



---

## מדוע בדיקות אפקטיביות אינן אפקטיביות? ומה הפתרון?

---

מאת :

ד"ר נועם פישמן, מומחה במחקר יישומי ובהערכת תכניות  
ד"ר עינת נוטע-קורן, מנהלת המחלקה



## מטרות המסמך

בגופים ציבוריים שונים, כולל מוסדות להשכלה גבוהה, נעשות התערבויות רבות, אולם, קשה עד בלתי אפשרי לבדוק את האפקטיביות שלהן.

במסמך זה נעסוק בנושאים הבאים:

1. [האתגרים במצב הקיים והמצב הרצוי לתפיסתנו](#)
2. [גישות מחקר מקובלות לבדיקת אפקטיביות של התערבויות באמצעות מחקרים מבוקרים, והיתרונות של גישות אלו](#)
3. [דיון בנושא אתיקה הכרוך במחקרים מבוקרים](#)
4. [המלצות לקריטריונים למצבים שבהם מומלץ ליישם מחקר מבוקר](#)
5. [המלצות לקריאה נוספת](#)
6. [דוגמאות להמחשה](#)

## בדיקת אפקטיביות של פרויקטים

### המצב הקיים

בגופים ציבוריים שונים, ובכלל זה במוסדות להשכלה גבוהה, קיימים פרויקטים הנמצאים בשלבים שונים של בשלות מבחינת היישום. במרבית המקרים הפרויקטים יוצאים לפועל מבלי שנלקח בחשבון בשלב התכנון הצורך במחקר הערכה. כתוצאה מכך, לא אחת, הבקשה לבדוק את אפקטיביות הפרויקט או את האימפקט שלו מגיעה בשלב מתקדם יחסית של התפתחות הפרויקט. על פי רוב, בשלב זה מגלים שאין בסיס השוואה מתאים ולכן אי אפשר להעריך בצורה מספקת את מידת ההשפעה על מדדים קשיחים, כגון הצלחה והתמדה בלימודים.

הבעיה העיקרית שמחקרי הערכה מנסים להתמודד עימה היא **הטיית בחירה** (selection bias). בתחום ההשכלה הגבוהה, הטיה זו נובעת מכך שמקבלי ההתערבות במסגרת הפרויקטים הם בדרך כלל סטודנטים שפנו ביוזמתם לקבלת ההתערבות או שהשיבו לפנייה של הפרויקט, לעומת רבים שקיבלו את אותה פנייה ולא השיבו. אי אפשר להשוות את מקבלי ההתערבות (השירות) לאלו שלא קיבלו את ההתערבות משום שקבלת ההתערבות עלולה להיות כרוכה במאפיינים חשובים של סטודנטים כגון: מוטיבציה, מחויבות ופרואקטיביות - בצד החיובי, לעומת נזקקות, תלות ו/או מצוקה חברתית, מנגד<sup>1</sup>.

בעיה נוספת קיימת במצבים שבהם מוגדרת מראש קבוצה שבה רוצים ליישם את ההתערבות, ומשלבי היישום הראשונים פונים לכולה, ללא עריכת מחקר חלוץ על מדגם מצומצם. גם במקרה זה, קשה בדיעבד למצוא קבוצת השוואה טובה דיה, שכן על פי רוב ההשוואה לעבר אינה תקפה עקב שינויים נוספים שחלים ואינם קשורים להתערבות.

<sup>1</sup> מאפיינים אלו כוללים מאפיינים נצפים (או נמדדים) כגון: מין, נקודות זכות שנרכשו, ציון ממוצע וכו', וכן מאפיינים בלתי נצפים (או בלתי נמדדים) כגון: מוטיבציה, מחויבות ופרואקטיביות. שיטות סטטיסטיות מסוימות מנסות למזער את ההבדלים במאפיינים הנצפים בין אלו שקיבלו את ההתערבות לבין אלו שלא קיבלו אותה, אך הן אינן יכולות להקטין את ההבדלים במאפיינים הבלתי נצפים. רק מערך של ניסוי מבוקר יכול לנטרל את ההבדלים בין הקבוצות הן במאפיינים נצפים והן במאפיינים בלתי נצפים.

