

בחינות בגרות מתוקשבות בישראל: פיתוח, תפעול והערכת הפעולה

טלי פרוינד talif@cet.ac.il יעקב שוורץ מיכל שמש חני שלטון תהילה סנד

מטח
(המרכז לטכנולוגיה חינוכית)

בתשנ"ט הוחל בפיתוח ובהפעלה של בחינות בגרות מתוקשבות בכמה מקצועות לימוד בנתיב הטכנולוגי. בתשס"ה החליט משרד החינוך ליישם את המהלך הזה בבחינה בגאוגרפיה, הנהוגה גם בנתיב העיוני וגם בנתיב הטכנולוגי. מרכז הבחינות של מטח מפתח עבור משרד החינוך את כל בחינות הבגרות המתוקשבות (הבחינה בגאוגרפיה מפותחת במכון סאלד, והתאמתה לסביבה מתוקשבת מבוצעת במרכז הבחינות של מטח), והוא מפתח גם כלים מתוקשבים המאפשרים לבתי הספר להיערך לבחינה. בקיץ תשס"ו נבחנו בבחינות מתוקשבות 2,909 תלמידים מ-133 בתי ספר.

המהלך של בחינת הבגרות המתוקשבת בגאוגרפיה לווה בשנים תשס"ה ותשס"ו בפעולה של הערכה מעצבת שביצעה יחידת ההערכה של מטח, ובמסגרתה נבדקו הישגי התלמידים שהשתתפו במהלך בהשוואה להישגי תלמידים שלא השתתפו בו. בנוסף, נבדקו עמדות התלמידים והצוות שהשתתפו במהלך זה כלפי הערכה בסביבה המתוקשבת ונערכו תצפיות במהלך הבחינה.

הנתונים מלמדים, כי ציוני הבגרות של הנבחנים באמצעות השאלון המתוקשב היו גבוהים מאלו של הנבחנים באמצעות שאלון נייר ועפרון, אולם היתרון שלהם נזקף לזכות הישגיהם הטרומיים (ציוני המגן) הגבוהים יותר ולא לזכות היבחנותם בדרך מתוקשבת.

התלמידים והצוות הביעו שביעות רצון כללית מההכנות לבחינה המתוקשבת. נרשמה שביעות רצון גם מהתמיכה הטכנית וממשק הבחינה. העמדות כלפי הוראה, למידה והערכה בסביבה מתוקשבת היו חיוביות בקרב המורים ומעורבות בקרב התלמידים.

מבוא

עם הרחבת השימוש בתקשוב בהוראה ובלמידה מתרחב גם השימוש בו לצורכי הערכה. כבר בשנת 1970 ניבא גרין ש"כיבוש המחשבים בתחום המבחנים הינו בלתי נמנע" (Green, 1970, עמ' 194). ואכן, היבחנות בסביבה מתוקשבת הופכת לעניין שגרת (Bennett, 2003), הודות ליתרונותיה הרבים.

אחד היתרונות הבולטים של הבחינה המתוקשבת הוא שאפשר לכלול בה סוגי פריטים ייחודיים, לרבות כאלה המודדים כישורים שקשה להעריכם באמצעות בחינת נייר ועיפרון (Bennett, 2003). יתרון נוסף מתייחס לבחינת מחשב מותאמת (computer-adaptive test), אותה ניתן להתאים למאפייני הנבחנים ולביצועיהם תוך כדי הבחינה, בהשוואה לבחינה מבוססת מחשב (computer-based test) שזוהה עבור כל הנבחנים (Bugbee, 1996). הבחינה המתוקשבת מאפשרת גם לחסוך בעלויות של הדפסה, אחסנה, הפצה, איסוף ואבטחה, העברתה אינה מצריכה ציוד מיוחד בשל יכולתו של

המחשב להציג גירויים דינאמיים כמו אודיו, וידיאו ואנימציה. ציינון אוטומטי המתאפשר בבחינה המתוקשבת מוזיל אף הוא את פעולת ההערכה ומקצר את המשך שלה (Bennett, 2001).

