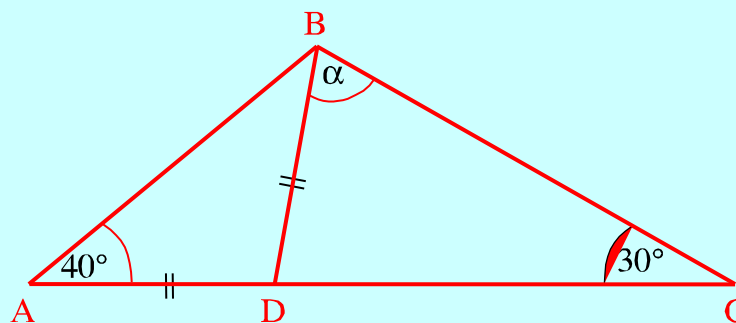
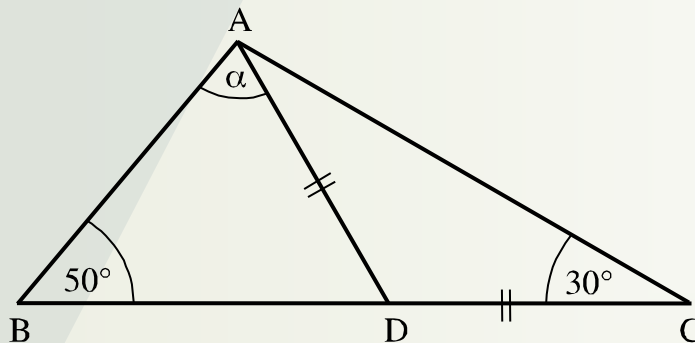
1.  $30^\circ$

2.  $50^\circ$

3.  $60^\circ$

4.  $70^\circ$

5.  $80^\circ$



1.  $30^\circ$

2.  $40^\circ$

3.  $50^\circ$

4.  $60^\circ$

5.  $70^\circ$

**שיעור**

**ההצלחה**

**גבוה ב- 8%**

**וממד**

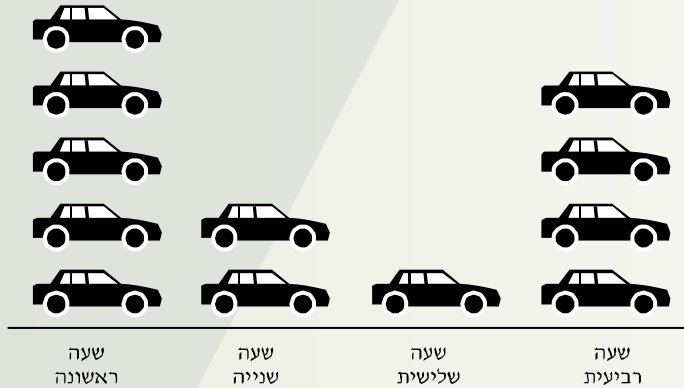
**האבחנה ירד**

**מ- 0.60**

**ל- 0.48**

# דוגמא מספר 4 - שינויי הקשר (מתמטיקה ה')

לפניכם דיאגרמה המתארת את מספר כלי-הרכב שעברו בצומת במשך 4 שעות.

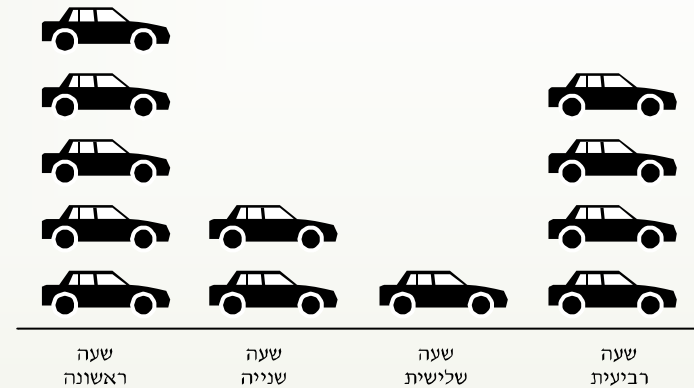


כל מכונית בדיאגרמה מייצגת 100 כלי-רכב.

א. באיזו שעה עברו בצומת 200 כלי-רכב?

לפניכם דיאגרמה המתארת את מספר כלי-הרכב שעברו בצומת במשך 4 שעות.

כל מכונית בדיאגרמה מייצגת 100 כלי-רכב



א. כמה כלי רכב עברו ברמזור בשעה הראשונה?

ב. בכמה גדול מספר כלי-הרכב שעברו בצומת בשעה הראשונה ממספר כלי-הרכב שעברו בצומת בשעה השנייה?

שינויי בקונטקסט יצר עליה של כ- 13% אחוז בשיעור המשיבים נכונה לסעיף ב' אשר הוא עצמו לא שונה

# שינויים במיקום ובהקשר של פרקי מבחן

# הספרות הפסיכומטרית

## מיקום של פריטים

מחקרים משנות ה-50 עסקו בהשפעה של סדר הפריטים (מהקל לכבד, מהכבד אל הקל) על ביצוע הנבחנים.

