אוניברסיטת בן-גוריון בנגב

הוראות לנבחן בצידו השני של הדף

אין לכתוב מעבר לשוליים משני צידי הדף

מס׳ כיתה אבר בנין פ

מס' נבחן 28421

1 7/7 הנדסת מערכות מידע מערכות מידע הנדסת מערכות מידע אחזור מידע וספריות דיגיטליות 1 03721440601402

לשימוש המרצה הבודק
יחידות ועשרות ומאות
1 10 1
- 8 9 2
1 11 3
-1 1 1 1 1
5
1 40 6
8
20 AM
ציון הבחינה
שם המרצה
חתימה
תאריך



אוניברסיטת בן-גוריון בנגב

- בהגיע הסטודנט למקומו ,עליו להניח את כרטיס הנבחן ותעודה מזהה עם תמונה על השולחן.
- חל איטור על הכנסת כל חומר לבחינה , למעט זה שהותר במפורש ע"י המורה וצויין ע"ג טופס הבחינה לגבי סטודנטים עם התאמות מיוחדות -ראו התייחסות נפרדת עזב הסטודנט את חדר הבחינה לאחר חלוקת השאלונים , דינו כדין" נבחן"
- אסור לנבחן לשוחח בזמן הבחינה ,או לעזוב את מקומו ללא נטילת רשות. סטודנט רשאי לבחור אם להשתמש בעט או בעפרון. במקרה שכתב הבחינה בעפרון והבחינה לא נסרקה ,לא יהיה זכאי לערער על הבחינה . אם הבחינה נסרקה ,יהיה רשאי לערעד על הבחינה ובתנאי שאיכות הסריקה תקינה .יש לכתוב בכתב ברור ,נקי וקריא ,על עמוד אחד של כל דף . אין לכתוב בשוליים ואין להשתמש בטיפקס . טיוטא תיכתב בצד ימין של המחברה זיש לתחומ אוחה רתנידרה מו אחור לחיליו דמים או האברה

בס"ד

372.1.4406 – איחזור מידע תשע"ד

16.02.2014 'סמסטר חורף מועד ב'

פרופ' ברכה שפירא, אביגיל פרדיס, יוסי בן-שלמה

משך המבחן : שעתיים וחצי

חומר עזר: מותר (לא מחשב נייד), מותר מחשבון

יש להחזיר את גיליון הבחינה. המבחן כולל 3 דפים. יש לענות על כל השאלות

שאלה 1 - 16% - יש לענות על הגיליון

לא נכון אי תלות בין ה terms נכון לא נכון אי תלות בין ה SVD 2% (כון לא נכון אי תלות בין במנשק מובן למשתמש הקצה (כון לא נכון אינדן במנשק מובן למשתמש הקצה (כון לא נכון אינדן במנשק מובן למשתמש הקצה (כון לא נכון אינדן באופן באופן 2% בדי לאפשר דינמיות של אינדקס ניתן להחזיק אינדקסים קטנים שיתעדכנו באופן שוטף ויתמזגו עם האינדקס המרכזי פעם בתקופה כדי למנוע עדכונים תכופים מדי של

האינדקס המרכזי. (נכוו לא נכון די להגיע לתוצאות טובות יותר מו 1.4 🗸

לנקוט crawler) כדי להגיע לתוצאות טובות יותר מבחינת הfreshness של דפים עדיף לרוצאות טובות יותר מבחינת הבחינת בגישה אחידה (נכון לא נכון בגישה פרופורציאנלית לתכיפות עדכון הדפים מאשר בגישה אחידה (נכון לא נכון

2% אחד ההבדלים בין מערכת המלצה למנוע חיפוש הוא שמערכת המלצה ממליצה על רלוונטיות של מוצרים (או מסמכים) למשתמש לעומת מנוע חיפוש שקובע את הרלוונטיות של המסמכים למשתמש נכון/לא נכון

4% נתונים שני מנועי חיפוש s2 ,s1 שונים המופעלים על שני מאגרים נפרדים, שחלק מהמסמכים בהם משותפים. שאילתא נשלחת לשני המנועים, כל מנוע מחזיר רשימה מדורגת של מסמכים. כדי למזג את הרשימה המוחזרת לרשימה מדורגת אחת מאוחדת מתוצאות שני המנועים מספיק ש:

א. יודעים שהאלגוריתמים של המנועים שווים,

ב. המנועים יחזירו את רמת הדמיון של כל מסמך לשאילתא

ג. את כמות המסמכים המשותפת בין המאגרים

ד. א+ג ה. א+ב

ו. א+ב+ג

אין דרך למזג את התוצאות

: אסטרטגיה של crawler כוללת החלטות על 4% אסטרטגיה של

א. תגובה להוראות בקבצי robots.txt

ב. תעדוף ובחירה של אתרים שונים בתור u

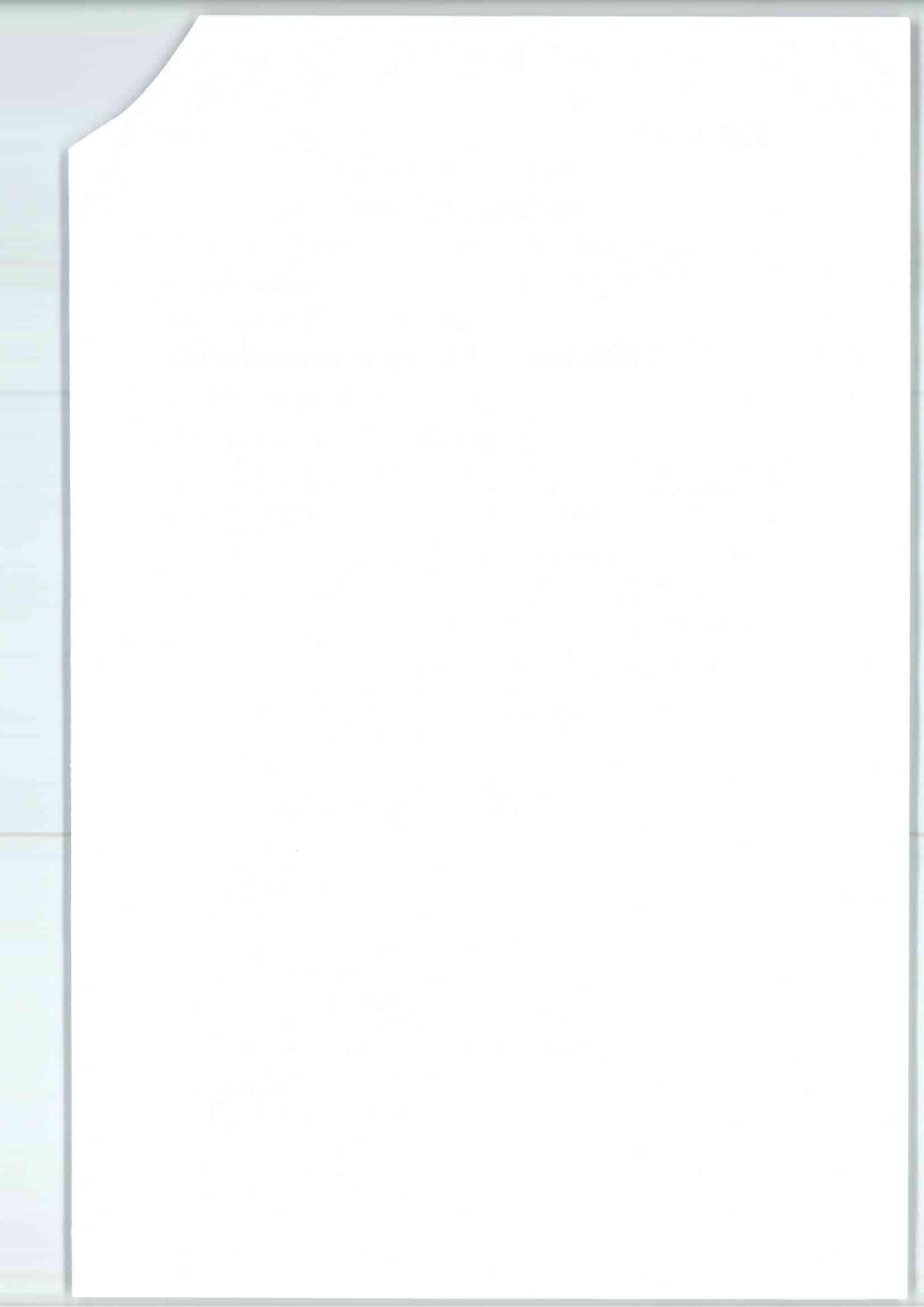
ג. שיטת אינדוקס דפים שה crawler מוריד

ד. שיטת הביזור של הדפים ב repository שהrawler מכין

עה. תכיפות עדכון של דפים באינדקס ע

ו. א+ב+ד

ז)א+ב+ה



2. 17% קיימים שני מסמכים d1, d2. ב-d1, d2 מילים ואילו ב-3000 d2 מילים. נתונה השאילתה "מאילתה שני מסמכים במסמכים: "who shot Lincoln". בטבלה המופיעה למטה מוצגות מספר ההופעות של חלק מהמילים במסמכים:

	7900				
	D1	D2			
who	4	7			
theatre	3	15			
show	6	9			
Lincoln	1	3			
play	2	2			
shot	8	1			

- א) 6% חשב את דירוגי המסמכים ביחס לשאילתה, תוך שימוש במודל השפה הבסיסי שהוצג בכיתה,
- ב) 6% חשב את דירוגי המסמכים ביחס לשאילתה ,תוך שימוש ב-KL divergence, קבע מי המסמך הטוב יותר על פי שיטה זו.
- ג) 5% הסבר מהי החלקה וציין שתי סיבות לשימוש בה בעת דירוג מסמכים באמצעות מודלי שפה אפר אבר אונין שתי סיבות לשימוש בה בעת דירוג מסמכים באמצעות מודלי שפה אבר אוני אונין שתי סיבות לשימוש בה בעת דירוג מסמכים באמצעות מודלי

