



מדינת ישראל
משרד החינוך

סיכום ממצאים של שתי שנות ניסוי במבחנים עם אינטרנט פתוח דו"ח מחקר

המנהל הפדגוגי אגף מו"פ ניסויים ויוזמות

אוקטובר 2022, תשרי תשפ"ג

תודה לשותפי המחקר:

מירב זרביב | מנהלת מערך המו"פ פיתוח ואישור המסלול
ד"ר עדית קמחי | ייזום, בנייה, תכלול ומחקר, מערך המו"פ
פרופסור חנה שחר | יעוץ, עיבוד נתוני מחקר וכתובה, מכללת גורדון
פרופסור סוגטה מיטרה | השראה, ליווי וייעוץ
נועה נוה | מנהלת מחלקת אקוסיסטם, מערך המו"פ
ריבי ארצי | מנהלת מחלקת מחקר, מערך המו"פ
רונית דרורי | ענבל ויסמן מרש | ד"ר ענת שפירא לביא | יפעת פילו | ניהול המסלול ופיתוח תוכן, מערך המו"פ
ד"ר מירי שליט | עיצוב מדיניות והובלת ממשקים יישומיים

בתי הספר המשתתפים בניסוי (לפי א"ב): אורט גוטמן, נתניה | אחד העם, פתח תקווה | תיכון אית"ן, סתריה
גזר | בויאר, ירושלים | דרור משלב, ברנקו וייס | הר טוב, ברנקו וייס | חט"ב בן צבי, פתח תקווה | חטיבה
צמודה ריינה, מקיף חדש | כוכבי המדבר, ברנקו וייס | כפר הנוער מאיר שפייה | מיתרים ברנקו וייס, מיתר |
מיתרים ברנקו וייס, רעננה | מקיף ד', אשדוד | קריית דרור, בני דרור | תבנית החינוך המוזהב, מעלה צבייה,
תיכון ניסויי, מסעדה

אוקטובר 2022, תשרי תשפ"ג

בשנת תשפ"א נשלח קול קורא אל בתי הספר הנמצאים ברשת מחוללי חדשנות, שהיא רשת של מומלצי מחוזות ושל מסיימי מסלולי פיתוח באגף מו"פ¹, להשתתפות בניסוי חלוץ שמטרתו להתנסות בכתיבה ובהערכה של שאלות מבחן שבמהלכו מתאפשרת גישה פתוחה למקורות דיגיטליים ואחרים, דהיינו, שאלות למבחנים עם אינטרנט פתוח. בנוסף, נשלח הפרסום גם לבתי ספר מחוץ לרשת שהביעו עניין בחדשנות פדגוגית. כ-150 בתי ספר הגיבו לקול הקורא ולאחר פעולות הדרכה והבהרה, נרשמו לניסוי עצמו 10 בתי ספר. הרקע ליוזמה זו הייתה הבנה שהמציאות משתנה באופן דרמטי ושאיננו לנו יותר אפשרות להמשיך לנהל את מערכת החינוך, הן ההוראה והן ההערכה אך בעיקר ההערכה שהיא מעין "תחנה סופית", באותו אופן מסורתי שבו היא מתנהלת במאתיים השנים האחרונות. המחיר של המשך דרכי ההערכה המסורתיות הוא גבוה ובלתי הפיך: חוסר התאמה של הבוגר למציאות המשתנה, הנצחת פערים וחוסר רלוונטיות של החינוך הציבורי.

רקע תיאורי קצר

חוקרים רבים ובולטים בעולם פירטו את הבעיות המורכבות הכרוכות בהערכה על-פי מבחנים אחידים (Berliner, 2015; Dieter-Meyer, 2014; Hargreaves, 2020; Jackson, Khavenson, & Chirkina, 2020;) (Menken, 2006; Nichols, Berliner, & Noddings, 2007). הביקורת עליהם ברורה ונוקבת, והיא כוללת את התחום הפדגוגי, החברתי והן את התחום המוסרי.

בתחום הפדגוגי נזכיר כאן שתי טענות. האחת היא שמבחנים אחידים מצמצמים מאוד את תכני הלימוד כיוון שזמן רב מוקדש לאימון ולתרגול כדי להצליח בהם, והדבר פוגע ברוחב השכלתם של התלמידים. הטענה השנייה מתייחסת לחרדת בחינות. חרדת בחינות הוא נושא שנחקר לא מעט, ובמהלך השנים פותחו כלי מחקר שנעשה בהם שימוש רב (ראו השאלון של Mandler & Sarason, 1952). ממצאים מראים שחרדת בחינות קשורה לטיב הביצועים דהיינו, שככל שרמת החרדה גבוהה יותר, איכות הביצועים יורדת (Putwain & Best, 2011), ויש לה השפעה מזיקה על הדימוי העצמי (DePhil, Brilot & Nettle, 2011). כל הפחתה במתח הקשור למבחנים עשויה לשחרר כוחות שיכולים להיות מופנים ליצירה, חשיבה ופתרון בעיות, שבונים בתורם אקולוגיה רצויה של למידה ואמון בין מורים לתלמידים.

בתחום החברתי הביקורת כנגד שאלות במבחנים אחידים, המלווה בממצאים אמפיריים רבים, היא שהם מגדילים את הפערים בין תלמידים חזקים לבין כאלה שאינם ופוגעים ביכולת ההתקדמות של תלמידים מבחינת ניווד חברתי (Jackson, Khavenson,) (Domina & Penner, 2017; Furuta Schofer & Wick, 2020; van Noord et al., 2019; & Chirkina, 2020). דומינה ופנר (2017) אף טוענים באופן נחרץ שבתי ספר מהווים מכונות מיון והסללה חברתית של תלמידים על ידי השימוש במבחנים אחידים. הגדלת פערים בין אוכלוסיות של תלמידים ידועה כנושאת עימה נזקים חברתיים וכלכליים לטווח ארוך בכל חברה דמוקרטית (Azzolini, Schnell, & Palmer, 2012; Santamaria, 2009; Strand, 2016; Van de Werfhorst, Mijs, & Jonathan, 2010).

בתחום המוסרי המבחנים גורמים לתלמידים, למורים ולבתי ספר, לרמות לגבי התוצאות בכל דרך אפשרית (Berliner, 2015; Shalberg, 2015). מכיוון שמערכת החינוך כפופה לאופי התחרותי של מטרות סטנדרטיות

¹ התנסות מחקרית זו נעשתה לבקשתה של מנהלת מערך המו"פ במשרד החינוך, הגב' מירב זרביב, במסגרת מחלקת המחקר של מערך המו"פ בראשותה של גב' ריבי ארצי, כאחד מתוך שישה צירי פעולה של רשת מוסדות חינוך מחוללי חדשנות 2.0, שהחל את פעולתו בדצמבר 2020, בניהולה של גב' רונית דרורי.

ושאלונים סטנדרטיים ומרבה בהשוואות בין תלמידים, בין כיתות, בין בתי ספר ובין מדינות, הרי בבתי ספר רבים קורה שתלמידים שידועים כמי שעלולים לפגוע במוצע מקבלים "עידוד" שלא להיבחן ביום המכריע של בחינות ארציות או לעבור לבתי ספר מקצועיים. אלה הם מסרים קשים מבחינת הערכים שהם מייצגים, והם עומדים בניגוד חריף להצהרות על חינוך לכבוד, ולקבלה והכלה של השונה, שנשמעות השכם והערב מפי קובעי מדיניות במערכות חינוך ציבוריות. הדרישה לסטנדרטיזציה יוצרת תחרותיות, לחץ ומתח אשר מגבילים ומסתירים את מלוא הפוטנציאל האנושי, הן האישי והן הקולקטיבי. הצורך להצליח במבחנים הסטנדרטיים יצר מציאות שבה שולטת חווית הישרדות. צורך ההישרדות של תלמידים גורם לחלקם לרמות כדי להשיג ציון גבוה. צורך ההישרדות של המורים, הנשפטים על הישגי תלמידיהם, גורם להם לסייע לתלמידים בדרכים שאינן תמיד כשרות (על-אף שהן אנושיות), וצורך ההישרדות של בתי ספר גורם למקרים רבים שבהם תלמידים שמסכנים ממוצע של בית ספר כלל אינם ניגשים לבחינות, ולא אחת גם לפרסום מסולף של ציונים על ידי בית הספר.

עקב השינויים במציאות, התחזקה בשלושה העשורים האחרונים המגמה להחלפת דרך ההערכה באמצעות מבחנים אחידים כדרך בלעדית, להערכה בדרכים חלופיות (Plucker, Callahan & Tomchin, 1996). מגמה זו קבלה דחיפות ותאוצה לאחר התפרצות מגפת הקורונה (Cairns, 2020), שבמהלכה השתנו סדרי עולם בכלל וסדרי העולם החינוכי בפרט. לאור השינויים הדרמטיים מפתיע הניסיון להחזיק בתרבות המבחנים הרגילים כמעט בכל מחיר. בשמה נעשים מאמצים למצוא דרכים יצירתיות לבלוש ולעקוב אחר תלמידים שנבחנים בבתייהם, לעיתים קרובות בזום, וכפי שכינה זאת מיטרה (Mitra, 2014; 2018), משמרים את הדינמיקה של 'שוטרים וגנבים'. אלא שלצד ניסיונות אלה יש ניסיונות חדשים אשר נשענים על ההבנות שהתפתחו בשלושה העשורים האחרונים לגבי הערכה אחרת, כזו שמכירה ביכולת ההתפתחות של אינטליגנציה, וכן, לגבי ההכרה שלבני אדם שונים יש אינטליגנציות שונות בתחומים שונים (Shearer, 2004). בעקבות הכרה זו פותחו כלי הערכה מגוונים שבעיקרם בודקים את טיבם של תהליכי למידה ורק בחלקם הקטן בודקים את תוצרי הלמידה (Weurlander et al., 2012). לצד אלה נכנס לתמונה בשני העשורים האחרונים האינטרנט שהפך כמעט את כל הידע הקיים בעולם לנגיש לכמעט כל אדם ולכל ילד, בכל זמן, עובדה שכמעט מייתרת את המורים כמקור הידע. תפקיד המורים כיום הוא, כפי שצוין בדו"ח המבקר האחרון, יותר ללמד את תלמידיהם כיצד להפיק את הידע הדרוש להם, וכיצד לבדוק את מידת מהימנותו.