✦ סקירה של Leary & Dorans (1985)

✦ מטה-אנליזה של Aamodt & McShane (1992)

מבחנים שפותרים בפריטים קלים, קלים יותר ממבחנים שפותרים בפריטים קשים. הדבר נמצא בעיקר במבחנים לחוצי זמן ופחות במבחני עוצמה; יותר במבחני כושר ופחות במבחני הישגים.

# סדר של פרקי הבחינה: ממצאים לא עקביים

## דוגמאות למחקרים:

נמצא אפקט מסוים לסדר (במבחני כושר) בעיקר בתחום המילולי (למשל, Mollenkopf, 1950; Flaughner et al., 1968).

Klein (1981) לא מצא השפעה כזו.

רפ ואללופ (1999) מצאו אפקט זניח למיקום הפרק במבחן הפסיכומטרי על הביצוע בו.

אין השפעה למיקום מטלת החיבור (התחלה או סוף) בבחינת ה-SAT על הביצוע בשאר הפרקים (Liu & Oh, 2003; Oh & Walker, 2007).

לא נמצא קשר בין אפקט המיקום לבין רמת היכולות של הנבחנים (רפ ואללופ, 1999; Leary & Dorans, 1985).

## ממצאים אמפיריים

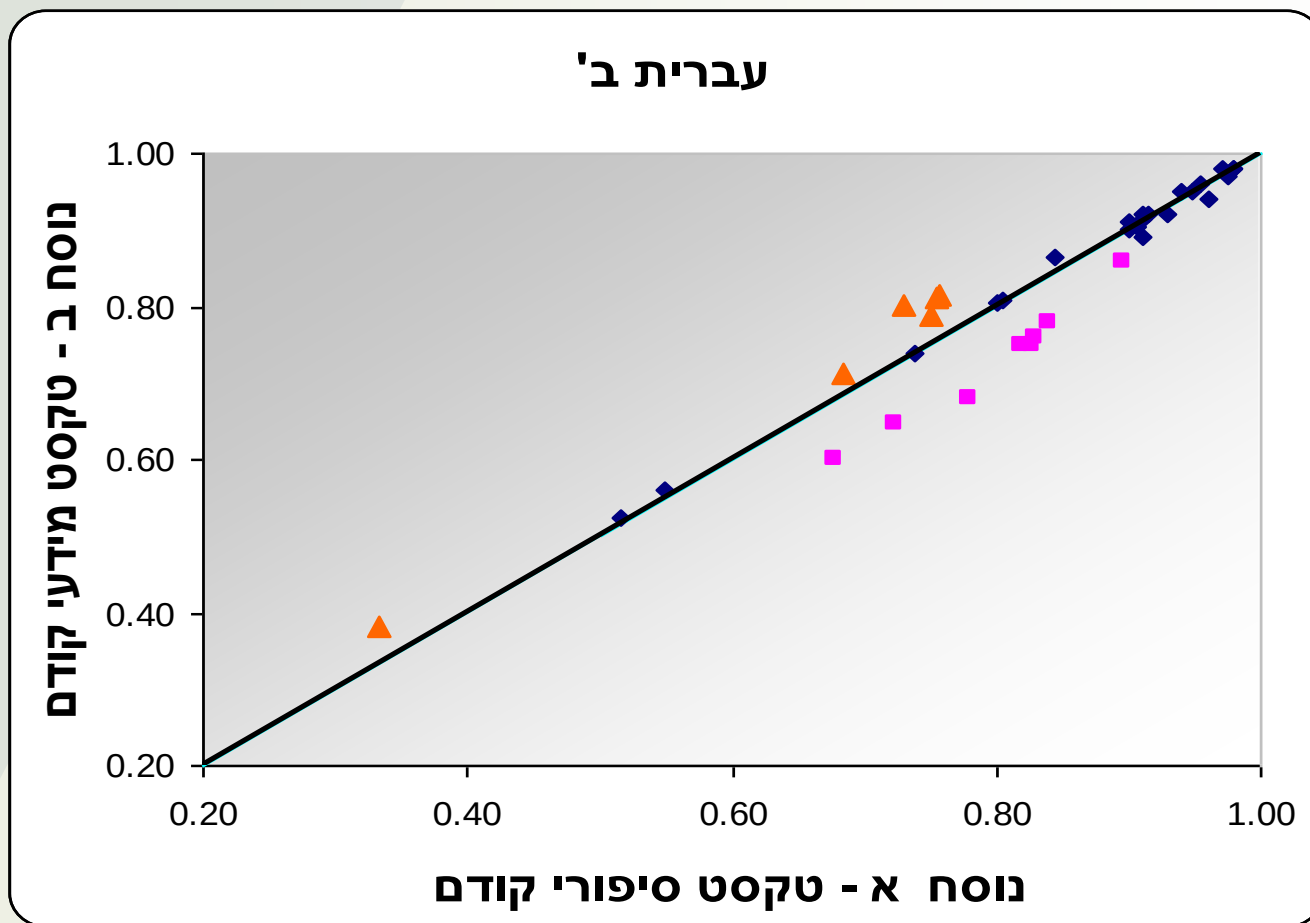
### המבחן בעברית לכיתה ב':