עם זאת, השימוש בבחינה מתוקשבת מעורר גם קשיים: תלות באינטרנט והצורך לפתח תוכנות מתאימות; חשש שהישגי הנבחנים יושפעו מגורמים טכניים, כמו קצב העברת הנתונים דרך האינטרנט, תכונות המסך וסוג מערכת ההפעלה; קשיי תחזוקה בתוך בתי הספר; וצורך באבטחת המידע (Bennett, 2001, 2003).

במקצוע שבו מועברת בחינה מתוקשבת לצד בחינת נייר ועיפרון יש להעמיד זה מול זה את שני הפורמטים ולהבטיח שביצועי הנבחנים לא יושפעו מההבדלים הטכניים ביניהם, כמו מספר הפריטים שמופיעים על צג המחשב לעומת מספרם על הדף המודפס, או אמצעי התגובה של הנבחנים – שימוש בעכבר או במקלדת לעומת שימוש בעיפרון לצורך סימון התגובה (Bennett, 2003).

תוצאות שונות יכולות להתקבל גם במהלך הצייון בהשפעת גורמים טכניים, כמו פורמט התוצר המוערך – תדפיס מחשב לעומת כתב יד. על-אף שצייון מקוון מקובל זה עשור במבחנים קובעים כמו GRE, GMAT ו-TOEFL, יש יחסית מחקר מועט בנושא (Bennett, 2003), והמחקרים שקיימים מניבים תוצאות שנויות במחלוקת (Zhang et al., 2003; Powers et al., 1994), בעיקר בנוגע למבחנים המורכבים מפריטים פתוחים.

הביצוע בבחינות מתוקשבות מושפע גם מאוריינות המחשב של הנבחנים ומהעמדות שלהם כלפי פורמט ההיבחנות, ומאחר שבעלי ניסיון בעבודה עם מחשב (בעיקר לצורך כתיבה) מגיעים להישגים גבוהים יותר בבחינות מתוקשבות, חשוב להבטיח שלכל הנבחנים בבחינות כאלה יהיה ניסיון בכך (Russell & Plati, 2001; Schaeffer et al., 1998). סוגיה נוספת שיש להתייחס אליה היא אופן ההשפעה של בחינות מתוקשבות על קבוצות נבחנים שונות שמוגדרות על פי מוצא, מגדר, שפת אם וכד' (Gallagher, Bridgeman & Cahalan, 2000).

למרות המגבלות וההסתייגויות הללו, השימוש בבחינות מתוקשבות יכול להוזיל את עלויות ההערכה, ליעיל אותה ולשפר את איכותה. על-אף שכיום איננו מוכנים עדיין להערכה מתוקשבת רחבת היקף בבתי הספר, כדאי להסתכל על הערכה זו במבט קדימה (Bennett, 2001).

תיאור הפרויקט

בתשנ"ט הוחל בפיתוח ובהפעלה של בחינות בגרות מתוקשבות בכמה מקצועות לימוד בנתיב הטכנולוגי (ביוטכנולוגיה, מדעי החיים והחקלאות, ואדריכלות). בתשס"ה החליט משרד החינוך להרחיב את המהלך בהשראת מגמות עולמיות וליישמו בבחינה בגאוגרפיה, הנהוגה בנתיב העיוני ובנתיב הטכנולוגי. הפרויקט נערך בשיתוף פעולה עם המפמ"רית לגאוגרפיה. לצורך המהלך נבחר שאלון אחד מבין השאלונים שמרכיבים את הבחינה בגאוגרפיה ברמה של 5 יחידות לימוד – "חבל ארץ שלא נלמד" (unseen).

כל שאלוני הבחינה, לרבות שאלון זה, פותחו במכון סאלד. התאמת השאלון לסביבה מתוקשבת נעשתה במטח.

לצורך ההפעלה של בחינות מתוקשבות מרכז הבחינות של מטח פיתח סביבה וכלים מתוקשבים המאפשרים לבתי הספר להיערך לבחינה: אתר אינטרנט (www.cet.ac.il/exams) ותוכנה לניהול בחינות. היערכות בתי הספר לקראת הבחינה כוללת התאמה של תשתית התקשוב, השתתפות הצוות בהשתלמות מכינה והיערכות לוגיסטית. התלמידים מתכוננים לבחינה באמצעות תרגול של בחינות בגרות משנים קודמות, שהוסבו לסביבה מתוקשבת ושולבו באתר מרכז הבחינות של מטח, ובאמצעות השתתפות בבחינות הדמיה המדמות באופן מלא את התהליך של בחינת הבגרות הרשמית.