האופן בו מבוגרים פועלים להשגת ידע בזמן קצוב הוא באמצעות האינטרנט. בין שהם עובדי היי טק, אנשי אקדמיה, אנשי ממשל, ממלאי תפקידים בסקטור הפרטי, בסקטור השירותים, באומנות או בתרבות, הדרך לבדיקה של ידע, ללמידה ולהצלבת מידע הוא כמעט ללא יוצא מן הכלל בעזרת האינטרנט (Sanger, 2010). מציאות זו משאירה את המבחנים האחידים במתכונתם הנוכחית כבועה אנכרוניסטית בזמן הזה. היא גם מעלה שאלות לגבי נחיצותם, הן מבחינת יעילותם והן מבחינת התאמתם לעולם שאליו יגיעו התלמידים, שבו יצטרכו אוריינות דיגיטלית מפותחת, שאינה קשורה למבחנים. בזמן שתפקיד המורה משתנה מהיותו מקור הידע להיותו מארגן ומנחה לומדים עצמאיים כפי שניתן לתמצת באמירה מפי המורה/מנחה: "למקומות אליהם אתם הולכים, אני בא ביחד איתכם", לשינוי זה אין כל זכר במבנה המבחנים והערכה, בסוג השאלות ובחוויות ההיבחנות (Mitra, 2020). ענין זה קבל התייחסות בדו"ח המבקר האחרון (מבקר המדינה דו"ח שנתי, 2021) האומר ש"יסיבת הלימודים הפיזית, הטכנולוגית והתיקשובית של בתי הספר... לא סיפקה באורח מיטבי את התנאים הנחוצים להקניה אפקטיבית של מיומנויות המאה ה-21 בכלל, והאוריינות הטכנולוגית והדיגיטלית בפרט". עוד מוסיף המבקר כי "על משרד החינוך לגבש תכנית אסטרטגית ייעודית להטמעת מיומנויות אלו, שתשמש מסגרת לפעילות מתואמת בין כל יחידות המטה כדי להסדיר את אופן ההטמעה של המיומנויות בתוכניות הלימודים ובהערכה ואת... האמצעים להרחבת הגמישות הפדגוגית בבתי הספר העל-יסודיים". אוריינות דיגיטלית היא כשירות מורכבת הכוללת מיומנויות חיפוש, מיון, סינון, והבנת התכנים, כמו גם הצלבה והשוואה של מקורות ושילוב תכנים. כל אלה, מצוין הדו"ח, הן מן המיומנויות החשובות הדרושות להתנהלות אישית מושכלת ולעולם התעסוקה.

כל אלה הולידו את הניסוי המתואר בהמשך.

מהן שאלות עם אינטרנט פתוח?

שאלות במבחנים עם אינטרנט פתוח (מא"פ) הן שאלות שבמהלך העבודה עליהן יש לנבחנים גישה פתוחה למקורות מידע ברשת ו/או אחרים, כפי שימצאו לנכון ויוכלו להיעזר בהם בזמן הנתון והמוגבל שנקבע לאירוע ההיבחנות. בשאלות אלה נשאלת שאלה שאין עליה תשובה אחת ואין עליה תשובה מן המוכן. השאלה מזמנת שילוב של כישורים (מיומנויות וידע) של חיפוש, מיון וסיווג מקורות, חשיבה מסדר גבוה ויכולות יישומיות ואינטגרטיביות בזמן קצר. השאלות שנשאלות הן באוריינטציה מעשית, הן רלוונטיות ובדרך כלל מעניינות בכך שהן מערבות את חייו של הנבחנים, והן מבקשות לזמן גם את ההיגיון הפנימי שלהם בפתרון בעיות, תכנון ובהשגת ידע נדרש בזמן מוגבל. הנחת העבודה במא"פ היא שהשאלה צריכה להיות שאלה רלוונטית לחיי הילדים, כלומר שיש סיכוי שהם או חבריהם או מי ממכריהם יידרשו למשימה זו או דומה לה. כמו כן, רצוי שהשאלות יהיו בסבירות גבוהה שלא התלמידים ולא מוריהם נתקלו בהן, או עסקו בהן עד כה. הכוונה היא שהשאלה תאפשר לנבחנים מרחב ליצירתיות וליישומי ידע המדמים את הדרכים בהם מבוגרים מנהלים תחקירים בחייהם השוטפים ומקבלים החלטות כגון לאיזה רופא לפנות, איזה מוצר להזמין או האם פרסום מסוים אמת או כזב. לדוגמה, אם תחום הלימוד הוא טריגונומטריה, השאלה יכולה להיות "אתם כותבים מדריך על מכשירים סלולריים. אנא כיתבו את הפרק: כיצד המכשיר הסלולרי שלכם יודע היכן אתם נמצאים בעזרת שתי פסקאות ושני שרטוטים?" (מיטרה, 2020). בדוגמה הזו מתקיימים התנאים של מעשיות ורלוונטיות. במקרה הזה, גם אם ישנה תשובה מן המוכן על כך במרשתת, הרי כאשר הנבחן נדרש למצוא אותה ולהסביר אותה בזמן קצר, נבחנת יכולתו לייצר מילת חיפוש מתאימה, לסנן ולאחר מכן המקורות את הרצוי, לקרא, להבין ולהסביר במסגרת המגבלה המבנית של שתי פסקאות ושני שרטוטים, שם גם מקומו להביא לידי ביטוי את ההיגיון והיצירתיות שלו.

ניסוי שנה ראשונה (מא"פ 1.0)

שיטה

ההכנה לביצוע הניסוי: בשלב ראשון נכתב מתווה מחקרי בשיתוף עם מומחים. בשלב שני גויסו בתי הספר המיועדים להתנסות. בתי הספר עברו מיפוי על פי מקצועות לימוד וחיטבות גיל. לאחר המיפוי נעשו פעולות הדרכה לצוותים החינוכיים שנרשמו כמשתתפים בניסוי. ההדרכה כללה שלושה מפגשים שבהם ניתנו הסברים לגבי הניסוי וכן כללים לכתיבת שאלות המבחן. בשלב השלישי התבקשו הצוותים בבתי הספר לכתוב את שאלות המבחן ואת מחוון ההערכה לכל מבחן ולשלוח אותם לצוות המארגן של הניסוי.

צוות המחקר קרא, העיר ותיקן בהתאם לאמות המידה שנקבעו מראש. על מנת להעריך את התאמת השאלות למא"פ פותח כלי הערכה ובו חמש אמות מידה שכל אחת נעה בין 0 (לא מתאימה) ל-2, (מתאימה ביותר). להלן מוצגות אמות המידה להערכת השאלות עם הפירוט של כל אחת מהן (זוהי הגרסה הראשונה ששימשה בשנה הראשונה וכונתה מאוחר יותר "מחוון שאלות מא"פ"):

1. זימון מעשה/חשיבה	2. רלוונטיות לבחנים/נבחנים	3. זימון ביקורת/דעה/עמדה	4. זימון מיומנות למידה יישומית	5. זימון אוריינות מידענית
התלמיד מתבקש לתכנן מעשה, לתאר מעשה, לחשוב ולמסגר סוגייה.	הסוגייה או משימת הלמידה מתייחסת למציאות החיים	התלמיד מתבקש להביא את קולו האישי	התלמיד מתבקש להביא ידע ולקשר אותו למשימת הלמידה/סוגייה	התלמיד מתבקש להיוועץ במקורות מידע על פי שיקול דעתו והערכתו ולמזג מידע רלוונטי

הניתוח של השאלות שהוצעו על ידי המורים נערך על-פי הקריטריונים לעיל. הלוח הבא מציג ניתוח ציונים של 24 שאלות במקצועות השונים, שנבחרו מתוך כלל השאלות שחוברו על ידי המורים. ככל שהשאלה נמצאה מתאימה יותר כך הציון גבוה יותר. הערה: ממוצע המקצוע הוא בעצם ממוצע של הציונים שקיבלו שתי שאלות באותו מקצוע (על ידי צוותים שונים) שהצוות המארגן העריך בנפרד ואחר כך חושב הממוצע.

לוח 2: הערכה מדגמית של 18 (מתוך 36) שאלות במקצועות השונים לגבי מידת ההתאמה של השאלות למא"פ

מקצוע	אמת מידה 1	אמת מידה 2	אמת מידה 3	אמת מידה 4	אמת מידה 5	סה"כ
תנך	2	2	2	2	2	10
תנך	2	2	2	1	1	8
אזרחות	2	2	1	2	1	8
אזרחות	2	2	2	2	2	10
מדעים	0	1	0	1	0	2
מדעים	0	1	0	1	0	2
היסטוריה	1	2	1	2	2	8
היסטוריה	2	2	2	2	2	10
ספרות	1	1	2	2	2	8
תנך	1	1	2	1	1	6
תנך	0	1	1	1	0	3
אזרחות	2	2	0	2	2	8
תקשורת	0	2	1	2	1	6
תקשורת	0	2	1	2	2	7
אנגלית	1	1	1	1	1	5
אנגלית	1	1	1	1	1	5
ספרות	1	1	1	2	2	7
ספרות	2	2	2	2	2	10

מלוח 2 ניתן לראות שהשאלות במקצועות היסטוריה ואזרחות קבלו בממוצע את ציוני ההתאמה הגבוהים ביותר (9), ואילו מקצוע המדעים קיבל את הציון הנמוך ביותר (2). השאלות בספרות קבלו בממוצע 8 נקודות, ותנך 6.75 נקודות.

ההנחיות למורים היו לחלק את הכיתה לשתי קבוצות אקראיות, לכתוב מבחן עם שתי שאלות, משך הזמן לכל שאלה היה 30 ד'. תחילה קבוצה א' ענתה על שאלה ראשונה רגילה בזמן שהקבוצה השנייה עונה של שאלה באינטרנט פתוח ולאחר חצי שעה, מתחלפים: קבוצה א' עונה על השאלה באינטרנט פתוח והקבוצה השנייה באינטרנט סגור, קרי שיטה רגילה.

עיבוד הנתונים: הניתוח משווה בין הציון שכל תלמיד קיבל בשאלה שענה עליה באופן רגיל ובין הציון שקיבל בשאלה שענה עליה באינטרנט פתוח. נערכו גם השוואות בין בנים ובנות, בין מקצועות לימוד שונים ובין קבוצות הגיל השונות. כמו כן נותחו משובי התלמידים והמורים ביחס לאיכות החוויה החינוכית.

סוג המחקר: משולב איכותני וכמותי. החלק הכמותי כולל את חישוב הציונים שקיבל כל תלמיד בשאלה הרגילה ובשאלה באינטרנט פתוח. החלק האיכותני כולל ניתוח תוכן של משובים בכתב של תלמידים,

שנתבקשו על ידי מוריהם לכתוב באופן חפשי את דעתם על ההתנסות בשני השאלות. מספר התלמידים שהשיבו היה 40, שהם כעשרה אחוזים מכלל התלמידים שהשתתפו. על אף המספר הנמוך עברו המשובים ניתוח תוכן, מתוך רצון ללמוד משהו לגבי האופן שבו תפסו התלמידים את ענין השימוש באינטרנט פתוח.

המשתתפים: 10 בתי ספר ו-33 כיתות, בחטיבות גיל שונות ובמקצועות שונים. הכתות שייכות למגזרים שונים, יהודי ממלכתי, ערבי וממלכתי-דתני.

ממצאים של שנת הניסוי הראשונה (מא"פ 1.0)

1. הבדלים בין התוצאות בשני סוגי השאלות

תחילה חושבו מתאמים בין ממוצע הציונים של השאלה עם א"פ לבין זו של השאלה במתכונת מבחן רגיל. החישוב העלה שיש ביניהם מתאם בינוני גבוה ($t = .47, p < .001$). כלומר ככל שתלמיד קיבל ציון גבוה במבחן אחד, הוא בדרך כלל קיבל ציון גבוה במבחן השני.