- ✦ הכתבה של מילים או צמדי מילים
- ✦ 5 פריטי פענוח
- ✦ 12 פריטים של הבנת נקרא ברמת משפט
- ✦ כתיבת סיפור
- ✦ 2 קטעי קריאה - סיפורי ומידעי. כל קטע מלווה בכ- 10 פריטים

### הבדל בין הנוסחים: סדר קטעי הקריאה

| מיקום | נוסח א | נוסח ב |
|-------|--------|--------|
| 1     | סיפורי | מידעי  |
| 2     | מידעי  | סיפורי |

# רמת קושי של שאלות משותפות לנוסח א' ונוסח ב'



טקסט  
מידעי

טקסט  
סיפורי

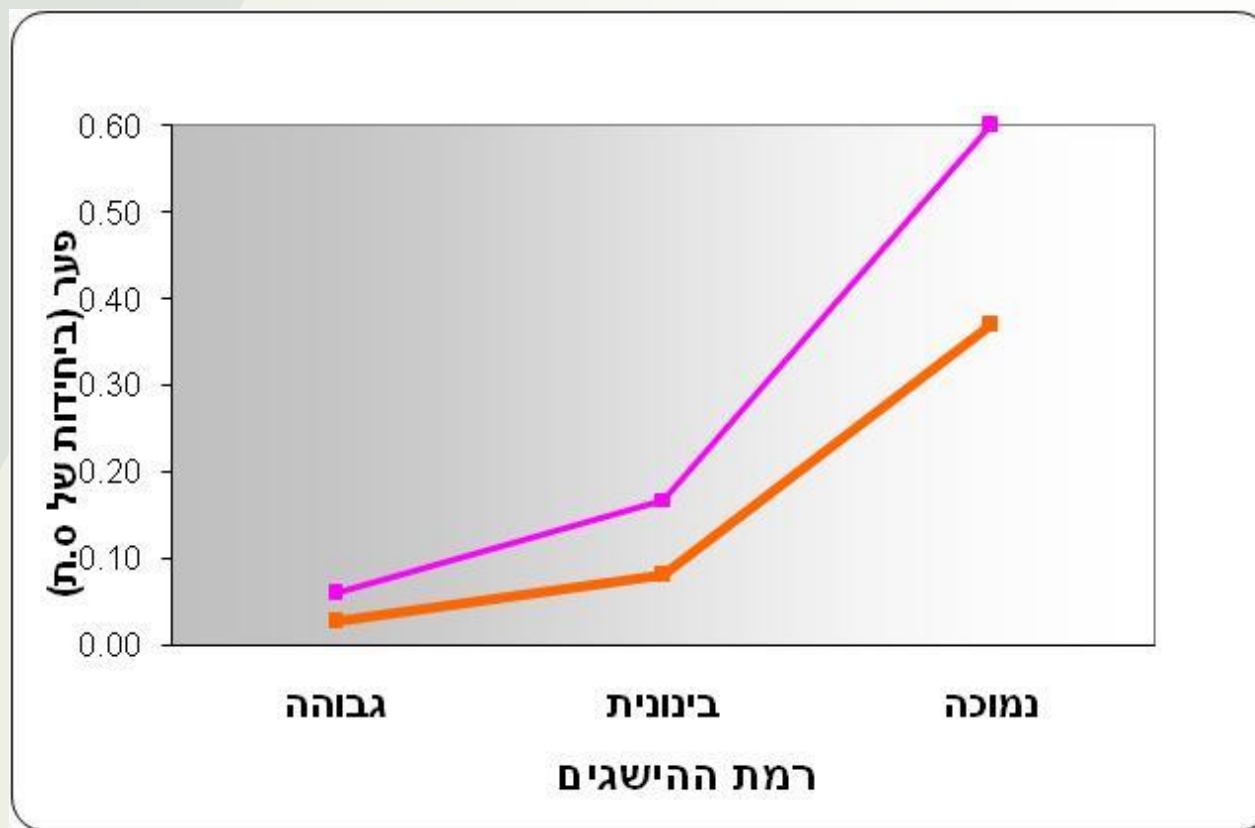
פריטים  
1-16

## ציונים בפרקים בכל נוסח והפער המתוקנן בין הנוסחים

| פער בין מיקום מוקדם למאוחר (ביחידות של ס.ת.) | נוסח ב           |       | נוסח א           |       |                           |
|----------------------------------------------|------------------|-------|------------------|-------|---------------------------|
|                                              | ממוצע (ס.ת.)     | מיקום | ממוצע (ס.ת.)     | מיקום |                           |
| 0.02                                         | 85.59<br>(12.75) | זהה   | 85.36<br>(13.12) | זהה   | פענוח, הבנת משפטים, כתיבה |
| 0.26                                         | 71.12<br>(30.09) | 2     | 78.33<br>(24.27) | 1     | טקסט סיפורי               |
| 0.18                                         | 68.85<br>(25.98) | 1     | 63.93<br>(29.45) | 2     | טקסט מידעי                |



# פערים בציונים בין הנוסחים לפי סוג הטקסט ורמת ההישגים



טקסט  
סיפורי

טקסט  
מידעי

## עברית לכיתה ב'- סיכום הממצאים

נראה כי קיימת השפעה למיקום (ראשון או שני) עבור כלל התלמידים ובמיוחד עבור תלמידים חלשים. ההשפעה באה לידי ביטוי הן בעליה ברמת הקושי והן בעליה בפיזור הציונים.