בתשס"ה השתתפו במהלך 38 בתי ספר (521 נבחנים), ובתשס"ו – 49 בתי ספר (840 תלמידים), גידול של 29% במספר בתי הספר ושל 61% במספר הנבחנים.

תיאור פעולת ההערכה

המהלך של בחינת הבגרות המתוקשבת בגאוגרפיה לווה בשנים תשס"ה–תשס"ו בהערכה מעצבת שביצעה יחידת ההערכה של מטח. שאלות ההערכה נגזרו מהצרכים של מזמיני ההערכה במשרד החינוך, והתייחסו לשלושה תחומים עיקריים:

- א. **תשומות:** קיומן של תשתיות טכנולוגיות וארגוניות החיוניות להשתתפות בתי הספר במהלך.
- ב. **תהליכים:** הכנת התלמידים והצוותים הבית ספריים לבחינה, והתנהלות הבחינה בפועל.
- ג. **תפוקות:** הישגי הנבחנים בבחינה המתוקשבת בהשוואה להישגי הנבחנים בבחינת נייר ועיפרון, עמדות לגבי למידה והערכה בסביבה מתוקשבת ועוד.

מערך המחקר בתשס"ה

כדי להשיב על שאלות ההערכה נבנה מערך מחקר שכלל שלוש קבוצות:

- **קבוצת הפרויקט:** 521 תלמידים מ-38 בתי הספר שהשתתפו במהלך ונבחנו באמצעות השאלון המתוקשב, המורים לגאוגרפיה ורכזי המחשוב באותם בתי ספר.
- **קבוצת השוואה I:** 621 תלמידים מאותם 38 בתי הספר שהשתתפו במהלך, אבל נבחנו באמצעות שאלון נייר ועיפרון. מתוך קבוצה זו נדגמה תת-קבוצה בת 127 תלמידים (תת-קבוצת השוואה I), שהתפלגות ציוני המגן שלה הייתה זהה להתפלגות ציוני המגן בקבוצת הפרויקט. מהלך זה נועד להתמודד בדרך סטטיסטית עם היעדר השמה אקראית של תלמידים לקבוצת הפרויקט, ותחת אילוץ זה – לאפשר השוואה לגיטימית של ביצועיהם בבחינה הרשמית עם ביצועי תלמידים בעלי מאפיינים דומים, שלא השתתפו בפרויקט.
- **קבוצת השוואה II:** 792 תלמידים מ-76 בתי ספר שלא השתתפו במהלך, ולהם מאפיינים דומים לאלה של בתי הספר שהשתתפו בו. קבוצת השוואה זו נבחרה מתוך כלל בתי הספר העל-יסודיים בישראל, שתלמידיהם נבחנו בתשס"ה בבחינת הבגרות בגאוגרפיה. הבחירה נעשתה בדגימת שכבות על-פי קריטריונים כמו: מגזר, סוג פיקוח, מדד טיפוח, גודל בית הספר וטווח גילאי התלמידים בו.

כלי ההערכה

- **מבחנים למדידת ההישגים של כל תלמידי ישראל שנבחנו בגאוגרפיה בקיץ תשס"ה.** הנתונים התקבלו ממשרד החינוך, וכללו את ציון המגן של התלמיד ואת הציון שקיבל בכל אחד משאלוני הבחינה בגאוגרפיה.

- **שאלונים** לתלמידים, למורים לגאוגרפיה ולרכזי המחשוב ב־38 בתי הספר שהשתתפו במהלך. השאלונים כללו שאלות סגורות ופתוחות, שעסקו בנתוני רקע של המשיבים, עמדות ותפיסות לגבי הבחינה המתוקשבת – יתרונית, חסרונותיה, התנהלותה בפועל וכיוצ"ב. בסך הכול השיבו על השאלונים 454 תלמידים, 35 מורים ו־31 רכזים.
- **דף תצפית** שימש לתיעוד ההתרחשויות במהלך הבחינה הרשמית ב־9 מבתי הספר שהשתתפו במהלך.