בשלב שני חושבו ההבדלים בין ממוצעי הציונים בשני סוגי השאלות. אלה מוצגים בלוח 3.

לוח 3: ממוצעים, סטיות תקן וערכי t של ההבדלים בין ממוצעי הציונים בשתי השאלות (N=376)*

t	ציון שאלה במבחן רגיל (N=376)	ציון שאלה עם א"פ (N=376)	M	SD	ציוני השאלות
6.42***	36.31 (11.44)	39.90 (9.40)			
					*** $p < .001$

(*תלמידים שנבחנו רק באחת מן השאלות הושמטו מן החישוב)

מן הלוח ניתן לראות שנמצא הבדל מובהק בממוצע הציונים בין שני סוגי השאלות ($M = 39.90, SD = 9.40$), $(M = 36.31, SD = 11.44)$, לטובת הציונים שהתקבלו בשאלות עם אינטרנט פתוח ($t = 6.42, p < .001$).

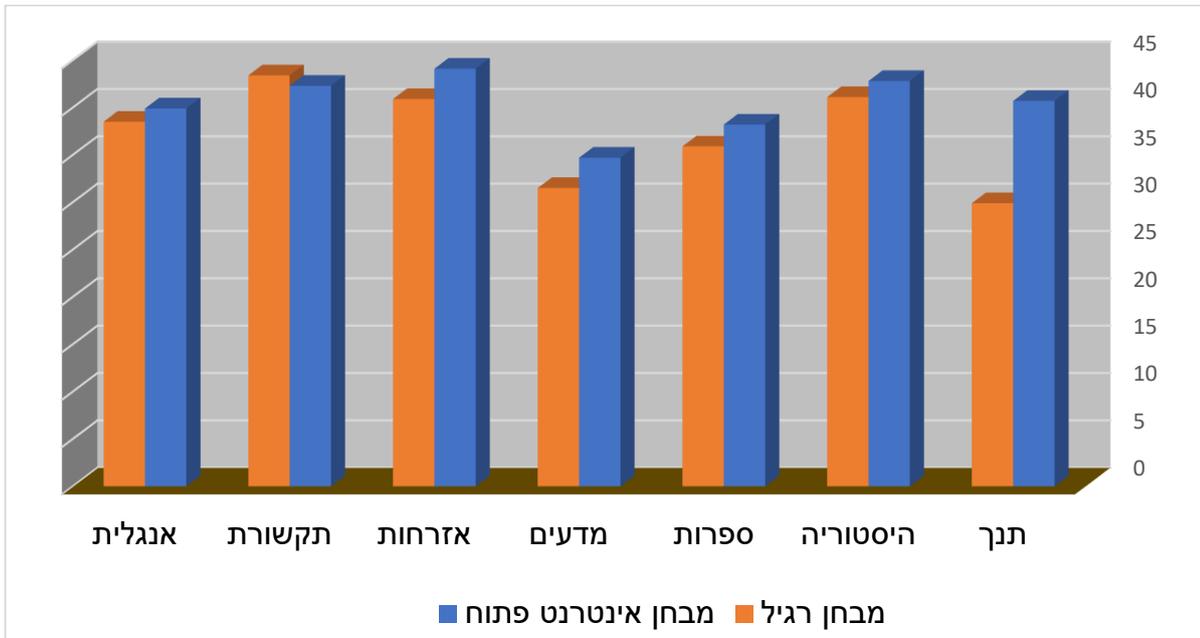
בלוח הבא מוצגים ההבדלים בין ממוצעי שתי השאלות בכל אחד ממקצועות ההוראה בנפרד

לוח 4: הבדלים בין ציון א"פ לשאלה רגילה, בכל אחד ממקצועות ההוראה שנבדקו (N=410)

מתאמים בין הציונים בשתי השאלות	אינטרנט סגור	אינטרנט פתוח	מקצוע ההוראה				
t	SD	M	SD	M	N		
.49***	9.01***	10.8	29.9	6.9	40.7	63	תנך
.65***	2.19*	9.6	41.1	8.2	42.8	89	היסטוריה
.46***	2.02*	8.6	35.9	8.6	38.2	61	ספרות
.26*	1.84	13.8	31.5	11.5	34.7	77	מדעים
.52***	2.17*	8.5	40.9	9.3	44.1	35	אזרחות
.33	-0.5	9.3	43.4	9.1	42.3	22	תקשורת
.73***	1.15	9.0	38.5	8.4	39.9	29	אנגלית
							*** $p < .001$ ** $p < .01$ * $p < .05$

מלוח 4 עולה כי: 1. במקצועות המדעים, התקשורת ואנגלית, לא נתגלו הבדלים בין שני סוגי השאלות. 2. ממוצע הציונים הגבוה ביותר בשאלה עם א"פ התקבל במקצוע אזרחות, והנמוך ביותר במדעים; 3. בשאלה הרגילה היה הציון הממוצע הגבוה ביותר במקצוע תקשורת, השני בהיסטוריה והשלישי באזרחות. הציון הנמוך ביותר בשאלה הרגילה היה בתנך. 4. בתנך נמצאו גם ההפרשים הגבוהים ביותר בין שני סוגי השאלות, לטובת א"פ.

התרשים הבא ממחיש את ההבדלים בין שני סוגי השאלות



תרשים 1: הבדלים בין ממוצעי הציונים בשני סוגי השאלות בכל אחד ממקצועות הלימוד (N=410)

ענה נבדק אם קיימים הבדלים בין בנות לבין בנים בשני סוגי השאלות. לוח 4 מציג את ממצאי הבדיקה.

לוח 5: מוצעים, סטיות תקן וערכי t של ההבדלים בין הציונים בשני סוגי השאלות בקרב הבנים והבנות בנפרד (N=376)

t	בנים (172=N)		t	בנות (204=N)		
	שאלה רגילה	א"פ		שאלה רגילה	א"פ	
5.15***	35.19	39.42	4.00***	37.26	40.30	ממוצע
	12.26	9.70		10.64	9.14	סטיות תקן

***p < .001

מלוח 5 עולה כי לא נתגלו הבדלים מובהקים בממוצעי הציונים בין הבנות לבין הבנים בשני סוגי השאלות.

2. איזה רמת לומדים "הרויחה" יותר משאלות עם אינטרנט פתוח?

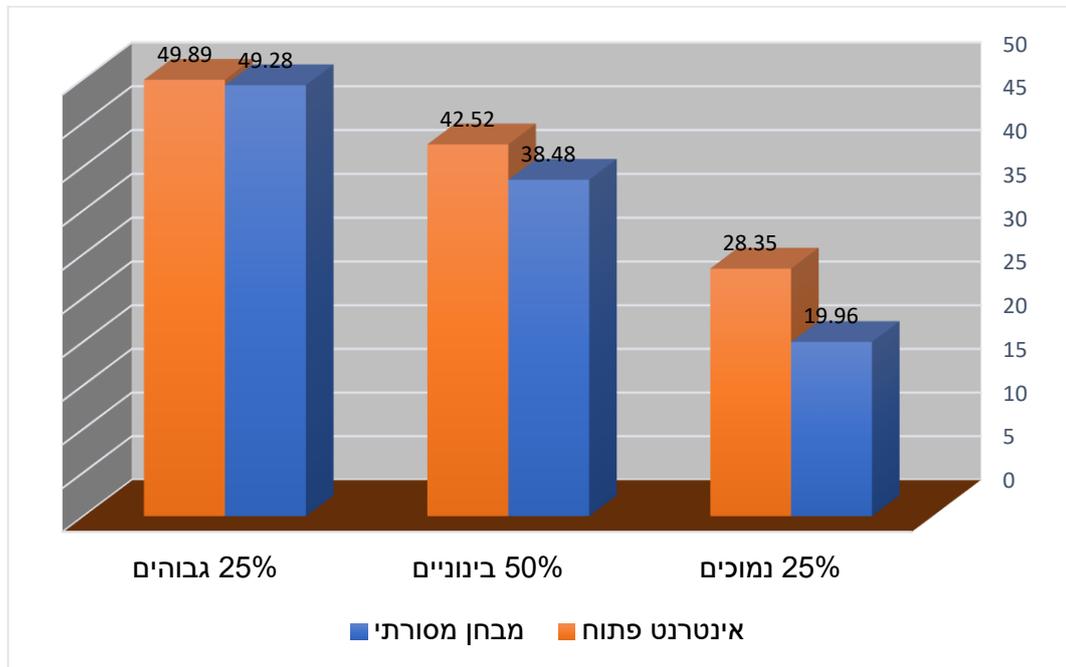
כדי לענות על השאלה אם יש קשר בין רמת הישגים של תלמידים לבין הפער בין הציונים בשני סוגי השאלות חושב הציון הממוצע של שתי השאלות כבסיס לקביעת רמת ההישגים. נערכה השוואה של ההבדלים בין ציוני השאלות בין שלוש רמות הישג: 25% בעלי ההישגים הנמוכים ביותר, 50% בעלי הישגים בינוניים, ו-25% בעלי הציונים הגבוהים ביותר. הלוח הבא מציג את הממוצעים, סטיות התקן וערכי t של ההבדלים בין שלוש רמות ההישגים בשני סוגי השאלות.

לוח 6 : ממוצעים, סטיות תקן וערכי t של ההבדלים בין שלוש רמות הישגים בשני סוגי השאלות (N=379).

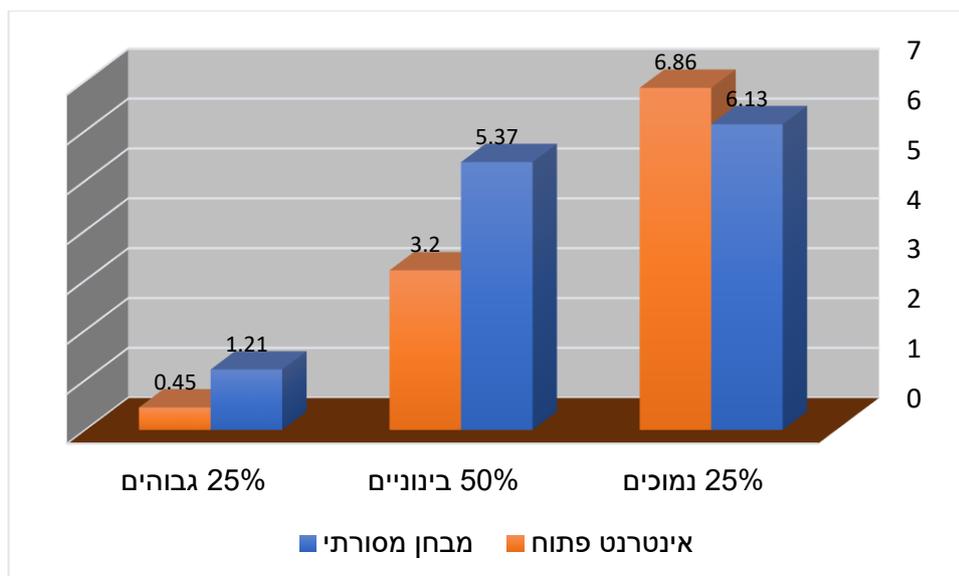
25% גבוהים			50% בינוניים			25% נמוכים			
t	מא"פ (92=N)	שאלה רגילה (85=N)	t	מא"פ (176=N)	שאלה רגילה (N=199)	t	מא"פ (125=N)	שאלה רגילה (95=N)	
4.74***	49.89	49.23	8.97***	42.52	38.48	9.40***	28.35	19.96	ממוצע
	.45	1.21		3.20	5.37		6.86	6.13	ס.ת.

