

סקירה בנושא הערכה ומדידה של בית הספר

מגישה : ד"ר אבישג אדרי

מוגש ליוזמה - מרכז לידע ולמחקר בחינוך, כחומר רקע לקבוצת העבודה בנושא מדידת שיפור בבתי הספר בישראל.

ספטמבר, 2019

מבוא

הערכה בחינוך היום לא מוגבלת רק להערכת הישגי תלמידים, אלא עוסקת גם בהערכת מורים ומנהלים, תוכניות לימודים וכללית האקלים החינוכי של מוסדות החינוך. גם מטרותיה אינן כוללות רק מיון וסיכום אלא גם נתינת משוב לצורך שיפור ומימוש זכות הציבור לדעת (נבו, 2003). מרבית המחקרים המתייחסים להערכת המערכת הלימודית מתמקדים בהשוואה בין בתי הספר בארץ ובעולם ולא מתייחסים כמעט לשיפור הישגי הסטודנטים והשוואתם לאורך השנים. מטרת הסקירה הנוכחית הינה להציג את המדדים שמטרתם לבחון שיפור בהישגי התלמידים לאורך השנים.

גישות שונות לקידום השיפור הבית ספרי :

למרות ששיפור בית ספרי נתפס כמפתח לשיפור ההישגים החינוכיים, האסטרטגיות בהן משתמשת המערכת החינוכית כדי לקדם שיפור בית ספרי, והתיאוריות שמחזקות את האסטרטגיות האלה שונות באופן משמעותי ממדינה אחת לאחרת. בקצה האחד אסטרטגיית התוצאות המבוססות על "דין וחשבון" של המערכת החינוכית באמצעות תמריצים הבאים לידי ביטוי בצורת תגמולים או סנקציות בהתאם לתוצאות המבחנים. בקצה השני אסטרטגיית המערכת המתפתחת שמבקשת להשפיע על השיפור באמצעות התמקדות בבניית היכולות ושינוי שיטות העבודה.

תוצאות המבוססות על דין וחשבון של המערכת : גישה זו הושאלה מעולם העסקים. כלומר הדרך ליצירת שיפור הינה באמצעות בירור מהן המטרות, פיתוח מדדים שיבחנו האם התוצאות הושגו ועידוד השגת המטרות באמצעות תמריצים משמעותיים. התפישה העומדת ביסוד גישה זו הינה שכאשר עובדים (במקרה הזה מורים) יודעים בדיוק מה מצופה מהם ומקבלים תגמול על כך – השיפור יקרה. מצד שני, אין כמעט התייחסות לאופן שבו השיפור יקרה ולדרכי ההוראה. חוסר המעורבות בדרך נחשבת כיתרון בגישה זו, שכן היא נותנת למורים חופש לבחור בכל דרך שנראית להם טובה לשיפור התוצאות. מעט מדינות אימצו את הגישה מבוססת התוצאות לשיפור בית הספר, ביניהן : הונגריה וארצות הברית. במדינות אלה מתייחסים לאיכות בית הספר בהתאם לתוצאות המבחנים הסטנדרטיים.

תרגול המבוסס על סיקור בית הספר : כאמור, רק מעט מדינות מתבססות על הגישה שפירטנו לעיל, רוב המדינות בודקות את בית הספר דרך תצפית ישירה על תהליכי הלמידה המתקיימים שם, כבסיס להערכת בית הספר ושיפור הלמידה בו לאורך זמן. הביקורת על בית הספר משמשת בעיקר כמשוב לבית הספר עצמו על הביצועים שלו והתחומים בהם עליו להשתפר. התפישה העומדת בבסיס הגישה הזו הינה שכאשר מזוהים תחומים הדורשים שיפור בבית הספר ובתי הספר מקבלים עידוד ותמיכה לתקן בהתאם – השיפור יקרה (Masters, 2012).

ברלינר (Berliner, 2018) טוען ששתי הגישות הנפוצות להערכת מורים הינן באמצעות ציוני מבחנים סטנדרטים ותצפיות. לגבי המבחנים הוא טוען שלא הוגן להעריך מורים לפי הציונים שתלמידיהם קיבלו, מכיוון שהמורים לא משפיעים באופן הגבוה ביותר על הישגי התלמידים, הרקע הסוציאקונומי ממנו מגיע התלמיד משפיע יותר מהמורה שיש לו. לטענתו, גם השיטה להערכת מורים באמצעות תצפיות יכולה להיות לא הוגנת משום שרמת האמינות שלה נמוכה. כמו כן הוא טוען שהמתאם בין שתי השיטות נמוך יחסית, וכששתי שיטות השואפות להעריך את אותו משתנה נמצאות במתאם נמוך, אחת מהן או שתיהן כשלו כנראה בהערכה.

במחקר שנערך בארצות הברית הראו החוקרים שקיימת הטיה בהערכת מורים באמצעות תצפיות, שההבדלים בהערכות לא נבעו מאיכות ההוראה אלא היו קשורים לנושאים שאינם בשליטת המורים כמו ההרכב המגדרי והאתני של אוכלוסיית התלמידים (Campbell & Ronfeldt, 2018).