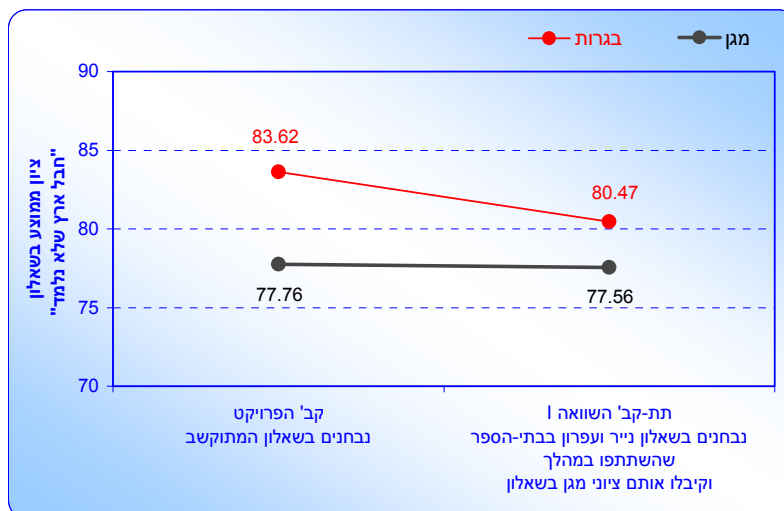
ממצאים

הישגי התלמידים

הממצאים, אשר הופקו באמצעות ניתוח שונות חד־כיווני (Scheffe), מלמדים על הבדל מובהק בין שלוש קבוצות התלמידים בהישגיהם בבחינת הבגרות הרשמית בשאלון "חבל ארץ שלא נלמד" (ממוצע 83.62, ס.ת. 13.48), שהיו באופן מובהק גבוהים מאלה של עמיתיהם בקבוצת הפרויקט I (ממוצע 75.92, ס.ת. 17.75) ובקבוצת השוואה II (ממוצע 74.48, ס.ת. 17.54).

גם בציוני המגן נמצא הבדל מובהק בין שלוש הקבוצות ($F=18.8$; $df=2$, 1931; $p\leq 0.0001$), עם יתרון לתלמידים בקבוצת הפרויקט (ממוצע 77.76, ס.ת. 13.19) לעומת התלמידים בקבוצת השוואה I (ממוצע 73.0, ס.ת. 14.62) ובקבוצת השוואה II (ממוצע 74.18, ס.ת. 12.83). דפוס זה חזר גם בציונים שהושגו בשאלונים אחרים בגאוגרפיה.

כדי לברר את המקור ליתרון של תלמידי הפרויקט, הושוו הישגיהם לאלו של התלמידים בתת־קבוצת השוואה I, שהתפלגות ציוני המגן שלהם הייתה דומה להתפלגות ציוני המגן של התלמידים בקבוצת הפרויקט.



ניתוח שונות דו־כיווני שכלל את קבוצת ההשתייכות (קבוצת הפרויקט / תת־קבוצת השוואה I) ואת סוג הציון (מגן / בגרות) מלמד, כי הפער בין ציוני הבגרות ובין ציוני המגן של התלמידים בקבוצת הפרויקט אינו שונה באופן מובהק מהפער המקביל אצל התלמידים בתת־קבוצת השוואה I.

אם כן, אפשר לומר שהציונים הגבוהים יותר של תלמידי הפרויקט בשאלון "חבל ארץ שלא נלמד" יכולים להיות מוסברים על־ידי מאפיינים ייחודיים של התלמידים, כמו היותם מלכתחילה בעלי

הישגים גבוהים יותר, ולא על-ידי משתני תהליך כמו צורת ההיבחנות (שאלון מתוקשב) או מאפיינים בית ספריים ייחודיים (הכנת התלמידים ו/או הסגל).

עמדות התלמידים והצוות

התלמידים והצוות הביעו שביעות רצון מההכנות לבחינה המתוקשבת, מהתמיכה הטכנית שקיבלו וממשק הבחינה, אך ביטאו דעות מעורבות לגבי איכות המפגש המתוקשבות וחלוקת המסך בין המפגש לטקסטים (ואכן מרבית התלמידים נעזרו גם באטלס המודפס במהלך הבחינה).