מלוח 6 עולה שההבדל בין שתי השאלות בקרב בעלי הישגים הנמוכים היה 8.34 נקודות לטובת שאלה עם א"פ, באותו אופן היה ההבדל בקרב בעלי הישגים בינוניים (4.04 נקודות) ואילו בקרב בעלי הישגים גבוהים לא היה הבדל. במקביל, ניתן לראות כי טווח הציונים בקבוצת בעלי הישגים הנמוכים היה הגבוה ביותר מבין שלוש הקבוצות, בשני סוגי השאלות.

התרשימים הבאים ממחישים את ההבדלים בין שלוש הקבוצות.



תרשים 2 : הבדלים בין שני סוגי השאלות בקרב שלוש רמות הישגים



תרשים 3 : הבדלים בסטיות התקן של ציוני השאלות בשלוש רמות הישגים

כדי לבדוק אילו משתני רקע מסבירים את הציונים בשני סוגי השאלות נערכה רגרסיה מרובה שלתוכה הוכנסו רמת ההישגים, מקצוע ההוראה ושכבת הגיל.

לוח 7 מציג את ממצאי הרגרסיה.

לוח 7 : רגרסיה מרובה להסבר הציונים בשני סוגי השאלות על-ידי משתני הרקע

מבחן במא"פ				
משתנים מסבירים	β	B	t	R ²
רמת הישגים	.76	9.88	22.57***	0.582
מקצוע ההוראה	-.16	-.73	-4.15***	0.013
שכבת גיל	.09	.65	2.18*	0.005

מבחן עם מבחן רגיל				
משתנים מסבירים	β	B	t	R ²
רמת הישגים	0.79	12.56	25.6***	0.636
מקצוע ההוראה	0.1	0.55	3.13***	0.014

מלוח 7 עולה שרמת ההישגים הכללית של התלמידים מסבירה למעלה מ- 50% מן השונות בציוני א"פ. עוד אחוז אחד מוסבר על ידי מקצוע ההוראה ותרומה קטנה נוספת הייתה של שכבת הגיל, שהוסיפה חצי אחוז להסבר הציונים. בשאלות רגילות לעומת זאת, הסבירה רמת ההישגים של התלמידים 63% מן השונות בציונים, ומקצוע ההוראה תרם עוד 1.4%. שכבת הגיל לא נכנסה כלל למשוואת הרגרסיה.

הבדלים בין שכבות הגיל

כדי לבדוק אם היו הבדלים בין שכבות הגיל השונות נערכה תחילה השוואה ביניהן, שממנה עלה כי בין שתי הכיתות הבוגרות (יא, יב) לא נמצאו הבדלים. לכן אוחדו שתי כיתות אלו לשכבת גיל אחת. בשלב שני נערך ניתוח שונות רב-משתני חד כיווני ANOVA כדי לבדוק את ההבדל בין כל השכבות, ואת ההפרשים בין שני השאלות בכל השכבות.

לוח 8 מציג את הממוצעים, סטיות התקן וערכי F של ההבדלים בין השכבות ואת ההבדל בציוני הפער בין שתי השאלות בכולן.

לוח 8: ממוצעים, סטיות תקן וערכי F של ציונים בשתי השאלות והפער ביניהם על פי שכבות גיל (N= 393)

F	יא+יב (N= 103)	י (N= 39)	ט (N= 173)	ח (N= 78)		
17.23***	44.27 (7.92)	39.53 (9.02)	36.50 (9.98)	41.05 (7.23)	ממוצע סטיות תקן	ציוני אפ
11.43***	40.65 (11.83)	40.04 (8.87)	34.85 (10.55)	31.93 (11.92)	ממוצע סטיות תקן	ציוני אס
12.85***	4.18 (8.73)	- .97 (9.21)	1.55 (11.39)	9.54 (10.38)	ממוצע סטיות תקן	הפער בין ציוני שתי השאלות

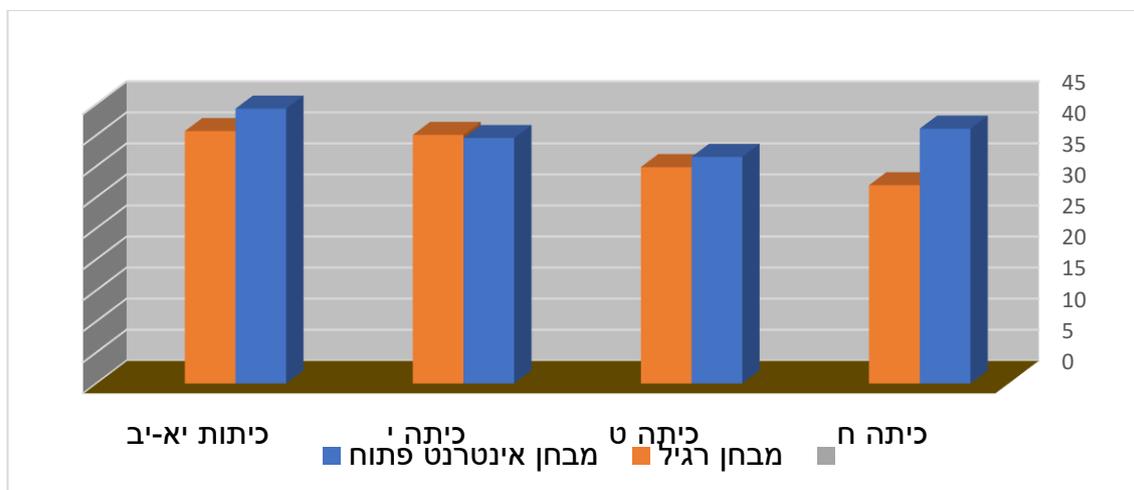
*** $p < .001$

עיון בלוח 8 מראה כי היו הבדלים מובהקים בציוני השאלות בין כל השכבות כמקשה אחת, וכן היה הבדל מובהק בהפרש בין שתי השאלות בכל השכבות. עתה נבדקה השאלה בין אילו שתי שכבות נמצאו ההבדלים הללו. לשם כך נעשו מבחני Post Hoc ונמצא כי:

1. בציוני השאלה עם אפ: ההבדל היה בין שכבת ח (M= 41.05, SD= 7.23) לבין שכבת ט (M= 36.50, SD= 9.98, $p < .001$), ובין לבין שכבת יא+יב (M= 44.27, SD= 7.92, $p < .008$).
2. בציוני השאלות עם אס: ההבדלים היו בין שכבת ח (M= 31.93, SD= 11.92) לבין שכבת י (M= 40.04, SD= 8.87, $p < .01$) ושכבת יא+יב (M= 40.65, SD= 11.83, $p < .001$). כמו כן, בין שכבת ט (M= 34.85, SD= 10.55) לבין שכבת י (M= 40.04, SD= 8.87, $p < .06$) ובין לבין שכבת יא+יב (M= 44.27, SD= 7.92, $p < .001$).
3. באשר לפער בין ציוני אפ לציוני אס: היו הבדלים בין שכבת ח (M= 9.54, SD= 10.38) לבין שכבת ט (M= 1.55, SD= 11.39, $p < .001$) לבין שכבת י (M= - .97, SD= 9.21, $p < .001$); ולבין שכבת יא+יב (M= 4.18, SD= 8.73, $p < .01$).

לסיכום: נמצאו הבדלים מובהקים בהפרשים בין ציוני שתי השאלות בין כל שכבה לבין כל שכבה אחרת, כאשר בשאלה הרגילה הציונים עולים בהדרגה ככל שעולה הגיל, ואילו בשאלה עם א"פ היה הציון גבוה יותר בשכבה הצעירה ביותר, והיה דומה להישגי השכבה הבוגרת.

תרשים 4 ממחיש את ההבדלים בין שכבות הגיל בשני סוגי השאלות.



תרשים 4: הבדלים בין שכבות הגיל בממוצעי הציונים בשני סוגי השאלות

שאלה נוספת שנבדקה הייתה האם יתגלו הבדלים בציוני השאלה עם א"פ בין השכבה הבוגרת (יא, יב) לבין השכבה הצעירה (ח). הסיבה לשאלה זו היא שבמציאות של חידושים טכנולוגיים המתרחשים במהירות עצומה, יתכן שככל שתלמידים צעירים יותר, תהיה עבורם הרשת טבעית יותר. לפיכך חושבו תחילה ההפרשים בין שני סוגי השאלות בכל אחת מן השכבות. נמצא שההפרש הממוצע בכיתות יא-יב היה נמוך יותר בשכבה הבוגרת ($M=4.2, SD=8.7$) לעומת השכבה הצעירה ($M=9.5, SD=10.4$), כלומר הבדל של 5.3 נקודות בין הפרשי הציונים של שתי שכבות הגיל לטובת השכבה הצעירה, שנמצא מובהק. כאמור, הממוצעים וסטיות התקן של שתי שכבות הגיל, יא ו-יב אוחדו כיון שלא נמצאו ביניהן הבדלים מובהקים בציונים בשתי השאלות.

הלוח הבא מציג את הממוצעים, סטיות התקן וערכי t של ההבדל בין השכבה הבוגרת לשכבה הצעירה בשאלה עם אפ.

לוח 9: ממוצעים, סטיות תקן וערכי t של ההבדל בין השכבה הבוגרת לשכבה הצעירה בשתי השאלות ($183=N$)

F	t	כיתות ח (80=N)	כיתות יא-יב (103=N)		
	-2.81**	41.0	44.8	ממוצע	אפ
		7.23	8.1	סטית תקן	
	-4.82***	31.92	40.64	ממוצע	אס
		11.92	11.8	סטית תקן	
	-3.60***	9.54	4.2	ממוצע	הפרש בין השאלות
8.26**		10.38	8.7	סטית תקן	

מלוח 9 ניתן לראות שתלמידי השכבה הבוגרת אמנם הצליחו באופן מובהק יותר מתלמידי השכבה הצעירה בשאלה עם א"פ, אך תלמידי השכבה הצעירה הראו שיפור יותר מכפול בין שתי השאלות בהשוואה לתלמידי השכבה הבוגרת.

סיכום הממצאים:

א. נמצא הבדל מובהק בין ממוצע הציונים בשאלה עם א"פ לבין הממוצע בשאלה הרגילה, לטובת א"פ, כשההפרש בין הממוצעים היה 3.8 נקודות;

ב. מתוך שמונה מקצועות לימוד שנבדקו, נמצא שבשלושה מהם, מדעים, תקשורת ואנגלית, לא נתגלו הבדלים בהישגים בין שני סוגי השאלות. בחמשת המקצועות האחרים היו הבדלים מובהקים בין שני סוגי השאלות.