אמידה לפי תמונת מצב

המבחנים המתוקננים המוצגים בחלק זה בודקים אמנם גם שיפור בין השנים, אולם ממוקדים יותר בהצגת תמונת מצב אודות כלל בתי הספר בארץ ובעולם ודירוג הרמה שלהם. המבחנים נותנים לנו תמונת מצב אודות המצב הלימודי בין בתי הספר בארץ ובין המדינות המשתתפות במבחן בעולם

מבחני הבגרות :

מבחני הבגרות הם היום המטרה הלימודית המרכזית, ואולי היחידה, של מדינת ישראל, המשפיעות על התלמיד לאורך כל שנות לימודיו (פרנק, 2017). בחינות הבגרות מהוות את קו הסיום של תהליך חינוך שתחילתו בגן הילדים והוא נמשך עד סוף כיתה יב'. לכן כל דיון המתייחס למטרות בחינות הבגרות צריך להתחיל במטרות של מערכת החינוך הישראלית (זוהר, 2010).

לפי נתוני משרד החינוך המופיעים באתר בחינת הבגרות צריכה לכלול לפחות 21 יחידות לימוד בהתאם לחלוקה הבאה : 3 מתמטיקה, 3 אנגלית, 2 תנ"ך (למוסלמים ודרוזים לימודי התרבות שלהם), 2 היסטוריה, 2 אזרחות, 2 ספרות, 2 לשון והבעה וכמו כן מקצוע מורחב אחד בהיקף של 5 יחידות.

זוהר (2010) טוענת במאמרה שבחינות הבגרות לא עברו כמעט כל שינוי בעשורים האחרונים. במבנה הנוכחי בחינות הבגרות בודקות הישגים הנמדדים באמצעות בחינות בכתב, בהן נבחן התלמיד בעיקר על ידע והבנה בסיסיים. בבחינות יש מעט מטלות הדורשות הבנה מעמיקה וחשיבה גבוהה. כמו כן בתעודת הבגרות אין התייחסות למטרות ערכיות, חברתיות או כישורי למידה.

מבחני המיצ"ב :

מבחני המיצ"ב (מדדי יעילות וצמיחה בית ספרית) הם חלק ממגמה עולמית של התרחבות השימוש במדדים כמותיים במטרה לעקוב אחר ההישגים הלימודיים ולבחון אם הם תואמים לנדרש. השימוש במבחני המיצ"ב התחיל ב-2002. החל משנת 2006 מבצעת אותם ר"מ (פניגר, 2010).

באתר הרשות הארצית למדידה והערכה בחינוך (רא"מ), מתוארים מבחני המיצ"ב כמערכת מדדים בית ספריים שמטרתה לספק מידע על תפקוד בית הספר במגוון תחומים. המיצ"ב כולל מבחני הישגים וסקרי

אקלים וסביבה פדגוגית. מבחני הישגים מתקיימים בכיתות ה' וח'. בכיתות ה' נבחנים התלמידים בשלושה נושאים : אנגלית, שפת אם ומתמטיקה. בכיתה ח' הם נבחנים בנושאים אלו ובנוסף גם במדעים. סקרי האקלים מועברים לתלמידים, למורים ולמנהלים וכוללים נושאים כמו : תקשורת הורים מערכת החינוך, מוטיבציה ללמידה, תחושת מוגנות ועוד.

בכתבה בהארץ מתאריך 01.01.2005 מוצגת טענתו של פרופ' דוד נבו, שכדי לדעת את רמת התלמידים בבית הספר אפשר להסתפק במדגם מייצג של מבחנים במוסדות החינוך. עריכת מבחנים בכל המערכת החינוכית משמשת כאמצעי הפחדה והטלת משמעת כלפי בתי הספר והמורים (אבדור, 2014).

קליגר (2009) טוענת, בהתאם להצהרות משרד החינוך, שהמיצ"ב נועד לשמש כלי להערכה מעצבת, כלומר מטרתו לתת לבתי הספר כלי לקידום עבודת בית הספר. במחקר שערכה בנושא, במסגרתו קיימה 30 ראיונות עם מנהלים ו-92 ראיונות עם מורים, עלה שלדעת רוב המנהלים והמורים, בדיקת הישגי התלמידים היא המטרה העיקרית של המיצ"ב. כמו כן רק מעט מורים ומנהלים ראו במיצב כלי לשיפור תהליכים. נראה אם כן כי יש פער בין מטרות המיצ"ב כפי שתכנן אותם משרד החינוך ובין תפישת המיצ"ב בקרב מורים ומנהלים (קליגר, 2009).

פניגר (2010) מתייחס ליתרונות ולחסרונות של מבחני המיצ"ב. הוא מתבסס על סקירת מחקרים שנערכו על מבחנים סטנדרטים בארצות הברית, וטוען שהמבחנים הסטנדרטים מעודדים רמאות ושהשיפור בהישגים זניח. הוא מביא את דברי פרופסור דיויד ברלינר הטוען כי עלייה בהישגים אינה מעידה בהכרח על למידה טובה יותר, לעתים אפילו להפך. לטענתו, אחת הדרכים הפשוטות בעזרתן יכולים בתי הספר לשפר את ציוני התלמידים הינה באמצעות לימוד אינטנסיבי של החומר הנלמד למבחן על חשבון לימוד של מקצועות אחרים. לסיכום דבריו מציע פניגר להפנות את המשאבים הגדולים המושקעים במבחני מיצ"ב למחקרי אורך שיבחנו תהליכים במערכת החינוך.

בניגוד לדעתו של פניגר, שטען כאמור לשיפור זניח בהישגים, בלס (2014) מציג את תוצאות מבחני המיצ"ב לשנים 2008 ו-2013 וטוען בהתאם לנתונים שמדובר בשיפור משמעותי של הציונים, הן בקרב דוברי הערבית והן בקרב דוברי העברית.