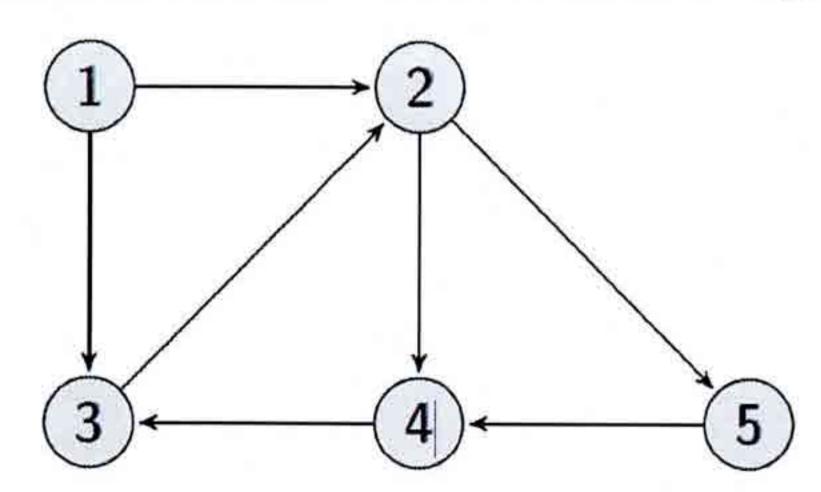
12% .3^J

: א. 4% מה יהיו ערכי ה hub של רשת שבה כל הקשתות דו-צדדיות, כלומר hub א. 4% מה יהיו ערכי ה

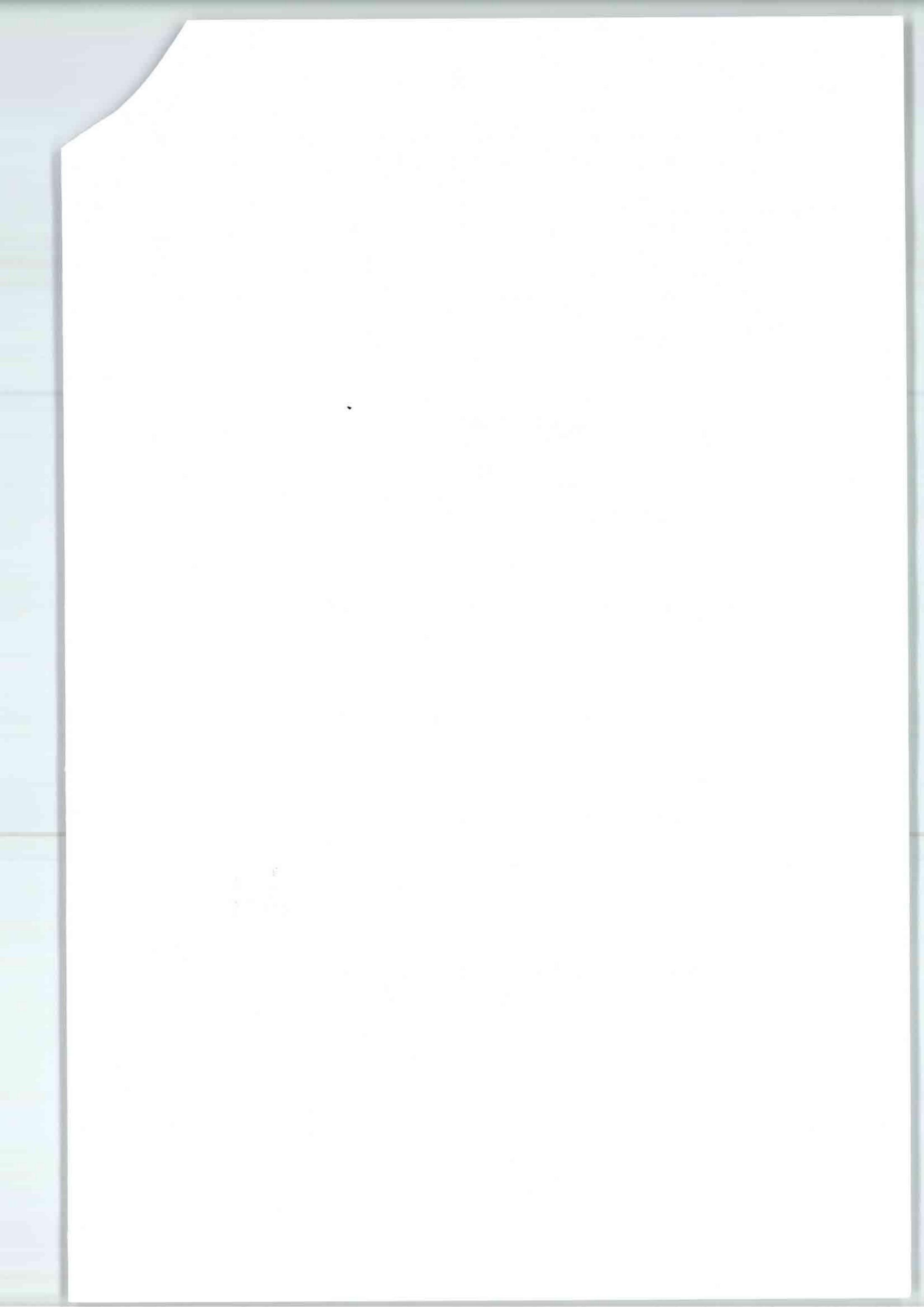
$$(u,v) \in E \Leftrightarrow (v,u) \in E$$

הצדק את תשובתך.

ב. 4% הצג את מטריצת המעברים עבור אלגוריתם Page rank לרשת הבאה:



ג. 4% מיהם שני הצמתים בעלי ה pagerank הנמוך ביותר (הסבר, ללא חישוב!)



(items) נתונה הטבלה הבאה של דירוגי משתמשים למוצרים שונים 4.4

	User1	User2	User3	User4	User5
Item1	5		4	4	4
Item2		5	5	3	1
Item3	2	ŧ\	1		2
Item4		1		3	
Item5	5				
Item6		5	4		

- Collaborative Filtering מא. 6% רוצים להמליץ למשתמש 5 על tem מה וtem מה 6 בגישת 6% רוצים להמליץ למשתמש 5 על משתמשים אילו משתמשים ישתתפו בחישוב? מה באלגוריתם המתחשב בדמיון בין משתמשים אילו משתמשים ישתתפו בחישוב? מה הניבוי של הציון שיתן משתמש 5 לtem מספר 6?
- ב. 7% רוצים להמליץ על למשתמש 2 על item מס 3, אם משתמשים בגישת הדמיון בין המוצרים? הסבר אילו דירוגים של אילו מוצרים משתתפים בחישוב.
 - שבכל ריצה של שאילתא הוא ירחיב את השאילתא עם % מפתח של מנוע חיפוש החליט שבכל ריצה של שאילתא הוא ירחיב את השאילתא עם מילים נוספות מתוך וויקיפדיה, שקשורות למילות השאילתא .
- א. 15% הצע ש<u>יטה למציאת מילים קשורות</u> למילות השאילתא בוויקיפדיה, במפורט תאר את האלגוריתם למציאת מילים קשורות והערכת מידת הקשר של מילה מסויימת בוויקיפדיה למילות השאילתא.
 - ב. 5% אילו מבני נתונים נוספים (מעבר לאינדקס ההופכי ולפוסטינג של המאגר של המנוע) כדאי למנוע להחזיק כדי לממש את הרעיון הנ"ל.
- 6/ 10% אלגוריתם דרוג של מנוע חיפוש נותן עדיפות למילים שהופיעו בכותרת, כלומר, אם כל המילים של השאילתא מופיעות בכותרת, המנוע מחליט שהמסמך הוא רלוונטי, ואז מחשב את רמת הרלוונטיות שלו, אחרת המסמך לא רלוונטי. תאר מבנה של אינדקס (מילון ופוסטינג) שייתן פתרון יעיל למנוע כזה.

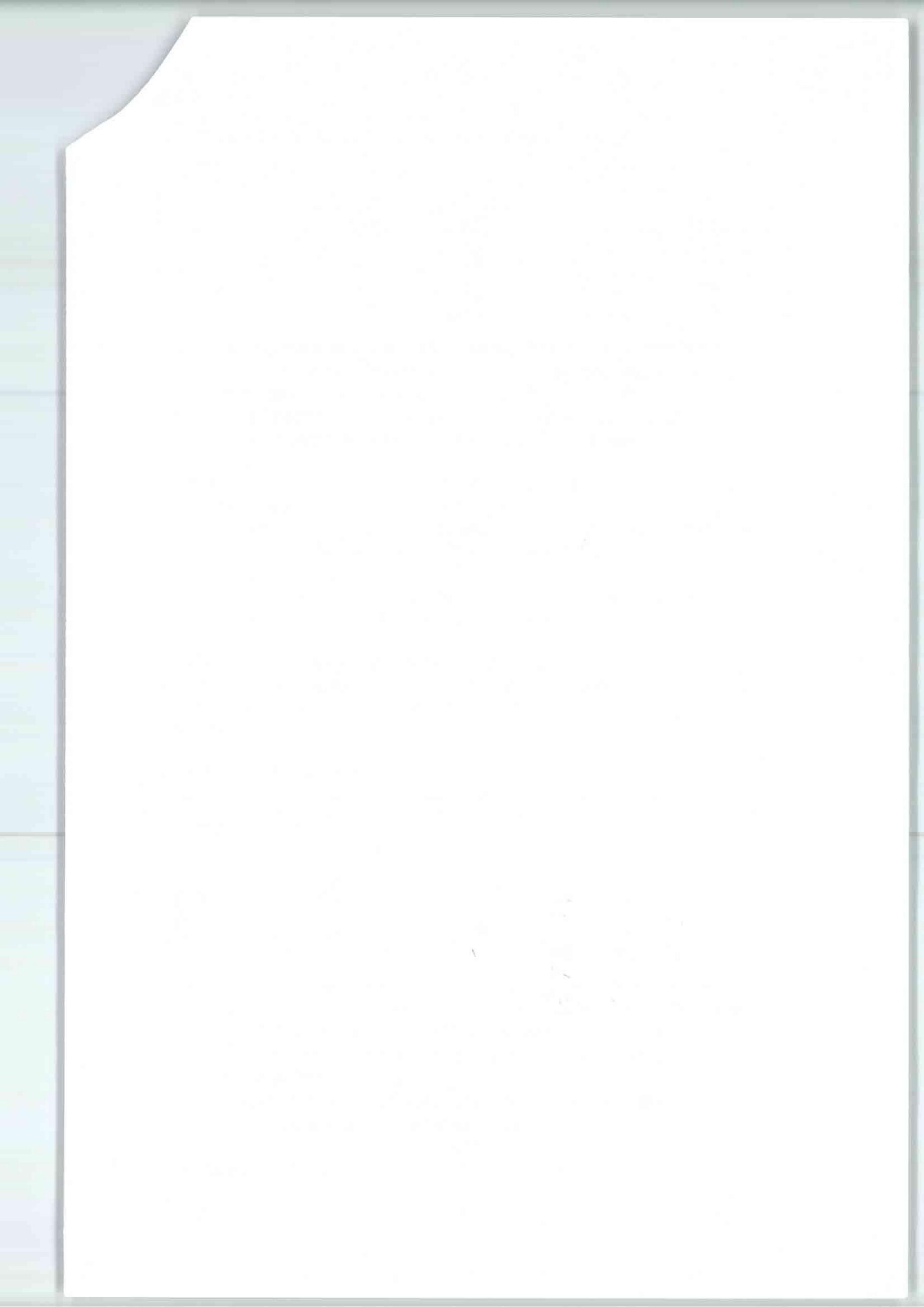
ובא: 12%. אגר המסמכים הבא: les flour eggs clove ielly