השפעת המיקום שונה בין סוגי הטקסטים שונים:

✦ כאשר הטקסט הקשה יותר (מידעי) הופיע אחרי הטקסט הקל יותר (סיפורי), הציון בו "ניזוק" בכחמישית ס.ת (0.18).

✦ כאשר הטקסט הקל יותר (סיפורי) הופיע אחרי הטקסט הקשה יותר (מידעי) הציון בו "ניזוק" יותר, כרבע סטיית תקן (0.26).

זאת בהנחה כי תלמידים בכיתה ב' משיבים על המבחן לפי הסדר. הנחה סבירה לתלמידים צעירים.

✦ פער גדול באחוז אי ההשבה בין מיקום קודם למיקום מאוחר:

✦ טקסט סיפורי: 4% לעומת 10%

✦ טקסט מידעי: 4% לעומת 8%

✦ השערה: יתכן שההקשר (קונטקסט) שבו מופיע הטקסט משפיע על הציון בו. כאשר הטקסט הקודם קל יותר, הדבר משפיע על המוטיבציה של הנבחנים ואז מידת ההצלחה בטקסט הקשה יותר עולה.

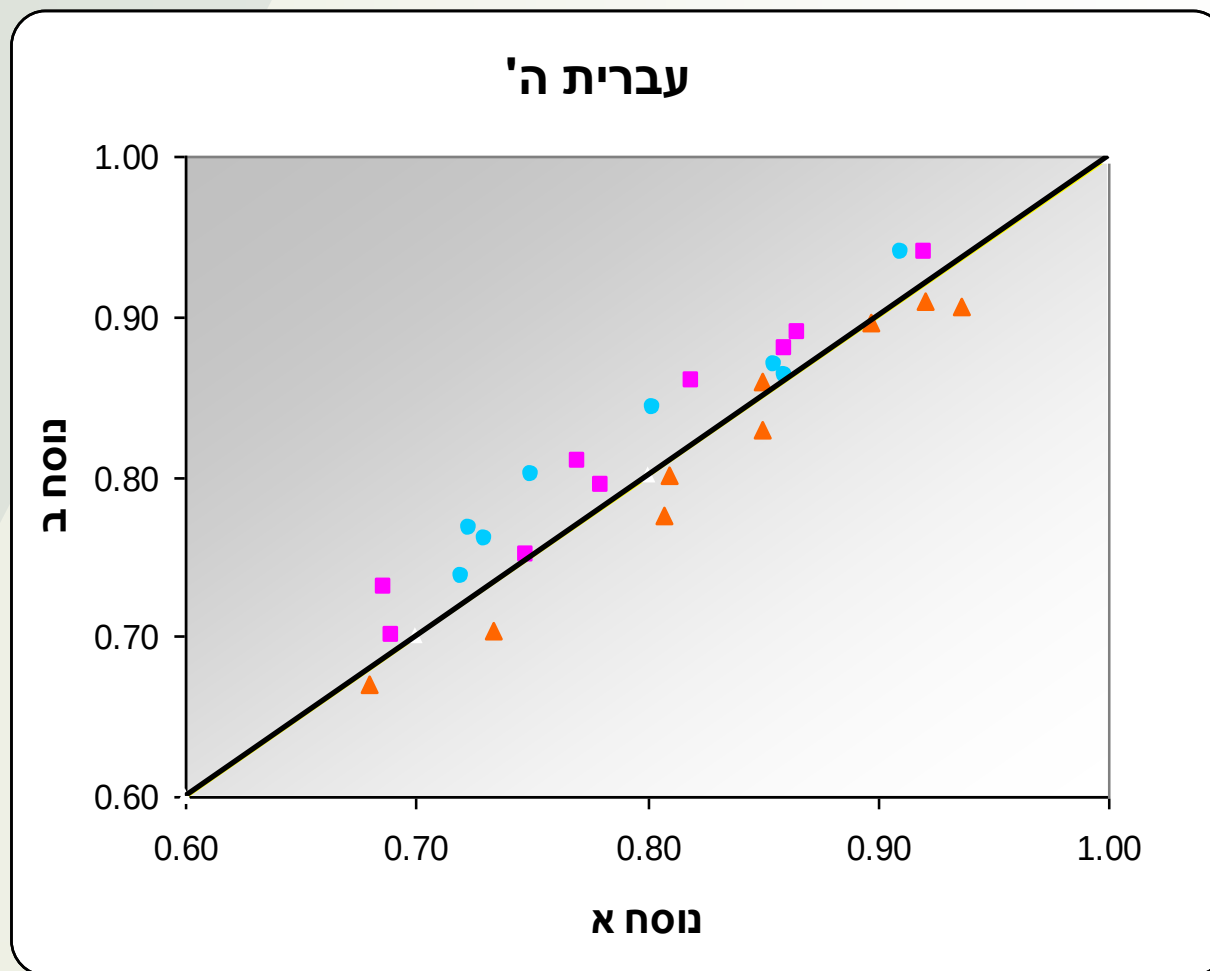
✦ השערה זו טעונה בדיקה והיא עקבית עם דיווחים בספרות שהצביעו על השפעת סדר הפריטים (מהקל אל הכבד או להפך) על רמת הקושי של המבחן.