מבחנים בינלאומיים :

באתר הרשות הארצית למדידה והערכה בחינוך (ראמ"ה) מצוין כי בשנים האחרונות ישראל משתתפת בארבעה מחקרים שונים הנערכים אחת למספר שנים. המחקרים מנוהלים ע"י שני ארגונים : OECD (הארגון לשיתוף פעולה ופיתוח כלכלי) שבמסגרתו ישראל משתתפת במחקרי PISA ו-TALIS, וארגון IEA (הארגון הבינלאומי להערכת הישגים בחינוך), שבמסגרתו ישראל משתתפת במחקרי PIRLS ו-TIMSS. מבחן PISA בודק את רמת האוריינות של תלמידים בגיל 15 בקריאה מתמטיקה ומדעים. השאלות הנכללות במחקר בוחנות ידע החיוני לחיים המעשיים תוך דגש על מיומנויות. המבחן נערך פעם בשלוש שנים. כמו כן הוא כולל גם סקר מורים וסקר מנהלים. מבחן TALIS בודק תפישות הורים ומורים בנוגע להוראה ולמידה. המידע נאסף באמצעות שאלונים המועברים למורים ולמנהלים בבית הספר. מבחן PIRLS נערך אחת לחמש שנים ובודק אוריינות קריאה בקרב תלמידי כיתות ד' תוך התמקדות במיומנויות הבנת הנקרא, השגת יעדי קריאה ועמדות כלפי קריאה. מבחן TIMSS בוחן את רמת השליטה של תלמידי כיתות ח' במתמטיקה ומדעים, תוך התאמה לתוכנית הלימודים הקיימת במדינה המשתתפת, המחקר נערך כל ארבע שנים.

קליגר (2013) בחנה את השפעת המבחנים הבינלאומיים על מערכת החינוך בישראל תוך התמקדות במקצוע מדעים שהוא מקצוע הנמדד במבחני הטיימס והפיזה. היא טענה במאמרה כי היתרונות של שילוב מבחנים בין לאומיים הינם בשיקוף הישגי התלמידים, העלאת רמת החשיבה, מיקוד נושאי הלימוד ושיפור המבחנים הארציים. לעומת זאת החסרונות שהיא תיארה הינם: ההכנה למבחנים, הגבלת האוטונומיה של המורים והלחץ הקיים על המורים לדאוג שתלמידיהם יעברו בציון גבוה.

בדו"ח מחקר שכתב קרנוי (Carnoy, 2015) על המבחנים הבינלאומיים, הוא סוקר את ארבעת הביקורות המרכזיות עליהם. בביקורת הראשונה הוא מציין שהמטרה של דירוג המדינות על פי ניקוד המבחן הממוצע הינו כדי לאפשר הסברים על איכות המערכות החינוכיות הלאומיות, אך הדירוג מטעה מכיוון שהוא מתבסס רק על הישגי המערכת ומתעלם מהרקע המשפחתי של התלמיד שהינו גורם משמעותי המשפיע על ההישגים. הביקורת השנייה מתייחסת לכך שתלמידים קיבלו במבחן בינלאומי אחד בחשבון ציונים גבוהים ואילו במבחן בינלאומי אחר ציונים נמוכים. זה מעלה את השאלה איזה מבין המבחנים הוא בעל תוקף נמוך יותר. הביקורת השלישית טוענת שהמבחנים הינם בעלי אחוז שגיאה גבוה בהשוואה למבחנים מתוקננים אחרים. בביקורת הרביעית הוא טוען שארגון ה-OECD הציג את מערכת החינוך בשנחאי כמודל לשאר העולם וכמייצגת את סין. אך המדגם לא מייצג אפילו את אוכלוסיית שנחאי הצעירה ובוודאי לא את סין. כמו כן בתי הספר בשנחאי אינם כוללים אוכלוסיית מהגרים. לפיכך היה צריך להוציא את ציוני שנחאי מחוץ לקבוצת ההשוואה ולהעלות תהיות לגבי האמינות של ה-OECD כארגון האחראי על השוואת הישגים בין מדינות בעולם (Carnoy, 2015). גם הרגריבס ובראון (Hargreaves & Braun, 2013) בדוח שערכו בארצות הברית בנושא הערכה, המליצו לצמצם את המבחנים וטענו שאין צורך לבחון כל ילד כל שנה.

אמידה לפי מודל צמיחה

בדוח מחקר שנכתב עבור היוזמה לשיפור בית הספר (RSI) באוסטרליה, במטרה לייעץ אודות הדרכים הטובות ביותר להעריך שיפור בית הספרי (Masters, 2012), מתואר כי מכיוון שמטרת בית הספר הינה לשפר את הישגי הסטודנטים, נראה הגיוני להעריך בתי ספר בהתאם לציונים שתלמידיהם מצליחים להשיג. בפועל, השימוש בהישגי הסטודנטים כדי להעריך את יעילות בתי הספר היא בעייתית משתי סיבות:

ראשית המבחנים נערכים בטווח מוגבל של תחומי למידה (למשל אוריינות וחשבון), כך שמדובר במאגר נתונים מוגבל שיש להסיק לפיו על רמת היעילות הבית ספרית. כמו כן ברגע שיש תמריצים שצריך להשיג דרך ציונים גבוהים זה מעוות את המיקוד על מאמצי בית הספר ומעודד התנהגות לא ראויה במטרה למקסם את ההישגים כמו למשל "העלמת" תלמידים שיכולים לפגוע בממוצע או סיוע בעת המבחן.