	milk	pepper	raisins	sugar	cinnamon	apples	flour	eggs	clove	jelly
d_1	X1	0	0	X1	0	1_	L	0	0	0
d_2	XX	75	0	21	0	04	0	0	七士	0
d_3	31	11	0	2 %	0	0	0	23	0	()
d_4	24	27	11	XL	21	O	21	十五	. 0-	()
d_5	23	0	23	0	1 +	0	81	2 5	ナキ	2 3
do	11	0	0 3	()	0	0	社上	12	0	21
d_7	21	24	0	0	十五	()	0	0	0	才上 之
d_{8}	0	0	33	2 }	0	7 4	0	×1	0	0

ב. 6% נניח שהמנוע החזיר את המסמכים d2,d4,d7 והמשתמש סימן את d2, d4 כרלוונטים ואת d7 כרלוונטים ואת d7 לא רלוונטי.

אם β=1, α=2, אם Rocchio חשב את השאילתא החדשה שתיוצר על פי אלגוריתם γ=0,β=1, α=2 אם γ=0,β=1 (יש להראות את הווקטור החדש של השאילתא)

בהצלחה! ברכה אביגיל ויוסי





אוניברסיטת בן-גוריון בנגב 🔏

הוראות לנבחן בצידו השני של הדף

אין לכתוב מעבר לשוליים משני צידי הדף

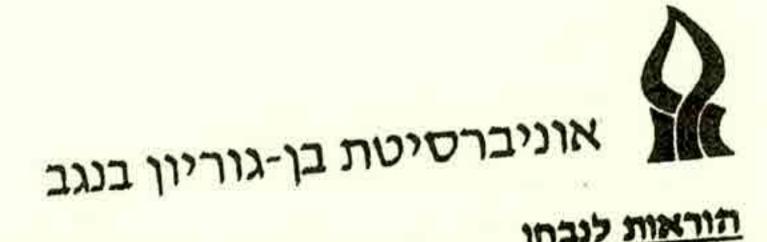
DO 238 מס' כיתה

מס' נבחן

מדבקה מדבקה

המחלקה גאיל גיק שנה _____ שנה _____ אריך בחינה ____ או , ב. מקצוע בחינה אמניר גיל ע

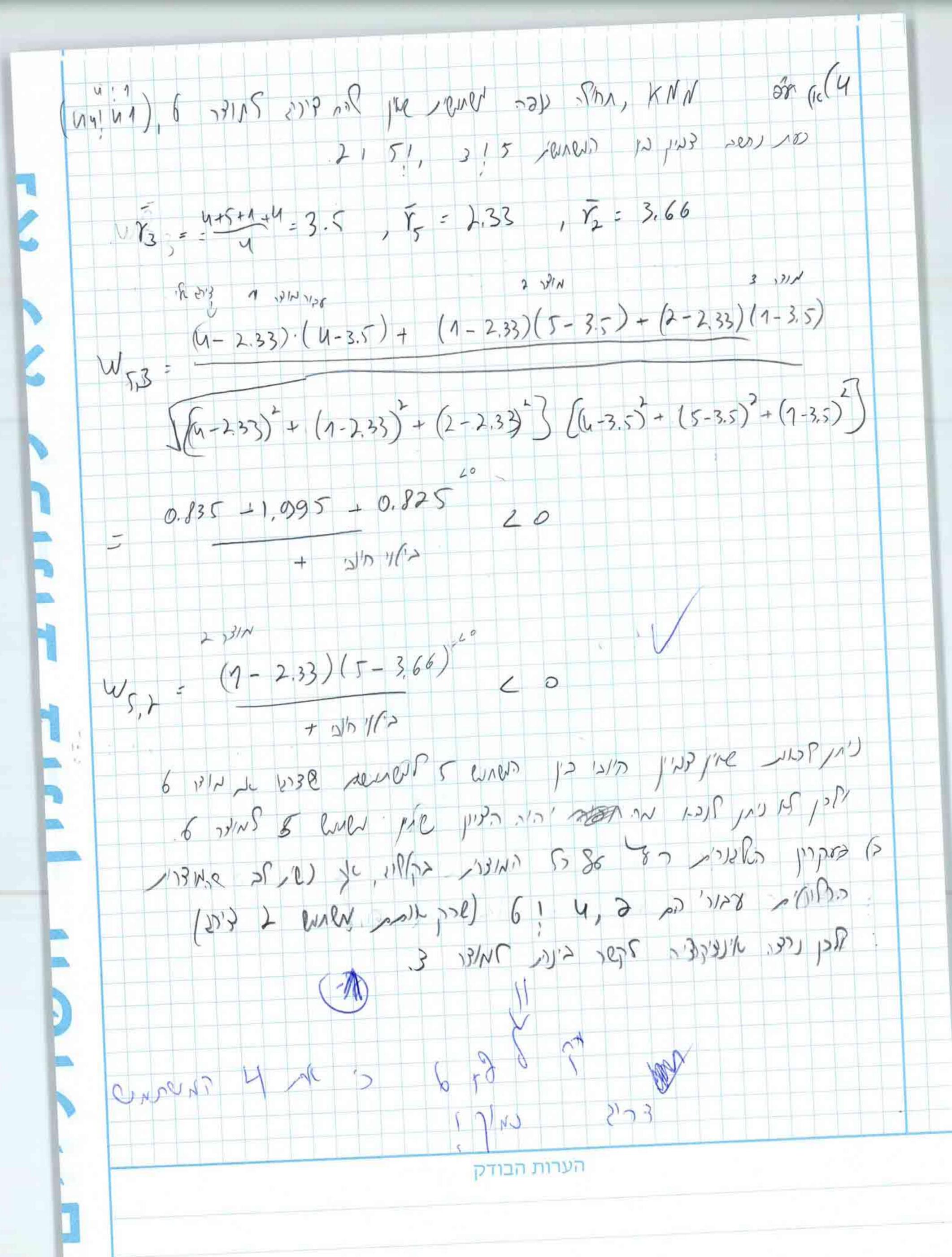
The state of the s	לשימוש המרצה הבודי	
	יחידות ועשרות ומאות	
	3	
	4	
	5	
	6	
	□	\prod
	9	
	ציון הבחינה	
	שם המרצה	
	חתימה	
	תאריך	