ג. לא נתגלו הבדלים מובהקים בממוצעי הציונים בין הבנות לבין הבנים, אך ההפרש בין שתי השאלות בקרב הבנים היה מעט גבוה יותר (4.23 נקודות) מזה של הבנות (3.0).

ד. בחלוקה לשלוש רמות הישגים נמצא שההתקדמות הגבוהה ביותר הייתה של 25% בעלי ההישגים הנמוכים ביותר (8.39 נקודות), על אף שפיזור הציונים בקבוצה זו היה הגבוה ביותר מבין שלוש הקבוצות בשתי השאלות.

ה. נמצאו הבדלים מובהקים בהפרשים בין ציוני שתי השאלות בין כל שכבת גיל לבין כל שכבה אחרת. ההפרש הגבוה ביותר בין שתי השאלות היה בשכבת כיתות ח. ואילו בשכבת כיתות י לא נמצאו הבדלים בין שתי השאלות. בנוסף, נמצא שההפרש בין שני סוגי השאלות בכיתות הבוגרות (יא, יב) היה 1.84 נקודות, בעוד שההפרש בכיתות הצעירות (ח) היה 10.1 נקודות לטובת א"פ.

ו. נמצא שהייתה התאמה בין מידת ההתאמה של השאלות עם א"פ לבין ממוצע הציונים במקצועות שבהן קבלו שאלות המבחן ציון גבוה. ככל שהשאלות התאימו יותר לשאלות עם א"פ כך השיגו בהן התלמידים ציונים גבוהים יותר.

ז. כל הממצאים מזמינים עיון מחקר נוסף תוך פיתוח מתווי מחקר משלימים נוספים.

ניתוח משוברים של תלמידים

חלק מן המורים שהשתתפו בניסוי בקשו מתלמידיהם משוב קצר בכתב לאחר השאלות. בסך הכול נתקבלו 40 משוברים כתובים, שכללו 59 היגדים. על-אף שמדובר במספר לא רב של משיבים, הרושם היה שיש טעם לשמוע את דבריהם. המשוברים עברו ניתוח תוכן שממנו עלו שבע תמות. לוח 10 מציג התמות שעלו מניתוח המשוברים ואת שכיחותה של כל תמה.

לוח 10 : התמות שעלו מניתוח משובי התלמידים ושכיחותן (N=59 היגדים)

שכיחות	דוגמאות להיגדים	התמה
13	אהבתי את פורמט המבחן ; אני חושבת שהמבחן הוגן וטוב ; ממש התחברתי לפורמט הזה כי היה הרבה יותר רגוע	השיטה והפורמט [של מא"פ] היו טובים, נוחים והוגנים, ולא מלחיצים
7	המבחן היה בסדר גמור, השאלות היו קלות יחסית ; זה יותר קל אני מרגישה יותר בטוחה בעצמי	המבחן היה די קל
5	הקושי מול המידע הבלתי מוגבל הקיים באינטרנט הוא לחפש את המידע הרלוונטי הקשור לשאלה עליה אנחנו עונים ולסנן את מה שלא רלוונטי ; באינטרנט היה כל כך הרבה מידע שלא תמיד ידעתי מה אני צריכה ומה לא,	הקושי בעבודות עם חומר פתוח הוא להפריד בין עיקר לטפל
4	הקושי שנתקלתי בו הוא במתן דוגמות אקטואליות, לקח לי זמן רב לחשוב על דוגמות וחלק מן הדוגמות שנתתי הן טריוויאליות ; הקשיים היו בעיקר מחשבה על מקרים אקטואליים ספציפיים שישמשו דוגמאות לשאלות. כמי שאינו מעורה באופן שוטף באקטואליה היום יומית במדינה לקח לי זמן להיזכר במקרים המתאימים לשאלות.	היה קושי לחשוב על דוגמאות אקטואליות [במבחן רגיל באזרחות]
15	היבחנות עם אינטרנט פתוח היא נכונה יותר לחיים בעידן שלנו ; הידע הוא אינסופי ולכן כדאי לחנך את הדור הצעיר לחשיבה ביקורתית, הטלת ספק והבנה מעמיקה [יותר] מאשר שינון שמוביל להבנה שטחית לזמן קצר בשביל המבחן ; זה ממש עוזר לי להתרכז ולא להילחץ ; האינטרנט המהווה מקור מידע גדול שעומד לרשות התלמידים, מאפשר להם ללמוד מעבר לחומר שהמורה מקנה להם במהלך השיעורים ; אני חושבת שכך יכולתי להתרכז בחקר ובהבנת החומר בניגוד לשאלות רגילים בהם אני מתבקשת לשנן מושגים ; השימוש באינטרנט מוריד את כמות המידע שצריך לזכור ובכך העבודה מתבססת יותר על הבנת החומר.	יתרונות בעבודה עם א"פ
4	אני מכירה את החומר בצורה טובה אין לי בעיה לנסח טוב תשובה ודווקא להיעזר באינטרנט החליש אותי מהבחינה שהוא האט את קצב כתיבת התשובה שלי בגלל שרציתי לכתוב כל מילה. במקום זה דפדפתי הרבה זמן באינטרנט ;	קשיים בעבודה עם א"פ
11	תהליך העבודה היה לא מוצלח, כיוון שהיה מדובר בחומר סגור, השאלות הסתמכו נטו על שינון של החומר ופחות על הבנה, דבר שלא כל כך התחברתי אליו ; הידע שלי לא השתנה מתחילת העבודה בחומר סגור ; אני חושבת שאם השאלות היו מאפשרות חומר פתוח	חסרונות של מבחן רגיל

	ומתבססות יותר על הבנה, הייתי לומדת ומבינה את החומר הרבה יותר לעומק וגם הייתי נהנית תוך כדי מתהליך החשיבה.	
--	-----------------------------------------------------------------------------------------------------------	--

דיון

הממצא הראשון מראה שיש מתאם גבוה בין הישגי התלמידים בשני סוגי השאלות. כלומר ככל שתלמיד קיבל ציון גבוה במבחן אחד, הוא בדרך כלל קיבל ציון גבוה במבחן השני. ממצא זה יכול לשמש עדות לכך שניתן לאמץ שיטת היבחנות עם א"פ בלי חשש לעיוות תוצאות. זה ממצא שעשוי להניח את דעתם של מקבלי החלטות וקובעי מדיניות בהערכה, בין שהם בית-ספריים, אזוריים, ארציים ואף בינלאומיים.

עם זאת, ההישגים בשאלה עם א"פ היו מעט יותר גבוהים. תלמידים הצליחו בה יותר וקבלו ציונים גבוהים יותר ב-3.8 נקודות בממוצע. זאת מעבר לגיל ולמקצוע הלימוד. מזה שנים רבות משננים בהכשרות מורים שהצלחות מחזקות ביטחון עצמי של ילדים, שהן מעלות את המוטיבציה ללמוד, ומנגד, שיש לכישלונות בבחינות מחירים בתחום האישי והכלכלי (Machin, McNally, & Ruiz-Valenzuela, 2020). ההצלחה של יותר תלמידים בשאלה עם א"פ, כפי שהתקבלה מן הממצאים נראית כמו כיוון רצוי הן מן הטעם של תחושת הרווחה של התלמידים והן מן ההיבט הכלכלי-חברתי (Orr & Foster, 2013).

עיון יותר מעמיק מראה כי במקצועות המדעים, התקשורת והאנגלית, לא היו הבדלים בין השאלות. במדעים, ניתן לראות שני ממצאים שחשוב להתייחס אליהם: א. סטיות התקן במקצוע זה היו הגבוהות ביותר מבין כל המקצועות, בשתי השאלות. פירוש הדבר שטווח הציונים במקצוע זה היה הגדול ביותר. טווח כזה אומר שקיים פער ניכר בין תלמידים שקבלו ציון מאוד נמוך לבין אחרים שקבלו ציון מאוד גבוה. פער כזה מסביר את העובדה שבמדעים לא היה הבדל מובהק בין שני סוגי השאלות. ב. שאלות המבחן עם א"פ במדעים שחברו על ידי המורים למדעים קבלו את הציון הנמוך ביותר מבחינת התאמתן לסוג מבחן כזה, שאולי מעיד על קושי של מורים למדעים לשנות מתכונת של שאלות מבחן. יתכן ששני הממצאים מעלים שאלה בסיסית הנוגעת לדרכי ההוראה במקצועות המדעים. ספרות המחקר בנושא הוראות המתמטיקה עוסקת בשני העשורים האחרונים באופן אינטנסיבי למדי בצורך לשנות גישה ותפיסות בסיסיות הנוגעות למקצועות הללו (Pengelley, 2011; Önder Bütüner, 2015), כאשר אחת ההצעות המרכזיות היא לימוד ההיסטוריה של המדעים, הן בהכשרת מורים למתמטיקה והן לתלמידים בכיתות. יתכן שהממצאים הקשורים למקצוע המדעים עשויים להיות פתח לחשיבה חדשה הן על הכשרת המורים במקצועות אלה והן לגבי דרכי ההוראה שלהם. באשר למקצוע האנגלית, יתכן שהדמיון בין שתי השאלות מעיד על כך שמבחן בשפה אינו בהכרח קשור לשימוש או אי-שימוש באינטרנט, ויש טעם בבדיקה נוספת של סוגיה זו. במקצוע תקשורת היה מספר הנבחנים הקטן ביותר (כיתה אחת) ויתכן שקיימים משתנים נוספים שהשפיעו על המבחן שלא נבדקו בניסוי זה.

ממצא מעניין נוסף הוא שלא נתגלו הבדלים מובהקים בממוצעי הציונים בין הבנות לבין הבנים, אף שההפרש בין שתי השאלות בקרב הבנים היה מעט גבוה יותר (4.23 נקודות) מזה ששל הבנות (3.0). לממצא זה יש משמעות לאור ספרות המחקר המדווחת בעשור האחרון על יתרון של בנות על בנים בלימודים, מעבר לגיל, מצב חברתי-כלכלי או מיקום גיאוגרפי (Hoff-Sommers, 2015; Zimbardo & Colombe, 2016). על אף העדויות על השימוש השונה של בנים ובנות ברשת מעידים הממצאים שכאשר מדובר במבחן בעל אופי שונה מן המקובל, הן הבנים והן הבנות הצליחו יותר והיה דמיון בהישגיהם.

אחד הממצאים החשובים בניסוי המתואר כאן קשור לציונים הגבוהים באופן משמעותי של תלמידים בעלי הישגים נמוכים יותר בשאלה עם א"פ. כמעט כל המשתתפים במדגם הצליחו יותר בשאלה זו, אך שיעור ההצלחה של תלמידים אלה היה הגבוה ביותר, ב-8.39 נקודות. יתכן שממצאים אלה מעידים על תופעה

מוכרת בתכניות התערבות איכותיות והיא שכולם משפרים הישגים ביחס לטווח השיפור שלהם – בעלי ההישגים הגבוהים משפרים מעט כי יש להם טווח קצר לשיפור ואילו בעלי ההישגים הנמוכים משפרים יותר כי יש להם טווח שיפור ארוך יותר.

