

סיכום תוצאות ניסוי מפע"ס 2
(מבחן פסיכומטרי על מחשב)

אביטל מושינסקי
המרכז הארצי לבחינות ולהערכה

יוני 2000

תוכן העיניינים

3	תמצית.....
4	מבוא.....
5	שיטה.....
7	תוצאות.....
7	תקלות בניסוי.....
7	יצוגיות המדגם.....
8	התפלגות הציונים בניסוי.....
10	ניתוח קבצי הנבחנים בבחינה הממוחשבת.....
12	זמני תגובה.....
16	דגימת פריטים.....
18	ציוני משתתפי הניסוי בבחינה הפסיכומטרית.....
25	הבדלים בין המינים.....
28	תוצאות המשובים.....
40	סיכום.....
42	מקורות.....

תמצית

ניסוי הגרסה השנייה של מפע"ים (בחינה מותאמת פסיכומטרית על מחשב) נערך במהלך שבועיים בסוף אוגוסט ובתחילת ספטמבר 99. הניסוי נערך במתכונתו של ניסוי מפע"ים 1 (הלר ומושינסקי, 2000). לניסוי זומנו נבחנים שנרשמו לבחינה הפסיכומטרית במועד אוקטובר 99. בניסוי השתתפו 363 נבדקים, שחולקו אקראית לקבוצת בחינה ממוחשבת ולקבוצת נייר ועפרון. ציוני הנבדקים בניסוי הושוו לציוניהם בבחינה הפסיכומטרית חודש לאחר מכן. ממצאי הניסוי דומים לממצאים שנתקבלו בניסוי מפע"ים 1. ממצאי הניסוי מראים כי בחינת מפע"ים שקולה לבחינת נייר ועפרון, וכי אופנות ההבחנות, או מין הנבחנים אינם משפיעים על הציון. הנבדקים בבחינה הממוחשבת הביעו שביעות רצון מן הבחינה וממידת בהירותה. תקלות שהופיעו בניסוי מפע"ים 1 טופלו, ולא חזרו על עצמן.

מבוא

בחינת מפע"ם (מותאמת פסיכומטרית ע"י מחשב) הינה בחינה פסיכומטרית ממוחשבת אדפטיבית המיועדת לנבחנים בתנאים מותאמים (במו"ת). הבחינה כוללת שתי גרסאות (1 ו-2). בכל גרסה שלושה תחומים (מילולי, כמותי ואנגלית). כל תחום מכיל כ-150 פריטים, הלקוחים ברובם משישה פרקים תפעוליים של בחינת מכפ"ל, המייצגים את כל סוגי הפריטים המופיעים בבחינה הפסיכומטרית ובתוספת השלמות של פריטים קשים ואשכולות.

דו"ח זה מתאר את תוצאות ניסוי מפע"ם 2. ניסוי זה נערך במתכונת ניסוי מפע"ם 1. תוצאות הניסוי עובדו על סמך התוכניות ששימשו בניסוי מפע"ם 1, והדוח נכתב אף הוא במתכונת הדוח המסכם של תוצאות ניסוי מפע"ם 1 (הלר ומושינסקי, 99). הבדלים קלים בין הניסויים יפורטו בהמשך במקום המתאים לכך בדוח.

מטרות הניסוי היו כדלקמן :

- ✓ בדיקה של תהליך ההיבחנות במבחן הממוחשב קודם להוצאתו לראשונה של מבחן פסיכומטרי ממוחשב תפעולי; כולל איתור תקלות אפשריות בזמן ההיבחנות או מצבים שלא נצפו מראש ולא התגלו בסימולציות או בהתנסויות שנערכו.
- ✓ בדיקת האינטראקציה בין הנבחנים והמבחן : בהירות ההוראות וההסברים, מידת העצמאות של הנבחנים בהפעלת המבחן ומידת שביעות רצונם של הנבחנים מן המבחן. בדיקת התאמת זמני התגובה שהוגדרו לפריטי המבחן לנבחנים.
- ✓ בדיקה של כיוול ציוני הבחינה הממוחשבת לציוני מכפ"ל.
- ✓ בדיקת מאפייני המבחן- העברת המבחן למדגם נבחנים, אשר אפשרה לעקוב אחר מאפייני המבחן : הציונים, אומדן הטעות, זמני תגובה, שיעורי דגימת הפריטים, אורך המבחן ומשך ההיבחנות.
- ✓ השוואת נתוני המבחן לנתוני הסימולציות, כולל השוואה של ההתפלגויות של אורך המבחן, טעות אומדן היכולת, ושיעורי דגימת הפריטים.
- ✓ השוואה בין המינים לגבי ביצועיהם במבחן הממוחשב, תוך השוואה של הבדלים אלו להבדלים בין המינים במבחן נייר ועפרון (להלן : ני"ע) בניסוי ובבחינת מכפ"ל.

שיטה

מדגם

הזמנה להשתתף בניסוי נשלחה לנרשמים לבחינה הפסיכומטרית במועד אוקטובר 99, אשר:

- ❖ נרשמו להבחן בירושלים (כ- 400 איש) ול- 900 מקרב הנרשמים להבחן בת"א, אשר נדגמו אקראית מתוך כ- 3500 נרשמים.
- ❖ נרשמו להבחן בבחינה בעברית.
- ❖ לא נבחנו קודם בבחינה הפסיכומטרית, וזאת כדי למנוע תוספת בציון של היבחנות חוזרת.

מקבלי ההזמנות נתבקשו להתקשר למרכז הארצי ולתאם מועד לניסוי. המתקשרים הוקצו אקראית לאחת משתי קבוצות הניסוי: הממוחשבת או הנייר ועיפרון. המבחן נערך בשני מקומות: בירושלים במרכז הארצי לבחינות בתאריכים 27/8, 29/8-31/8, 1/9, 3/9 (סי"ה 4 ימי חול של שני מחזורי נבחנים, ושני ימי שישי של מחזור נבחנים אחד), ובתל-אביב באוניברסיטת תל-אביב בתאריכים 2/9, 5/9-9/9 (סי"ה 6 ימי חול של שני מחזורי נבחנים). נבחנים בודדים התקשרו והביעו רצון להשתתף בניסוי על אף שלא קיבלו הזמנות. נבחנים אלו שובצו למקומות שנתרו פנויים, אלא אם נבחנו קודם לכן, והוקצו אקראית לאחת משתי קבוצות הניסוי: הממוחשבת או הנייר ועיפרון. חלוקה זו לימי ניסוי בירושלים ובתל-אביב נעשתה לאור הפקת הלקחים מניסוי מפע"ם 1 שבו נמצא כי לאור גודלו של מאגר האנשים בתל-אביב קל שם יותר למלא את הכיתות. לפיכך הוסף יום ניסוי בתל-אביב והורד יום ניסוי בירושלים. סה"כ השתתפו בניסוי 363 נבחנים, מתוכם 181 (50%) נבחנו בבחינה הממוחשבת ו- 182 (50%) נבחנו בבחינת ני"ע. שיעורי ההשתתפות בניסוי היו דומים לניסוי מפע"ם 1 בו השתתפו 352 נבחנים, מתוכם 184 (52%) נבחנו בבחינה הממוחשבת ו- 168 (48%) נבחנו בבחינת ני"ע.

מכשירים:

קבוצה ממוחשבת - 15 מחשבים ניידים (Compaq Armada 1590DT) של המרכז הארצי לבחינות והערכה עליהם הותקנה בחינת מפע"ם 2, דיסקטים של גיבוי, טפסי משוב, דפי נוסחאות, תעודות ציון. קבוצת ני"ע - חוברות של נוסחי מכפ"ל 77 ו- 78, תשובונים, טפסי משוב, דפי נוסחאות, תעודות ציון.

מערך:

המשתתפים בניסוי חולקו אקראית לשתי קבוצות, קבוצת ני"ע וקבוצת הבחינה הממוחשבת. באופן אקראי, ניתן למחצית (90 נבחנים כ- 50%) מנבחני קבוצת ני"ע נוסח 77, וליתר (92 נבחנים כ- 50%) נוסח 78 (ללא החלק"בים). ציוניהם של משתתפי הניסוי בבחינה הפסיכומטרית מאוקטובר 99 אותרו והשוו לציוני המשתתפים בניסוי בקבוצת הבחינה הממוחשבת ובקבוצת הני"ע. מערך מחקר זה מאפשר להעריך את ההשפעה של היבחנות חוזרות, ולהשתמש בבחינת מכפ"ל כעוגן, אשר מאפשר לבדוק את טיב הציונים בבחינת

מהלך:

הנבחנים לא ידעו מראש לאיזה קבוצה ישובצו, אך עקב העברת אינפורמציה בלתי נמנעת מן הניסוי הקודם ומן היום השני לניסוי בין משתתפי הניסוי הנוכחי, הם ידעו כי חלקם יבחנו במחשב. לנבדקים לא ניתן לעבור בין הקבוצות, אף אם ביקשו. קבוצות הניי"ע והבחינה הממוחשבת נבחנו בחדרים נפרדים. בכל חדר נבחנו עד שניים עשר נבחנים. בכל יום בחינה היו שני מועדים: בוקר – 9:00, וצהרים ב- 13:00.

נבחני בחינת הניי"ע קיבלו באופן אקראי את נוסח 77 או נוסח 78 של הבחינה הפסיכומטרית. הם החלו וסיימו את הבחינה יחדיו. זמני הבחינה נשמרו ע"י בוחנים מיומנים באופן זהה לחלוקת הזמנים בבחינה הפסיכומטרית. בתום הבחינה נתבקשו הנבחנים למלא טופס משוב לבחינת ניי"ע (ראה נספח). נבדקי קבוצת ניי"ע המתינו כ- 15 דקות בממוצע, אגב מילוי טפסי המשוב, עד שחושב עבורם הציון ידנית (עיי "שבולונות") ובסיוע תוכנית מחשב שנכתבה במיוחד לשם כך.

נבחני הבחינה הממוחשבת נבחנו בבחינת מפע"ים 2. הם החלו את הבחינה יחדיו, אולם משך הבחינה היה שונה מנבחן לנבחן. בתום הבחינה, אך קודם לקבלת הציון, התבקשו הנבחנים למלא את שאלון המשוב הממוחשב (ראה נספח). עם תום מילוי שאלון המשוב קיבלו הנבחנים את ציונם על גבי צג המחשב.

יש להדגיש כי הן נבחני הקבוצה הממוחשבת והן נבחני קבוצת ניי"ע מילאו את שאלון המשוב קודם לקבלת הציון על מנת שהערכותיהם לא יושפעו מהציון שקיבלו במבחן.

תוצאות

תקלות בניסוי

תוכנות - בניסוי מפעיים 1 שלושה קבצי נבחנים לא נכתבו כלל (הקבצים היו בגודל 0KB ושלושה קבצי נבחנים נוספים היו חסרים חלקית). תקלות אלו קרו בשל סיום לא תקין של תוכנית מחולל המבחנים, ובשל אי סגירה ואריזה מלאה של אותם הקבצים. בעקבות ניסוי מפעיים 1 הוכנסו השינויים הבאים בתוכנת מחולל המבחנים : 1. סגירה תקינה של תוכנית מחולל המבחנים עם תום המבחן. 2. אריזה וסגירה של קבצי הפלט לאחר כל סוג פריט, כך שגם אם ישנה נפילה פתאומית של המחשב, הרשום בקובץ הנבחן אינו נמחק. תיקונים אלו הוכיחו את עצמם, ובניסוי מפעיים 2 כל קבצי הנבחנים נרשמו בצורה מלאה ללא חסרים. פסיכומטריקה – ביומו הראשון של הניסוי אחד הנבחנים מצא שגיאה בפריט השוואה כמותית. בבדיקה נמצא כי הפריט הועתק מן המקור באופן שגוי וכי הדבר לא נתגלה בכל ההגהות שנעשו. הפריט תוקן, ובהמשך הניסוי לא נתגלו תקלות נוספות. מחשבים נישאים – כונן הגיבוי של אחד המחשבים התקלקל באחד מימי הניסוי. באחד המועדים המחשב לא חובר כראוי לחשמל. המחשב נכבה כאשר אזלה הסוללה. לאחר החיבור לחשמל, המחשב הודלק וחזר לפריט בו הופסק המבחן.

יצוגיות המדגם

על מנת להעריך את יצוגיות המדגם נערכה השוואה בין ציוני משתתפי הניסוי לבין ציוני כלל הנבחנים במועד אוקטובר 99.

טבלה מס' 1: השוואת ציוני מכפ"ל מועד אוקטובר 99 של משתתפי הניסוי לציוני כלל הנבחנים במועד

נבחני מועד אוקטובר 99 (n=14,419)		נבחני הניסוי (n=363)		
ממוצע	ס.ת.	ממוצע	ס.ת.	
107.54	19.70	111.45	19.63	מילולי
112.50	18.82	111.61	18.68	כמותי
113.34	21.10	113.38	22.61	אנגלית
561.23	98.47	567.91	98.69	כללי

עיון בטבלה 1 מעלה כי ציוני קבוצת הניסוי דומים מאוד לציוני כלל הנבחנים במועד. הבדל של 4 נקודות בקירוב לטובת קבוצת הניסוי קיים בתחום המילולי, אך זהו הבדל קטן לפי Cohen (1988). ולכן נראה כי ניתן להכליל מנתוני המדגם לכלל אוכלוסיית הנבחנים (בשל ההבדל הגדול בין שתי הקבוצות לא נעשו מבחני מובהקות).

התפלגות הציונים בניסוי

תחילה יוצגו נתונים סטטיסטיים אודות ציוני הניסוי. בטבלה 2 מוצגים הציונים שהתקבלו בשתי קבוצות הניסוי, בטבלה 3 מוצגת השוואה בין ציוני הנבחנים בנוסחים 77 ו-78 של בחינת הניייע

טבלה מס' 2: התפלגויות ציוני התקן בניסוי

תחום	מילולי	כמותי	אנגלית	כללי
ציוני התקן במבחן מפע"ם 2 (n=181)				
ממוצע	109.52	111.33	111.87	561.37
ס.ת.	18.38	20.09	21.96	98.18
טווח	72-148	58-143	58-150	315-769
ציוני התקן בבחינת נייע בניסוי (n=182)				
ממוצע	113.37	111.89	114.87	574.99
ס.ת.	20.66	17.23	23.21	98.86
טווח	68-150	73-146	61-150	355-759

עיון בטבלה מעלה כי הציונים בבחינת נייע ובבחינה הממוחשבת דומים. ההבדל המקסימלי הוא של כחמישית סטיית תקן בין הקבוצות בתחום המילולי לטובת קבוצת הניייע. ההבדל בציון הכולל הוא קטן ועומד על כשמינית סטיית תקן בלבד (ההבדלים כולם אינם מובהקים $p > 0.05$).