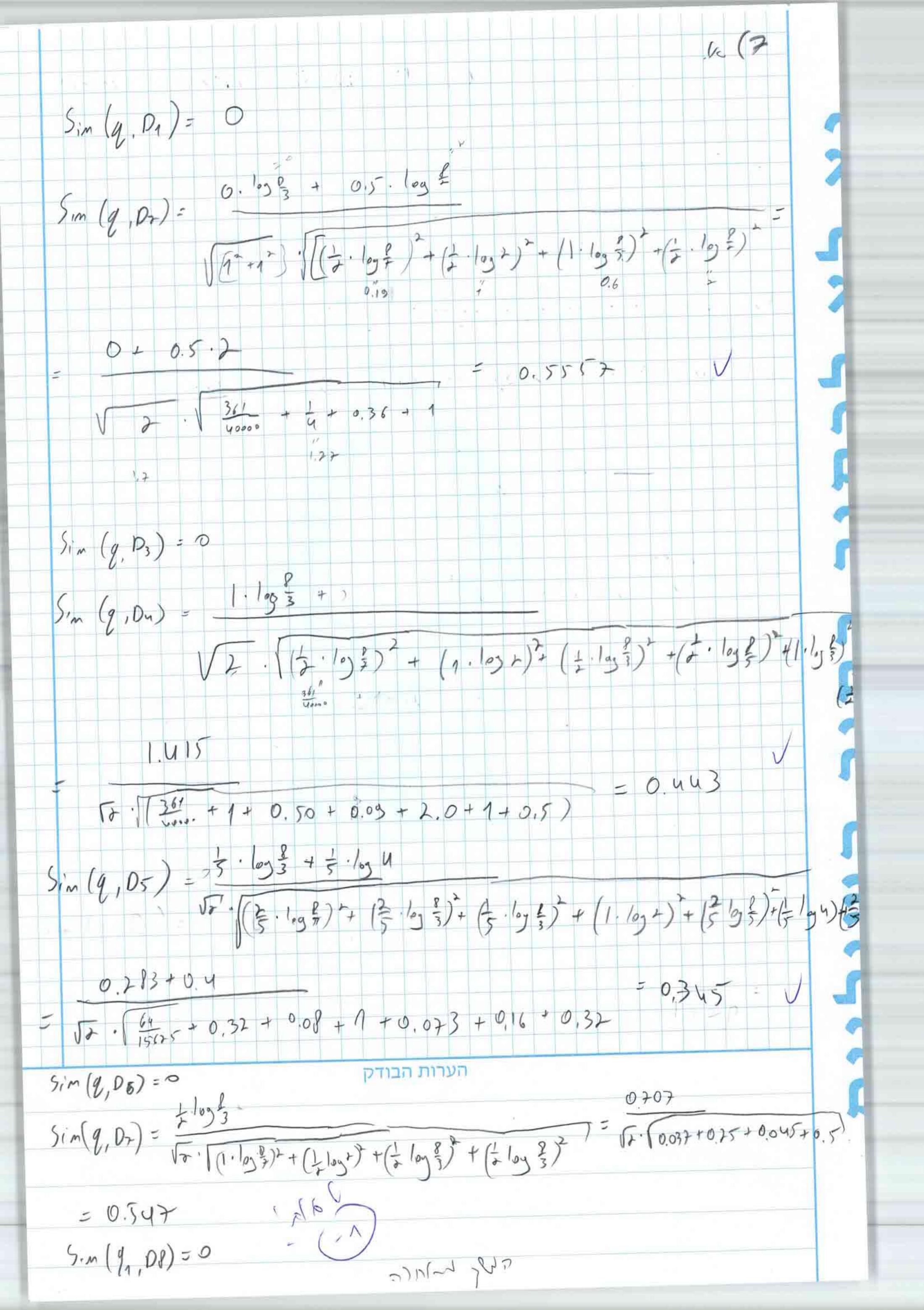


הוראות לנבחן

- בהגיע הסטודנט למקומו ,עליו להניח את כרטיס הנבחן ותעודה מזהה עם תמונה על השולחן.
- חל איסור על הכנסת כל חומר לבחינה , למעט זה שהותר במפורש ע"י המורה וצויין ע"ג טופס הבחינה לגבי סטודנטים עם התאמות מיוחדות -ראו התייחסות נפרדת עזב הסטודנט את חדר הבחינה לאחר חלוקת השאלונים , דינו כדין" נבחן" בבחינה.
- .4 אסור לנבחן לשוחח בזמן הבחינה , או לעזוב את מקומו ללא נטילת רשות. סטודנט רשאי לבחור אם להשתמש בעט או בעפרון. במקרה שכתב הבחינה בעפרון והבחינה לא נסרקה ,לא יהיה זכאי לערער על הבחינה . אם הבחינה נסרקה ,יהיה רשאי לערעד על הבחינה ובתנאי שאיכות הסריקה תקינה .יש לכתוב בכתב ברור ,נקי וקריא ,על עמוד אחד של כל דף . אין לכתוב בשוליים <u>ואין להשתמש בטיפקס</u> . טיוטא תיכתב בצד ימין של המחברת זיש למחוק אותה בהעברת קו . אסור לתלוש דפים מו המחברת.

7/2 Ins or A:(x)=H:(x)= CM (18/16/ MM - MTM = nxn להוקונה הימחגית nx1 A - 12 H 3 0718 S 1101 GD 100 DD 100 PR(1) = 0 AB BALLED 13 PR(s) . (1) 100 000 5 8 77,0 710 5057 46 1 8 860 1237 71-6 18/M , 22431 1. HD CIN 211/1/1/ 164 16 (((())) // 1/ 1/6) הערות הבודק



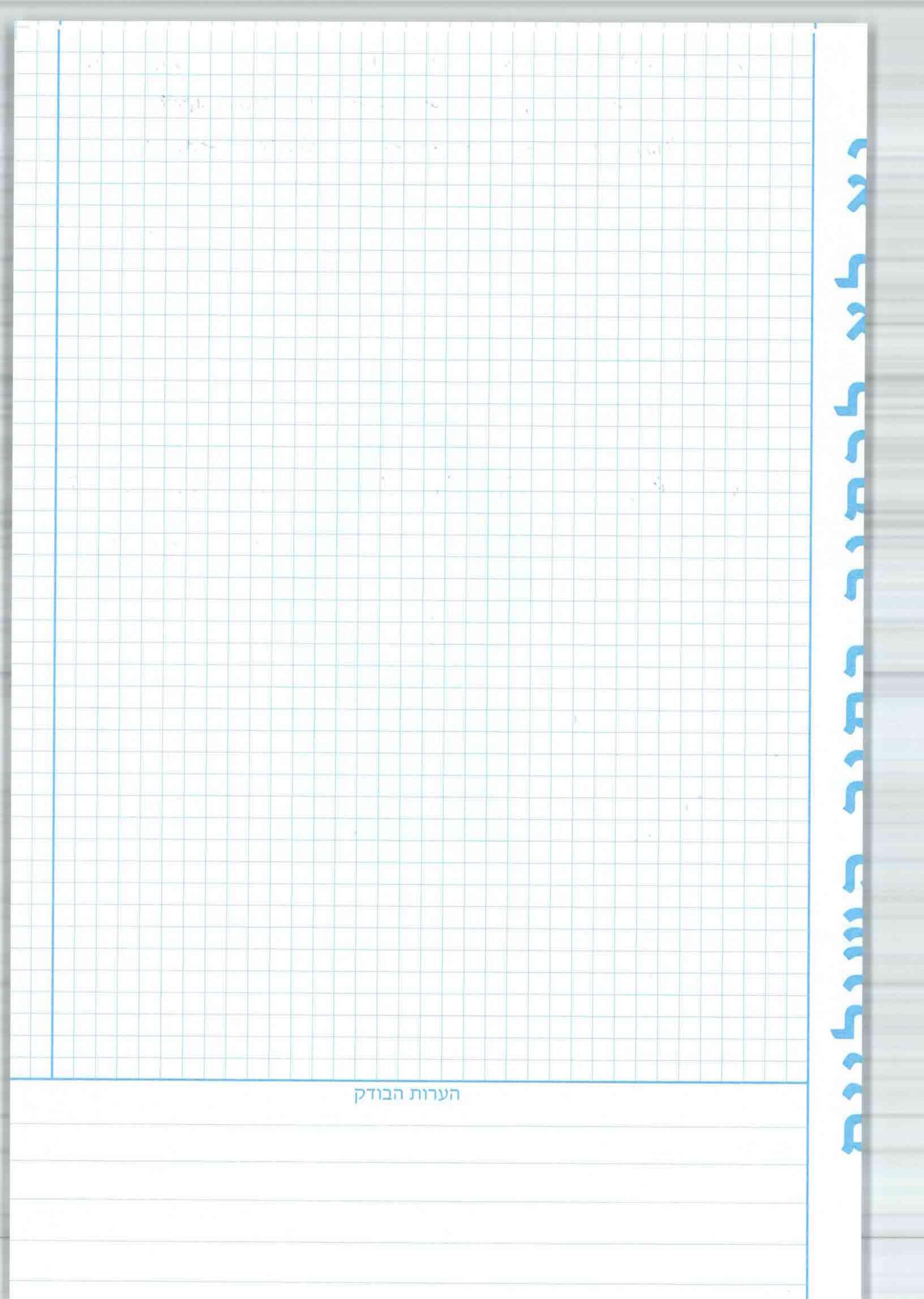


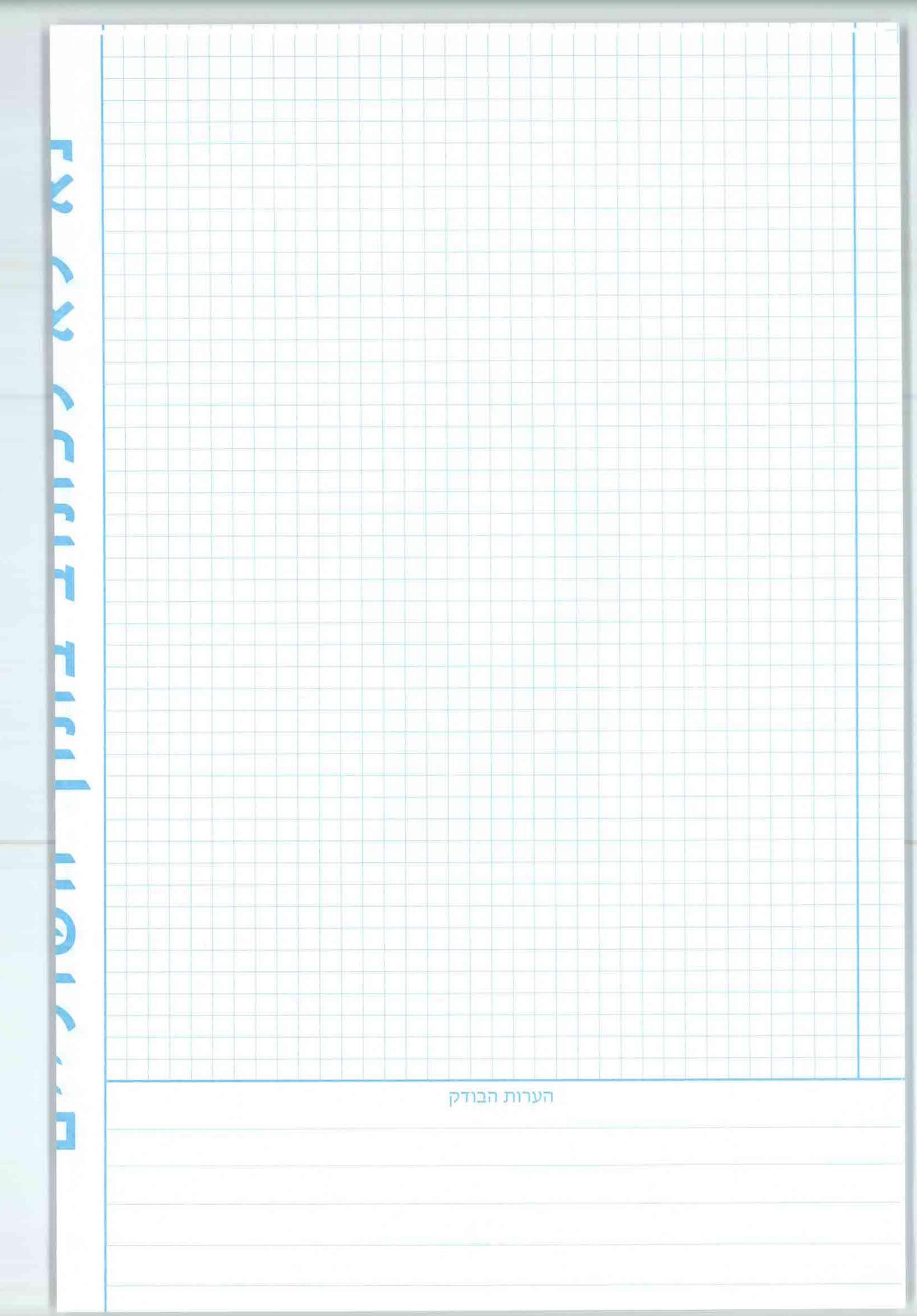
D5, D4, D2, D2 :112512 //2021,7 (0,0,0,0,1,0,0,0,1,0) 0, (1, 1, 0, 2,0,0,0,0,1,0) (1,2,1,1,2,0,2,1,0,0) 10 1015 15 E /"UTN 165 . コルカン = (0,0,0,0,2,0,0,0,0) + (1,1.5, 1,5,1,0,1,1,1,2,0) = 1= (1,1,5, \(\frac{1}{2},1,5,\), \(\frac{1}{2},\), \(\frac{1},\), \(\frac{1}{2},\), \(\frac{1}{2},\), \(\frac{1}{2},\), \ הערות הבודק