# המבחן בעברית לכיתה ה'

מבנה המבחן:

| מיקום | נוסח א | נוסח ב |
|-------|--------|--------|
| 1     | מידעי  | סיפורי |
| 2     | סיפורי | מידעי  |
| 3     | כתיבה  | שימושי |
| 4     | שימושי | כתיבה  |

# רמת הקושי של שאלות משותפות לנוסח א' ולנוסח ב'



טקסט  
סיפורי

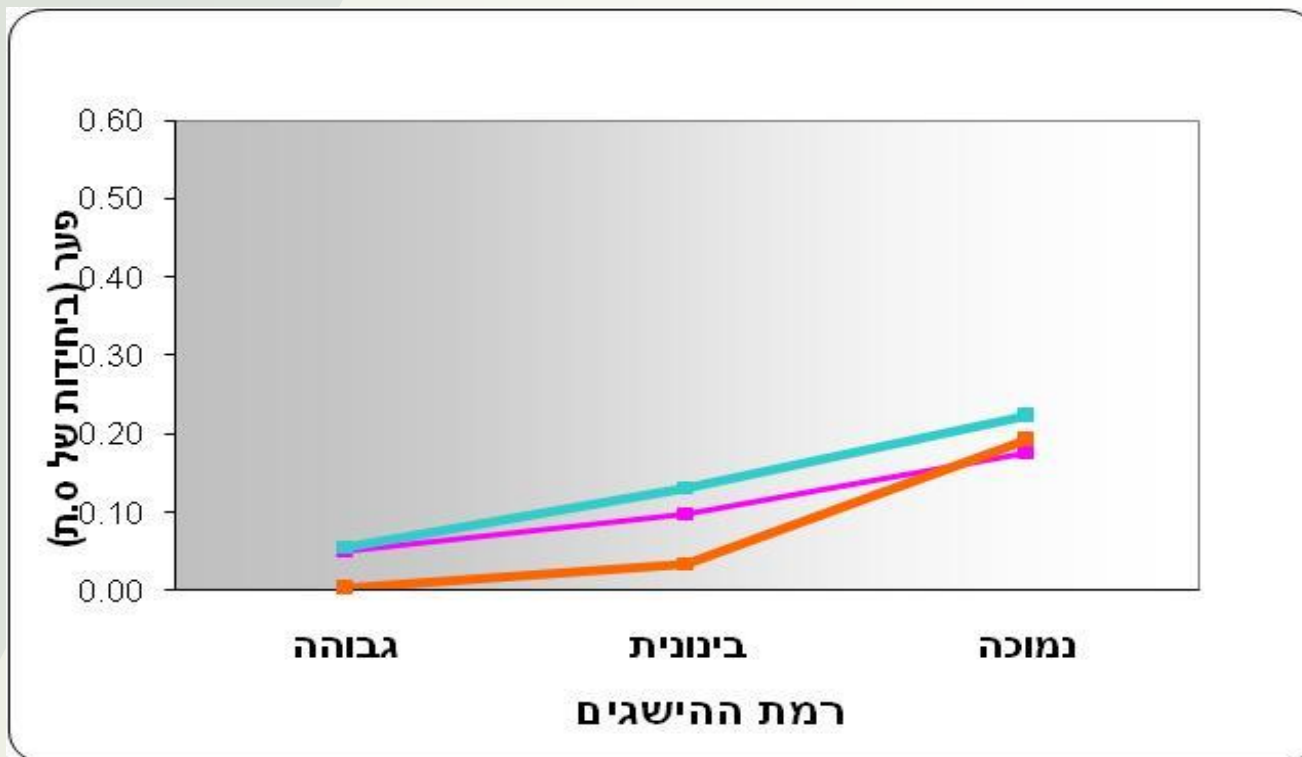
טקסט  
שימושי

טקסט  
מידעי

# ציוני הפרקים בכל נוסח והפער המתוקנן בין הנוסחים- עברית ה'

| פער בין מיקום מוקדם למאוחר (ביחידות של ס.ת.) | נוסח ב           |                   | נוסח א           |                   |             |
|----------------------------------------------|------------------|-------------------|------------------|-------------------|-------------|
|                                              | ממוצע (ס.ת.)     | מיקום             | ממוצע (ס.ת.)     | מיקום             |             |
| 0.09                                         | 81.98<br>(19.36) | 2                 | 83.71<br>(17.69) | 1                 | טקסט מידעי  |
| 0.09                                         | 80.44<br>(19.08) | 1                 | 78.67<br>(20.35) | 2                 | טקסט סיפורי |
| 0.12                                         | 81.55<br>(19.84) | 3<br>(לפני כתיבה) | 79.04<br>(21.85) | 4<br>(אחרי כתיבה) | טקסט שימושי |