במבחן נייע נבחנו 182 נבחנים: מתוכם 90 בנוסח 77 והיתר (92 נבחנים) בנוסח 78. בטבלה מספר 3 מוצגת השוואה של ציוני הנבחנים בנוסח 77 לעומת ציוני הנבחנים בנוסח 78.

טבלה מס' 3: השוואת ציוני הנבחנים בנוסח 77

לעומת נוסח 78

מבחן t		נוסח 78 (n=92)		נוסח 77 (n=90)		
ערך *t	df	ס.ת.	ממוצע	ס.ת.	ממוצע	תחום
-0.85	180	21.92	114.66	19.33	112.04	מילולי
-0.54	180	17.63	112.57	16.87	111.19	כמותי
-0.50	180	23.77	115.73	22.71	114.00	אנגלית
-0.86	180	103.31	581.20	94.24	568.64	כללי

* כל ההבדלים נמצאו לא מובהקים ברמת $p < 0.05$.

ניתן לראות כי בכל התחומים ובציון הכללי הנבחנים בנוסח 78 הנם בממוצע בעלי ציונים מעט יותר גבוהים

מאשר הנבחנים בנוסח 77, אולם ההבדלים בציונים אינם מובהקים סטטיסטית ברמת $p < 0.05$. יש לציין כי בניסוי מפע"ם 1 היו הבדלים לא מובהקים לטובת נוסח 77, כלומר, ההבדלים אינם עקביים לטובתו של נוסח מסוים, וממילא אינם עקביים.

בטבלה 4 מוצגות ההתפלגויות תטה במבחן הממוחשב: ציון התטה הבייסיאני המשמש בקביעת הציון, וציון

תטה לפי אמדן "הנראות המקסימלית" המחושב לצרכי מחקרים עתידיים. למעשה אין מדובר באומדן "נראות

מקסימלית" במובן המקובל, אלא בשילוב של שתי שיטות: המבחן עצמו, חישוב הציונים ודגימת הפריטים

מבוססים על האומדן הבייסיאני; בשלב שני, בקביעת הציון הסופי מחושב אומדן נראות מקסימלית עבור וקטור

התשובות של הנבחן, ללא שימוש בהתפלגות אפריורית.

טבלה מס' 4: התפלגויות ציוני התטה בקבוצת המבחן הממוחשב

(n=181)

ציון תטה – בייסיאני			
אנגלית	כמותי	מילולי	
0.45	0.53	0.46	ממוצע
1.00	1.09	1.10	ס.ת.
-2.10-2.38	-2.28-2.49	-1.83 -2.83	טווח
ציון תטה - אומדן נראות מקסימלית			
0.49	0.58	0.54	ממוצע
1.12	1.26	1.29	ס.ת.
-2.62-2.99	-2.92-2.99	-2.17-2.99	טווח

ניתן לראות מטבלה 4 כי יש דמיון בין שני סוגי האומדנים. ממצא זה אינו מפתיע, משום שכאמור אין מדובר באומדן נראות מקסימלית רגיל, אלא במבחן המתנהל עפ"י האומדן הבייסיאני. ממוצע ציוני התטה הבייסיאני מעט נמוך יותר מהממוצע של ציוני התטה של נראות מקסימלית בכל שלושת התחומים, וזאת משום שהאמידה בקצוות (בקרב בעלי היכולת הגבוהה והנמוכה) קרובה יותר לממוצע, ולעומת זאת שונות אומדן הנראות המקסימלית גבוהה יותר מאותה סיבה. תוצאות אלו עולות בקנה אחד עם תוצאות סימולציות שיטות הכיול השונות (הלר ורוין, 2000).

ההפרש בין ממוצעי התחום המילולי מעט גדול יותר מבתחומים האחרים, וכן גדול מבמפע"ם 1 (הלר ומושינסקי, 99, ע"מ 9). יש לבחון את טיב הציון המילולי גם ביחס לשאר הנתונים הנוגעים לו בהמשך. באשר לפיזור הציונים, הרי שכצפוי אומדני נראות מקסימלית בעלי פיזור רב יותר (סטיית תקן וטווח גדולים יותר) מאשר האומדן הבייסיאני כפי שנמצא גם בניסוי מפע"ם 1.

ניתוח קבצי הנבחנים בבחינה הממוחשבת

בבחינה הממוחשבת נבחנו 181 איש. בטבלה 5 מוצגים נתונים כלליים אודות בחינת מפע"ם 2 בחלוקה לתחומים השונים. בטבלה מוצגים מספר השאלות הממוצע במבחן, אומדן הטעות - השונות הפוסטרירורית, ומספר הנבחנים שערו הטעות הסופי שחושב עבורם היה גבוה מהסף שנקבע ("נבחנים שלא התכנסו").

טבלה מס' 5: נתונים אודות המבחן

תחום	מילולי	כמותי	אנגלית
מס' נבחנים	181	181	181
מס' שאלות ממוצע (טווח השאלות)	37.59 (32-38)	30.28 (28-33)	22.51 (21-28)
% נבחנים שקיבלו מספר שאלות מינימלי*	0	41%	62%
% נבחנים שקיבלו מספר שאלות מקסימלי**	79%	35%	13%
ממוצע השונות הפוסטריורית – אומדן הטעות*** (ס.ת.)	0.14 (0.02)	0.11 (0.03)	0.08 (0.02)
נבחנים שמבחינהם לא התכנסו- (שערך הטעות שלהם לא ירד מתחת לסף)	68 (37.6%)	41 (22.7%)	17 (9.4%)

* מספר שאלות מינימלי: במילולי 31 שאלות, בכמותי 28 שאלות, באנגלית 21 שאלות.

** מספר שאלות מקסימלי: במילולי 38 שאלות, בכמותי 33 שאלות, באנגלית 28 שאלות.

*** ערך הסף של הטעות הוא: במילולי 0.13, בכמותי 0.10 ובאנגלית 0.08.

השוואה של נתונים אלו לתוצאות סימולציה של 1,500 נבחנים מהתפלגות נורמלית (מושניסקי ורון, 98) מצביעה על נתונים דומים לגבי התחום המילולי באשר למספר השאלות הממוצע 37.5, ולממוצע השונות הפוסטריורית - 0.14. אך יש לציין כי אחוז הנבחנים שציוניהם לא התכנסו בניסוי גבוה באופן משמעותי מבסימולציות - 38% לעומת 19%. עם זאת יש לזכור שהסימולציות נערכו על 1,500 איש שיכולתם מתפלגת נורמלית באופן מושלם. כמו כן ניתן לראות כי גם בסימולציות וגם בניסוי רוב רובם של הנבחנים מקבלים את מירב השאלות בתחום המילולי, כלומר ההתכנסות במפעיים 2 בתחום המילולי היא בעייתית, ואינה תלויה בתוצאות הניסוי דווקא. ככל הנראה הדבר קשור בעיקר לכך שהתחום המילולי אינו לגמרי חד ממדי, שכן בעיית אי הנורמליות של האוכלוסיה אינה באה לידי ביטוי בתחומי האנגלית והחשיבה הכמותית.

השוואת הנתונים בין ניסוי מפעיים 2 ובין הסימולציות דומים גם בתחום הכמותי: מספר שאלות ממוצע של 30.4, ממוצע שונות פוסטריורית של 0.10 ואחוז הנבחנים שלא התכנסו- 17.2%. בתחום האנגלית מספר השאלות הממוצע וממוצע השונות הפוסטריורית דומים (22.9 ו-0.08, בהתאמה), ואחוז הנבחנים שמבחינהם לא התכנסו שנמצא בסימולציות (8%) היה דומה לזה שנמצא בניסוי (9.4%).

בניסוי מפעיים 1 נתוני הניסוי היו דומים לנתוני הסימולציות, למעט ביחס לאחוז הנבחנים שלא התכנסו באנגלית (בניסוי אחוז זה היה גבוה פי שניים מבניסוי). ניתן אפוא להכליל ולומר כי נתוני שני הניסויים דומים מאוד לנתוני הסימולציות למעט אחוזי התכנסות בתחום זה או אחר, המושפעים ככל הנראה מגודל המדגם ומהשונות שלו, אשר אינם מתפלגים בניסוי או במציאות באופן נורמלי מושלם כמו בסימולציות. שוב, אם כן ניתן לראות כי ממצאי

הסימולציות אמינים וניתן להסתמך עליהם בבניית מבחנים, וכך שהנתונים על בסיסם נבנתה גרסת מפע"ם 2 התאמתו, ולפיכך אכן ניתן לקבל את גרסת המבחן כפי שנבנתה ללא שינוי. עם זאת ייתכן שיש לייצר התפלגות יכולות הדומה יותר למציאות, ולאו דווקא להתפלגות נורמלית.

זמני תגובה

נתונים כלליים

בחלק זה, יוצגו נתונים אודות זמני התגובה במבחן מפע"ם. בטבלה 6, מוצגים פרקי הזמן הממוצעים שארכו חלקי השונים של מבחן מפע"ם. סטיות התקן מצוינות בסוגריים.

טבלה מס' 6: ממוצעי הזמנים בדקות במבחן מפע"ם 2

n=181

טווח	אחוז מתוך תהליך ההבחנות כולו	ממוצע מפע"ם 1 (ס.ת.)	ממוצע מפע"ם 2 (ס.ת.)	(הזמנים מוצגים בדקות)
75.57-228.32	100%	144.31 (36.09)	136.58 (31.74)	תהליך ההבחנות כולו
22.91-74.93	31%	44.38 (12.91)	42.58 (10.67)	פתרון שאלות המבחן בתחום המילולי
27.42-97.53	39%	56.37 (15.82)	53.75 (14.79)	פתרון שאלות המבחן בתחום הכמותי
11.60-56.25	22%	30.55 (8.95)	30.00 (8.72)	פתרון שאלות המבחן בתחום האנגלית
3.57-22.50	8%	12.84 (6.21)	10.25 (4.19)	הזמן שהוקדש לקריאת הוראות

הערות הבהרה לטבלה 6:

1. זמן פתרון שאלות במבחן כולל גם את זמן קריאת הקטעים הרלוונטיים (קטעי הבנת הנקרא וקטעי הגיון בתחום המילולי וקטעי הבנת הנקרא בתחום האנגלית).
2. הזמן הממוצע לפתרון שאלות המבחן אינו כולל את זמן קריאת ההסברים לשאלות.
3. הזמן הממוצע לתהליך ההיבחנות כולו, כולל מלבד זמן קריאת ההוראות וזמן לפתרון של שאלות, גם את זמני חישוב הציונים, דגימת שאלות והכנת מסכים לתצוגה.

ניתן לראות כי פרק הזמן הארוך ביותר שנמשך מבחן היה כשלוש שעות ושלושת רבעי (כ- 2 סטיות תקן מעל הממוצע), ופרק הזמן הקצר ביותר שנמשך מבחן היה כשעה ורבע. משך הזמן הממוצע למבחן היה שעתים ורבע (במפעים 1 - כשעתיים ו- 24 דקות). מתוך זה, כ- 39% מהזמן מוקדש לפתרון השאלות בתחום הכמותי, 31% מהזמן מוקדש לפתרון השאלות בתחום המילולי ו- 22% מהזמן מוקדש לפתרון השאלות בתחום האנגלית. לקריאת ההוראות נבחנו מקדישים בממוצע 8% מהזמן. חלוקת זמן זו זהה לחלוקת הזמן שנתקבלה בניסוי מפעים 1.

התפלגות זמני תגובה לפי סוג פריט

פרקי הזמן שהוקדשו לסוגי הפריטים השונים מוצגים עבור כל תחום, בטבלאות 7א', 7ב' ו- 7ג'. זמן הביצוע לפריט נמדד מרגע הופעת הפריט על המסך ועד לאישור התשובה לפריט על ידי הנבחן בעזרת המקש Enter, או לחלופין, עד לתום הזמן המוקצב לשאלה. בטבלאות מוצגים הממוצע וסטיות התקן של זמן התשובה, הממוצע וסטיות התקן של אחוז ניצולת הזמן, השוואה של הממוצעים וסטיות התקן בתשובות נכונות לעומת תשובות שגויות, המתאם בין נכונות התשובה (נכון/לא נכון) וזמן התגובה. בנוסף, מוצג מספר (אחוז) הפריטים בהם מנוצל פחות מ- 75% מהזמן. הנתון האחרון מוצג משום שבבניית המבחן הייתה כוונה לא להגביל (או לפחות לא להלחיץ) את הנבחנים בזמן, הזמן שנקבע היה נדיב ביותר ואמור היה להספיק לנבחנים כדי להשיב ולהיות בטוחים בתשובתם. זאת בין השאר כפיצוי על חוסר היכולת לחזור אחורה.