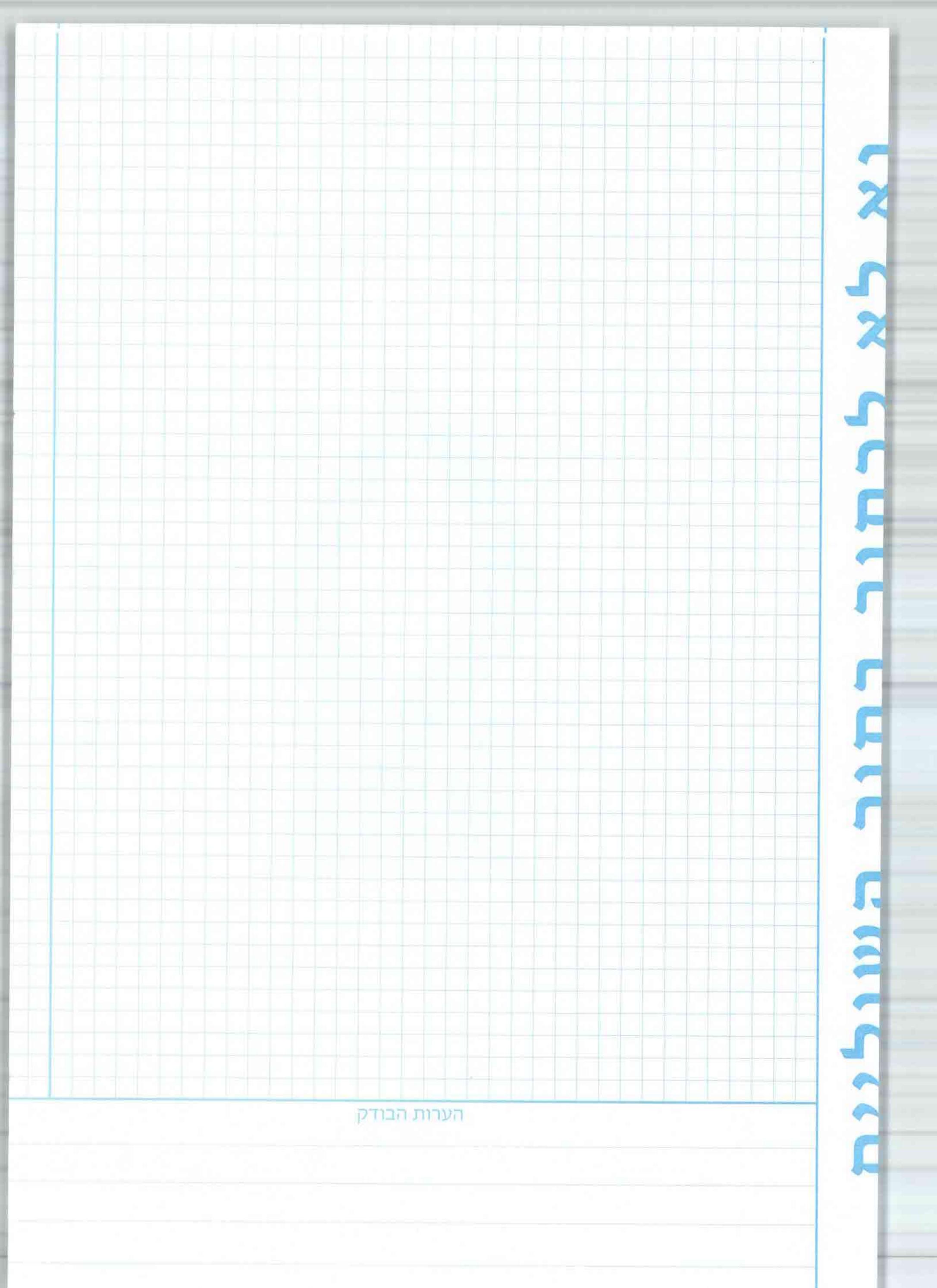
7 MB 26/5 120 203 1128 334 -15 16/6 55 60 (26) 9 (1/2 1/2) 2/2 Men 1 / 1/2 / 12/25 M' 61/3 (3) 1/20/1), 20/26 / 18 SER 200 320 05 pel , more e some e sono se sono con sono se Cost (150) 100/ 100/ (150) 0 mass p1 6000 cus 13 110 mod rom ph P("who shot linder") - P(wholm) > p (shoot In) + g(Grows In) = 4 = 25000 + 25000 prevot zera probubility : shows p(uld) ka mas & 2) prevent cases where terms with (relatively) high probability "diminate" the netribal presess - pipipa Dar will KL (9,0)= - & P(w/9). log(w/d) (1) 5 (2) 1), 1 _ ~ 237) הערות הבודק

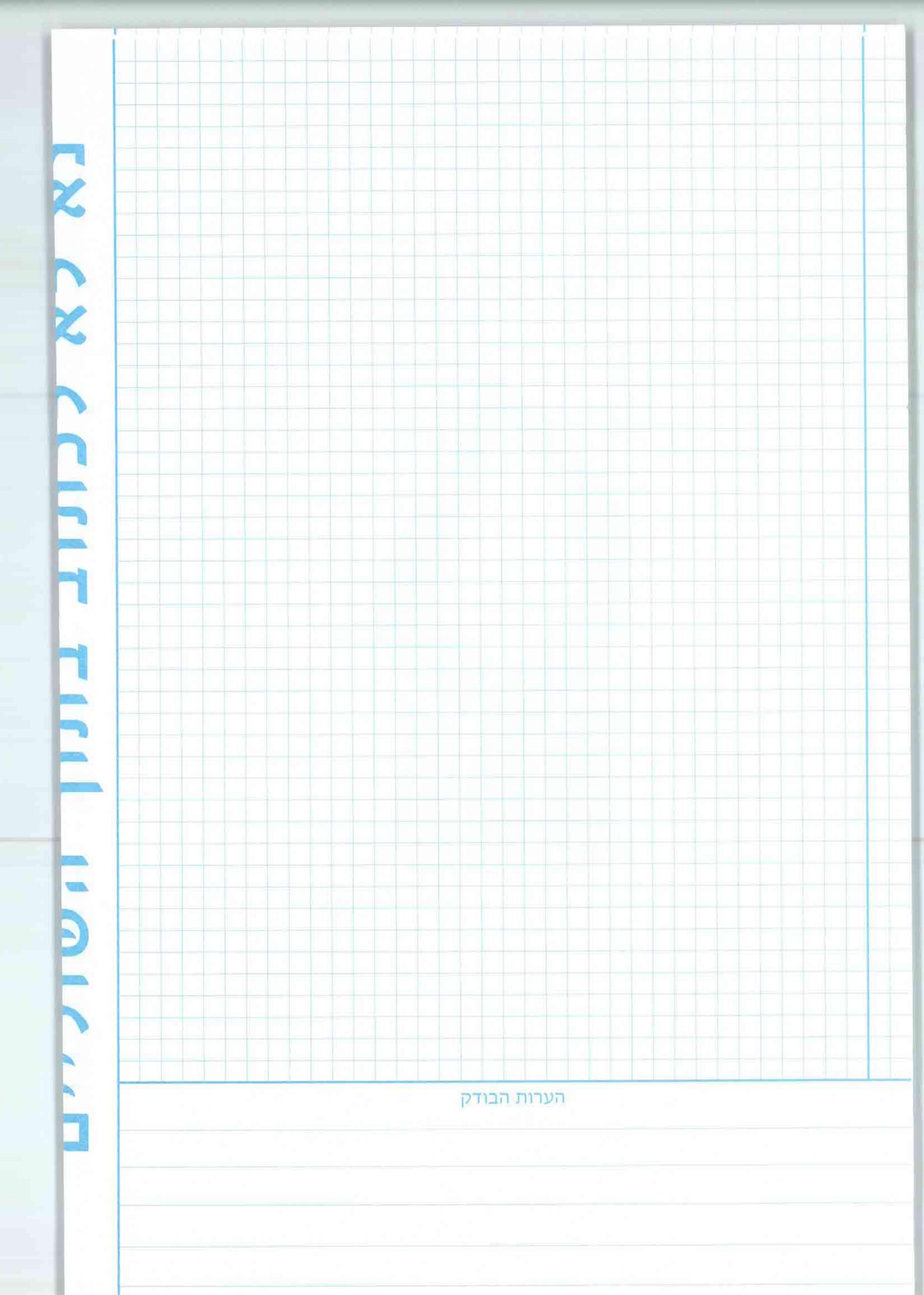
173/1/1 Miker Mus - ON -1805 TOE YOR 1365 - gull on while were my fine - A COSCI M PW PW POND -128'LEIN - Puis = ternser torn i A Mel, -mos will kastern spare willing in (U) tom 2 2 1/2 (6/20) (1) 4/1) Les (1/20) 100) 200) Phi ple 1874, and will will allen some inform 100 (5 137), PI 13'MP, 1310) D'UEN IL 1810, D'UN MATKEN 9/10/20 1.20/10 2/10 SIDI (1/1) PICIE NOTO 18e 1910) (1) forms N-1/2 205-1 51/2 * (() ()) M'EL - 60 - 97/0'(1 / 1/2) SAD MONEY "MY ((())) (()) * 2000 MUNA 2016 -PW B 20 100/20 1/20 Je 6 SHC -PINI THE ISE NEWS - SHOWN NEW 1'D gris 1821 - 2.00 mones (100m) 200 1800 1000 000 000 9 = 307, 6 = 1007) , 2012 GIN MICHIN WOON B 19 11 all one onight form sing, on 15 1/1/2 (1817) - (100) promp (7), (7) 10) 3 2018/11 7 /13 (270 JUNG 1NIC (101) (10 JUS) -188/18-1000 DO