# פערים בציונים בין הנוסחים לפי סוג הטקסט ורמת ההישגים



בכל אחד מהטקסטים קיימת מגמה דומה:  
מיקום הטקסט משפיע יותר על תלמידים בעלי הישגים **נמוכים**  
(כחמישית סטיית תקן) ביחס לתלמידים בעלי הישגים **בינוניים** (-0.04  
0.10 סטיית תקן) או **גבוהים** (0 עד 0.05 סטיית תקן).

# המבחן בעברית לכיתה ח'

מבנה המבחן:

| מיקום | נוסח א | נוסח ב |
|-------|--------|--------|
| 1     | טיעוני | מידעי  |
| 2     | כתיבה  | טיעוני |
| 3     | מידעי  | שימושי |
| 4     | שימושי | כתיבה  |

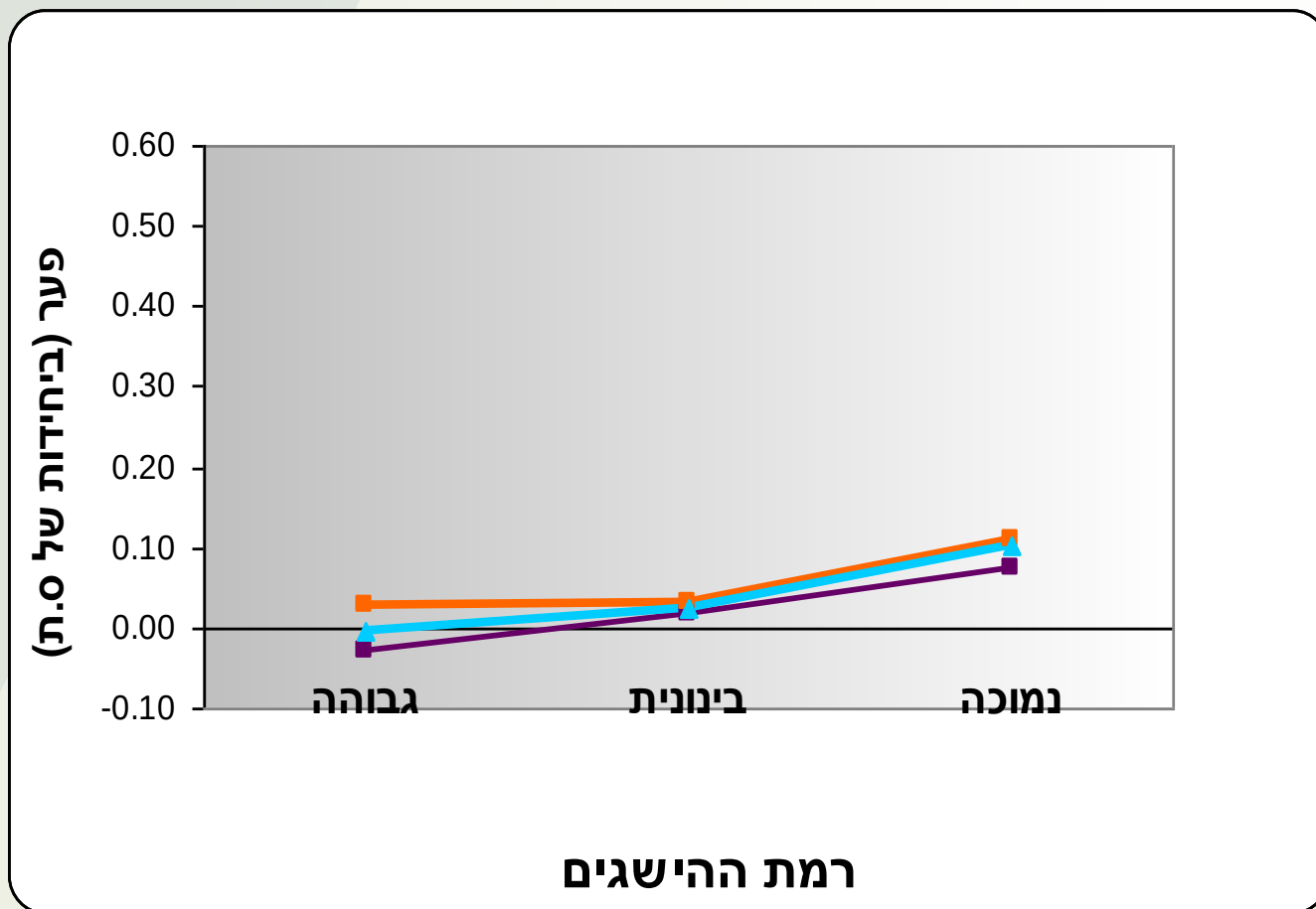


# ציוני הפרקים בכל נוסח והפער המתוקנן בין הנוסחים- עברית ח'

| פער בין מיקום מוקדם למאוחר (ביחידות של ס.ת) | נוסח ב           |       | נוסח א           |       |             |
|---------------------------------------------|------------------|-------|------------------|-------|-------------|
|                                             | ממוצע (ס.ת)      | מיקום | ממוצע (ס.ת)      | מיקום |             |
| 0.02                                        | 64.72<br>(25.13) | 2     | 65.18<br>(24.60) | 1     | טקסט טיעוני |
| 0.06                                        | 79.00<br>(16.70) | 1     | 77.96<br>(17.82) | 3     | טקסט מידעי  |