טבלה מס' 7א': ממוצע זמן תגובה בשניות לפריטי מילולי לתשובות נכונות ושגויות

סוג פריט (הזמן המוקצב בשניות)	השלמת משפטים (120) n=1086	אנלוגיות (90) n=1917	הגיון בדיד (180) n=724	אשכול הגיון (240) n=543	הבנת הנקרא (240) n=905	שורשים (180) n=724	מלים וביטויים (60) n=904
ממוצע זמן תגובה (ס"ת)	76.75 (26.67)	39.34 (21.50)	98.98 (42.81)	102.52 (53.69)	84.62 (50.14)	63.90 (41.36)	20.92 (14.65)
ממוצע אחוז ניצולת הזמן (ס.ת.)	64 (22)	44 (24)	55 (24)	43 (22)	35 (21)	36 (23)	39 (24)
מספר (אחוז) הפריטים בהם מנוצל פחות מ- 75% מהזמן	741 (68%)	1673 (87%)	559 (77%)	484 (89%)	849 (94%)	665 (92%)	829 (92%)
ממוצע זמן תגובה לתשובה נכונה (ס.ת.)	73.39 (25.88)	35.54 (45.76)	94.31 (110.04)	97.67 (50.91)	80.03 (46.64)	55.77 (35.40)	16.38 (12.83)
ממוצע זמן תגובה לתשובה שגויה (ס.ת.)	83.97 (26.94)	45.71 (23.36)	120.12 (48.85)	112.45 (65.28)	105.80 (60.37)	106.83 (52.33)	28.62 (14.34)
מתאם בין נכונות התשובה וזמן תגובה	-0.18	-0.23	-0.17	-0.15	-0.19	-0.42	-0.41

באופן כללי נראה כי הזמן לפתרון השאלות בתחום המילולי הוא מספק, בממוצע הנבחנים מנצלים 35%-64% מהזמן המוקצב לפריט, וב- 94%-68% מפריטי התחום המילולי עושים הנבחנים שימוש ב- 75% מהזמן או פחות. יוצאי דופן הם פריטי השלמת משפטים, בהם מנצלים הנבחנים בממוצע 64% מהזמן, רק ב- 68% מן הפריטים מנצלים הנבחנים פחות מ- 75% מהזמן, ויתכן שיש לשקול הוספת זמן לפריטים אלו. במיוחד לאור נתון זה נמצא גם בניסוי מפעיים 1 וגם בניסוי CPETSE (רפ, רוני וכהן, 1996). יש להזכיר כי הזמן שהוקצב להשלמת משפטים ולהגיון בדיד היה זהה בניסוי מפעיים לזה שהונהג בהעברה הניסיונית של ה"קומפיוטרי" (180 שניות בהגיון בדיד ו- 120 שניות בהשלמת משפטים), וכי אף רפ ואחרים (1996) המליצו על תוספת זמן של לפחות דקה בהקצבת הזמנים לפריטים אלו. נראה כי לאור ממצאי ניסוי מפעיים 1 ו- 2 יש לשוב ולהמליץ על תוספת של דקה לפריטי השלמת משפטים.

עוד ניתן לראות כי בכל התחומים הממוצע (ופיזור) זמן התגובה של הנבחנים שהשיבו נכונה על פריט קטן מזה של הנבחנים ששגו בפריט. כל המתאמים בין נכונות התשובה וזמן התגובה, בפריטי המילולי, מובהקים סטטיסטית. בפריטי שורשים, מילים וביטויים ואנלוגיות יש אפקט גדול של ידיעת/אי ידיעת התשובה על זמן התגובה (מתאמים של -0.42, -0.41, -0.23, בהתאמה). אפקטים חזקים אלו נובעים מהמרכיב של "תובנה" הקיים בפריטי השורשים, וממרכיב הידע הקיים בפריטי מילים וביטויים ואנלוגיות. גם נתונים אלו דומים מאוד לנתונים שנתקבלו בניסוי מפעיים 1.

טבלה מס' 7: ממוצע זמן תגובה בשניות לפריטי כמותי

לתשובות נכונות ושגויות

שוג פריט (הזמן המוקצב בשניות)	אלגברה	גיאומטריה	גרפים וטבלאות	השוואה כמותית
n=2765 (240)	n=905 (240)	n=799 (240)	n=1086 (240)	
ממוצע זמן תגובה (ס"ת)	104.76 (61.95)	128.44 (64.75)	92.87 (55.65)	87.64 (55.48)
ממוצע אחוז ניצולת הזמן (ס.ת.)	44 (25)	54 (27)	40 (27)	37 (23)
מספר (אחוז) הפריטים בהם מנוצל פחות מ- 75% מהזמן	2,367 (86%)	684 (76%)	648 (90%)	992 (91%)
ממוצע זמן תגובה לתשובה נכונה (ס.ת.)	100.29 (57.12)	122.36 (58.80)	88.89 (53.21)	88.30 (55.36)
ממוצע זמן תגובה לתשובה שגויה (ס.ת.)	115.92 (68.20)	141.14 (74.19)	115.62 (63.18)	86.46 (55.76)
מתאם בין נכונות התשובה זמן תגובה	-0.12	-0.14	-0.20	*0.02

* המתאם אינו מובהק ברמת $p < 0.05$

נראה כי הזמן המוקצב לפתרון השאלות בתחום הכמותי מספק (לפי רפ, רוני וכהן, 1996). הנבחנים מנצלים בממוצע 37%-54% מהזמן המוקצב לפריט, וב- 91%-76% מפריטי הכמותי עושים הנבחנים שימוש ב- 75% מהזמן או פחות. עוד ניתן לראות כי בפריטים הכמותיים ממוצע הזמן שמקדישים נבחנים לתשובות שגויות עולה על הזמן שהם מקדישים לתשובות נכונות. כל המתאמים בין נכונות התשובה וזמן התגובה מובהקים סטטיסטית, למעט פריטי השוואה כמותית. בפריטי אלגברה, גיאומטריה, גרפים וטבלאות יש אפקט של ידיעת/אי ידיעת התשובה על זמן התגובה (מתאמים של -0.12, -0.14, -0.20, בהתאמה). בשאלות מסוג ההשוואה הכמותית ממוצע הזמן המוקדש לתשובות שגויות דומה מאד לזה המוקדש לתשובות הנכונות. נתונים אלו דומים מאוד גם הם לנתונים שנתקבלו בניסוי מפע"ם 1.

טבלה מס' 7ג: ממוצע זמן תגובה בשניות לפריטי אנגלית לתשובות נכונות ושגויות

סוג פריט	ניסוח מחדש	השלמת משפטים	הבנת הנקרא
(הזמן המוקצב בשניות)	(240)	(120)	(240)
ממוצע זמן תגובה (ס"ת)	98.39 (48.27)	49.46 (25.51)	72.98 (46.95)
ממוצע אחוז ניצולת (ס.ת.)	41 (20)	41 (21)	30 (20)
מספר (אחוז) הפריטים בהם מנוצל פחות מ- 75% מהזמן	1,339 (93%)	1,580 (92%)	869 (96%)
ממוצע זמן תגובה לתשובה נכונה (ס.ת.)	93.74 (44.95)	44.61 (23.88)	67.50 (43.94)
ממוצע זמן תגובה לתשובה שגויה (ס.ת.)	111.27 (54.48)	59.07 (25.93)	94.00 (52.05)
מתאם בין נכונות התשובה זמן תגובה	-0.16	-0.27	-0.23

* כל המתאמים מובהקים ברמת מובהקות $p < 0.05$.

נראה כי הזמן המוקצב לפתרון שאלות בתחום האנגלית מספק. הנבחנים מנצלים בממוצע 41%-30% מהזמן המוקצב לפריט, וב 96%-92% מפריטי האנגלית עושים הנבחנים שימוש ב- 75% מהזמן או פחות. גם בפריטי האנגלית ממוצע הזמן שמקדישים נבחנים לתשובות שגויות עולה על הזמן שהם מקדישים לתשובות נכונות. כל המתאמים בין נכונות התשובה וזמן התגובה מובהקים סטטיסטית. בפריטי השלמת משפטים יש אפקט גדול של ידיעת/אי ידיעת התשובה על זמן התגובה (מתאם של -0.27). נתונים אלו דומים מאוד לנתונים שנתקבלו בניסוי מפע"ם 1 (רי ע"מ 15).

הקריטריון המומלץ לקביעת הזמן המוקצב לפריט של רפ ואחרים (1996) היה: "זמן פריט יקבע כך ששלושה רבעים מן התגובות לסוג הפריט יעשו בפחות מכמחצית הזמן המוקצב לסוג השאלה". לפי קריטריון זה

היו במפע"ם 1 ארבעה סוגי פריטים בעייתיים: בתחום המילולי - השלמת משפטים, והגיון בדיד, ובתחום הכמותי - גיאומטריה ואלגברה. בניסוי הנוכחי אותם סוגי הפריטים בעייתיים, אם כי יש לציין שבמידה שונה. בפריטי אלגברה על 86% מהפריטים עונים הנבחנים במהלך 75% מהזמן המוקצב, בפריטי גיאומטריה על 76% מהפריטים עונים הנבחנים במהלך 75% מהזמן המוקצב, בפריטי השלמת משפטים על 68% מן הפריטים עונים הנבחנים במהלך 75% מהזמן המוקצב, ובפריטי הגיון על 77% מן הפריטים עונים הנבחנים במהלך 75% מהזמן המוקצב. מנתונים אלו עולה שוב כי עיקר הקושי הוא בפריטי השלמת משפטים, ובמידה פחותה יותר בפריטי הגיון בדיד וגיאומטריה. בכל סוגי הפריטים נבחנים העונים תשובה לא נכונה עושים שימוש בזמן רב יותר מאשר המשיבים נכונה. פיזור הזמנים של השוגים אף הוא רב יותר מאשר של המשיבים נכונה. ממצאים אלה נמצאו גם בניסוי מפע"ם 1 ובהעברה הניסיונית של מבחן ה"קומפיומטרי" שם נמצא כי בכל סוגי הפריטים היה זמן התשובה ארוך יותר בתשובות השגויות מאשר בתשובות הנכונות והפיזור גדול יותר. השערות באשר לסיבות לכך וחקירה מעמיקה יותר של כל סוג פריט בנפרד ייבדקו בעתיד.

דגימת פריטים

בחלק זה, נבדקה שכיחות הדגימה של פריטי המאגר במבחנים. בטבלאות א', ב', ג', מוצגים נתונים אודות התפלגות החשיפה של פריטי המבחן בחלוקה לתחומים.

טבלה מס' א': דגימת פריטים בתחום המילולי

סוג פריט	השלמת משפטים	אנלוגיות	הגיון בדיד	אשכול הגיון	הבנת הנקרא	שורשים	מלים וביטויים
מס' פריטים	33	43	28	21	35	16	38
מס' פריטים שנדגמים פחות מ- 2.5%, כולל אלו שאינם נדגמים	1 (3%)	1 (2%)	2 (7%)	0	0	0	3 (8%)
אחוז דגימה ממוצע לפריט* (ס.ת.)	18% (13%)	25% (18%)	14% (12%)	14% (10%)	14% (8%)	25% (13%)	13% (10%)
מספר פריטים הנדגמים למעלה מ- 40% (אחוז)	2 (6%)	7 (16%)	1 (4%)	0	0	2 (13%)	0

* מתוך הפריטים שנדגמו לפחות פעם אחת.

עיון בטבלה מעלה כי רוב רובם של הפריטים בתחום המילולי נדגמו, וכי נעשה שימוש בבנק הפריטים כולו. אחוז הדגימה הממוצע לפריט נע בין 2.5%-13%, עם סטיית תקן של 10%-18%. בכל סוגי הפריטים ניצול המאגר הוא משביע רצון ויותר מכך, למעט באנלוגיות שם ישנה דגימה גבוהה של כ- 16% מן הפריטים. הסיבה לכך היא שכאמור רוב רובם של הנבחנים בתחום המילולי אינם מתכנסים ולפיכך מקבלים שני סבבי אנלוגיות, דבר הגורם למיצוי מלא של מאגר פריטי האנלוגיות.

טבלה מס' 8ב: דגימת פריטים בתחום הכמותי

סוג פריט	אלגברה	גיאומטריה	גרפים וטבלאות	השוואה כמותית
מס' פריטים	77	34	28	43
מס' פריטים שנדגמים פחות מ- 2.5%, כולל אלו שאינם נדגמים	4 (5%)	4 (12%)	4 (14%)	6 (14%)
אחוז דגימה ממוצע לפריט* (ס.ת.)	20% (15%)	15% (14%)	14% (11%)	14% (14%)
מספר פריטים הנדגמים למעלה מ- 40% (אחוז)	8 (10%)	2 (6%)	0	2 (5%)

* מתוך הפריטים שנדגמו לפחות פעם אחת.

עיון בטבלה מעלה כי אחוז הדגימה הממוצע של פריטי חשיבה כמותית נע בין 14%-20%, עם סטיית תקן של 15%-11%. כ- 19%-14% מכלל פריטי התחום הכמותי נדגמים מעט מדי (עבור פחות מ- 2.5% מהנבחנים) או יותר מדי (עבור יותר מ- 40% מהנבחנים), אך זהו גם האחוז המשוער של דגימת פריטים וניצולם לפי הסימולציות.

טבלה מס' 8ג: דגימת פריטים בתחום האנגלית

סוג פריט	השלמת משפטים	ניסוח מחדש	הבנת הנקרא
מס' פריטים	71	38	60
מס' פריטים שנדגמים פחות מ- 2.5%, כולל אלו שאינם נדגמים	3 (4%)	2 (5%)	15 (25%)
אחוז דגימה ממוצע לפריט* (ס.ת.)	13% (9%)	21% (12%)	8% (8%)
מספר פריטים הנדגמים למעלה מ- 40% (אחוז)	1 (1%)	3 (9%)	0

* מתוך הפריטים שנדגמו לפחות פעם אחת.

עיון בטבלה מעלה כי אחוז הדגימה הממוצע לפריט אנגלית נע בין 8%-21%, עם סטיית תקן של 12%-8%. רבע מפריטי הבנת הנקרא נדגמו מעט מדי, וכעשירית מפריטי ניסוח מחדש נדגמים יתר על המידה. בסייה ניצול מאגר הפריטים משביע רצון, במיוחד לאור גרסאות קודמות של אמיריים בהם כשליש מן הפריטים לא נדגמו כלל, ואחוז גבוה נדגם יתר על המידה.

לסיכום, אחוז הדגימה הממוצע לפריט סביר, ודומה לנתונים שנתקבלו בניסוי מפעיים 1.