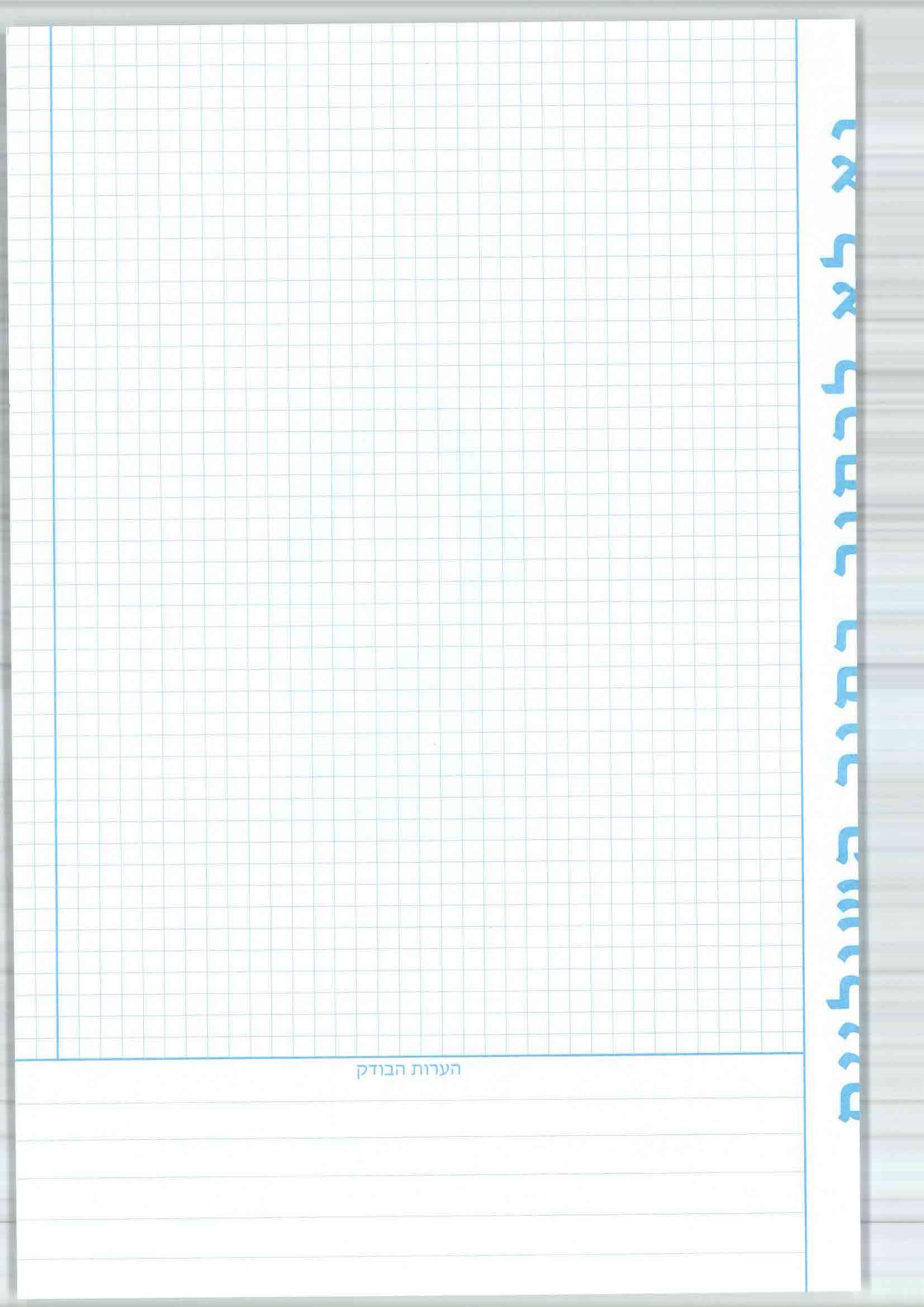
p1845 -840 -8401 Roll 118 (5) 2017 -118 (5) (5) 95/50 (18 11 N/3'N/ (18) MOIS DE181. (DOD 1879) 12 /310) 157 toras 13 2 5 (2000 NOWN) 1-1-078/11/ 1/33 /11/62 of south 133 11/2/69/2 (11/2/60) 21/1 1/1/6 416,9p 22/2 22/10 13/10/10/10/2010 120/20120 Ce10535 01610) Rap 210, 1000 080 Allean, 115 des 11000 1100 100/61 MMC/11 (co) (DOC NUM, 1/0, ping), (Doc Num, 1/0, ping). ((0= 5) ((Le)) 19/4 CAU DOC > 6, 100) 20 22/20 (180) 1000 URD MIRA ph/ 56000 1018' KID, 10/02 8'0/N 1600 ph, term K 01.0 MRIC HONE HOLD (4000 FON LOLD KIC WEN, 1) הערות הבודק