יתכן שיש בממצא זה עדות על יכולת טובה למדי של שימוש ברשת בקרב תלמידים בעלי הישגים נמוכים יותר, שמשותווה כנראה ליכולת זו בקרב התלמידים החזקים אך אינה באה לידי ביטוי בשאלות מסורתיות שמבוססות על שינון ידע.

הסבר נוסף כאמור יכול להיות שטווח ההשתפרות של תלמידים בעלי הישגים נמוכים (על פי השיטה המסורתית) עשוי להיות מושפע לטובה כאשר משנים עבורם את אופן הבחינה ואת הגישה ללימודים, מה שהוכח במחקרים לא מעטים שבדקו שינויים בשיטות הוראה ולמידה (שחר 2011 : 133-155). עם זאת, חשוב לשים לב לכך שעל-אף השיפור, ציוני ההישגים בקרב הרבעון הנמוך אינם משקפים יכולות אקדמיות גבוהות, וכן, שסטיות התקן בקבוצה זו היו הגבוהות ביותר, שפירושו הטרוגניות רבה בתוך הקבוצה. ענין זה מזמין העמקה נוספת.

ההבדלים שנמצאו בהפרשים בין ציוני שתי השאלות בין כל שכבת גיל לבין כל שכבה אחרת מראים שההפרש הגבוה ביותר בין שתי השאלות היה בשכבת כיתות ח. ואילו בשכבת כיתות י לא נמצאו הבדלים בין שתי השאלות. בנוסף, נמצא שההפרש בין שני סוגי השאלות בכיתות הבוגרות (יא, יב) היה 1.84 נקודות, בעוד שההפרש בכיתות הצעירות (ח) היה 10.1 נקודות לטובת א"פ. ענין זה דורש בדיקה נוספת.

מבחינה מתודולוגית עולה צורך ללמד את המורים כיצד להעריך מבחנים בא"פ כי רבים העידו על קושי בנושא מאחר שמא"פ מסמן מעבר מפרדיגמה של בחינה של ידע שנלמד לפרדיגמה של בחינה של הכושר להשיג את הידע הנדרש שלא נלמד. יש לתת את הדעת במחקר הבא להביא את המורים לעבודה עם כלי מוסדר ובה בעת לתקף את כלי ההערכה של מא"פ.

ממצאים של שנת הניסוי השנייה (מא"פ 2.0)

חמישה עשר בתי ספר על-יסודיים ביקשו להיכנס לפרויקט בשנתו השנייה. מהם היו שמונה בתי ספר שהשתתפו בשנה הראשונה ושבעה חדשים.

השיטה

המשתתפים: 517 תלמידים מכל בתי הספר שהשתתפו בשנת הניסוי השנייה התבקשו לענות על מבחן רגיל במקצוע שעסק בחומר הלימודים שלהם עם שתי שאלות מהן יכלו לבחור אחת, ובזמן אחר במבחן עם אינטרנט פתוח עם שתי שאלות מהן יכלו לבחור אחת.

המורים: 38 מורים השתתפו בניסוי. כדי לאפשר למורים לחבר שאלות מבחן שיתאימו לחיפוש תשובות במקורות דיגיטאליים ואחרים, שופר המחווון על חמשת מדדיו, שמאפשר למורה לבחון את מידת ההתאמה של השאלה לסוג החדש של שאלות. בנוסף, חוברו אמות מידה שלפיהן יוכלו המורים להעריך את איכותן של תשובות התלמידים. גם בשנה השנייה התקיימו מפגשי הדרכה והסבר ובמהלכם המורים העלו את השאלות לאתר וקבלו משובים, הערות ותיקונים. רק לאחר מכן הם קיימו את המבחנים.

הלוח הבא מציג את הפיתוח של מחווון השאלות למא"פ, קרי "מחווון השאלות".

1. השאלה מזמנת אוריינות מידענית: חיפוש ואיסוף מידע נדרש ממקורות דיגיטליים ו/או אנושיים תוך הערכת מהימנותם, לצד ציון, זיהוי ואפיון המקורות הנבחרים.
2. השאלה מזמנת כושר למידה והבנה: חיבור בין ידע ומידעיים חדשים לידע קודם.
3. השאלה מערבת את ההיגיון הפנימי: מזמנת ביטוי שלו בצורת רעיון יצירתי, הבעת ביקורת, הבעת דעה, נקיטת עמדה הכוללים דיווח עצמי, רפלקטיבי ועובדתי, ביחס לציון על מה אלה מתבססים.
4. השאלה רלוונטית ללומדים: מתייחסת למציאות החיים של הלומד/ת ומנוסחת בשפה מדוברת.
5. השאלה מזמנת חשיבה המשגתית, תכנונית, יישומית או פתרון בעיות: דורשת מהלומד/ת לבחור בידע, גישה, ערכים ותובנות לצורך קידום מעשה/פעולה/פתרון בעיה.

הלוח הבא מציג את אמות המידה להערכת התשובות, קרי, מחוון הערכה של מא"פ.

כשירות	למה הכוונה?	ביצוע בסיסי	ביצוע מתקדם	ביצוע מיטבי
1. אוריינות מידענית	התשובה מסתמכת על אוריינות מידענית: - חיפוש ואיסוף מידע נדרש ממקורות דיגיטליים ו/או אנושיים - הערכת מהימנות המקורות - ציון, זיהוי ומא"פיון המקורות הנבחרים.	התשובה מסתמכת על מקורות מידע מועטים ומציינת אותם	התשובה מסתמכת על כמה מקורות מידע מוצלבים כדל להשיג מהימנות, ומציינת אותם	התשובה מסתמכת על מגוון מקורות מידע מוצלבים, בדוקים ואיכותיים, ומציינת אותם
2. כושר למידה עצמאית	מציגה יכולת חשיבה ולמידה עצמאית - מוצגים תכנים רלוונטיים ועדכניים - התכנים החדשים מקושרים לידע תחומי בסיסי	התשובה מציגה ידע מבלי לקשר אותו באופן ברור לשאלה	התשובה מציגה ידע ומקשרת אותו לשאלה	התשובה מציגה ידע בסיסי ומידע עדכני, מקשרת ביניהם, ובכך מרחיבה, מעמיקה ומדייקת את מסגרת ההתבוננות על השאלה
3. היגיון פנימי ברור המיוצג בתשובה	התשובה כוללת את ההיגיון האישי הפנימי: - מוצגת עמדה אישית המבוססת על מודעות עצמית - מוצגת זווית אישית ביקורתית/ יצירתית / חדשנית על הנושא	התשובה כוללת התייחסות כללית לשאלה ללא קול אישי, או, תוך הבעת דעה אך מבלי לציין על מה זה מתבססת	התשובה כוללת התייחסות אישית או הבעת עמדה, תוך ציון על מה זה מתבססת	התשובה כוללת התייחסות אישית מנומקת תוך ציון על מה היא מתבססת, וזווית יצירתית ו/או ביקורתית על הנושא
4. כושר שפתי	התשובה בהירה ומקדמת הבנה: - מנוסחת באופן בהיר - מחברת חלקים לטקסט שוטף	יש קשר רופף בין המשפטים והפסקאות בתשובה	יש קשר המשפטים והפסקאות בתשובה ויש שימוש מתאים במילות קישור	הטקסט מאורגן ומוביל את הקורא/ת בצורה ברורה ועניינית לנושא לנשוא עד לקבלת מסר בהיר

ניתן לראות שההבדל הגדול ביותר לטובת מא"פ היה בקרב 25% של התלמידים בעלי הציונים הנמוכים ביותר. ברבעון השני – תלמידים שבין 25-50% - לא נמצא הבדל בין שני סוגי המבחנים. ברבעון השלישי נמצא פער מובהק לטובת מבחן רגיל, וברבעון הגבוה ביותר נמצא הפער הגדול ביותר בהישגים לטובת מבחן רגיל.

2. הבדלים בין בתי ספר שבהם רמת ההישגים הממוצעת בכיתות שנבחנו בחינות רגילות הייתה גבוהה לבין אלה שבהם הייתה רמה ממוצעת נמוכה

ביקשנו לדעת האם הפערים שראינו קודם באים לידי ביטוי גם ברמה של בתי הספר ככלל. עיון בממוצעי הציונים במבחנים הרגילים בבתי הספר השונים הראה שהיו הבדלים משמעותיים ביניהם. עובדה זו העלתה שאלה אם ההבדלים בין ציוני שני המבחנים יהיו שונים בבתי ספר שממוצע ציוניהם במבחן הרגיל היה גבוה בהשוואה לבתי ספר שממוצע ציוניהם במבחן הרגיל היה נמוך. לצורך זה חולקו בתי הספר לשתי קבוצות, חמישה בתי ספר שבהם היו הציונים הגבוהים ביותר בממוצע במבחן הרגיל, ושבעה בתי ספר שבהם היו ציונים נמוכים.

הלוח הבא מציג את הממוצעים, סטיות התקן וההבדלים הסטטיסטיים בין שני סוגי המבחנים בבתי הספר שציוני התלמידים שנבחנו במבחן הרגיל היו הנמוכים ביותר.

לוח 2 : ממוצעים, סטיות תקן וערכי t של ההבדלים בציוני שני השאלות בחמישה בתי ספר שממוצעייהם היו הנמוכים ביותר

t	מבחן רגיל (161=N)		מא"פ (163=N)		בית ספר
	סטיות תקן	ממוצע	סטיות תקן	ממוצע	
	16.04	27.48	12.41	33.04	9 איתן גזר
	7.73	27.96	12.01	21.17	10 ברנקו וייס ב"ש
	12.15	28.76	7.95	37.75	3 מאיר שפייה
	10.55	34.05	9.80	38.05	12 מיתרים רעננה
	12.07	35.46	11.61	41.81	7 מקיף ד אשדוד
4.55***	12.99	30.01	12.38	34.89	ממוצע כללי

מלוח 2 עולה שבבתי ספר אלה היה ממוצע הציונים במא"פ גבוה ב-4.88 נקודות מזה של מבחן רגיל, וההבדל נמצא מובהק.

הלוח הבא מציג את הממוצעים, סטיות התקן וההבדלים הסטטיסטיים בין שני סוגי המבחנים בבתי הספר שציוני תלמידיהם היו הגבוהים ביותר.