ציוני משתתפי הניסוי בבחינה הפסיכומטרית

מתוך 363 הנבחרים בניסוי אותרו 344 (כ- 95%) שניגשו לבחינה במועד אוקטובר 1999 של הבחינה הפסיכומטרית. שיעור זה של ניגשים מתוך נרשמים מעט גבוה מהמקובל בבחינה פסיכומטרית רגילה, ואף מעט גבוה מאחוז המשתתפים בניסוי מפע"ים 1 שניגשו לבחינה הפסיכומטרית (92%). ממצא זה אינו מפתיע שכן מי שניגש לניסוי היה מן הסתם בעל כוונות רציניות לגשת לבחינה במועד אוקטובר, במיוחד לאור העובדה שאוכלוסיית המשתתפים בניסוי הייתה כזו שנרשמה למעלה משישה שבועות טרם הבחינה. נבחרים אלו תכננו הרשמתם מראש ומן הסתם התכוונו לבחינה והתכוונו לגשת אליה. להלן סיכום ציוני הנבדקים בניסוי מפע"ים 2 במועד אוקטובר 99. כן מופיעים ציוני נבדקי מפע"ים 1 לשם השוואה.

טבלה מס' 9א: ציוני משתתפי ניסוי מפע"ים 2 בבחינת מכפ"ל

מועד אוקטובר 1999

(n=344)

תחום	מילולי	כמותי	אנגלית	כללי
ציון תקן				
ממוצע	116.97	119.39	117.30	602.77
ס.ת.	19.78	18.21	21.05	97.96
טווח	65-150	66-150	70-150	332-797

טבלה מס' 9ב: ציוני משתתפי ניסוי מפע"ים 1 בבחינת מכפ"ל

מועד דצמבר 1998

(n=323)

תחום	מילולי	כמותי	אנגלית	כללי
ציון תקן				
ממוצע	115.34	116.75	116.41	592.04
ס.ת.	18.58	17.74	21.14	94.51
טווח	61-150	69-150	64-150	350-780

ניתן לראות מן הטבלה כי ממוצעי שתי קבוצות הניסוי דומים זה לזה. ההפרשים בין הממוצעים אינם מובהקים

(חשיבה מילולית - $t=-1.10$, $df=665$, $Pr(t)=0.27$, חשיבה כמותית - $t=-1.89$, $df=665$,

$Pr(t)=0.06$, אנגלית - $t=-0.54$, $df=665$, $Pr(t)=0.59$) הבדלים בין שתי הקבוצות צפויים לאור העובדה

ששתי קבוצות נדגמו באופן בלתי תלוי במועדים שונים.

מתוך 344 הנבחנים 173 (50%) נבחנו בבחינה הממוחשבת, ו-171 (50%) נבחנו בבחינת הניי.ע. מקרב 19 הנבחנים

שלא ניגשו למכפ"ל 8 (42%) נבחנו בבחינה הממוחשבת ו-11 (58%) נבחנו בבחינת ניי.ע. להלן סיכום ציוני קבוצת

הנבחנים שלא ניגשו למכפ"ל בהשוואה לאלו שניגשו.

טבלה מס' 10: ציוני הנבחנים בניסוי שניגשו למכפ"ל

בהשוואה לאלו שלא ניגשו לבחינה

מבחן t		לא נבחנו במכפ"ל (n=19)		נבחני מכפ"ל (n=344)		קבוצה
ערך *t	df	ס.ת.	ממוצע	ס.ת.	ממוצע	תחום
*-3.63	361	15.90	95.79	19.47	112.31	מילולי
*-4.55	361	12.37	93.11	18.45	112.63	כמותי
*-2.98	361	23.60	98.47	22.30	114.20	אנגלית
*-4.38	361	75.85	473.68	97.25	573.12	כללי

* הבודל מובהק - $p<0.05$

ניתן לראות כי באופן עקבי משתתפי הניסוי שנבחנו במכפ"ל הינם בעלי ציונים גבוהים יותר מאלו שלא נבחנו

במכפ"ל. כל ההבדלים מובהקים סטטיסטית. ייתכן כי לציונים בניסוי הייתה השפעה על החלטת הנבחנים בניסוי לגשת

לבחינת מכפ"ל. ציוני קבוצת הניסוי שניגשה למכפ"ל גבוהים ב-100 נקודות בממוצע בהשוואה לציוני 19 הנבחנים שלא

ניגשו למכפ"ל. הציון הממוצע של 19 משתתפי הניסוי שלא ניגשו למכפ"ל הוא 474 וסטיית התקן היא 76, כלומר

התפלגות ציוניהם היתה נמוכה במיוחד, ואפשר כי נתון זה עמד לנגד עיניהם כשהחליטו שלא לגשת לבחינה

הפסיכומטרית כחודש לאחר מכן. עובדה זו מלמדת כי קבוצת משתתפי הניסוי שניגשה למכפ"ל לא היתה קבוצה מקרית

מקרב משתתפי הניסוי (וזאת בניגוד לניסוי מפע"ם 1 שם לא היו הבדלים מובהקים סטטיסטית בין קבוצת הנבדקים

שניגשה למכפ"ל לקבוצת הנבדקים שלא ניגשה למכפ"ל). עם זאת עדיין ניתן להתייחס לקבוצה שניגשה למכפ"ל

כקבוצה ייצוגית לקבוצת נבחני אוקטובר 99, בשל ההתפלגות הדומה של ממוצע ציוני שתי הקבוצות כפי שתואר בטבלה

מספר 2.

בטבלה מס' 11 מוצגים ציוני משתתפי הניסוי שניגשו למכפ"ל במועד אוקטובר 99 לפי חלוקה לקבוצות הניסוי

(מפע"ם 2 או ניי.ע). ובטבלה 11 מוצגים לשם השוואה אותם נתונים כפי שנתקבלו בניסוי מפע"ם 1. ניתן לראות כי

הציון הכללי של קבוצת הניי.ע במפע"ם 2 גבוה יותר מן הציון הכללי של קבוצת הניי.ע במפע"ם 1. גודל האפקט קטן

(0.2), לפי Cohen (1988).

טבלה מס' 11א: ציוני משתתפי הניסוי בבחינת הניסוי (ניייע / מפעיים 2) שניגשו למכפ"ל בחלוקה לקבוצות הניסוי

מבחן t		מפעיים 2 (n=173)		ניייע (נוסח 77 ו-78) (n=171)		בחינה
ערך t	df	ס.ת.	ממוצע	ס.ת.	ממוצע	תחום
*2.26	342	18.42	109.97	20.25	114.69	מילולי
0.32	342	19.87	112.31	16.92	112.95	כמותי
1.53	342	21.74	112.38	22.77	116.05	אנגלית
1.64	342	97.77	564.58	96.24	581.75	כללי

* הבדל מובהק - $p < 0.05$

טבלה מס' 11ב: ציוני משתתפי הניסוי בבחינת הניסוי (ניייע / מפעיים 1) שניגשו למכפ"ל בחלוקה לקבוצות הניסוי

מבחן t		מפעיים 1 (n=165)		ניייע (נוסח 77 ו-78) (n=158)		בחינה
ערך t	df	ס.ת.	ממוצע	ס.ת.	ממוצע	תחום
-0.51	321	17.13	111.56	19.14	110.53	מילולי
*-2.31	321	18.10	114.28	17.52	109.70	כמותי
-0.52	321	22.08	113.32	21.84	112.04	אנגלית
-1.39	321	92.93	574.33	92.24	559.99	כללי

* מובהק ברמת $p < 0.05$.

ההבדלים בין קבוצת הניייע וקבוצת המחשב בניסוי מפעיים 2 אינם מובהקים סטטיסטית פרט להבדל בתחום המילולי העומד על כרבע סטיית תקן. הבדלים בין שתי קבוצות ניסוי יכולים לנבוע משני מקורות לפחות: האחד, הבדלים קטנים בין שתי הקבוצות אשר לא נוטרלו ע"י ההקצאה המקרית לשתי קבוצות; השני, החלק המילולי של בחינת מפעיים 2 מעט קשה יותר מאשר בחינת הניייע ויש לתקן מעט את כיוול הציונים. על מנת לבחון את המקור הראשון מוצגת בטבלה הבאה השוואה של ציוני שתי הקבוצות בבחינת מכפ"ל מועד אוקטובר 99 בטבלה 12א. בטבלה 12ב מופיעים אותם נתונים כפי שנתקבלו בניסוי מפעיים 1 לשם השוואה. ניתן לראות כי ההפרשים בציון הכללי בין שתי הקבוצות בשני מועדי ההיבחנות נותרו דומים.

טבלה מס' 12א: ציוני משתתפי ניסוי מפע"ם 2 במכפ"ל מועד אוקטובר
בחלוקה לקבוצות הניסוי

מבחן t		מפע"ם 2 (n=173)		ניי"ע (נוסח ו-77) (n=171)		קבוצה
ערך t	df	ס.ת.	ממוצע	ס.ת.	ממוצע	תחום
1.29	342	20.16	115.61	19.37	118.35	מילולי
0.98	342	19.38	118.43	16.94	120.35	כמותי
1.44	342	20.57	115.68	21.45	118.94	אנגלית
1.35	342	100.80	595.67	94.76	609.95	כללי

טבלה מס' 12ב: ציוני משתתפי ניסוי מפע"ם 1 במכפ"ל מועד דצמבר
בחלוקה לקבוצות הניסוי

מבחן t		מפע"ם 1 (n=165)		ניי"ע (נוסח ו-77) (n=158)		קבוצה
ערך t	df	ס.ת.	ממוצע	ס.ת.	ממוצע	תחום
-1.32	321	18.73	116.67	18.28	113.95	מילולי
-1.51	321	17.40	118.21	18.02	115.23	כמותי
-0.95	321	21.39	117.50	20.87	115.27	אנגלית
-1.46	321	95.37	599.55	93.26	584.20	כללי

* מובהק ברמת $p < 0.05$.

בטבלה 12א ניתן לראות כי ממוצעי קבוצת הניסוי שנבחנה בניי"ע גבוהים מאלו של קבוצת הבחינה הממוחשבת בניסוי מפע"ם 2, אולם הבדלים אלו לא נמצאו מובהקים. ההבדלים בין הקבוצות קטנים מאד (מתחת לסף 0.2 של אפקט נמוך עבור הפרשים בין שתי קבוצות שמציע Cohen (1988)). כלומר ההבדל המובהק בין שתי הקבוצות בתחום המילולי, שהופיע בניסוי, אינו מופיע במכפ"ל, אך הכיווניות נשמרת. גם מהשוואת ציוני משתתפי הניסוי לכלל נבחני אוקטובר 99 עולה כי ציוני קבוצת הניי"ע גבוהים בהשוואה לממוצע כלל הנבחנים. נראה אפוא כי לפחות חלק מההבדל בין שתי הקבוצות מקורו בהבדל אמיתי ביכולת המילולית של קבוצת הניי"ע. ההבדל בין שתי קבוצות הניסוי בחשיבה כמותית שהופיע בניסוי מפע"ם 1 אינו חוזר על עצמו בניסוי זה. העובדה שההבדלים בין הקבוצות בתחום מסויים אינם עקביים מעבר לשני הניסויים, או מעבר לאופנות ההיבחות מלמדת כי עצם ההיבחות במחשב לעומת ני"ע אינה הסיבה להבדלים אלו. להלן יוצגו ההבדלים בין הישגי הנבחנים בבחינת הניסוי (ניי"ע או מחשב) והישגיהם בבחינת מכפ"ל מועד אוקטובר 1999. בטבלה מס' 13 מוצג ניתוח של ההפרשים בין בחינת מכפ"ל לבין בחינת מפע"ם 2, ובטבלה מס' 14 מוצג ניתוח של

ההפרשים בין בחינת מכפ"ל לבין בחינת הנייט בניסוי (נוסחים 77 ו-78). השוואה של ממוצעי ההפרשים של מפע"ם 2 מול אלו של בחינת הנייט יכולה ללמד על מידת המקבילות של שני המבחנים.

טבלה מס' 13: ניתוח ההפרשים בציונים בין מכפ"ל מועד אוקטובר 99 ומפע"ם 2 (n=173)

ממוצע הפרשי הציונים (ס.ת.)	טווח הפרשי הציונים	ציון מכפ"ל גבוה יותר	שוויון (הפרש של עד נקודה)	ציון מכפ"ל נמוך יותר	מילולי
(ס.ת.)		שכיחות נבחנים	שכיחות נבחנים	שכיחות נבחנים	
5.64 (9.84)	-29 - (-25)	129 (74.6%)	6 (3.5%)	38 (22.0%)	מילולי
6.12 (10.80)	-31 - (-21)	121 (69.9%)	9 (5.2%)	43 (24.9%)	כמותי
3.30 (9.06)	-28 - (-25)	106 (61.3%)	19 (11.0%)	48 (27.7%)	אנגלית
31.09 (39.31)	-123 - (-109)	140 (80.9%)	2 (1.2%)	30 (17.3%)	כללי

ממוצע הציונים בבחינת החוזרת (מכפ"ל) גבוהים מאלו שבבחינה הראשונה, מפע"ם 2. ההבדלים הממוצעים בתחומים נעים בין 3-6 נקודות, שהם כשליש עד שתי חמישיות סטיית תקן. בציון הכללי ממוצע ההבדלים עומד על 31 נקודות. עוד ניתן לראות כי בציון הכללי כ- 80% מהנבחנים קיבלו ציון גבוה יותר בבחינת מכפ"ל מאשר במפע"ם 2, וכ- 17% מהנבחנים קיבלו ציון גבוה יותר בבחינת מפע"ם 2 מאשר במכפ"ל.