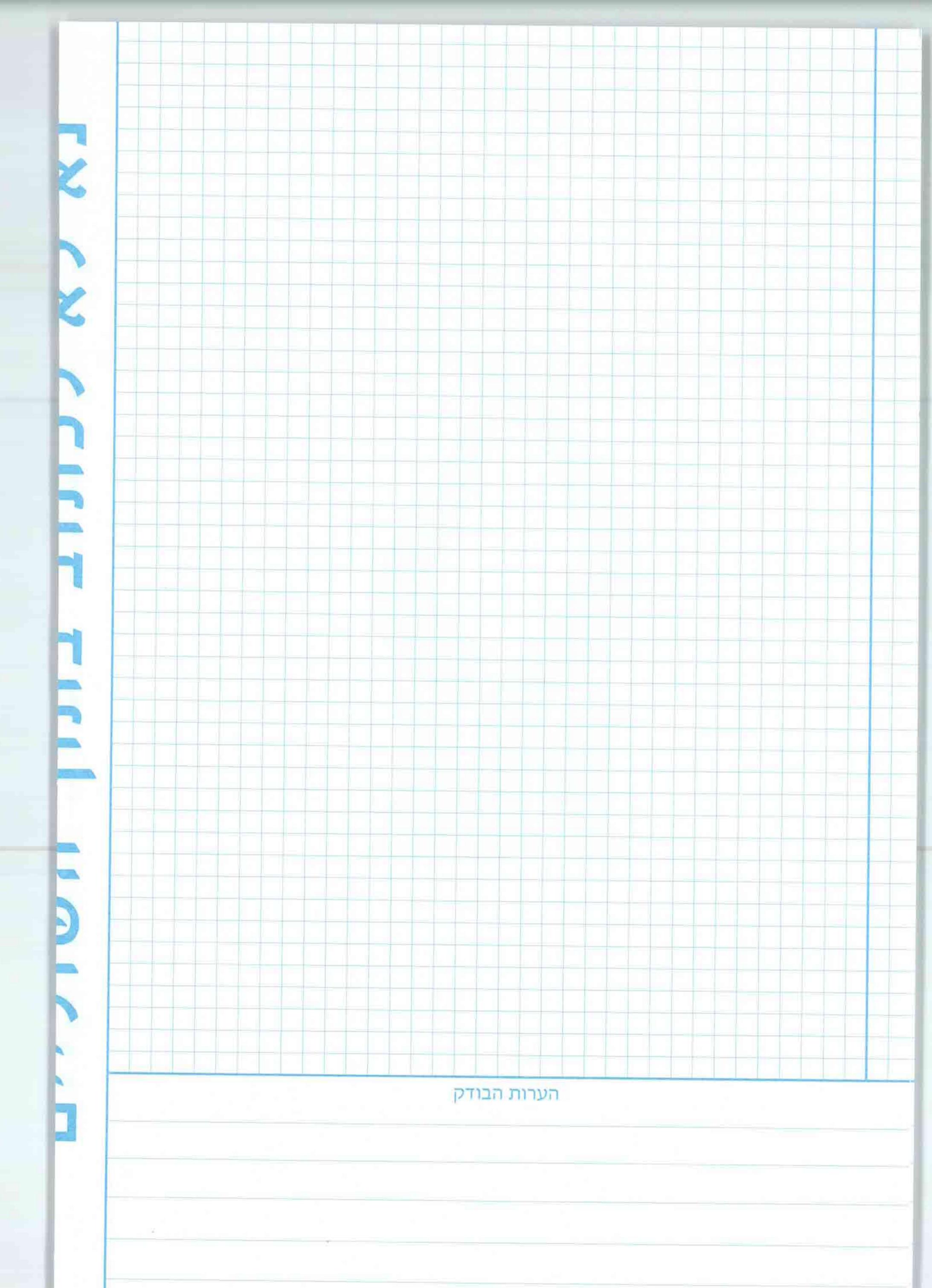


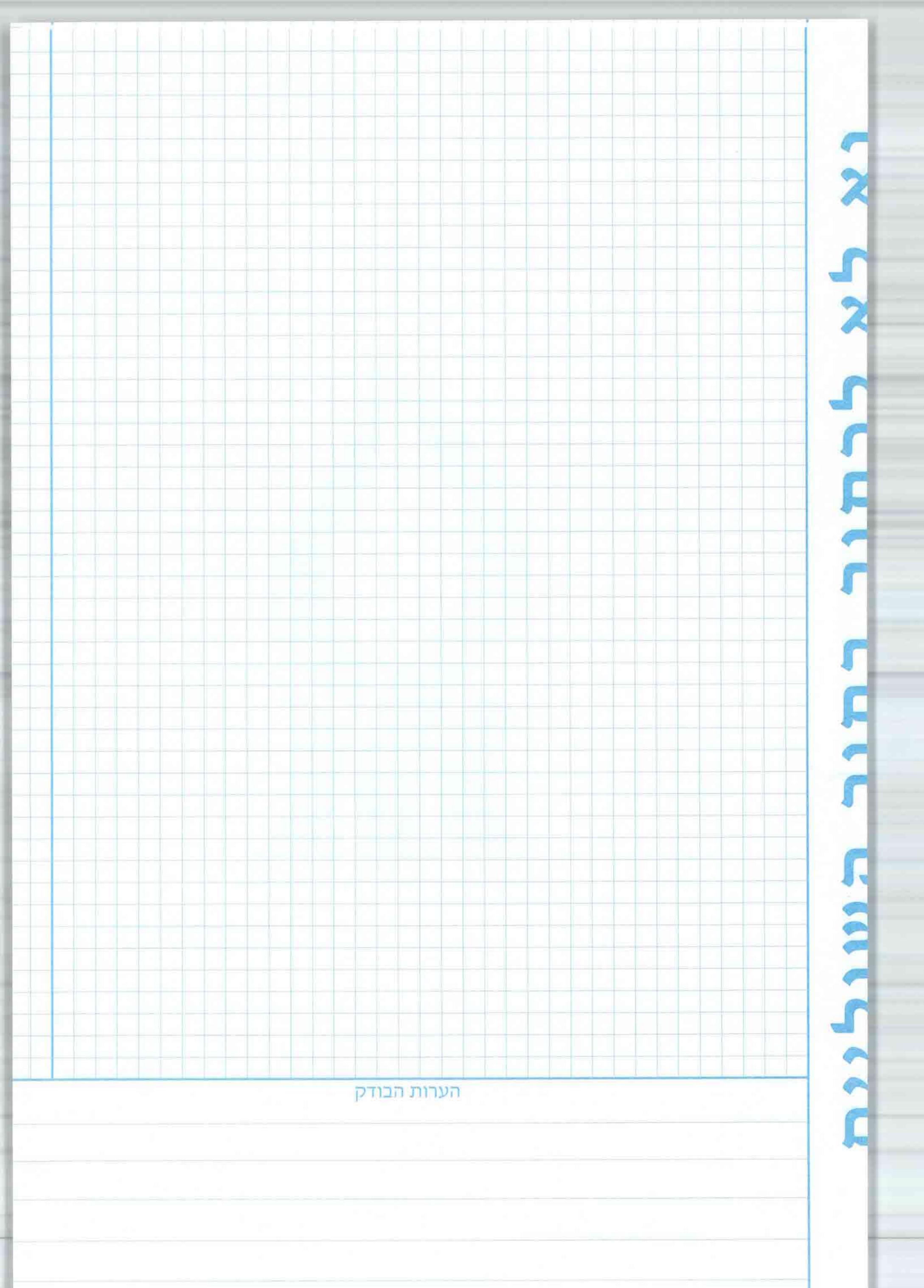


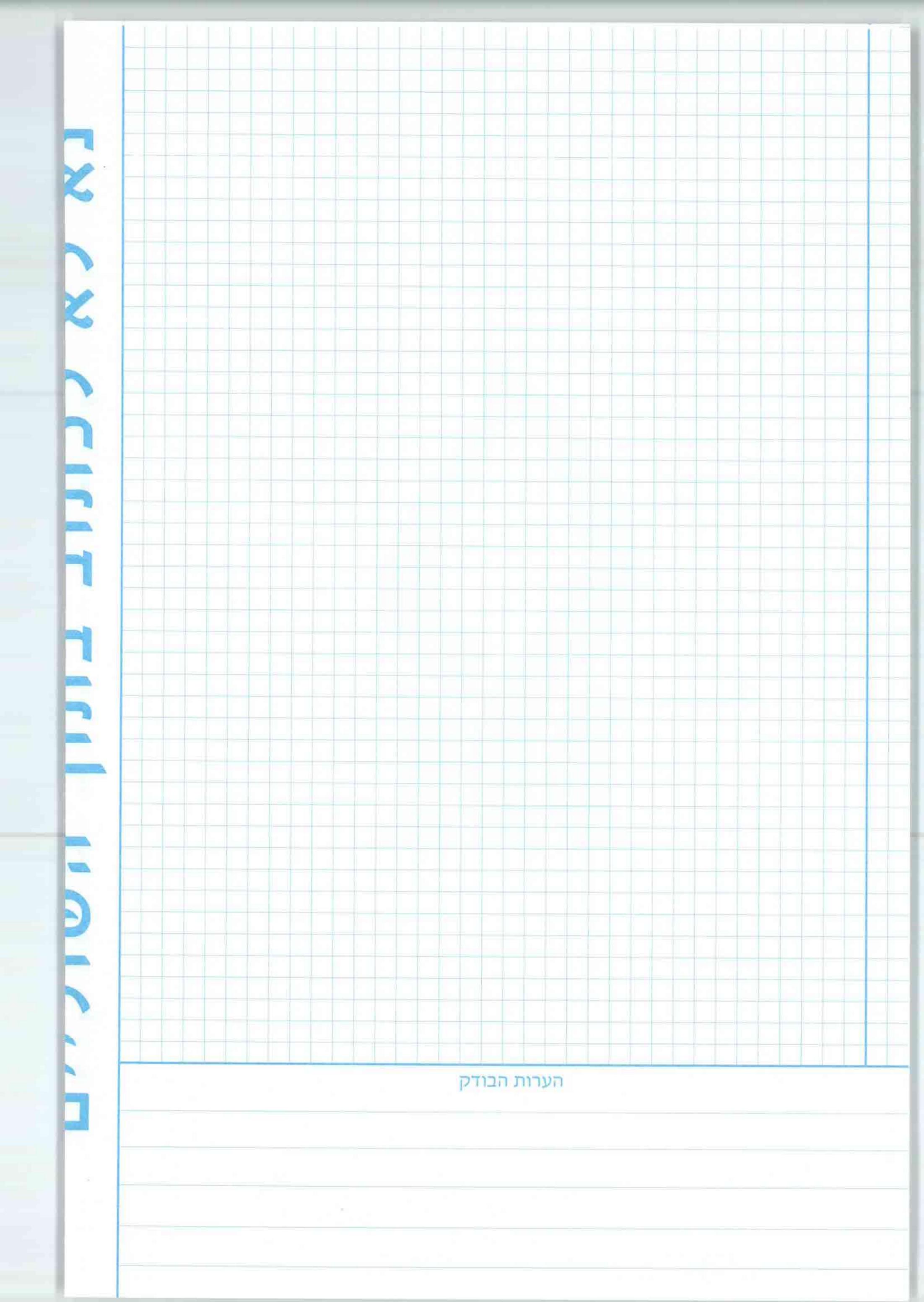


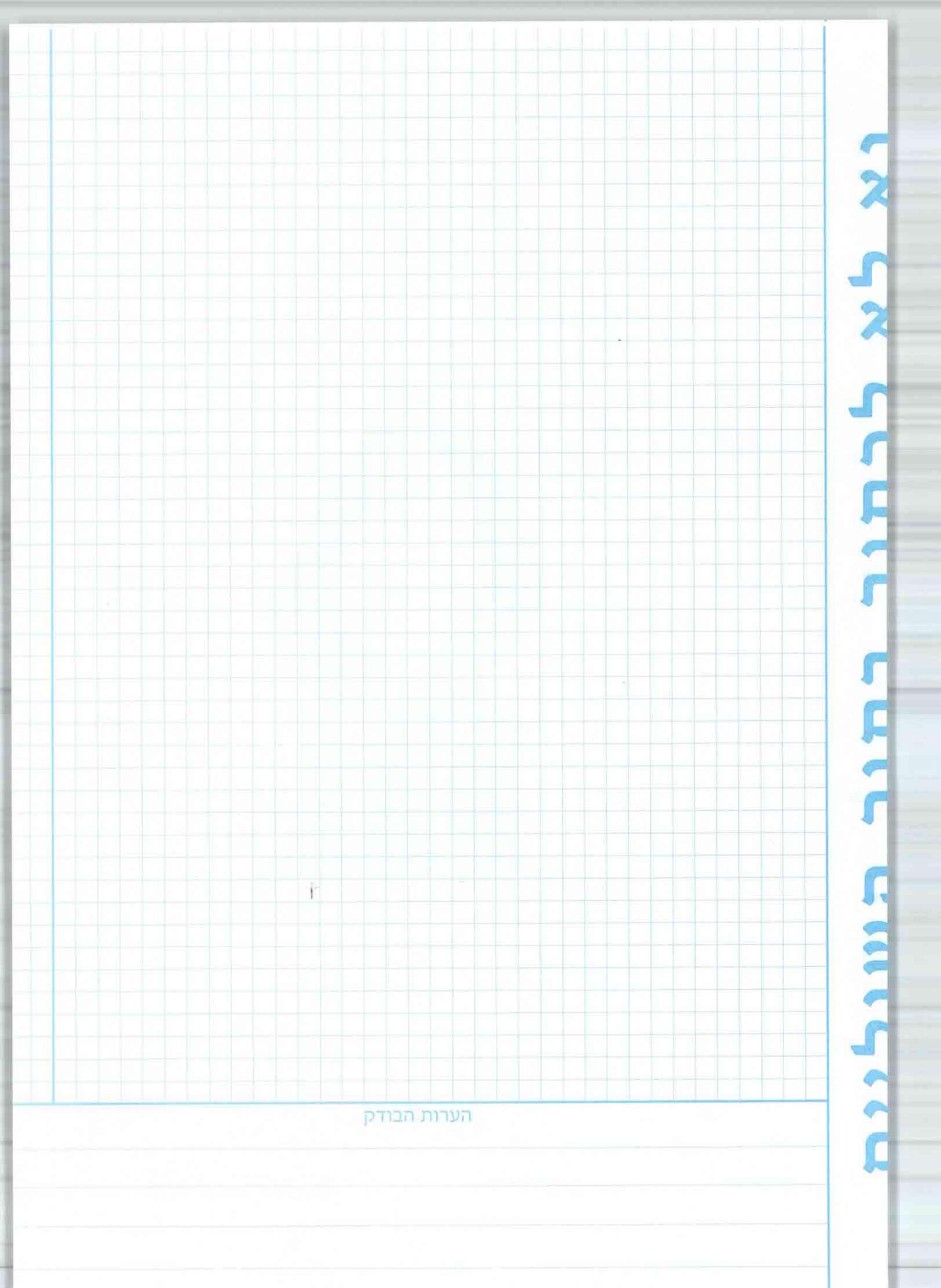


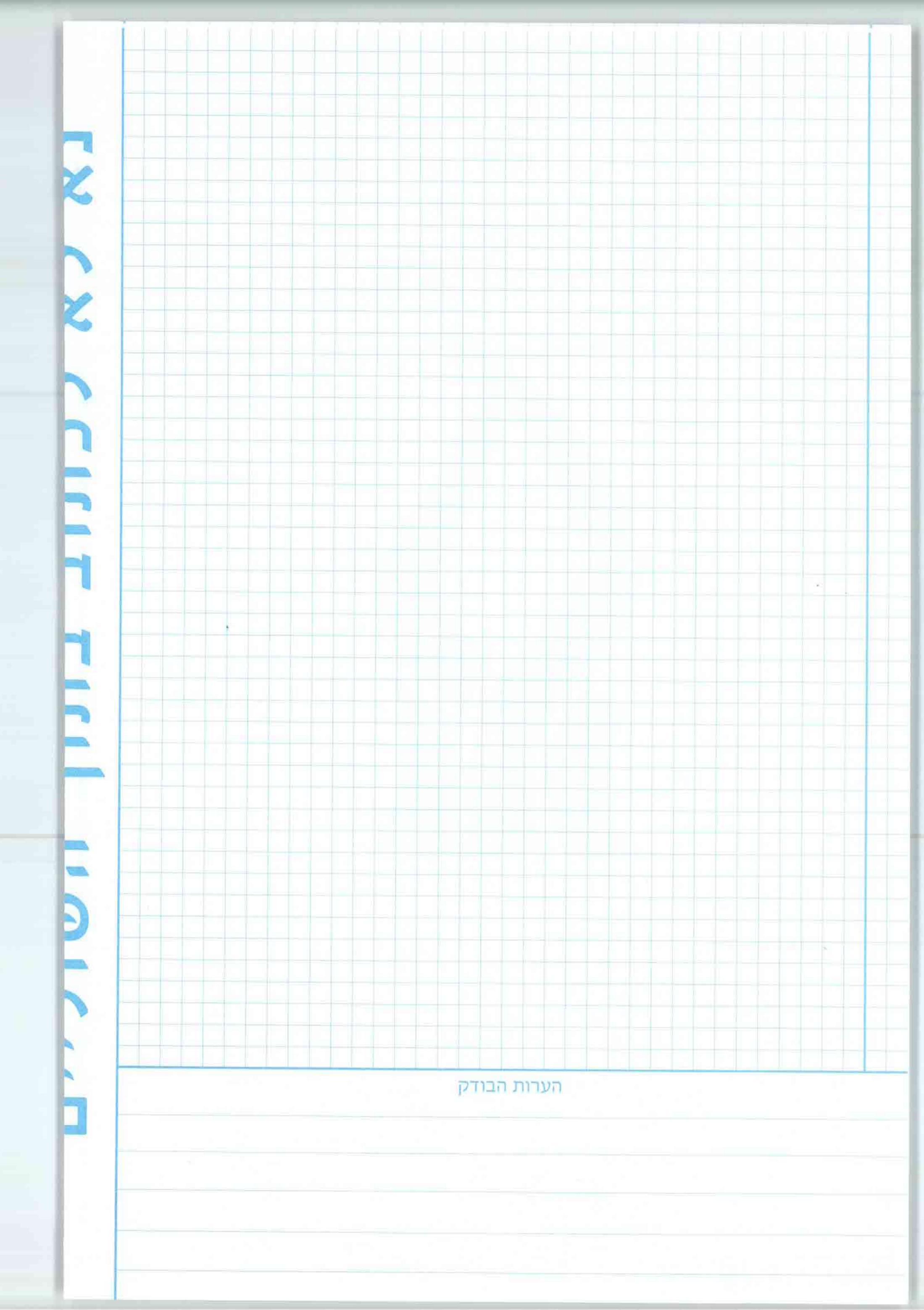