מרבית המורים ביטאו עמדות חיוביות כלפי הוראה, למידה והערכה בסביבה מתוקשבת, מעל מחציתם אמרו שהבחינה המתוקשבת הניעה אותם לשלב יותר את המחשב במקצוע הגאוגרפיה, וכשליש דיווחו על שילוב רב יותר של המחשב ביתר מקצועות הלימוד. מרבית המורים חשו שההשתתפות בבחינה המתוקשבת תורמת למעמדם המקצועי וליוקרת המקצוע, והעידו על תחושת מסוגלות עצמית בהכנת תלמידיהם לבחינה המתוקשבת.

רק כמחצית מהתלמידים ביטאו עמדות חיוביות כלפי הוראה, למידה והערכה בסביבה מתוקשבת, אך היו תמימי דעים לגבי יכולתם לתפקד בסביבה זו. כשליש ציינו שהבחינה המתוקשבת הופכת את מקצוע הגאוגרפיה לנחשב יותר, ומרביתם הביעו תחושת מסוגלות עצמית בביצוע הבחינה.

בעקבות ההערכה

חלק מממצאי ההערכה יושמו כבר בשנת הלימודים תשס"ו: בוצע שדרוג משמעותי בממשק הבחינה ובמפגש המתוקשבות, והופקו לקחים לגבי הכשרת הצוות הבית ספרי. יתר הממצאים יוכלו לכוון לשינויים ולשיפורים נדרשים בתכנית ובדרכי הפעלתה ולסייע בהתוויית המדיניות של משרד החינוך לגבי בחינות בגרות מתוקשבות בגאוגרפיה ובמקצועות אחרים.

ביבליוגרפיה

- Bennett, R. E. (2001). How the Internet Will Help Large-Scale Assessment Reinvent Itself. *Education Policy Analysis Archives*, 9(5), Retrieved September 10, 2004, from <http://epaa.asu.edu/epaa/v9n5.html>
- Bennett, R.E. (2003). *Online assessment and the comparability of score meaning*. ETS, Princeton, NJ.
- Bugbee, A. C., Jr. (1996). The equivalence of paper-and-pencil and computer-based testing. *Journal of Research on Computing in Education*, 28, 282-299.
- Gallagher, A., Bridgeman, B., & Cahalan, C. (2000). The effect of computer-based tests on racial/ethnic, gender and language groups. ETS Research Report, 00-8, Retrieved May 10, 2005, from <http://www.ets.org/portal/site/ets/menuitem.c988ba0e5dd572bada20bc47c3921509/?vgnextoid=4ff0af5e44df4010VgnVCM10000022f95190RCRD&vgnnextchannel=d33246f1674f4010VgnVCM10000022f95190RCRD>
- Green, B., Jr. (1970). Comments on tailored testing. In W. H. Holtzman (Ed.), *Computer-assisted instruction, testing and guidance*. New York: Harper and Row.
- Powers, D. E., Fowles, M. E., Farnum, M. and Ramsey, P. (1994). Will they think less of my handwritten essay if others word process theirs? Effects on essay scores of intermingling handwritten and word-processed essays. *Journal of Educational Measurement*, 31, 220-233.
- Russell, M., & Plati, T. (2001). Effects of computer versus paper administration of a state-mandated writing assessment. *Teachers College Record*, Retrieved August 4, 2006, from <http://www.tcrecord.org>.
- Schaeffer, G.A., Bridgeman, B., Golub-Smith, M.L., Lewis, C., Potenza, M.T., & Steffen, M. (1998). Comparability of Paper-and-Pencil and Computer Adaptive Test Scores on the GRE General Test. ETS Research Report 98-38 Retrieved August 5, 2006, from <http://ftp.ets.org/pub/gre/002.pdf>
- Zhang, Y. Powers, D.E, Wright, W. and Morgan, R. (2003). Applying online scoring network (OSN) to advanced placement program (AP) tests. Retrieved July 29, 2006, from <http://www.ets.org/research/researcher/RR-03-12.html>