טבלה מס' 14: ניתוח ההפרשים בין הציונים במכפ"ל מועד אוקטובר 99 ובחינת נייט בניסוי (n=171)

ממוצע הפרשי הציונים (ס.ת.)	טווח הפרשי הציונים	ציון מכפ"ל גבוה יותר	שוויון (הפרש של עד נקודה)	ציון מכפ"ל נמוך יותר	מילולי
(ס.ת.)		שכיחות נבחנים	שכיחות נבחנים	שכיחות נבחנים	
3.66 (9.31)	-25 - (-23)	105 (61.4%)	18 (10.5%)	48 (28.1%)	מילולי
7.40 (9.03)	-36 - (-19)	138 (80.7%)	10 (5.8%)	23 (13.5%)	כמותי
2.89 (8.21)	-26 - (-21)	98 (57.3%)	17 (9.9%)	56 (32.7%)	אנגלית
28.19 (32.74)	-117 - (-68)	135 (78.9%)	2 (1.2%)	34 (19.9%)	כללי

גם בקבוצת הניייע, בממוצע הציונים בבחינת מכפ"ל גבוהים מאלו שבניסוי. ההבדלים הממוצעים בתחומים נעים בין 3-7 נקודות, שהם שליש עד שני שלישים סטיות תקן. בציון הכללי ממוצע ההבדלים עומד על 28 נקודות, שהם כרבע סטיית תקן. עוד ניתן לראות כי בציוני התחומים בין 60% ל-80% מהנבחנים קיבלו ציון גבוה יותר בבחינת מכפ"ל מאשר בבחינת הניייע בניסוי. בציון הכללי כ-80% מהנבחנים קיבלו ציון גבוה יותר במכפ"ל מאשר בבחינת הניייע. השיפור במכפ"ל גבוה מעט יותר בניסוי מפע"ם 2 בהשוואה לניסוי מפע"ם 1. ייתכן כי קבוצת הניגשים לניסוי הפיקה הפעם לקחים בצורה מלומדת יותר מבפעם הקודמת, וייתכן כי ההבדל הוא מקרי. לשם השוואה נערכה בדיקה של ההפרשים בין היבחנויות חוזרות בבחינת מכפ"ל. ממדור ציונים נתקבלו ציוניהם של הנבחנים בבחינת מכפ"ל אוקטובר 99, אשר נבחנו בעבר בבחינת מכפ"ל נוספת, בצירוף ציוניהם בבחינת מכפ"ל קודמת (שאינה בהכרח בחינת מכפ"ל הראשונה שלהם). בדיקה של ההפרשים שם מצביעה על הבדל ממוצע בשלושת התחומים של כמחצית סטיית תקן (9-10 נקודות, עם סטיית תקן של 10-11 נקודות), ובציון הכללי הפרש ממוצע של כ-55 נקודות (הבדל של למעלה ממחצית סטיית תקן) עם סטיית תקן של 42 נקודות. ניתן לראות כי ההפרשים בין היבחנויות חוזרות בניסוי הינם כמחצית מן ההפרשים שנמצאו אצל הנבחנים החוזרים בבחינת מכפ"ל דצמבר 98. יחד עם זאת יש להדגיש את פער הזמנים הקצר של חודש בין הבחינה בניסוי והבחינה במכפ"ל, לעומת פערי זמן גדולים בהרבה בין היבחנויות חוזרות בבחינת מכפ"ל (כך, פער הזמן המינימלי ההכרחי בין שתי היבחנויות במכפ"ל עמד על כעשרה חודשים). הפרש דומה לזה שהתקבל בניסוי בתחום האנגלית, ניתן למצוא אצל נבחנים חוזרים בבחינת המיון הממוחשבת אמיר"ם 4, שם עמד ההפרש הממוצע על 3-4 נקודות (ראה שפר, רפ, רוני 1998). פער הזמן הממוצע בין ההבחנות הראשונה והשנייה בבחינת אמיר"ם זו, היה קצר יותר מאשר הפער בין הבחינות החוזרות במכפ"ל, ועמד על פחות משישה חודשים.

השיפור בהיבחנות השנייה (מכפ"ל אוקטובר 99) בהשוואה להיבחנות הראשונה (הניסוי) צריך להיות זהה עבור שתי הקבוצות בניסוי, קבוצת המבחן הממוחשב וקבוצת הניייע. השוואה של ההפרשים הממוצעים שהוצגו לעיל, בין מכפ"ל למפע"ם 2 ובין מכפ"ל לבחינת הניייע בניסוי, מצביעה על דימיון רב בין ההפרשים הממוצעים, והינה עדות נוספת לכך שבחינת מפע"ם 2 מקבילה לבחינת הניייע בניסוי ולתקינותו של תהליך הכיול.

טבלה מס' 15: הקשר בין הציון בניסוי לבין הציון

במכפ"ל מועד אוקטובר 99

מפע"ם 2 ציון תקן (n=173)	קבוצת ניייע ציון תקן (n=171)	
0.87	0.89	מילולי
0.85	0.86	כמותי
0.91	0.93	אנגלית
0.92	0.94	כללי

* כל המתאמים מובהקים - $p < 0.05$

הממצאים המופיעים בטבלה 15 הינם מעודדים למדי. ניתן לראות כי הקשר בין בחינת מפע"ם 2 לבין מכפ"ל מועד אוקטובר 99 הוא גבוה, והוא דומה למהימנות שבין מבחן למבחן חוזר. מתאם של 0.92 בין הציון הכללי במפע"ם 2 ובין הציון הכללי במכפ"ל מצביע על כך שמבחנים אלו מקבילים. המהימנות היא מהימנות של מבחן חוזר. יתר על כן, המתאמים של מפע"ם 2 מול מכפ"ל דומים מאוד לאלו של בחינת הני"ע בניסוי עם מכפ"ל. לסיכום חלק זה ניתן לומר כי הנתונים מצביעים על כך שבחינת מפע"ם 2 הינה גירסה מקבילה לבחינת מכפ"ל. מתאמי בחינת מפע"ם 1 עם ציוני מכפ"ל גבוהים למדי (0.92 של הציון הכללי במפע"ם 2 עם הציון הכללי בבחינת מכפ"ל) ודומים מאוד לאלו של בחינת הני"ע בניסוי עם ציוני מכפ"ל. הממוצעים וסטיות התקן של הפרשי הציונים בין מפע"ם 2 למכפ"ל דומים מאוד לאלו של הפרשי הציונים בין בחינת הני"ע בניסוי ומכפ"ל. נמצא שהנתונים שנתקבלו בניסוי מפע"ם 2 דומים מאוד לנתונים שנתקבלו בניסוי מפע"ם 1, וכי ההבדלים שנמצאו בהשוואה כמותית בניסוי הראשון לא חזרו על עצמם בניסוי זה. לעומת זאת בניסוי השני נמצא הבדל מובהק בין הקבוצות בחשיבה מילולית, הבדל אשר הופיע בניסוי, אך לא במכפ"ל, אף כי הכיוונית נשמרה. ניתן להסיק אפוא בבטחה כי אופנות ההיבחנות אינה משפיעה על הציון הפסיכומטרי באופן משמעותי, וכי ההבדלים בתחום זה או אחר נובעים בעיקר מהבדלים מקריים בין קבוצות הניסוי.

הבדלים בין המינים

נושא נוסף שנבדק בניסוי היה קיומם של הבדלים בין המינים במבחן הממוחשב. זיהוי מינם של נבחנים נעשה על פי דיווח עצמי שלהם כחלק משאלוני המשוב, ולא על בסיס דיווח עצמי של הנבחנים בעת הרשמתם לבחינה הפסיכומטרית כפי שנעשה בניסוי מפעיים 1. לפיכך, ידוע לנו מינם של 353 מתוך 361 הנבדקים בניסוי מפעיים 2 (98%). שיעור זה של נבחנים המדווחים על מינם גבוה במידה ניכרת מהמקובל – לרוב כ- 85% מן הנבחנים מדווחים על מינם (פבלוב ואורן, 1999). בטבלה מס' 16 מוצגת התפלגות הנבחנים עפ"י מין וקבוצת המבחן.

טבלה מס' 16: התפלגות נבחנים עפ"י קבוצה ומין

קבוצת נייע	גברים	נשים	סה"כ
קבוצת נייע	89 (51%)	86 (49%)	175 (47%)
קבוצת בחינה ממוחשבת	95 (53%)	83 (47%)	178 (53%)
סה"כ	184 (52%)	169 (48%)	353

כפי שניתן לראות שיעור הגברים (שזוהו) בניסוי עולה במעט על שיעור הנשים (52%, 48%, בהתאמה). זאת לעומת, יחס הפוך של נבחנים (מזוהים) בשנת 98 במכפ"ל: 45% גברים ו-55% נשים. ניתן גם לראות כי בשל ההקצאה המקרית לקבוצות פרופורצית הנשים והגברים בשתי הקבוצות (נייע, בחינה ממוחשבת) דומה למדי. בטבלה מס' 17 מוצגת השוואה של ציוני הנבחנים בתחומים השונים ובציון הכללי בבחינת מפעיים 2 ובבחינות מכפ"ל נוסחים 77 ו-78 עפ"י קבוצת ניסוי (ממוחשב, נייע) ומין.

טבלה מס' 17: השוואת ציוני הנבחנים בניסוי בחלוקה לקבוצת הנייר ועפרון ולקבוצת המחשב ולמין הנבדקים

מבחן F	משתנים		ממוחשב				נייע				מין
	אXמ	מינ	אופנות	SD	M	SD	M	SD	M	SD	
0.21	2.25	*5.02	18.32	107.84	18.49	111.99	18.32	107.84	20.53	115.85	מילולי
2.50	*23.34	0.11	21.00	105.73	16.96	118.26	18.51	109.79	14.67	116.12	כמותי
0.47	*3.90	2.75	22.99	108.92	22.18	115.39	23.85	114.79	21.81	117.91	אנגלית
1.28	*10.39	2.86	98.60	540.75	92.47	586.19	101.6	571.35	89.59	593.13	כללי

* הפרש מובהק - $p < 0.05$

ניתן לראות כי הן במבחן הממוחשב והן במבחן הנייע בשלושת התחומים ובציון הכללי, גברים הינם בעלי ציון ממוצע גבוה יותר מנשים (בחשיבה כמותית ובאנגלית הפרש מובהק. בתחום המילולי גורם האופנות מובהק סטטיסטית. הבדל זה בין קבוצת הנייע למחשב נדון בפרק הקודם. בדיקה של אפקט האינטראקציה בין אופנות

ההיבחות למין הנבדקים אינה מובהקת באף לא אחד מן התחומים. ממצא זה חשוב משום שהוא מראה כי במבחן הממוחשב לא נמצאו הבדלים אחרים בין המינים מאשר בבחינת ני"ע.
 בטבלה מס' 18 מוצגת השוואה של ציוני הנבחנים במכפ"ל מועד אוקטובר 99 בתחומים השונים ובציון הכללי עפ"י קבוצת ניסוי (ממוחשב, ני"ע) ומין.

טבלה מס' 18: השוואת ציוני הנבחנים במכפ"ל מועד אוקטובר 99 בתחומים השונים עפ"י מין ואופנות היבחות

מבחן F			ממוחשב				ני"ע				מין
אפקטים			נשים (n=79)		גברים (n=90)		נשים (n=79)		גברים (n=85)		
מ	מי	אופנות	SD	M	SD	M	SD	M	SD	M	
0.39	2.84	1.47	19.68	113.19	20.55	118.17	18.75	117.29	19.82	119.57	מילולי
1.00	*23.17	1.07	19.21	112.92	17.75	124.01	17.69	117.04	14.84	124.29	כמותי
0.08	*4.58	2.13	19.57	112.66	21.25	118.21	21.64	116.77	21.21	121.02	אנגלית
0.61	*10.92	1.83	96.90	574.18	100.18	617.11	98.21	597.54	88.27	624.02	כללי

* אפקט מובהק ברמת $p < 0.05$

העיון בציוני מכפ"ל של שתי הקבוצות מעלה כי:

ניתן לראות כי בדומה לתמונה שהצטיירה מנתוני הניסוי הן בקבוצת המבחן הממוחשב והן בקבוצת מבחן הני"ע בשלושת התחומים, באופן עקבי גברים הינם בעלי ציון ממוצע גבוה יותר מנשים. אפקט ראשי זה של מין נמצא מובהק בשני תחומים ובציון הכללי (אפקט בינוני בחשיבה כמותית – 0.5, אפקט בינוני בציון הכללי – 0.35, ואפקט קטן באנגלית – 0.25), ובתחום המילולי לא נמצא הבדל מובהק. אופנות ההיבחות בניסוי כמשפיעה על ציון מכפ"ל לא נמצאה מובהקת באף לא אחד מן התחומים. גם בבדיקה זו לא נמצא אפקט אינטראקציה מובהק סטטיסטית.

התמונה שהוצגה לעיל לגבי הבדלים בין המינים בציונים אינה יחודית למדגם זה. בטבלה 19 מוצגים ציוני הנבחנים בעברית בשנת 1999 בתחומים השונים עפ"י מין, הלקוחים מנתוני מדור מחקר. יש לציין כי 17.6% מן הנבחנים לא ציינו את מינם, ולפיכך אינם כלולים בנתוני הטבלה.

טבלה מס' 19 : השוואת ציוני הנבחנים בעברית במכפ"ל שנת 1999
בתחומים השונים עפ"י מין

מכפ"ל שנת 99 – עברית				
נשים (n=21,328)		גברים (n=17,766)		מין
SD	M	SD	M	תחום
19.7	106.4	19.2	111.2	מילולי
18.3	106.6	17.8	117.5	כמותי
21.9	107.5	22.1	116.4	אנגלית
99.4	538.7	96.2	584.2	כללי

מן הנתונים עולה כי קיימים הבדלים לא מבוטלים בין המינים. כך, בתחום הכמותי הבדל בינוני של כמעט

שני שליש סטיית תקן (מתאם של 0.30 בין מין וציון בתחום הכמותי), בתחום האנגלית הבדל בינוני של כשתי חמישיות סטיית תקן (מתאם של 0.20) ובתחום המילולי הבדל בינוני של כרבע סטיית תקן (מתאם של 0.13). בציון הכללי הבדל של כמחצית סטיית תקן (מתאם של 0.24). נתונים אלו דומים להבדלים הבינוניים שהתקבלו במבחן הממוחשב בניסוי: למעלה ממחצית סטיית תקן הבדל בכמותי, באנגלית כשתי חמישיות סטיית תקן, במילולי הבדל של כמעט מחצית סטיית תקן, ובציון הכללי הבדל של למעלה ממחצית סטיית תקן, ומחזקים את העובדה שהמבחן הממוחשב אינו משפיע על המינים באופן שונה, כך שההבדלים בין המינים הם כמו במבחן נייר ועיפרון.