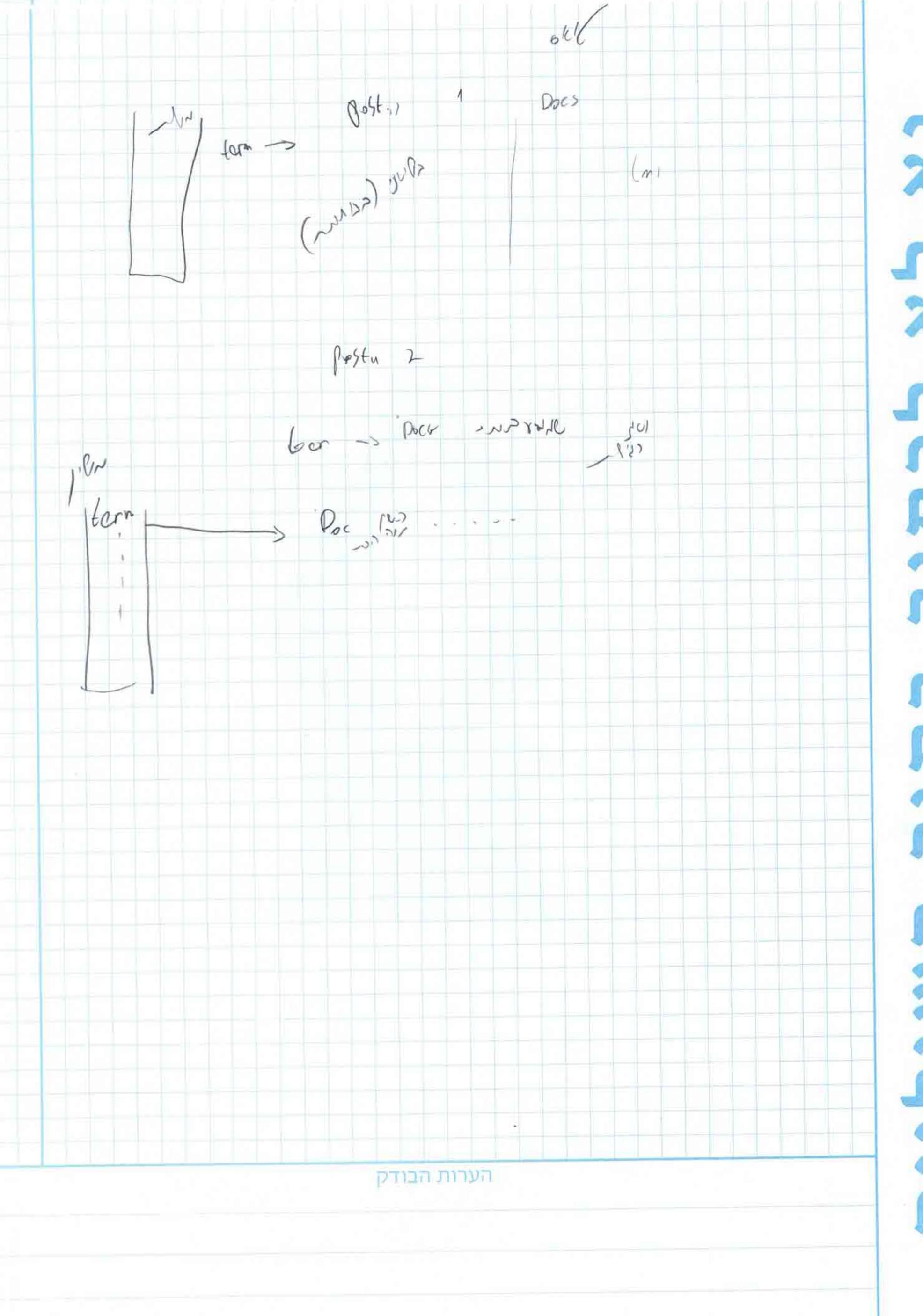












715/ 31/11/ [112 Jaan 1030(6) (Cm) (cm) PR(1) = 0 PR(2) = = = PR(3) PR(3) = pR(n) PR(4) = PR(5) + = PR(2) הערות הבודק PR(5) = = = PR(2) 425

3×3 +3×9 (3+3-3)/ (3) 81