הנקודה השניה הינה שהישגי סטודנטים מושפעים גם מגורמים שאינם בשליטת בית הספר. אחת הדרכים להתגבר על הבעיה הינה שימוש במדדים שפחות מושפעים מהגורמים שאינם בשליטת בית הספר, בניסיון להעריך את תרומת בית הספר בהתאם לנסיבות. למשל בחינת הבדלי הציונים בין השנה השלישית לשנה החמישית, במקום ציון הממוצע של קבוצת המחזור בשנה מסוימת של בית הספר (Masters, 2012).

במחקר שערכו סאן ועמיתיו (Sun, Penner & Loeb, 2017) הם בחנו את תרומת מענקי השיפור שנתנה הממשלה לבתי הספר החלשים ביותר במדינה. במחקר הנוכחי הם אספו נתונים לאורך כמעט עשור כדי

לבחון את השפעת התקציב על בית הספר מעבר לשלש שנות המענק. לצורך איסוף נתונים הם לקחו נתונים מהמחוז המאוחד של סאן פרנסיסקו (SFUSD) על תלמידים מורים ובתי ספר בין השנים 2005-2013 ובדקו נוכחות תלמידים, הישגי תלמידים ושימור מורים במערכת. ממצאי המחקר מראים על שיפור בהישגי התלמידים ובתחומים נוספים בשנה השלישית וזה תואם את ממצאי הסקירה שערכו, על כך ששינויים לוקחים זמן ומופיעים בהדרגה. בנוסף, הם מצאו במחקרם שיש ירידה בהיעדרויות לא מוצדקות מבית הספר, שימור מורים מקצועיים ופיתוח התמקצעות מורים.

מודל של הערכה לפי ערך מוסף (VAM- Value Added Model)

מודל ערך מוסף הוא מודל סטטיסטי המציג אומדנים לגבי תרומת בית הספר לצמיחה בהישגי התלמידים. המטרה של השימוש במודל הינה לאפשר השוואה הוגנת של הישגי התלמידים בין בתי הספר השונים, כשהנחת היסוד הינה שבתי הספר מכילים תלמידים מאוכלוסיות שונות וזה יכול להשפיע על רמת בית הספר (Meyer & Christian, 2008). המודלים המשמשים היום בבית הספר להערכה בעלת ערך מוסף חייבים להבטיח שההפרשים בין ערכי המשתנה יהיו שווים. כלומר שההפרש בין עשר נקודות לעשרים נקודות יהיה שווה להפרש שבין עשרים נקודות לשלושים נקודות. חשוב להקפיד על כך כדי שנוכל להסיק מסקנות מדויקות ככל האפשר אודות הנתונים (Ballou, 2009).

במחקר של מקאפי ועמיתיו (McCaffrey, Sass, Lockwood & Mihaly, 2009) הם בחנו את השפעת המורים על הישגי התלמידים, ושילבו את מודל ערך מוסף עם מדדים אחרים להערכה. הם טענו שמודל ערך מוסף פחות מדויק להערכה קצרת טווח ושכדאי לשלב אותו ביחד עם אמצעי הערכה נוספים כמו תצפית למשל. קודל (Koedel, 2009) בחן בעבודתו את ההשפעה של מורים בתיכון על הישגי התלמידים במבחני קריאה באמצעות מודל ערך מוסף. הממצאים הראו שאיכות המורים לאנגלית השפיעה באופן משמעותי על הישגי התלמידים במבחני הקריאה, ולעומת זאת מורים למתמטיקה יכולים להשפיע על ביצועי הקריאה אך התוצאות לכך פחות ברורות. לגבי התוקף של המדד הוא טוען שהתוצאות מעורבות. לגבי ההשפעה של המורים לאנגלית אין הוכחות להטיה, אך לגבי ההשפעה של המורים למתמטיקה קיימים רמזים להטיה. לסיכום הוא טוען שאמנם כדאי להשתמש במודל ערך מוסף בזירות, אך בסך הכל התוצאות מראות ששיטת ערך מוסף יכולה להיות כלי יעיל להערכת ההשפעה של המורים.

מודל ערך מוסף לפי הקשר (CVA – Contextual Value Added).

הנושא של שיפור בית ספרי הוא משמעותי במדיניות החינוכית של אנגליה, מגוון מדדים לשיפור בית הספר פותחו במטרה להעלות את הסטנדרטים של בתי הספר באנגליה, אחד המדדים החשובים והמרכזיים ביותר להערכה הינו מדד ערך מוסף (VAM). זמינות הנתונים הלאומיים של כל תלמיד מאז 2002, אפשרו את הפיתוח של מדד ערך מוסף לפי הקשר (CVA) (Ray, Evans & McCormack, 2008).

ישנם יתרונות וחסרונות למדד CVA בהערכת השיפור הבית ספרי. למשל הם יכולים לשלוט במשתנים המשפיעים על תלמידי בית הספר ולמפות אותם בהתאם לצורך, אך מצד שני גם משתנים אלה יכולים להיות אנדוגוניים בחלקם. הם מעמידים את בתי הספר על בסיס שווה יותר, כך שהערכים המוספים יכולים

לספק אינדיקטור הוגן יותר לאפקטיביות בית הספר, אך מצד שני ייתכן שיש הורים או אנשים אחרים שאינם מעוניינים לערוך השוואות על בסיס זה (Ray, Evans & McCormack, 2008).