לסיכום, לגבי ההבדלים בין המינים מצאנו כי בממוצע גברים היו בעלי ציון גבוה באופן מובהק מזה של הנשים במפע"ם 2. בציון הכללי היה הבדל של כמחצית סטיית תקן. דפוס זה דומה מאד לזה שהתקבל בנתוני כלל הנבחנים במכפ"ל בשנת 99. בציון משתתפי הניסוי בבחינת מכפ"ל (שדיווחו על מינם בבחינה) נמצאו הבדלים קטנים יותר בין המינים (הבדלים בינוניים של שליש סטיית תקן בציון הכללי, וכשתי חמישיות סטיית תקן בתחום הכמותי), ובתחום המילולי ההבדל בין המינים אף לא היה מובהק סטטיסטית (הבדל של כמעט שליש סטיית תקן). אפקט אינטראקציה בין מין וקבוצה (מבחן ממוחשב, ני"ע) לא נמצא מובהק בניסוי. נראה, אפוא, כי אין לחשוש לגבי הבדלים בין המינים בביצוע במבחן הממוחשב יותר מאשר במבחן ני"ע רגיל.

תוצאות המשובים

לאחר תום הניסוי, טרם קבלת הציון, נתבקשו הנבדקים למלא שאלון משוב אודות הבחינה, והרגלי השימוש במחשב (ר' נספח). כל אחת משתי קבוצות הניסוי קיבלה שאלון מעט שונה. מרבית השאלות היו זהות עם שינוי קטן בהתאם לאמצעי בו נבחנו הנבדקים (ראה בהמשך). קבוצת הניסוי הממוחשבת קיבלה מעט יותר שאלות אודות נושאים שונים שרצינו לקבל אודותם משוב (כגון טיב הגרפיקה, ההסברים וכד'). ההיענות למילוי השאלון הייתה מלאה. שאלוני המשוב היו זהים לחלוטין לשאלוני המשוב בניסוי מפע"ם 1 למעט העובדה שבניסוי זה הנבדקים נתבקשו לרשום את מס' תעודת הזהות שלהם, דבר שאפשר את הצלבת נתוני המשוב עם ציוניהם בניסוי ובמכפ"ל. ציון המין נכתב הפעם כחלק מן המשוב, ולא הוסף בכתב יד, דבר אשר שיפר את הנתונים לגבי מין הנבדקים בצורה ניכרת. ר' טבלה מספר 20.

טבלה 20 - התפלגות הנבדקים לפי מין בניסוי מפע"ם 1 בהשוואה לניסוי מפע"ם 2

מין	ניסוי מפע"ם 1	אחוזים	ניסוי מפע"ם 2	אחוזים
גברים	145	42	184	51
נשים	118	34	169	47
לא ציינו את מינם	80	24	8	2
ס"ה	343	100	361	100

נתונים הנוגעים למין הנבדקים בניסוי ולשכיחות השימוש במחשב

בטבלאות 21.1, 21.2 ו- 21.3 מופיעים נתונים לגבי שכיחות השימוש במחשב בקרב נבדקי הניסוי. נראה כי בקבוצת הניסוי הממוחשבת אחוז הנבדקים שאינם משתמשים במחשב מעט גבוה (כ- 5%) מבקבוצת ניסוי נייע. אך בסך הכל הנתונים המופיעים כאן דומים לנתוני הרגלי השימוש במחשב שנאספו בשנת 95 מנבחני במו"ת ומנבחני דצמבר 95 (מושינסקי, טננבאום, רפ ורונו, 1997): 22% השתמשו במחשב באופן יומיומי, 16% השתמשו במחשב פעם עד פעמיים בשבוע, 29% השתמשו במחשב פעם עד פעמיים בחודש ו- 33% לא השתמשו כלל במחשב. על אף שנתונים אלו נאספו לפני מספר שנים, ועל אף שבשנים אלו השימוש במחשב נעשה נפוץ ושכיח במידה ניכרת, הנתונים לא השתנו במידה ניכרת. זהו ממצא מפתיע, אך הוא מצדיק את פיתוח המבחן ובנייתו באופן שאינו מצריך כל ידע קודם במחשב, ואינו מניח ידע באשר לשימוש בעכבר. מנתוני שני הניסויים ומהסקר שנערך בקרב נבחני במו"ת ניתן ללמוד כי כשליש מן הנבחנים אינו משתמש במחשב באופן שוטף, וכי שכיחות זו יציבה יחסית לאורך הזמן, על אף השימוש ההולך וגובר לכאורה במחשב.

טבלה 21.1 - קבוצת ניסוי - מחשב - שכיחות השימוש במחשב

שכיחות השימוש במחשב	N	אחוזים
שימוש יומיומי	49	27
פעם פעמיים בשבוע	38	21
פעם פעמיים בחודש	30	17
אינו משתמש במחשב	63	35
ס"ה	180	100

טבלה 21.2 - קבוצת ניסוי - ני"ע - שכיחות השימוש במחשב

שכיחות השימוש במחשב	N	אחוזים
שימוש יומיומי	48	27
פעם פעמיים בשבוע	37	21
פעם פעמיים בחודש	40	22
אינו משתמש במחשב	53	30
ס"ה	178	100

טבלה 21.3 - כלל המשתתפים בניסוי - שכיחות השימוש במחשב

שכיחות השימוש במחשב	N	אחוזים
שימוש יומיומי	97	27
פעם פעמיים בשבוע	75	21
פעם פעמיים בחודש	70	20
אינו משתמש במחשב	116	32
ס"ה	358	100

בטבלאות 22.1, 22.2 ו- 22.3 מופיעים נתונים באשר לשכיחות השימוש במחשב בקרב גברים ונשים. יש הבדל מובהק בשכיחות השימוש במחשב בין גברים ונשים, ר' טבלה 22.3 - ($\chi^2=20.43$, $DF=3$, $p<0.05$). בניסוי מפע"ם 1 לא נמצא הבדל מובהק בין גברים לנשים במידת השימוש במחשב, אך כאמור יש לזכור כי כרבע מן הנבדקים לא ציינו אז את מינם. נתונים אלו דומים לנתוני הסקר שנערך בשנת 95 לגבי הרגלי השימוש במחשב, שם שכיחות השימוש במחשב בקרב גברים הייתה גבוהה יותר מאשר בקרב נשים. ההבדל בין גברים לנשים הוא משמעותי. אחוז הנשים שאינן משתמשות כלל במחשב גבוה כמעט פי שניים מאחוז הגברים שאינם משתמשים כלל במחשב.

טבלה 22.1 - קבוצת ניסוי – מחשב – שכיחות השימוש במחשב בקרב גברים ונשים

שכיחות השימוש במחשב	גברים (אחוז מתוך הכלל) [אחוז מתוך הגברים]	נשים (אחוז מתוך הכלל) [אחוז מתוך הנשים]
שימוש יומיומי	35 (20%) [37%]	14 (8%) [17%]
פעם פעמיים בשבוע	23 (13%) [24%]	14 (8%) [17%]
פעם פעמיים בחודש	13 (7%) [14%]	17 (10%) [20%]
אינו משתמש במחשב	24 (13%) [25%]	38 (21%) [46%]
ס'יה	95 (53%) [100%]	83 (47%) [100%]

טבלה 22.2 - קבוצת ניסוי – נייע - שכיחות השימוש במחשב בקרב גברים ונשים

שכיחות השימוש במחשב	גברים (אחוז מתוך הכלל) [אחוז מתוך הגברים]	נשים (אחוז מתוך הכלל) [אחוז מתוך הנשים]
שימוש יומיומי	30 (17%) [34%]	16 (13%) [25%]
פעם פעמיים בשבוע	19 (11%) [21%]	14 (11%) [22%]
פעם פעמיים בחודש	21 (12%) [24%]	11 (8%) [18%]
אינו משתמש במחשב	19 (11%) [21%]	22 (17%) [35%]
ס'יה	89 (51%) [100%]	63 (49%) [100%]

טבלה 22.3 - כלל המשתתפים בניסוי - שכיחות השימוש במחשב בקרב גברים ונשים

שכיחות השימוש במחשב	גברים (אחוז מתוך הכלל) [אחוז מתוך הגברים]	נשים (אחוז מתוך הכלל) [אחוז מתוך הנשים]
שימוש יומיומי	65 (18%) [35%]	31 (9%) [18%]
פעם פעמיים בשבוע	42 (12%) [23%]	31 (9%) [18%]
פעם פעמיים בחודש	34 (10%) [19%]	35 (10%) [21%]
אינו משתמש במחשב	43 (12%) [23%]	72 (20%) [43%]
ס'יה	145 (52%) [100%]	169 (48%) [100%]

משתנה נוסף שבדק את השימוש במחשב היה האם הנבדקים נבחנו בעבר במבחן ממוחשב. בטבלאות 23.1 ו-23.2 מופיעים הנתונים באשר למשתנה זה. אין הבדל משמעותי בין קבוצות הניסוי ביחס למשתנה זה. ובס'יה כמחצית מן הנבדקים נבחנו בעבר במבחן ממוחשב. מקרב אלו שציינו היכן עולה כי רובם נבחנו במבחן ממוחשב בצבא. בטבלה 23.4 מופיעים נתונים לגבי משתנה זה בקרב גברים ונשים. יש הבדל מובהק בין מספר הגברים שנבחנו בעבר במבחן ממוחשב לבין מספר הנשים שנבחנו בעבר במבחן ממוחשב ($\chi^2=5.49, DF=1, p<0.05$), ויתכן כי הדבר קשור לכך שהמבחן הממוחשב שרוב הנבדקים התנסו בו התקיים בצבא.

בסך הכל הנתונים מעידים כי לגברים התנסות קודמת רבה יותר עם מחשב בהשוואה לנשים הן בשימוש שוטף במחשב, והן בהתנסות קודמת עם מבחן ממוחשב. נתונים אלו מחזקים את העובדה שאכן צריך היה לבנות את המבחן כך שלא נדרש ידע קודם בשימוש במחשב כדי להיבחן בו. כמו כן מעודד לראות שעל אף ההבדלים בין גברים

לנשים. אופנות ההיבחנות לא נמצאה כמשתנה משמעותי המשפיע על הציון, ואף לא האינטרקציה בין אופנות ההיבחנות לבין משתנה המין.

טבלה 23.1 - קבוצת ניסוי - מחשב - שכיחות הנבחנים בעבר במבחן ממוחשב

אחוזים	שכיחות	
51	92	נבחן בעבר במבחן ממוחשב
49	88	לא נבחן בעבר במבחן ממוחשב

טבלה 23.2 - קבוצת ניסוי - ני"ע - שכיחות הנבחנים בעבר במבחן ממוחשב

אחוזים	שכיחות	
50	88	נבחן בעבר במבחן ממוחשב
50	89	לא נבחן בעבר במבחן ממוחשב

טבלה 23.3 - כלל הנבדקים בניסוי - שכיחות הנבחנים בעבר במבחן ממוחשב

אחוזים	שכיחות	
50	180	נבחן בעבר במבחן ממוחשב
50	177	לא נבחן בעבר במבחן ממוחשב

טבלה 23.4 - כלל הנבדקים בניסוי - שכיחות הנבחנים בעבר במבחן ממוחשב בקרב גברים ונשים

נשים	גברים	
74 (44%)	103 (56%)	נבחן בעבר במבחן ממוחשב
95 (56%)	80 (44%)	לא נבחן בעבר במבחן ממוחשב
169 (100%)	183 (100%)	ס"ה

משתנה נוסף שיכול היה לגרום להבדלים בין נבדקים בניסוי היה התנסות מוקדמת

בי"קומפיוטרי". תוכנה דומה מאוד למפע"ם, והתנסות בה ובמיוחד תרגולה לפי חלוקת הזמנים השונה (מתן זמן רב לכל פריט בנפרד) הנהוגה בה, יכולה להפחית חרדה ולשפר את הביצוע במבחן הממוחשב. נתונים אלו מופיעים בטבלאות 24.1 ו-24.2 ו-24.3. למרבה הצער, משתנה זה אינו משמעותי וקרוב לודאי שלא גרם להבדלים בין הנבדקים, שכן רק קומץ (כ- 4%) התנסה בתוכנה. זוהי עלייה של אחוז אחד בלבד בהשוואה לנתוני ניסוי מפע"ם 1.

טבלה 24.1 - קבוצת ניסוי - מחשב - שכיחות ההתנסות ב"קומפיוטרי"

אחוזים	שכיחות	
4	7	התנסה ב"קומפיוטרי"
96	172	לא התנסה ב"קומפיוטרי"

טבלה 24.2 - קבוצת ניסוי - ני"ע - שכיחות ההתנסות ב"קומפיוטרי"

אחוזים	שכיחות	
4	7	התנסה ב"קומפיוטרי"
96	171	לא התנסה ב"קומפיוטרי"

טבלה 24.3 - כלל הנבדקים בניסוי - שכיחות ההתנסות ב"קומפיוטרי"

אחוזים	שכיחות	
4	14	התנסה ב"קומפיוטרי"
96	343	לא התנסה ב"קומפיוטרי"

נתונים הנוגעים להתרשמות הנבדקים מהבחינה הממוחשבת ומטיבה בהשוואה לבחינת נייר ועיפרון

בנוסף למשתתפים הללו נתבקשו הנבדקים להשיב לשאלות הנוגעות למידת שביעות רצונם מן המבחן שעברו

ביחס למגוון משתתפים על סולם הנע בין 1 ל- 4. ניסוחי השאלות השונים, הממוצעים בקרב כל אחת מהקבוצות,

מספר המשיבים, וההבדל בין הקבוצות מופיעים להלן.