| 0.04                                        | 70.89<br>(22.56) | 3     | 69.82<br>(23.78) | 4     | טקסט שימושי |

הערה: הציונים מבוססים על הפריטים המשותפים בין שני הנוסחים

# פערים בציונים בין הנוסחים לפי סוג הטקסט ורמת ההישגים



טקסט  
מידעי

טקסט  
טיעוני

טקסט  
שימושי

## סיכום הממצאים בעברית

◀ מהנתונים עולה כי קיימת השפעה למיקום ולהקשר של הפרקים במבחן. השפעת המיקום וההקשר הלכה וירדה עם העלייה בגיל.

כיתה ב' - השפעה משמעותית

כיתה ה' - השפעה מתונה

כיתה ח' - השפעה קטנה

✦ יתכן כי הדבר קשור באורך / בעומס של המבחן באופן שהוא יחסי לקבוצת הגיל.

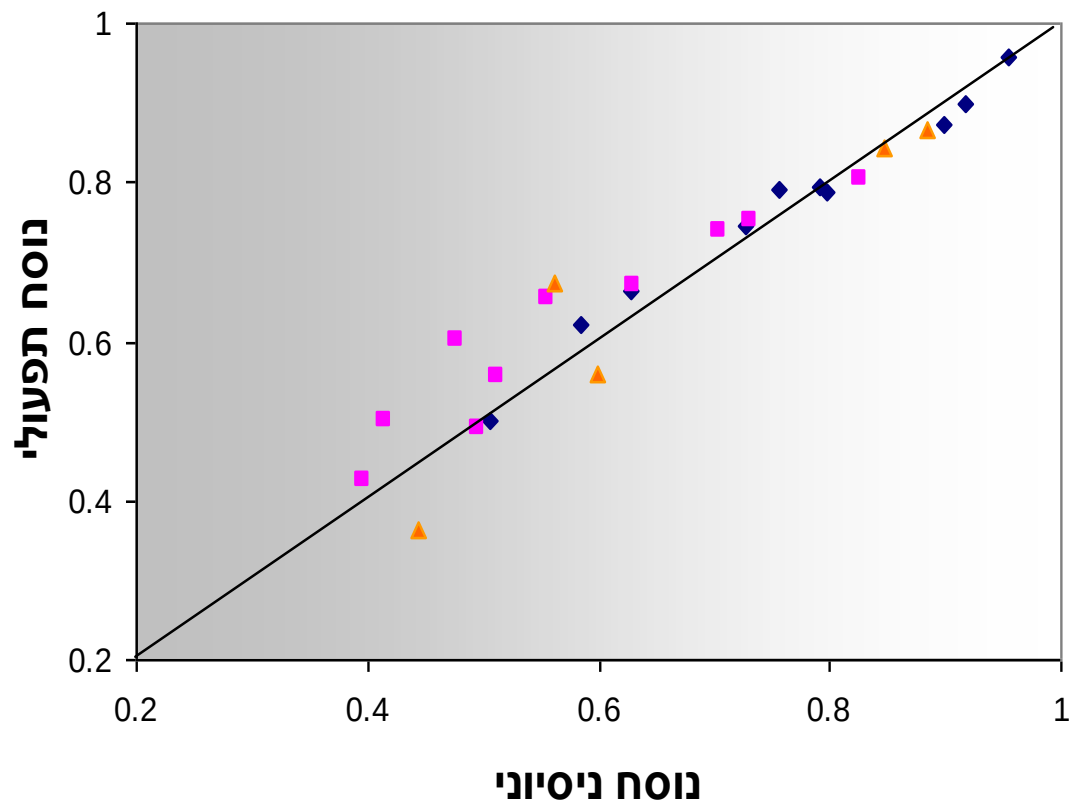
◀ השפעת המיקום וההקשר גדולה יותר בקרב תלמידים בעלי רמת הישגים נמוכה, וזאת בכל דרגות הכיתה.

✦ יתכן כי הדבר קשור בהתעייפות רבה יותר של תלמידים ברמת הישגים נמוכה.

## המבחן במתמטיקה לכיתה ה'

| נוסח תפעולי | נוסח ניסיוני |
|-------------|--------------|
| שלמים       | שלמים        |
| שברים       | גיאומטריה    |
| גיאומטריה   | שברים        |

## מתמטיקה לכיתות ה'



שלמים

שברים

גיאומטריה

הפריטים בנושא של שברים הפכו לקלים יותר בנוסח התפעולי, כאשר פריטי הגיאומטריה עברו לסוף המבחן. תמיכה נוספת להשפעת מיקום והקשר.