מדידה רבת משתנים

מאפיין שכיח של גישות להערכת מורים שפותחו בחמש השנים האחרונות בארצות הברית זה שילוב של מדידת ביצועי התלמידים עם עבודת המורים. בקולורדו, מודל הצמיחה של קריאה, כתיבה ומתמטיקה הוא מרכיב מפתח שיכול להשתלב במערכת הערכת המורים וממנו נוצר ה SGP (Student growth percentile). ישנן דרכים שונות להשתמש בנתונים אלה, אחת הדרכים הינה בשימוש במוצע או חציון של אחוזי צמיחת הסטודנטים ה-MGP (Mean or median grow percentile) ממדד ה-MGP יעילה כבסיס להערכת מורים, כשהמידע שהיא מספקת משולבת עם מידע נוסף אודות ביצועי הסטודנטים (Diaz-Bilello & Briggs, 2014).

ניתן להשתמש ב SGP לחישוב ממד בית ספרי (על בסיס ממוצע או חציון), המשקף כמה התלמידים בבית הספר שיפרו את הישגיהם בהשוואה לתלמידים אחרים (במדינה, במחוז וכדומה) שהיו בעלי הישגים התחלתיים דומים. מדד זה יכול להיות בטווח של 0-99; כאשר טווחים של 0-33% משקפים שיפור קטן; 33-66% שיפור טיפוסי ו-67-99 שיפור גדול¹.

מדד נוסף לבחינת שינויים/שיפור בהישגים של בתי ספר, המבוסס גם על השוואה להישגים קודמים הוא PROGRSS 8. זהו מדד חדש בבריטניה לבחינת שיפור של בתי ספר, כך שכל בית ספר מקבל ציון שנע בין 1 ל-1. כאשר 1 משקף שיפור ו-1 משקף, שבית הספר לא עומד בסטנדרט המצופה ממנו בהינתן ההישגים הקודמים של תלמידיו. מדד זה מבוסס על 8 מקצועות הנלמדים בחטיבה העליונה ("מקביל לבגרות בישראל").

בן-דוד הדר פיתחה את המודל התיאורטי של שיפור בראיה התפלגותית. לשם כך, הוגדר שיפור לאורך עקומת השיפור בהישגים (השב"ה) (BenDavid-Hadar, 2011). עקומת השב"ה מתארת את הקשר שבין שיפור ברמת ההישגים לבין צמצום פערים בהישגים. קשר זה מתואר כעקומת U הפוכה. כלומר, הקשר אינו ליניארי. מודל תיאורטי זה אושש ברמה האמפירית במחקר נוסף שלה (BenDavid-Hadar, 2009).

בנוסף לפיתוח התיאורטי משנת 2011, בן-דוד הדר פיתחה מודל הקצאה משפר המורכב מפיצוי על משתני רקע מוחלשים יחד עם פרסים על שיפור. לשם כך אמדה את השיפור על בסיס נתוני אורך ברמת התלמיד בישראל. מודל זה לקח בחשבון את תרומתם של הישגים קודמים (בבחינות המיצ"ב) בנוסף לתרומתם של משתני הרקע של התלמיד (למשל השכלת הורים), פרופילים של מורים (למשל השכלת המורים) ומאפייני בית הספר, להסברת השונות בהישגים בבחינות הבגרות. מודל זה כלל את כל בחינות הבגרות על כל סמלי

¹ יישום של רעיון זה ברמת בתי ספר בישראל (לא מחושב על בסיס תלמידים אלא על בסיס דירוג בתי ספר) הראה כי שליש מבתי הספר שיפרו את עצמם בזכאות לבגרות, כאשר נמצאו הבדלים לפי מגזרי חינוכיים, במידה שבתו ספר שיפרו עצמם בין השנים 2014 ו 2017 (Addi-Racah and Sal-Man, 2018)

המקצועות השונים ברמות (יחידות הלימוד) השונות של כל מקצוע בבגרות בישראל ברמת התלמיד. התרומה של שיפור לאחוז השונות המוסברת עומדת על כ- 25% (BenDavid-Hadar, 2018).

בן דוד הדר (Ben-David Hadar, 2015) טוענת במאמרה שהתוכניות להערכת החינוך (EETs) מתמקדות רק בפן אחד של מערכת החינוך (למשל הישגי הסטודנטים) ומשתמשות בשיטות הערכה ליניאריות (למשל מבחנים סטנדרטים) כדי להעריך שיפור, ולכן הן פחות אפקטיביות. למשל, בית ספר יכול להגיע להישגים לימודיים גבוהים אך "לשלם על כך מחיר" בפגיעה ברווחה הנפשית של התלמידים או בהפחתת פעילויות חברתיות. ניתן לטעון אמנם שהרבה מטרות חינוכיות לא ניתנות למדידה, ועדיין קיים קונצנזוס בקרב אנשי חינוך שלמערכת החינוך יש יותר משני היבטים הניתנים למדידה.