-1-----2-----3-----4-
 מסכים בהחלט לא מסכים
 כלל

<u>ממוצע ני"ע (ס"ת) n</u>		<u>ממוצע מחשב (ס"ת) n</u>	
(1) (מסכי) ההוראות וההסברים היו ברורים			
ns	178	3.83 (0.40)	180 3.89 (0.41)
(2) בבחינה רגילה הייתי מצליח לענות על השאלות טוב יותר			
ns	178	2.25 (0.93)	180 2.29 (1.07)
(3) השאלות שנשאלתי היו קשות עבורי			
s	178	2.44 (0.78)	180 2.24 (0.71)
(4) סימון התשובות (בעזרת לוח המקשים) היה נוח			
s	178	2.82 (1.13)	180 3.57 (0.79)
(5) השאלות הוצגו בצורה ברורה ונוחה			
s	178	3.37 (0.49)	180 3.74 (0.55)
(6) בחינה באמצעות (מחשב) נייר ועיפרון אינה הוגנת כלפי			
s	178	1.35 (0.73)	179 1.67 (0.96)
(7) התקשיתי להתרכז, עקב הצפייה הממושכת במסך המחשב / התקשיתי להתרכז לכל אורך הבחינה, שכן היא נמשכה זמן רב			
s	178	2.67 (0.49)	180 1.97 (1.06)
(8) מבחן ממוחשב מעניק זמן רב לכל שאלה, ומפצה על קריאה איטית -			
-	-	-	179 3.69 (0.62)
(9) בחינה באמצעות מחשב קשה יותר מבחינה המועברת בנייר ועיפרון -			
-	-	-	176 1.92 (0.97)
(10) קטעי הקריאה והתרשימים הוצגו בצורה שאינה מאפשרת קריאה נוחה -			
s	178	1.85 (0.94)	179 1.58 (0.88)
(11) העובדה שלשאלה מוקצב זמן משלה, מאפשרת להתרכז בה בשקט -			
-	-	-	179 3.68 (0.69)
(12) אני סבור שבבחינה רגילה הייתי מוציא ציון נמוך יותר -			
-	-	-	176 2.10 (0.91)
(13) העובדה שמשך העיון במסכי ההסבר אינו מוגבל בזמן / מוגבל בזמן, הקלה / הקשתה עלי -			
-	178	2.37 (1.14)	179 3.44 (0.89)

ממוצע מחשב (ס"ת) n ממוצע ני"ע (ס"ת) n t

(14) בחינה ממוחשבת מתחשבת באנשים הקוראים לאט יותר מהממוצע –

3.58 (0.73) 179 - - -

(15) לו הייתי נבחן בבחינה רגילה עם תוספת זמן, ציוני היה טוב יותר

2.84 (1.08) 176 3.57 (0.87) 178 s

-1-----2-----3-----4-
מסכים בהחלט לא מסכים כלל

הממוצעים וסטיות התקן שמתוארים לעיל דומים מאוד לנתוני המשובים שנתקבלו בניסוי מפע"ם 1. מן הנתונים עולה כי שביעות הרצון של הנבדקים בניסוי מן המבחן הממוחשב הייתה גבוהה. שאלות 1,4,5,10 הנוגעות לגרפיקה של המבחן לבהירותו ולנוחות תפעולו זכו כולן לציוני שביעות רצון גבוהים. השאלות הייחודיות למבחן הממוחשב ביחס לחלוקת הזמן ולהוראות שאינן מוגבלות בזמן זכו אף הן לציונים גבוהים של שביעות רצון (שאלות 8,11,13). באשר להשוואה בין קבוצת המחשב לקבוצת הנייר ועיפרון, הרי שהשוואה זו מעט בעייתית, שכן מערך הבדיקה הוא בין נבדקים ולא בתוך נבדקים, אך עם זאת ניתן לשער כי כל נבחני הקבוצה הממוחשבת יודעים בדיוק מהו מבחן הנייר ועיפרון, ויכולים להשתמש בו כנקודת התייחסות. נמצא כי כמעט בכל המשתנים הייתה עדיפות למבחן הממוחשב, אך יש לציין כי ההבדלים בממוצעים בין שתי הקבוצות קטנים, וכי גודלי האפקט אינם גדולים. חשוב לציין במיוחד כי המבחן הממוחשב אינו נתפס כקשה יותר, או כמפלה לרעה ולא הוגן בהשוואה למבחן נייר ועיפרון (שאלות 2,3,6,9,12), וזאת על אף חוסר ניסיונם של הנבדקים בשימוש במחשב, ועל אף שהנבחנים מקבלים במרבית הזמן שאלות שקושיין נע סביב היכולת של הנבדקים. במבחן ממוחשב נבחן חלש מקבל בממוצע מבחן קל יחסית למבחן נייר ועיפרון, ואילו נבחן חזק מקבל בממוצע מבחן קשה יחסית. נבחנים בכל הרמות מקבלים שאלות הקרובות לרמת היכולת שלהם. זאת בניגוד למבחן נייר ועיפרון שם הנבדקים מקבלים שאלות קלות מאוד, אך גם קשות מאוד ביחס ליכולתם. גם המשוב שנתקבל מן השאלות שבדקו את מידת התאמת המבחן לנבחנים עם ליקויי למידה היה חיובי (שאלות 7,8,14,15) הנבדקים התרשמו כי המבחן הממוחשב מעניק זמן רב לכל שאלה ומפצה על קריאה איטית. בקבוצת הניסוי הממוחשבת הקושי להתרכז לא היה גדול יותר מאשר בקבוצת הני"ע, אלא אף להיפך, קבוצת הני"ע דיווחה על קושי גדול יותר בריכוז מפאת משך הבחינה ללא הפסקות. קבוצת הניסוי ני"ע סברה כי אילו היתה מקבלת תוספת זמן בבחינה רגילה ציונה היה גבוה יותר, ואילו קבוצת נבדקי המחשב היתה משוכנעת בכך הרבה פחות. בסך הכל הנתונים המוצגים נראים מעודדים. הבחינה הממוחשבת בהירה ונוחה לתפעול, היא נתפסת כהוגנת ולא קשה יותר ממבחן נייר ועיפרון, וכן כבחינה המתאימה לנבחנים שקריאתם איטית. מכיוון שבניסוי זה הנתונים לגבי מין הנבדקים היו מלאים, ניתן היה לבדוק השפעתם של משתנה המין ומשתנה השימוש במחשב על מידת שביעות הרצון מן המבחן. נתונים אלו מתוארים בטבלה 26.

טבלה מס' 26: נתוני המשובים (ממוצע וסטיות תקן) לפי אופנות, מין ומידת השימוש במחשב

נייץ				מחשב				גורם	קבוצה
נשים		גברים		נשים		גברים			
לא	משתמשת	לא	משתמשים	לא	משתמשת	לא	משתמשים	בניתוח	
משתמשת	(n=48)	משתמשים	(n=66)	משתמשת	(n=42)	משתמשים	(n=67)	שונות	
(n=31)		(n=19)		(n=37)		(n=23)		$F < 0.05$	
3.94	3.73	3.90	3.85	3.81	3.90	3.78	3.96	קבוצה x שימוש	(מסכי) ההוראות וההסברים היו ברורים
(0.25)	(0.54)	(0.32)	(0.36)	(0.62)	(0.29)	(0.60)	(0.20)		
2.13	2.19	2.63	2.27	2.35	2.17	2.47	2.34	-	בבחינה רגילה הייתי מצליח לענות על השאלות טוב יותר
(0.93)	(0.92)	(1.07)	(0.87)	(1.08)	(1.08)	(1.16)	(1.04)		
2.52	2.40	2.64	2.36	2.38	2.24	2.39	2.15	שימוש	השאלות שנשאלתי היו קשות עבורי
(0.89)	(0.64)	(0.68)	(0.82)	(0.59)	(0.69)	(0.18)	(0.72)		
2.94	2.65	2.95	2.85	3.41	3.67	3.47	3.60	קבוצה	סימון התשובות (בעזרת לוח המקשים) היה נוח
(1.21)	(1.02)	(1.27)	(1.09)	(0.96)	(0.65)	(0.95)	(0.76)		
3.29	3.52	3.42	3.32	3.51	3.81	3.82	3.76	קבוצה מין x שימוש	השאלות הוצגו בצורה ברורה ונוחה
(0.82)	(0.68)	(0.61)	(0.75)	(0.73)	(0.46)	(0.39)	(0.56)		
1.29	1.29	1.05	1.44	1.89	1.46	1.91	1.61	קבוצה	בחינה באמצעות מחשב / נייר ועיפרון אינה הוגנת כלפי
(0.64)	(0.71)	(0.23)	(0.79)	(1.10)	(0.78)	(1.16)	(0.90)	קבוצה x שימוש	
2.77	2.85	2.79	2.55	2.27	1.83	1.87	1.90	ניסוחים שונים לשתי הקבוצות)	התקשיתי להתרכז, עקב הצפייה הממושכת במסך המחשב / התקשיתי להתרכז לכל אורך הבחינה, שכן היא נמשכה זמן רב
(1.06)	(1.05)	(1.08)	(0.95)	(1.07)	(1.01)	(1.10)	(1.03)		
-	-	-	-	3.78	3.76	3.78	3.58	-	מבחן ממוחשב מעניק זמן רב לכל שאלה, ומפצה על קריאה איטית
				(0.42)	(0.62)	(0.52)	(0.73)		
-	-	-	-	1.84	2.00	2.07	1.91	-	בחינה באמצעות מחשב קשה יותר מבחינה המועברת בנייר ועיפרון
				(1.07)	(0.85)	(1.20)	(0.93)		

נייץ				מחשב				גורם	קבוצה
נשים		גברים		נשים		גברים			מין
לא	משתמשות	לא	משתמשים	לא	משתמשות	לא	משתמשים		שימוש במחשב
משתמשות (n=31)	(n=48)	משתמשים (n=19)	(n=66)	משתמשות (n=37)	(n=42)	משתמשים (n=23)	(n=67)	מובחן בניתוח שונות $F < 0.05$	
2.03 (1.17)	2.00 (1.01)	2.05 (0.97)	1.68 (0.75)	1.95 (1.03)	1.64 (0.98)	1.52 (0.67)	1.41 (0.78)	קבוצה מין	קטעי הקריאה והתרשימים הוצגו בצורה שאינה מאפשרת קריאה נוחה
-	-	-	-	3.76 (0.55)	3.76 (0.53)	3.61 (0.89)	3.56 (0.81)	-	העובדה שלשאלה מוקצב זמן משלה, מאפשרת להתרכז בה בשקט
-	-	-	-	2.11 (0.86)	2.18 (1.07)	1.91 (0.90)	2.08 (0.85)	-	אני סבור שבבחינה רגילה הייתי מוציא ציון נמוך יותר
2.10 (1.08)	2.56 (1.24)	2.84 (1.02)	2.18 (1.12)	3.54 (0.69)	3.21 (1.09)	3.52 (0.73)	3.43 (0.91)	(ניסוחים שונים לשתי הקבוצות)	העובדה שמשך העיון במסכי ההסבר אינו מוגבל בזמן / מוגבל בזמן, הקלה / הקשתה עלי
-	-	-	-	3.81 (0.40)	3.68 (0.65)	3.65 (0.65)	3.37 (0.87)	מין	בחינה ממוחשבת מתחשבת באנשים הקוראים לאט יותר מהממוצע
3.48 (0.96)	3.60 (0.84)	3.95 (0.23)	3.46 (0.92)	2.65 (1.21)	2.93 (1.05)	3.09 (1.08)	2.86 (1.02)	קבוצה מין שימוש	לו הייתי נבחן בבחינה רגילה עם תוספת זמן, ציוני היה טוב יותר

כמו כן בניסוי זה התאפשר לדעת מה הם הציונים בתלות בכל אחד מן המשתתפים: קבוצת ניסוי, מין ומידת השימוש במחשב. נתונים אלו מופיעים בטבלה 27. נתונים אלו הכרחיים על מנת להבין חלק מן הנתונים אשר מוצגים בטבלה 26. בטבלה 27 ניתן לראות את ציוני הנבדקים בחלוקה לשלושת המשתתפים הבלתי תלויים שנבדקו: משתנה אופנות הבחינה שנמצא לא מובהק (רי בפרק אודות ציוני שתי הקבוצות), משתנה המין שנמצא מובהק, אף כי לא באופן שונה מן המקובל במבחן נייר ועיפרון (רי בפרק אודות ההבדלים בין המינים) ומשתנה השימוש במחשב. מנתוני טבלה 26 עולה כי הגורמים הבאים היו שונים באופן מובהק בין קבוצת הניסוי הממוחשבת ובין קבוצת הנייץ: סימון התשובות היה נוח יותר לקבוצת המבחן הממוחשב, השאלות הוצגו בצורה ברורה יותר לקבוצה הממוחשבת, קטעי הקריאה והתרשימים הוצגו בצורה שאינה נוחה – קבי הנייץ ועפרון הסכימה עם טענה זו

יותר מאשר הקבוצה המחשב, וכאמור קבוצת המחשב פחות נטתה להסכים עם הטענה כי תוספת זמן בבחינה רגילה היתה משפרת את ציונה. משתנים אלו מתייחסים בעיקר לתפעול הבחינה, ולנוחות ההיבחות בה, ובכולם שביעות הרצון של קבוצת המחשב היתה גדולה יותר, דבר המחזק את ההתאמה של מבחן זה לנבחנים בתנאים מותאמים. נמצאה אינטראקציה בין קבוצת ניסוי לבין שימוש במחשב בשאלות הבאות:

- ההוראות וההסברים היו ברורים יותר בקבוצת המחשב למשתמשים במחשב, בעוד בקבוצת הניייע הם היו ברורים יותר ללא משתמשים במחשב. ומכל מקום ממוצעי שאלה זו היו גבוהים מאוד, כלומר היתה הסכמה כי ההוראות היו ברורות.
 - נבדקים שאינם משתמשים במחשב בקבוצת המחשב סברו שהבחינה הוגנת יותר בהשוואה למשתמשים בקבוצת המחשב, ואילו בקבוצת הניייע הלא משתמשים סברו שהיא פחות הוגנת בהשוואה למשתמשים. ההסבר לכך ניתן בטבלה 27, ממנה עולה כי ציוני הלא משתמשים במחשב נמוכים באופן מובהק מציוני המשתמשים במחשב. לפיכך הלא משתמשים במחשב בקבוצת המחשב קיבלו שאלות קלות יחסית למה שהם רגילים, ולפיכך הבחינה נראתה להם הוגנת יותר. נתון זה מסביר גם את העובדה שמשתנה השימוש במחשב יצא מובהק בשאלה: השאלות שנשאלתי היו קשות עבורי. נראה כי הלא משתמשים במחשב נטו יותר להסכים עם קביעה זו, מכיוון שרמתם נמוכה יותר.
- משתנה המין נמצא מובהק בשתי שאלות – נשים יותר נטו להסכים עם הקביעה כי קטעי הקריאה והתרשימים הוצגו בצורה שאינה מאפשרת קריאה נוחה, וכן נטו יותר להסכים עם הקביעה כי בחינה ממוחשבת מתחשבת באנשים הקוראים לאט יותר מהממוצע.
- אינטראקציה בין מין ושימוש נמצאה בשאלות הבאות:

- נשים שאינן משתמשות במחשב סברו שהשאלות הוצגו בצורה פחות ברורה בהשוואה לנשים המשתמשות במחשב, ואילו גברים שאינם משתמשים במחשב סברו כי השאלות הוצגו בצורה יותר ברורה בהשוואה לגברים המשתמשים במחשב.
- לעומת זאת נשים המשתמשות במחשב סברו כי תוספת זמן בבחינה רגילה היתה משפרת את ציון בהשוואה לנשים שאינן משתמשות במחשב, ואילו בקרב גברים, הלא משתמשים במחשב סברו שציונם היה משתפר בבחינה רגילה עם תוספת זמן בהשוואה לגברים המשתמשים במחשב.