לוח 3 : ממוצעים, סטיות תקן וערכי t של ההבדלים בציוני שני השאלות בשבעה בתי ספר שבהם ממוצע הציונים במבחן הרגיל היו הגבוהים ביותר (216=N)

t	מבחן רגיל		מא"פ		בית ספר
	סטיות תקן	ממוצע	סטיות תקן	ממוצע	
	7.06	40.80	14.62	27.15	14 דרור
	8.47	41.48	10.73	40.27	11 הר טוב
	7.99	41.82	6.18	37.48	2 בן צבי

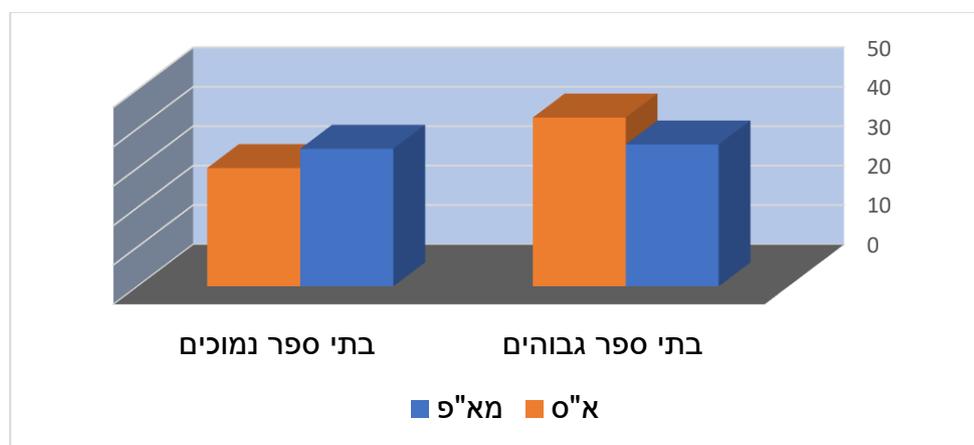
	6.31	42.40	6.45	41.10	13 בויאר
	5.72	43.31	7.80	26.55	8 התבנית
	6	43.51	8.16	37.19	4 אחד העם
	5.30	46.33	7.03	37.67	6 ריינה
	9.90***	7.14	42.81	10.45	35.99
					ממוצע כללי

א. מלוח 3 נראה שבבתי ספר אלה היה ממוצע ציוני המבחן הרגיל גבוה יותר ב-6.82 נקודות, וההבדל נמצא מובהק.

ב. בתי הספר אורט גוטמן (1), כוכבי המדבר (15) ומסעדה (5) היו באמצע מבחינת ציוניהם ולכן לא נכנסו לאחת הקטגוריות. באורט גוטמן היה שיפור של 5.44 נקודות לטובת מא"פ, בכוכבי המדבר הייתה ירידה של 3.23 נקודות במא"פ לטובת מבחן רגיל, ובמסעדה הייתה עליה של 4.56 נקודות לטובת מא"פ. כלומר, מבין שלושה בתי ספר שהיו באמצע מבחינת ציוניהם במבחן הרגיל, בשניים חל שיפור במבחן עם אינטרנט פתוח, ואילו באחד מהם היו הציונים גבוהים יותר במבחן הרגיל.

התרשים הבא ממחיש את ההבדלים בין שתי הקבוצות של בתי הספר.

תרשים 2: הבדלים בין בתי ספר שציוניהם היו הנמוכים ביותר לבין אלה שציוניהם היו הגבוהים ביותר, בשני סוגי המבחנים



3. הבדלים בין מקצועות שבהם הייתה רמת הישגים גבוהה לבין אלה עם רמת הישגים נמוכה

השאלה הבאה שעלתה הייתה האם יש הבדל בין שני סוגי המבחנים במקצועות שונים? כדי לענות על שאלה זו חולקו שבעת המקצועות לשתי קבוצות: קבוצת המקצועות שבהם היו הציונים במבחן הרגיל הגבוהים ביותר, וקבוצת המקצועות שבהם היו הציונים במבחן הרגיל הנמוכים ביותר.

הלוח הבא מציג את הממוצעים, סטיות התקן וערכי t של ההבדל בין שני השאלות במקצועות שקבלו את הממוצעים הגבוהים ביותר במבחן רגיל.

לוח 4: ממוצעים, סטיות תקן וערכי t של שני השאלות במקצועות שקבלו את הממוצעים הגבוהים ביותר במבחן רגיל (276=N)

t	מבחן רגיל		מא"פ		מקצוע
	סטיות תקן	ממוצע	סטיות תקן	ממוצע	
5.01***	6.31	42.40	6.45	41.1	תנך
	8.57	42.03	10.29	37.72	אזרחות
	7.24	41.59	6.63	39.36	ביולוגיה
	8.47	41.48	10.73	40.27	מקרא
	8.24	41.93	9.55	38.76	ממוצע כללי

מעיון בלוח 4 אפשר לראות שבמקצועות אלה היה הבדל של 3.17 נקודות לטובת מבחן רגיל, וההבדל מובהק.

בלוח הבא מוצגים ממוצעים, סטיות תקן וערכי t של שני השאלות במקצועות שקבלו את הממוצעים הנמוכים ביותר במבחן רגיל.

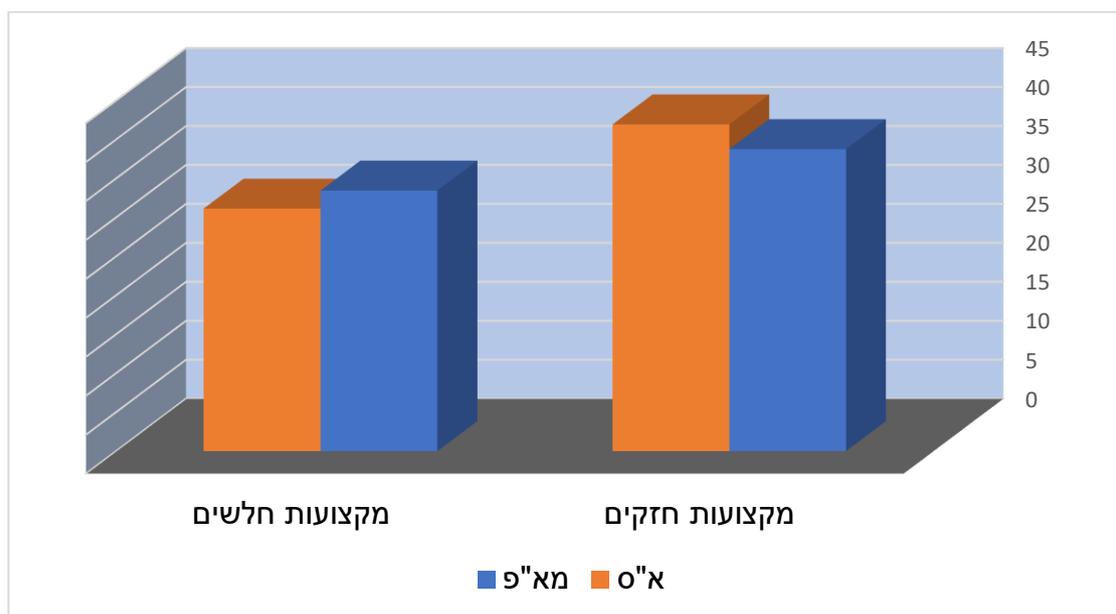
לוח 5: ממוצעים, סטיות תקן וערכי t של שני השאלות במקצועות שקבלו את הממוצעים הנמוכים ביותר במבחן רגיל (N=179)

t	מבחן רגיל		מא"פ		מקצוע
	סטיות תקן	ממוצע	סטיות תקן	ממוצע	
2.23*	16.04	27.48	12.41	33.04	פיסיקה
	9.51	30.79	13.98	29.23	מדעים
	12.38	33.25	11.66	35.59	שפה ולשון
	12.99	31.12	12.67	33.44	ממוצע כללי

מלוח 5 ניתן לראות שבמקצועות אלה היה הבדל של 2.32 נקודות לטובת מא"פ, וההבדל מובהק. בהיסטוריה נמצא הבדל של 5.85 נקודות לטובת מא"פ, ואילו בעברית לדוברי ערבית נמצא הבדל של 4.95 נקודות לטובת המבחן הרגיל.

התרשים הבא ממחיש את ההבדלים בין המקצועות שקבלו את הציונים הגבוהים ביותר לבין אלה שקבלו את הציונים הנמוכים ביותר בשני סוגי המבחנים.

תרשים 3: הבדלים בין מקצועות שקבלו את הציונים הגבוהים ביותר ("חזקים") לבין אלה שקבלו את הציונים הנמוכים ביותר ("חלשים") בשני סוגי המבחנים.



לסיכום:

1. בחלוקה לארבעה רבעונים של רמות הישג נמצא שהרבעון הנמוך ביותר שיפר את הישגיו במא"פ ב-8.2 נקודות בהשוואה למבחן רגיל, ברבעון השני לא נמצא כל הבדל בין השאלות, ואילו ברבעונים השלישי והרביעי נמצאו הבדל לטובת מבחן רגיל.
2. בבתי ספר שבהם היו ציוני המבחן הרגיל הנמוכים ביותר, היה ממוצע הציונים בהם נמוך ב-12.8 נקודות מהממוצע של קבוצת בתי הספר שבהם היו ציוני המבחן הרגיל גבוהים. בתי ספר אלה שיפרו את ציוניהם במא"פ באופן מובהק והפער ביניהם לבין בתי הספר עם הציונים הגבוהים הצטמצם ל-1.1 נקודות בלבד.
3. בבתי ספר שבהם היו הציונים הגבוהים ביותר במבחן הרגיל, הייתה ירידה מובהקת של 6.8 נקודות במא"פ.
4. באופן דומה בארבעה מקצועות שבהם היו הציונים הגבוהים ביותר במבחן הרגיל הייתה ירידה מובהקת בציונים במא"פ, ואילו במקצועות שבהם היו הציונים הנמוכים ביותר במבחן הרגיל חל שיפור ניכר בציוני המא"פ. הפער בין המקצועות עם הציונים הנמוכים לבין אלה עם הציונים הגבוהים במבחן הרגיל היה 10.81 נקודות, והוא הצטמצם ל-5.32 נקודות במא"פ. שיפור זה היה מובהק סטטיסטית.

דיון

אחד הממצאים החשובים בניסוי המתואר כאן קשור לציונים הגבוהים שקיבלו במא"פ של תלמידים שנחשבים חלשים בשיטת ההערכה הרגילה (בעלי הישגים נמוכים יותר במבחן הרגיל). בקרב הרבעון של תלמידים שקבלו ציונים נמוכים ביותר במבחן הסגור היה שיפור מובהק של 8.19 נקודות במא"פ. ברבעון הבינוני-נמוך לא נמצא הבדל בין שני סוגי המבחנים. ברבעון הבינוני-גבוה נמצא הבדל מובהק של 4.72 נקודות לטובת המבחן הסגור, וכך גם בקרב הרבעון שציוניו היו הגבוהים ביותר במבחן רגיל, שבו הייתה ירידה מובהקת של 7.35 נקודות. יתכנו כמה הסברים לממצאים אלה: ראשית, שטווח ההשתפרות של תלמידים "חלשים" עשוי להיות מושפע לטובה, יותר מאשר של תלמידים "חזקים", כאשר משנים עבורם את אופן הבחינה ואת הגישה ללימודים.