לאור הני"ל, מציעה בן דוד הדר גישה חדשנית רבת פנים, ולא ליניארית (NOMA-nonlinear multifaceted approach). לצורך בניית המודל היא בנתה מחקר שבחן את ההיבטים השונים של מערכת החינוך. המחקר התקיים בשני שלבים. בשלב הראשון התקיים סקר מקדים בקרב 50 אנשים חינוך שנשאלו אודות תוצאות החינוך, כך נאספו התוצאות הכי חשובות לחינוך. בשלב השני התבקשו 200 מומחים ואנשי חינוך לדרג את התוצאות לפי רמת חשיבותם. המודל המוצע מבוסס על התוצאות שקיבלו ציונים גבוהים. תוצאות הדירוג הינן (לפי הסדר מהחשוב ביותר להכי פחות חשוב): ערכים, הישגים לימודיים, בריאות ויעילות עצמית (קיבלו ביחד את המקום השלישי והרביעי), מעורבות חברתית, מוטיבציה להצליח, אופטימיות, אלימות, מודעות עצמית, שביעות רצון, מנהיגות, חדשנות, שיטות הוראה ויצירתיות.

היזמה לשיפור ההוראה של קרן גייטס ניסתה לשפר את הישגי התלמידים, במיוחד הישגי תלמידים המסיימים תיכון, וכן את הנוכחות של תלמידים מרקע סוציו אקונומי נמוך במכללות, באמצעות הרחבת נגישותם להוראה יעילה. הערכת התוכנית החלה ביולי 2010 ונמשכה שש שנים, היא התבססה על ראיונות עם המנהלים, סקרים שנתיים של מורים ומנהיגי בית ספר, חקר מקרים של בתי ספר, ניתוח רשימות מהאתר, ציוני המבחנים של התלמידים וסקירה של מסמכים רלוונטיים. היזמה כשלה, המקומות בהם פעלה היזמה ישמו גישות חדשות של הוראה אפקטיבית אך לא השיגו את המטרות מבחינת שיפור ההישגים או העלאת שיעור המסיימים את התיכון בקרב אוכלוסיות חלשות. חלק מההסברים לכשלון היזמה מתייחסים להשפעת גורמים חיצוניים או פרק זמן קצר מכדי להציג תוצאות (Stecher, 2018).

במחקר שבא לבחון את התוקף של מבחן הערכת אוצר מילים באנגלית כשפה שניה ללומדים צעירים בשוודיה (YLVAT), בדקו החוקרים קשר בין המבחן להערכת אוצר מילים לבין ציוני מבחן לאומי בהבנת הנקרא והנשמע באנגלית, הערכת הלומדים את המבחן והערכת המורים את הלומדים. הממצאים הראו על קשר חיובי משמעותי בין המבחן להערכת אוצר מילים לבין ציוני המבחן בהבנת הנקרא והנשמע, וכן בין המבחן להערכת אוצר מילים לבין הערכת הלומדים את המבחן והערכת המורים את אוצר המילים של הלומדים (Sylvén & Sundqvist, 2016). אמנם במחקר הנוכחי המטרה הייתה לבחון את התוקף של מבחן הערכת אוצר מילים, אולם השילוב של מספר מבחנים באותו נושא עם הערכת מורים ותלמידים יכול להעלות את אמינות המבחנים המשווים שכפי שראינו יכולים להתאים לשמש כמדד לשיפור. מחקר זה יכול לספק תשובה חלקית לחסרונות שתוארו לגבי המבחנים הסטנדרטים (עידוד לרמאות, פגיעה באוטונומיה של המורים וכן הלאה).

דוגמאות למודלים לשיפור בית ספרי בעולם

מסגרת מעגלית לשיפור הישגי התלמידים (FISO)

באתר² המחלקה לחינוך של מדינת ויקטוריה מתוארת מסגרת לשיפור הישגי הסטודנטים (FISO- Framework for improving student outcomes) הכוללת מעגל בן ארבעה שלבים. השלב הראשון כולל הערכה ואבחון, בשלב השני יש להציב מטרות בהתאם לסדרי העדיפויות של בית הספר, בשלב השלישי יש לבנות תוכנית שיפור שנתי ובשלב הרביעי יש ליישם את התוכנית ולפקח על תוצריה. השלב הראשון שכולל כאמור הערכה ואבחון משתמש בנתוני מדדי השיפור כדי להשוות ולמדוד את הישגי התלמידים. מדדי השיפור כוללים מגוון נתונים המתייחסים להיבטים השונים של בתי הספר. אמצעי השיפור בשילוב מרכיבים אחרים של המסגרת לשיפור הישגי הסטודנטים (FISO) יכולים לספק לבית הספר את הנתונים הנדרשים לתכנון ויישום יעיל של המסגרת.

תוכנית לשיפור החינוך בערים :

המרכז לחינוך ציבורי בחר 18 ערים בארצות הברית כדי לעקוב אחר שיפור החינוך שלהן. המרכז לחינוך ציבורי בדק את יישומן של שלוש מטרות :

1. מערכת החינוך נמצאת בשיפור מתמיד
2. לכל התלמידים יש גישה להשכלה איכותית גבוהה.
3. אסטרטגיית החינוך מושרשת בקהילה.

נתוני המחקר התבססו על 195 ראיונות עם מנהיגים וקבוצות, ניתוח אתרי אינטרנט ומסמכי מדיניות, סקרים עם 3,200 הורים שנערכו בשמונה ערים, נתונים על ביצועי בית הספר לפי משרד החינוך, נתונים על רישום תלמידים לקורסים מתקדמים במתמטיקה ממאגר הנתונים של האגודה לזכויות האזרח ונתונים דמוגרפיים של העיר.