# **שינויים בסדר מסיכים בשאלות**

## **מסוג רב-ברירה**

## שאלת המחקר

◀ האם ערבול סדר אפשרויות התשובה משפיע על ביצוע התלמידים?

◀ לפי Attali & Bar-Hillel (2001) קיימת תופעה של Edge Aversion אשר גורמת לנטייה של נבחנים "לחפש" את התשובה הנכונה באמצע.

# הליך

- המחקר מתבסס על 68 פריטים במבחן "מאה מושגים במורשת בציונות ודמוקרטיה", מהשנים תשס"ו-תשס"ח.
- המבחן במאה מושגים כלל בכל שנה 45 שאלות מסוג רב-ברירה.
- לצורך המחקר, נבחרו רק שאלות שסדר אפשרויות התשובה שלהן היה שונה בין נוסח א' לנוסח ב'.
- כל נוסח מבחן הועבר בכל שנה לכ-13,000 נבחנים, דוברי עברית.
- נבדק ההפרש בין שיעור ההצלחה לשאלה בשני הנוסחים:  
$$(רחוק מהגזע) - (קרוב לגזע) = הפרש$$



הפרש בשיעור ההצלחה הגדול מ 1.3% נחשב **לפער חיובי**

הפרש בשיעור ההצלחה הקטן מ (-1.3%) נחשב **לפער שלילי**

**נקודת החתך שנבחרה מגדירה את הגבול העליון של רווח הסמך הרחב ביותר ברמת ביטחון של 95%**

**חושבו אחוזי המקרים שבהם:**

**(א) לא נמצא פער בין הנוסחים**

**(ב) נמצא פער חיובי** - אחוז המשיבים גבוה יותר כאשר התשובה הנכונה מוקמה "גבוה" יותר (קרוב יותר לגזע).

**(ג) נמצא פער שלילי** - אחוז המשיבים גבוה יותר כאשר התשובה הנכונה מוקמה "נמוך" יותר (רחוק יותר מהגזע).

## הממצאים

- ב- 41% מהמקרים לא נמצא פער בין ביצוע התלמידים בשני הנוסחים.
- ב- 40% מהמקרים השאלה הפכה לקלה יותר ככל שהתשובה הנכונה מוקמה "גבוה" יותר.
- ב 19% מהמקרים השאלה הפכה לקלה יותר ככל שהתשובה הנכונה מוקמה "נמוך" יותר.

# סיכום ומסקנות: האם שינויים קלים הם באמת קלים? השלכות על פיתוח מבחנים

✦ לעיתים שינויים שנראים למתבונן כשינויים קלים, שאינם משנים את מהות המדידה - מביאים לידי שינויים של ממש בביצוע התלמידים.

✦ נראה כי לשינויי מיקום והקשר של פריטים או פרקי מבחן יש השפעה על הביצוע ובמיוחד בקרב תלמידים צעירים ותלמידים חלשים.

• הועלתה השערה כי הדבר קשור ברמת הקושי של המבחן / פרקי המבחן. ככל שמבחן קשה יותר לנבחן, כך למיקום ולהקשר של הפריטים/פרקים קיימת יותר השפעה.

✦ לשינוי סדר מסיחים בשאלות רב-ברירה יש השפעה על רמת הקושי של השאלה, כאשר שאלות שבהן התשובה ממוקמת קרוב יותר לגזע הופכות לקלות יותר.

יש לקחת בחשבון גורמים אלה כאשר הופכים גרסת מבחן ניסיונית לתפעולית

# דוגמאות למחקרי המשך

השפעה של שינויי עריכה  
קלים/ ניסוח על הביצוע?

בדיקת ההשפעות השונות בקרב קבוצות אוכלוסייה  
שונות? (בנים/בנות; לקויי למידה; דוברי  
עברית/דוברי ערבית)

מה בין השפעת מיקום  
להקשר?

בחינת השפעת  
השינויים על  
מדדי האבחנה

השפעת  
שינויים  
שונים  
כפונקציה  
fe משתנה  
ראת  
הקושי fe  
הפריטים/ה  
סקסט

הבחנה בין תלמידים  
שלא השיבו לשאלות  
לבין תלמידים שטענו  
בתשובות

מדוע שאלות  
רב-ברירה שבהן  
התשובה  
הנכונה מוקמה  
גבוה יותר, הן  
קלות יותר?  
מהי  
אסטרטגיית  
ההשבה בה  
נוקטים  
התלמידים?

מחקר שיטתי בנושא סדר מסיחים

## בהמשך

יש להמשיך ולמפות במחקריים שיטתיים את סוגי השינויים שהם "קלים" לעומת שינויים שהם משמעותיים.

קטגוריזציה כזו תסייע לצוות הפיתוח בהחלטה אלו שינויים הם לגיטימיים לביצוע בשלבי פיתוח שונים ואלו לא.

# תודה!