טבלה מס' 27: נתוני המשובים לפי קבוצת הניסוי, המין ומידת השימוש במחשב

קבוצה		מחשב				ניייע		
מין		גברים		נשים		גברים		נשים
שימוש במחשב	ציון (ס.ת.)	לא	משתמשים	לא	משתמשים	לא	משתמשים	משתמשים
		529.00	605.82	514.81	563.60	544.05	607.26	583.46
(79.54)	(88.83)	89.26	(101.78)	(94.18)	(83.73)	(103.89)	(96.66)	

בניתוח שונות תלת-גורמי שבדק את הציון הכללי בניסוי בתלות בשלושת המשתנים הבלתי תלויים : קבוצת ניסוי, מין, שימוש במחשב נמצא כי הציון תלוי במשתנה המין ($F=11.09, Pr(F)=0.0010$), תלות זו תוארה והוסברה בפרק אודות ההבדלים בין המינים. כמו כן הציון תלוי במשתנה השימוש במחשב ($F=23.89, Pr(F)=0.0001$). הציון אינו מושפע מקבוצת הניסוי, או מכל אחת מן האינטראקציות בין שלושת משתנים אלו. ניתן להסביר את העובדה שציוני המשתמשים במחשב גבוהים יותר בכך שקבוצה זו אינה מקרית – היא בעלת יכולת אינטלקטואלית גבוהה יותר, עושה שימוש רב יותר במחשב ובאינטרנט ואולי אף בצבא משרתת ביחידות בהן נעשה שימוש במחשב. כלומר, לא השימוש במחשב הוא זה שהשפיע על הציון, אלא הרקע של קבוצה זו. הבדלים אלו נמצאו בשתי קבוצות הניסוי.

לבסוף, לנבדקים התאפשר לתת הערה כלשהי חיובית או שלילית, או להציע דבר מה באשר למבחן. סיכום ההערות שהתקבלו בכל אחת מקבוצות הניסוי וחלוקתן לפי תוכן מופיע בטבלאות 28.1 ו- 28.2. חלק מן הנבדקים העירו יותר מהערה אחת, וחלקם לא העירו כלל.

טבלה 28.1 - קבוצת ניסוי – מחשב – הערות פתוחות

שכיחות	הערה
45	הערה חיובית באשר למבחן ממוחשב
18	מוטב היה לאפשר לחזור לשאלות קודמות
16	המבחן שונה ממבחן נייר ועיפרון
15	הערה חיובית באשר להעדר לחץ זמן
15	קושי בעבודה עם הגרפים והתרשימים על הצג
14	קושי עם דפי הטיוטה והמעבר מהצג לדף
11	קושי עם שאלות בגיאומטריה על הצג
8	הערה שלילית לגבי מבחן ממוחשב
8	מוטב היה לאפשר לצבור זמן בין שאלות
3	מבחן ממוחשב מצריך הכנה
3	היו בעיות עם ניסוח השאלות
3	בעיות בהעברת המבחן
9	אחר
168	ס"ה הערות

ניתן לראות מן הטבלה כי הנבדקים ראו את יתרונו הגדול של המבחן בהעדר לחץ זמן, ואת חסרונו הגדול של המבחן בקושי לעבור בין הצג לבין דף הטיוטה, ובמיוחד בשאלות גרפיים וגיאומטריה בהן התרגלו לסמן בעיפרון

על גבי חוברת הבחינה בזמן פתרון השאלות. קושי זה בלט גם בניסוי, שבו פעמים רבות הנבדקים נראו מנסים לסמן קו דמיוני על הצג, או השתמשו בעיפרון כסרגל כדי לחבר בין נקודה לנקודה. קושי זה קשה יהיה לפתור בעתיד.

טבלה 28.2 - קבוצת ניסוי – נייר ועיפרון – הערות פתוחות

שכיחות	ההערה
11	הערה שלילית באשר ללחץ זמן
13	בעיות בהעברת הבחינה (כיסאות לא נוחים וכד')
12	הערה שלילית באשר להעדר הפסקות
6	הערה חיובית על הניסוי
3	הערה שלילית כלשהי (המבחן אינו מנבא טוב וכד')
3	סימון לא נוח של התשובות
3	הערה שלילית על מבחן ממוחשב
10	המבחן ארוך מדי
1	הערה חיובית על המבחן
1	קושי עם גרפים ותרשימים
1	הערה חיובית על מבחן ממוחשב
10	אחר
74	ס"ה הערות

בקבוצת הנייר ועיפרון היו הרבה פחות הערות, ועל כך אין להתפלא שכן במבחן זה אין חידוש מיוחד. מן הטבלה עולה כי עיקר הקושי של מבחן הנייר ועיפרון הוא סביב לחץ הזמן בבחינה, אורכה והעדר ההפסקות. נתונים אלו דומים לנתונים שנתקבלו בניסוי מפע"ם 1.

לסיכום, מנתוני המשובים עולה כי הנבדקים היו שבעי רצון מן הבחינה הממוחשבת, וכי משתני השימוש במחשב ומין הנבדקים לא השפיעו בצורה משמעותית על שביעות רצון זו.

סיכום

ניסוי מפע"ם 2 נועד לבדוק את ההיבטים הבאים של המבחן: תקינות המבחן, שביעות הרצון של הנבחנים מן המבחן, המאפיינים הפסיכומטריים של המבחן (בהשוואה לסימולציות), כיול הציונים של המבחן לציוני מכפ"ל והשוואה בין המינים לגבי ביצועיהם במבחן הממוחשב. כללית, נתוני ניסוי מפע"ם 2 דומים מאוד לנתונים שנתקבלו מניסוי מפע"ם 1. להלן פירוט של הממצאים העיקריים של הניסוי.

העברת הבחינה. ההפעלה של הבחינה עברה באופן חלק. בניסוי זה לא נתקבלו קבצי נבחנים בגודל 0KB ולא נתקבלו קבצי נבחנים חסרים לאחר שנעשה תיקון במחולל המבחנים כך שנעשית סגירה של הקבצים לאחר כל שלב.

שביעות הרצון. ניתוח תוצאות המשוב מעלה כי הנבחנים שבעי רצון מן הבחינה הממוחשבת מפע"ם 2. הם מעריכים אותה כבהירה ונוחה לתפעול, ותופסים אותה כהוגנת ולא קשה יותר ממבחן ני"ע וכמתאימה לנבחנים שקריאתם איטית.

מאפיינים פסיכומטריים. המאפיינים הפסיכומטריים של המבחן (מספר שאלות ממוצע, ממוצע שונות פוסטרירית, אחוז הנבחנים שלא התכנסו וכו') היו דומים מאוד לאלו שנתקבלו בסימולציה של 1,500 נבחנים בעלי יכולת המתפלגת נורמלית על סולם תטה. זאת, למעט אחוז גבוה מעט של נבחנים שלא התכנסו בניסוי בתחום חשיבה מילולית. ממצאים אלו מעידים על כך שהסימולציות שעל בסיסם נבנתה בחינת מפע"ם 2 קיבלו ביסוס אמפירי.

יצוגיות המדגם. בניגוד לניסוי מפע"ם 1, ממוצע ציוני קבוצת הנבדקים שלא ניגשה למכפ"ל (19 נבחנים מתוך 161 משתתפי הניסוי) בניסוי זה היה נמוך באופן מובהק (הבדל של כ-100 נקודות) משל הקבוצה שניגשה למכפ"ל. כלומר, הקבוצה שניגשה למכפ"ל לאחר הניסוי לא היתה קבוצה מקרית. עם זאת מהשוואת ציוני הקבוצה שניגשה למכפ"ל לכלל האוכלוסיה שניגשה למועד זה עולה כי ממוצעי שתי הקבוצות בכל התחומים מאוד דומים, ולפיכך יצוגיות המדגם אינה נפגעת בצורה משמעותית.

זמני תגובה. נתוני זמני התגובה שנתקבלו בניסוי זה דומים מאוד לנתוני זמני התגובה שנתקבלו בניסוי מפע"ם 1. בשלושת התחומים הנבחנים מנצלים בממוצע כשליש עד מחצית מן הזמן המוקצב לשאלה, למעט שני סוגי פריטים בתחום המילולי: השלמת משפטים והיגיון בדיד בהם משתמשים הנבחנים במרבית הפריטים ביותר מ-75% מהזמן המוקצב לפריט. נתונים אלו בצירוף נתונים דומים שהתקבלו בניסוי מפע"ם 1 ובהעברה ניסיונית של ה"קומפיוטרי" מחייבים לשקול הוספת זמן בפרקים אלו.

דגימת פריטים. אחוז הדגימה הממוצע של הפריטים בשלושת התחומים סביר (בתחום המילולי בין 13%-25%, בתחום החשיבה הכמותית בין 14%-20% ובתחום האנגלית בין 8%-21%), והוא דומה לזה שהתקבל בסימולציות שנערכו קודם לעריכת הניסוי, וכך לנתוני ניסוי מפע"ם 1.

כיוול הציונים. לגבי הכיוול הרי שהנתונים הינם מעודדים למדי. מתאמי בחינת מפע"ם 2 עם ציוני מכפ"ל גבוהים למדי (0.92 של הציון הכללי במפע"ם 2 עם הציון הכללי בבחינת מכפ"ל) ודומים מאד לאלו של בחינת הניי"ע בניסוי עם ציוני מכפ"ל. ממוצעי וסטיות התקן של הפרשי הציונים בין מפע"ם 2 למכפ"ל דומים מאד לאלו של הפרשי הציונים בין בחינת הניי"ע בניסוי ומכפ"ל. נראה כי תהליך הכיוול השיג את מטרתו.

ציונים. הבדל מובהק בציונים בניסוי לטובת קבוצת הניי"ע, נמצא בתחום המילולי. אולם, ההבדל קטן וכיוונית נשמרת גם במכפ"ל, כך שכנראה מקורו בהבדלים אמיתיים בין קבוצות הניסוי. ההבדל בציון הכמותי בין קבוצות הניסוי שהופיע בניסוי מפע"ם 1 אינו חוזר על עצמו בניסוי זה.

הבדלים בין המינים. בניסוי מצאנו כי בממוצע גברים היו בעלי ציון גבוה באופן מובהק מזה של הנשים במפע"ם 2. דפוס זה דומה מאד לזה שהתקבל בנתוני כלל הנבחנים במכפ"ל בשנת 99. הבדל מובהק בין המינים נמצא גם בציוני מכפ"ל של משתתפי הניסוי. אפקט אינטראקציה בין מין וקבוצה (מבחן ממוחשב, ניי"ע) לא נמצא מובהק בניסוי. בסה"כ נראה כי אין לחשוש לגבי הבדלים בין המינים בביצוע במבחן הממוחשב יותר מאשר במבחן ניי"ע רגיל.

שימוש במחשב. כמו בניסוי מפע"ם 1 וכמו בסקר נבחני במו"ת משנת 95 נמצא כי כשליש מן הנבחנים אינם משתמשים כלל במחשב. נתון זה מפתיע, אך הוא עקבי ולא השתנה לאורך השנים על אף התחזיות. נתון זה מחזק את בניית המבחן כמבחן שאינו מצריך ניסיון קודם במחשב.

נמצא כי ציוני הנבחנים שאינם משתמשים במחשבים נמוכים באופן מובהק מציוני הנבחנים המשתמשים במחשב. נתון זה מאפיין את שתי קבוצות הניסוי בשני תנאי ההיבחנות, והדבר מיוחס לרקע שונה של הנבחנים ולביטוי של מימוש הפוטנציאל השונה שלהם.

בסה"כ מעידים ממצאי הניסוי על היותו של מפע"ם 2, מבחן מהימן ותקף המקביל למבחן ניי"ע ולגרסת מפע"ם 1.

מקורות

- הלר, ד. ורון, ט. (1998). תהליך כיוול ציוני מבחן מפעיים לציוני מכפ"ל לקראת העברה ניסיונית של מבחן מפעיים נוסח 1.
- הלר, ד. ומושינסקי, א. (2000). סיכום תוצאות ניסוי מפעיים 1 (מבחן פסיכומטרי על מחשב). (דוח מס' 256). ירושלים: מרכז ארצי לבחינות והערכה.
- מושינסקי, א., טננבאום, מ., רפ, י., רון, ט. (1997). הרגלי השימוש במחשב בקרב הנבחנים בתנאים מותאמים. (דוח מס' 72). ירושלים: מרכז ארצי לבחינות והערכה.
- מושינסקי, א., רון, ט. (1998). בחינת מפעיים (מותאמת פסיכומטרית ע"י מחשב) המיועדת לנבחנים בתנאים מותאמים נוסחים 1, 2, מבנה הבחינה וניתוח תוצאות הסימולציות. (דוח מס' 85). ירושלים: מרכז ארצי לבחינות והערכה.
- פבלוב, י., אורן, כ. (1999). דו"ח סטטיסטי לשנת 1998.
- רפ, י., רון, ט., כהן, י. (1996). ניתוח זמני ביצוע פריטים בהעברה הניסיונית של CPETSE. דו"ח טכני 52א'.
- שפר, ל., רפ, י., רון, ט. (1998). מבחן אמיר"ם אנגלית מיון לרמות ממוחשב גרסאות 1-4, מבנה המבחן והתפלגויות הציונים. (דוח טכני מס' 86). ירושלים: מרכז ארצי לבחינות והערכה.
- Cohen, J. (1988). Statistical Power Analysis for the Behavioral Sciences (2nd ed.) Hillsdale, NJ: Erlbaum.