קיימים כנראה 18 דוחות כמספר הערים, כל דוח מחקר פותח בפסקה קצרה אודות העיר בה הוא עוסק. החלק הראשון בדוח מתייחס לאתגרים העומדים בפני המקום. למשל בדוח המחקר של בוסטון (Center On Reinventing Public Education, 2018a) אחד האתגרים הוא לדאוג שלמשפחות תהיה האינפורמציה הנדרשת כדי לבחור את המערכת החינוכית המתאימה להן, ואילו בדוח המחקר של שיקגו (Center On Reinventing Public Education, 2018b) אחד האתגרים הוא שיפור אפשרויות התחבורה לבתי הספר.

החלק השני בדוח מבקש להאיר על נושאי חינוך המתרחשים באיזור. למשל בדוח המחקר של בוסטון (Center On Reinventing Public Education, 2018a) הם מתארים שבתי ספר קתולים, צארטרים ומחוזיים יצרו שיתופי פעולה לצורך שיפור העבודה של המורים, ואילו בדוח המחקר של פילדלפיה (Center On Reinventing Public Education, 2018d) התייחסו בחלק הזה לאימון מנהיגי בית ספר לצורך שיפור הישגי הסטודנטים וגיוס מורים.

² <https://www.education.vic.gov.au/school/teachers/management/improvement/Pages/improvement-measures.aspx>

החלק השלישי מתייחס לשינויים במערכת בשלושה תחומים. כאשר כל תחום מתחלק למספר נושאים וכל נושא מקבל ציון – קיימים ארבעה ציונים אפשריים – מעט, מתפתח, טוב וראוי לשבח:

1. האם המערכת החינוכית ממשיכה להשתפר? שאלה זו מתחלקת לשאלות הבאות: האם יש לבתי הספר את סוגי המנהיגים שהם צריכים? האם לבתי הספר יש את סוגי המורים שהם צריכים? האם המימון מתחלק באופן שוויוני בין התלמידים?

2. האם לתלמידים יש גישה לחינוך איכותי? שאלה זו מתחלקת לשאלות הבאות: האם היצע בית הספר מספק מערך של מודלים? האם התחבורה מאפשרת לכל משפחה לבחור בית ספר שיתאים לה גם אם הוא לא במרחק הליכה? האם העיר מנהלת את בתי הספר באופן אסטרטגי? האם תהליך ההרשמה עובד עבור המשפחות? האם למשפחות יש את המידע שהן צריכות ואת הידע איך להשתמש בו?

3. האם אסטרטגיית החינוך נעוצה בקהילה? שאלה זו מתחלקת לשאלות הבאות: האם מגוון של קבוצות עוסקות בחינוך? האם העיר משתפת משפחות בהחלטות החינוכיות שקשורות אליהן? האם מערכת החינוך מגיבה למושבי הקהילה? האם יש קואליציה חזקה של תמיכה במערכת החינוך?

כאמור, כל נושא וכל תחום מקבלים ציון. למשל בדוח מחקר על פילדלפיה הנושא - האם התחבורה מאפשרת למשפחות בחירה בין בתי הספר – קיבל ציון ראוי לשבח מכיוון שהמחוזות מספקים לכל התלמידים הגרים במרחק הליכה של יותר מ-1.5 מייל מבית הספר תחבורה בחינם לכל בית ספר שיבחרו. אך התחום גישה לחינוך איכותי קיבל רק ציון טוב מכיוון שרק שאלה נוספת קיבלה ציון ראוי לשבח, שאלה שלישית קיבלה ציון טוב ושתי שאלות נוספות קיבלו ציון מתפתח (Center On Reinventing Public Education, 2018d).

החלק הרביעי בדוח המחקר מתייחס להישגי התלמידים, הוא כולל סיכום קצר של הישגי העיר והצגת דיאגרמות המתארות כל אחד מהנושאים אליהם התייחסו בהישגי התלמידים: שיעור המסיימים את בית הספר בין השנים 2011-2015, ציוני המתמטיקה לאורך השנים 2011-2015, ציוני הקריאה לאורך השנים 2011-2015, האם בכל שכונה יש בית ספר שמשפק השכלה איכותית לתלמידים ממיצב נמוך? מהו אחוז המשתתפים בקורסים מתקדמים במתמטיקה בהשוואה בין הקבוצות לבנים, היספנים, שחורים ואחרים?

למשל בדוח מחקר על העיר ניו יורק (Center On Reinventing Public Education, 2018c) נכתב כי תלמידים ממיצב סוציאקונומי נמוך מצליחים מעט יותר במבחני הערכה סטנדרטים מאשר תלמידים אחרים באותו מיצב במדינה, ושיעורי ההצלחה במתמטיקה וקריאה השתפרו ביחס למדינה. שיעורי המסיימים את בית הספר נותרו מפגרים אחר המדינה ב-10%. רוב התלמידים נרשמים לשיעורים מתקדמים במתמטיקה בשיעור דומה לאחוזם היחסי באוכלוסיית בית הספר, אם כי זה לא נכון לגבי כל הקבוצות האתניות.