6 מורים שהשתתפו בניסוי והוזמנו לקבוצת דיון והפקת לקחים, תמכו בנקודה זו, ובמילותיהם: *תלמידים טובים רגילים לקרא חומר, והם מקובעים יותר על דרך הבחינה הרגילה; ככל שהילד נטוע יותר בפרדיגמה הישנה, כך קשה לו יותר לשנות פרדיגמה; לחלשים, שהבעיה שלהם זו קריאת חומר, חסר בדרך כלל ידע, ופתאום הם יכולים לקרוא ולהשיג יותר מידע; יש לדברים גם תמיכה מחקרית בעבודות שבדקו שינויים בשיטות הוראה ולמידה (שחר 2011: 155-133). חשוב לשים לב לכך שעל-אף השיפור, ציוני ההישגים בקרב הרבעון הנמוך אינם משקפים יכולות אקדמיות גבוהות, וכן, שסטיות התקן בקבוצה זו היו הגבוהות ביותר, שפירושו הטרורגניות רבה בתוך הקבוצה. לצד אלה, חשוב לזכור כי שנים רבות ידוע שהצלחות מחזקות ביטחון עצמי של ילדים, שהן מעלות את המוטיבציה ללמוד, ומנגד, שיש לכישלונות בבחינות מחירים בתחום האישי והכלכלי (Machin, McNally, & Ruiz-Valenzuela, 2020). כלומר, יתכן שההצלחה של תלמידים אלה עשויה לחזק את ביטחונם ולהעלות את הנעתם להישגים טובים יותר. לפיכך, ההצלחה של התלמידים בעלי ההישגים הנמוכים במבחן במא"פ, כפי שהתקבלה מן הממצאים נראית כמו כיוון רצוי הן מן הטעם של תחושת הרווחה של התלמידים והן מן ההיבט הכלכלי-חברתי (Orr & Foster, 2013).*

הממצאים שמראים שיפור בממוצע הציונים במא"פ בבתי ספר עם הישגים נמוכים מחזקים את הסברה ששינוי אופן ההערכה תורם לאוכלוסיות תלמידים חלשות יותר, גם ברמה הבית ספרית ולא רק ברמת הפרט. ענין זה דורש מחקר נוסף.

הממצאים חוזרים על עצמם גם כאשר בוחנים את ההבדלים בין המקצועות השונים, דהיינו, ככל שממוצע הציונים במקצוע נמוך יותר, כך היה שיפור במא"פ לעומת מבחן רגיל, ולהפך. אחד ההסברים האפשריים

לכך הוא שדווקא תלמידים בעלי הישגים גבוהים, בין אם הם שייכים לבית ספר זה או אחר, ובין אם הם חזקים במקצוע זה או אחר, הם אלה שהסתגלו באופן הברור ביותר לשיטה הקיימת של מבחן רגיל שעיקרו שינון, והשינוי באופן הבחינה ואופן ההערכה דורש מהם לחשוב מחדש על מושגי ההצלחה שלהם עצמם (Sampaio & Leite, 2021), ואכן אלה היו גם ההסברים שניתנו על ידי קבוצת מורים שהוצגו בפניה הממצאים ואשר התבקשה להסבירם.

מסקנות מתוצאות המחקר בשתי השנים

כאשר בוחנים את ממצאים בשתי השנים נראה שעולות שלוש מסקנות עיקריות:

1. שתלמידים בעלי הישגים נמוכים בשיטת ההערכה הרגילה הראו באופן עקבי שיפור משמעותי בהישגיהם כאשר נבחנו עם אינטרנט פתוח. שיפור זה מתקבל מעבר להבדלים במקצועות הלימוד וברמת השאלות שניתנו להם. משום ההבדלים באיכות ובאופי השאלות שניתנו לתלמידים בשתי שנות הניסוי יש להתייחס לנתונים במידה של זהירות. עם זאת, הממצא של ציונים גבוהים יותר אצל בעלי הישגים נמוכים גורף למדי.
2. ממצא חשוב נוסף הוא צמצום הפערים בין תלמידים בעלי הישגים גבוהים לאלה בעלי הישגים הנמוכים על פי שיטת ההערכה הרגילה. אף שהפערים בין תלמידים ברמות הישגים נשארים משמעותיים, האפשרות לצמצמם על ידי שינוי שיטת ההערכה היא בעלת חשיבות רבה. שכן, המגמה הלאומית לצמצום פערים עודנה בסדר עדיפות גבוה, ונראה ששינוי אופן ההערכה פועל בכיוון הרצוי.
3. נקודה חשובה היא המסקנה שכולנו, כולל המורים, בטוחים שילדים יכולים להשתמש באינטרנט לצורך הפקת מידע, שהרי הם שקועים ברשת שעות אין ספור. הניסוי העלה שאוריינות מידענית אינה בהכרח זהה לשימוש רב ברשת, ויש ללמד את הילדים ולאמנם ביכולת להפיק מידע אמין ברשת. כלל המורים שהשתתפו בניסוי ציינו זאת.

המלצות לשנה שלישית של הניסוי

1. להשוות את ציוני התלמידים במא"פ לציון הממוצע שלהם במקצוע, ללא צורך לקיים מבחן רגיל ייעודי.
2. חשוב ביותר לקיים ראיונות עומק עם מורים נוספים כדי לתקף את מחוון ההערכה של מא"פ, וגם כדי להמשיך ולהציג בפניהם את הממצאים ולשמוע מן הזווית המקצועית שלהם כיצד הם מסבירים אותם.
3. לבדוק את הדרוג של תלמידים על טווח הציונים, ולהשוותו לדרוג שהם מקבלים במא"פ.
4. ליצור תוכנית לימודים רצינית, עקבית ומפורטת, כוללת שיעורים ייחודיים, לאימון ילדים במידענות רשת.

מראי מקום

דו"ח מבקר המדינה (2021), דו"ח שנתי 71 פרק על מערכת החינוך.

שחר, ח', (2011) קונסטרוקטיביזם בחינוך: הוראה, הערכה ומחקר. אבן יהודה: רכס אקדמי

Azzolini, D., Schnell, P., & Palmer, J., (2012) Educational Achievement Gaps between Immigrant and Native Students in Two "New" Immigration Countries: Italy and Spain in Comparison, The ANNALS of the American Academy of Political and Social Science, 643(1), 46-47. <https://doi.org/10.1177/0002716212441590>

Berliner, D., (2015) The many facets of PISA. *Teacher College Record*, 117(1), 1-20.

Cairns, R. (2020) Examinations tested by Covid-19: An opportunity to rethink standardized senior secondary examinations. *Prospects*.
<https://doi.org/10.1007/s11125-020-09515-9>.

Dephil, M., Brilot, B., Nettle, D. (2011). Anxiety: An Evolutionary Approach. *Canadian Journal of Psychiatry*, 56(12),707-715.

Dieter-Meyer, H., (2014) The OECD as pivot of the emerging global education accountability regime: How accountable are the accountants? *Teacher College Record*, 116(9), 1-20.

Domina, T., & Penner, A., (2017) Categorical inequality: Schools as sorting machines, *Annual Review of Sociology*, 43(1), 311-330.

Furuta, J., Schofer, E., & Wick, S., (2020) The effects of high-stakes educational testing on enrollments in an era of hyper-expansion: Cross-national evidence 1960-2010. *Social forces*.

Hargreaves, A. (2020) Large-scale assessments and their effects: The case of mid-stakes examinations in Ontario. *Journal of Educational Change* 21, 393–420.
<https://doi.org/10.1007/s10833-020-09380-5>

Hoff-Sommers, C., (2015) *The war against boys: How misguided politics are harming young men*. New York: Simon and Schuster Inc.

Jackson, M., Khavenson, T., Chirkina, T., (2020) Raising the Stakes: Inequality and testing in the Russian education system, *SOCIAL FORCES*, 98 (4), 1613–1635, <https://doi.org/10.1093/sf/soz113>

Machin, S., McNally, S., & Ruiz-Valenzuela, J., (2020) Entry through the narrow door: The costs of just failing high stakes examinations. *Journal of Public Economics*, 190.

Mandler, G.; Sarason, S. B. (1952). A study of anxiety and learning. Journal of Abnormal and Social Psychology. 47: 166–173.

Menken, K., (2006) Teaching to the exam: How No Child Left Behind impacts language policy, curriculum, and instruction for English language learners, *Bilingual Research Journal*, 30(2), 521- 546.

Mitra, S. (2014) The future of schooling: Children and learning at the edge of chaos. *Prospects* 44, 547–558. DOI: 10.1080/15235882.2006.10162888

Mitra, S., (2018) New systems for children's learning: Changes required in education. In K. Koutsopoulos & Y. Kotsanis (Eds.) *Handbook of research on*

educational design and cloud computing in modern classroom settings. IGI Global publishers.

Mitra, S. (2020). Children and the internet: Learning, in the times to come. *Journal of Learning for Development*, 7(3), 286-305. Retrieved from <https://jl4d.org/index.php/ejl4d/article/view/445>

Nichols, S., Berliner, D., & Noddings, N. (2007) *Collateral Damage: How High-Stakes Testing Corrupts America's Schools*. Boston: Harvard Education Press.

Önder Bütüner, S., (2015) Impact of using history of mathematics on students' mathematics attitude: A meta-analysis study. *European Journal of Science and Mathematics Education*, 3(4), 337-349.

Orr, R., & Foster, S. (2013). Increasing student success using online quizzing in introductory (majors) biology. *CBE life sciences education*, 12(3), 509–514. <https://doi.org/10.1187/cbe.12-10-0183>

Pengelley, D. (2011). Teaching with primary historical sources: should it go mainstream? Can it? In V. Katz & C. Tzanakis (Eds.), *Recent Developments on Introducing a Historical Dimension in Mathematics Education* (pp. 1-8). MAA Notes 78. Washington: The Mathematical Association of America.

Plucker, J., Callahan, C., & Tomchin, E., (1996) Where art thou, multiple intelligence? Alternative assessment for identifying talent in ethnically diverse and low-income students. *Gifted Child Quarterly*, 40(2), 81-91, <https://doi.org/10.1177/001698629604000205>

Putwain D., Best, N. (2011). Fear appeals in the primary classroom: effects on test anxiety and test grade. *Learning and Individual Differences*, 21 580-584.

Sahlberg, P., (2015) *Finnish Lessons 2.0: What Can the World Learn from Educational Change in Finland?* (2nd ed.) New York: Teachers College Press.

Sanger, L., (2010) Individual knowledge in the internet age. *EDUCAUSE Review* 45(2), 14-24. <http://creativecommons.org/licenses/by-nc-sa/3.0/>

Sampaio, M., & Leite, C., (2021) Relationships between the assessment of school quality and social justice, *Educational Research*, 63:1, 133-146, DOI: [10.1080/00131881.2020.1857654](https://doi.org/10.1080/00131881.2020.1857654)

Santamaria, L., (2009) Culturally responsive differentiated instruction: Narrowing gaps between best pedagogical practices benefiting all learners. *Teachers College Record* 111(1), 214–247.

Strand, S. (2016). Do some schools narrow the gap? Differential school effectiveness revisited. *Review of Education*, 4(2), 107–144. doi: 10.1002/rev3.3054

Van de Werfhorst, H., Mijs, G., Jonathan J. (2010). Achievement inequality and the institutional structure of educational systems: A comparative perspective. *Annual Review of Sociology* 36: 407–28.

Weurlander, M., Söderberg, M., Scheja, M., Hult, H., & Wernerson, A., (2012) Exploring formative assessment as a tool for learning: students' experiences of different methods of formative assessment, *Assessment & Evaluation in Higher Education*, 37(6), 747-760, DOI: 10.1080/02602938.2011.572153

Zimbardo, P., & Colombe, N., (2016) *Man interrupted: Why young men are struggling and what can we do about it?* MA.: Conari Press.