סיכום

ממד לשיפור ההוראה	תיאור	יתרונות	חסרונות
אמידה לפי תמונת מצב	מבחנים סטנדרטים המשווים הישגי תלמידים	שיקוף הישגי התלמיד ובית הספר, שיפור הישגי התלמידים (לפי חלק מהדיעות).	מעודד רמאות, השקעה במיקוד החומר על חשבון נושאים אחרים, לחץ על המורים.
אמידה לפי מודל צמיחה	מעקב אחר ההישגים של בית ספר מסוים לאורך מספר שנים.	הפחתת ההשפעה של גורמים חיצוניים על ההישגים (למשל הרקע הסוציאקונומי של התלמידים)	לא ניתן לקבל תמונה רחבה של המצב בכלל בתי הספר במדינה.
מודל ערך מוסף (VAM)	חישוב מגמת הצמיחה הצפויה והשוואתה לצמיחה בפועל	שימוש מושכל ומדויק יותר בנתונים.	פחות מדויק להערכה קצרת טווח
מודל ערך מוסף בהקשרו (CVA)	אומדן הערך המוסף תוך ניכוי משתני הרקע השונים	מסייע להשוואה הוגנת ומדויקת יותר של מודל ערך מוסף.	משתני הרקע יכולים להיות אנדוגניים בחלקם.
מדידה רבת משתנים	שימוש במגוון רחב של מדדים לשיפור בית הספר.	תמונה רחבה ומדויקת יותר של המצב בבית הספר.	דורש משאבים גדולים יותר. אין התייחסות לשיפור לאורך שנים.

מקורות :

- אבדור, שי (2014). על הערכת איכות ההוראה. *קולות*, 7, 23-29.
- בלס, נ' (2014). *מגמות בהתפתחות מערכת החינוך*. ירושלים : מרכז טאוב.
- זוהר, ע' (2010). התחדשות בחינת הבגרות הישראלית : רבולוציה או אבולוציה. *עיונים בחינוך*, 3, 158-174.
- נבו, ד' (2003). הערכה כדיאלוג : הצעה לשינוי מדיניות ההערכה של בית הספר. בתוך י' דרור, ד' נבו ור' שפירא (עורכים), *תמורות בחינוך : קווים למדיניות החינוך בישראל לשנות האלפיים (עמ' 315-330)*. תל אביב : רמות.
- פניגר, י' (2010). האם מבחני המיצ"ב טובים או רעים למערכת החינוך שלנו? *הד החינוך*, פ"ד (4), 22-24.
- פרנק, א' (2017). מ"תעודת בגרות" ל"תעודת גמר" ומיון הוגן לאקדמיה. *ביטאון מכון מופת*, 60, 94-99.
- קליגר, א' (2009). בין המיצ"ב למצב : תפיסות מנהלים ומורים את מטרות המיצ"ב לעומת המטרות המוצהרות, *דפים*, 47, 142-184.
- קליגר, א' (2013). מבחנים בינלאומיים כסטנדרטים למערכת החינוך, *מעוף ומעשה*, 15, 79-108.
- Ballou, D. (2009). *Test scaling and value added measurement*. American Educational Finance Association.
- BenDavid-Hadar, I. (2009). Excellence and equality in Israeli schools' achievements: Policy implementation. *Megamot*, 46(3), 356-379. (In Hebrew).
- BenDavid-Hadar, I. (2015). Conceptualizing a multifaceted approach to evaluate improvement in educational outcomes. *Curriculum and Teaching*, 30(1), 37-50.
- BenDavid-Hadar, I. (2018). Funding education: Developing a method of allocation for improvement. *International Journal of Educational Management*, 32(1), 2-26.
- BenDavid-Hadar, I., & Ziderman, A. (2011). A new model for equitable and efficient resource allocation to schools: The Israeli case. *Education Economics*, 19(4), 341-362.
- Berliner, D. (2018). Between Scylla and Charybdis: Reflections on and problems associated with the evaluation of teachers in an era of metrification. *Education Policy Analysis Archives*, 26(54), 1-24.
- Campbell, S. L., Ronfeldt, M. (2018). Observational evaluation of teachers: Measuring more than we bargained for? *American Educational Research Journal*.
DOI: 10.3102/0002831218776216
- Carnoy, M. (2015). *International test score comparisons and educational policy*. Stanford University, National Education Policy Center.
- Center On Reinventing Public Education. (2018a). *Boston citywide education progress report*.
- Center On Reinventing Public Education. (2018b). *Chicago citywide education progress report*.

- Center On Reinventing Public Education. (2018c). *New York City citywide education progress report*.
- Center On Reinventing Public Education. (2018d). *Philadelphia citywide education progress report*.
- Diaz-Bilello, E.K., & Briggs, D.C. (2014). *Using Student Growth Percentiles for educator evaluations at the teacher level*. Center for Assessment.
- Hargreaves, A. & Braun, H. (2013). *Data-Driven Improvement and Accountability*. Boulder, CO: National Education Policy Center.
- Koedel, C. (2009). *An empirical analysis of teacher spillover effects in secondary school*. National Center On Performance Incentives.
- Masters, G. N. (2012). *Measuring and rewarding school improvement*. Commonwealth Department of Education, Employment and Workplace Relations.
- McCaffrey, D. F., Sass, T. R., Lockwood, J. R., & Mihaly, K. (2009). *The inter temporal variability of teacher effect estimates*. National Center on Performance Incentives.
- Meyer, R. H., & Christian, M. S. (2008). *Value added and other methods for measuring school performance*. National Center on Performance Incentives.
- Ray, A., Evans, H., & McCormack, T. (2008). "The use of national value added models for school improvement in English schools", *Revista de Education*, 348, 47-66.
- Sun, M., Penner, E.K., & Loeb, Susanna (2017). Resource- and approach-driven multidimensional change: three-year effects of school improvement grants. *American Educational Research Journal*, 54(4), 607-643.
- Stecher, B.M. et.al. (2018). *Improving teaching effectiveness: Final report: The intensive partnerships for effective teaching through 2015–2016*. Santa Monica, CA: RAND Corporation.