## המצב הרצוי

אנו מנסים לקדם מספר פתרונות אפשריים:

1. ניסוי מבוקר
2. מערך AB
3. שימוש במודל ניבוי<sup>2</sup>

במסמך זה נעסוק בשני הפתרונות הראשונים בלבד.

### ניסוי מבוקר: מה הוא? וכיצד מבצעים אותו?

קבלת ההתערבות בניסוי מבוקר נקבעת על ידי הגרלה אקראית. כל הפרטים באוכלוסיית היעד (למשל, סטודנטים) משתתפים בהגרלה כך שחלק מהם מוקצים לקבלת ההתערבות (קבוצת ההתערבות) וחלק מהם מוקצים לאי-קבלת ההתערבות (קבוצת הביקורת). באופן זה, קבוצת ההתערבות אמורה להיות דומה לקבוצת הביקורת בכל דבר - למעט בקבלת ההתערבות. אומדים את ההשפעה של ההתערבות באמצעות השוואה בין שתי הקבוצות במדדי התוצאה שייקבעו (למשל, הישגים/התמדה). פער לטובת קבוצת ההתערבות, יצביע על השפעה חיובית של ההתערבות.

ניסוי מבוקר נחשב ל"תקן הזהב" (gold standard, ראו איור 1) במערכי מחקר הערכה, ואחד היתרונות הבולטים שלו הוא פשטותו.

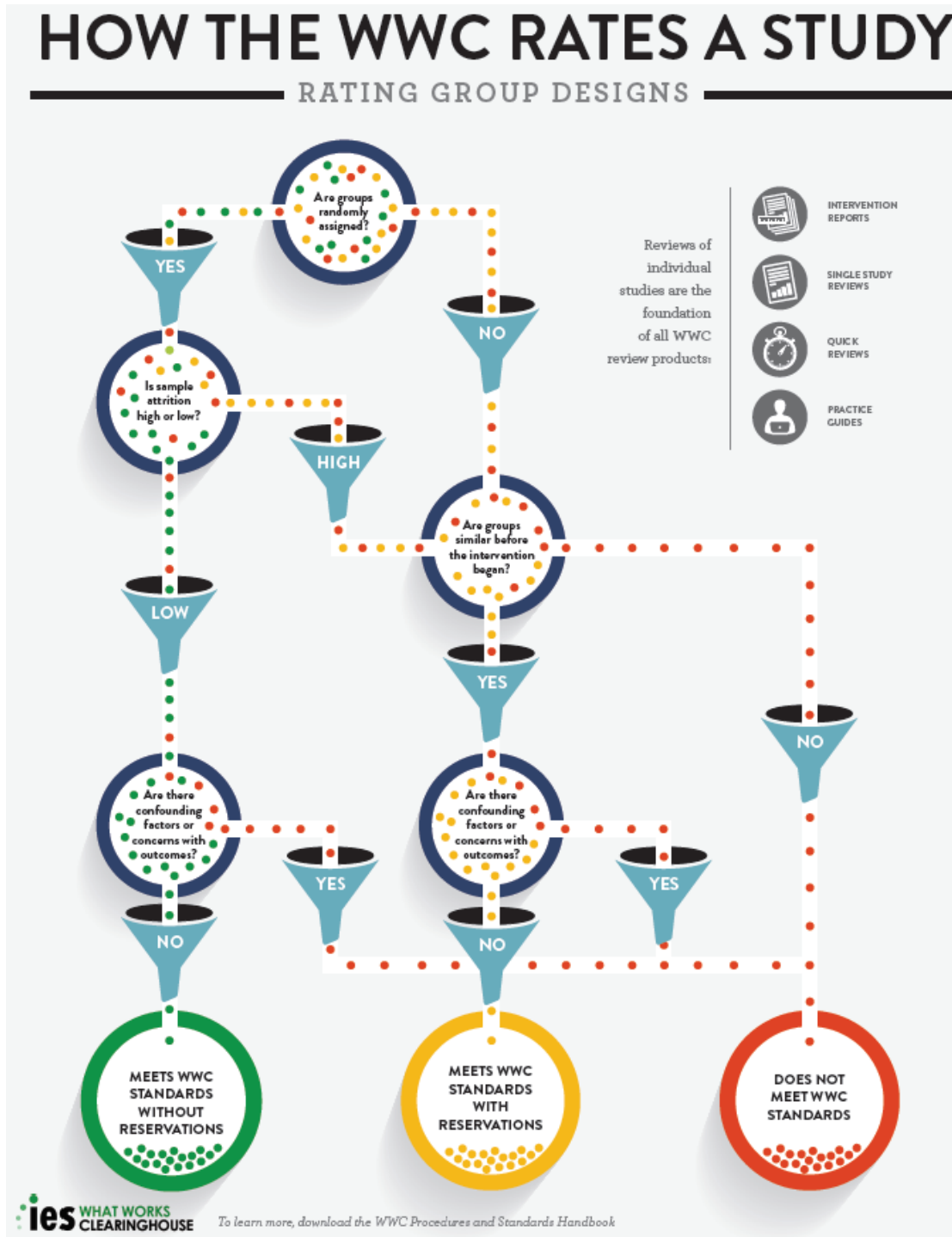
באיור סווגו מערכי מחקר שונים לאחת משלוש קטגוריות של איכות:

- 1) עמידה בסטנדרטים מחקרניים ללא הסתייגויות
- 2) עמידה בסטנדרטיים מחקרניים עם הסתייגויות
- 3) אי-עמידה בסטנדרטיים מחקרניים.

כפי שאפשר לראות, רק מערך הכולל הקצאה אקראית (כלומר ניסוי מבוקר) יכול לקבל את הסיווג האיכותי ביותר של עמידה בסטנדרטים ללא הסתייגויות.

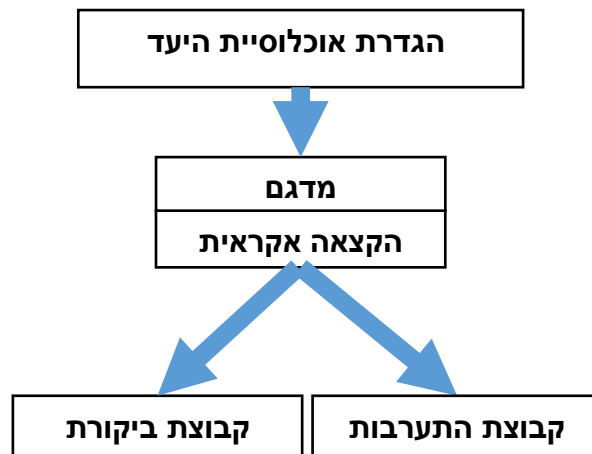
<sup>2</sup> מודל ניבוי מאפשר לזהות קבוצות עם מאפיינים מסוימים (למשל, קבוצות בסיכון לנשירה) ולמקד התערבויות ייעודיות להן. בהמשך, תוצרי המודל יכולים לשמש כלי נוסף לצד ניסוי מבוקר, לבדיקת האפקטיביות הלכה למעשה.

איור 1. תרשים עץ מה-What Works Clearinghouse של משרד החינוך בארה"ב



איור 2 מראה כיצד מבצעים ניסוי מבוקר. אפשר לראות שהמערך אינו מורכב מדי. זהו המערך העיקרי המשמש לצורך בדיקת יעילות של התערבויות רפואיות, כגון תרופות וטיפולים (ראו למשל את הדוח של Innovation Israel, [חדשנות בישראל: תמונת מצב 2018-2019, עמודים 66-69](#)). אפשר כמובן לערוך ניסוי מבוקר עם יותר משתי קבוצות - ניסוי שבו יש כמה קבוצות התערבות שונות וקבוצת ביקורת. בתחום החינוך וההשכלה, מחקרים מבוקרים עדיין נפוצים פחות. יחד עם זאת, התחום קיבל [תאוצה גדולה](#) בשנים האחרונות, כדרך להתמודד עם [הקושי להגיע למסקנות תקפות](#) על בסיס נתונים תיאוריים וקורלטיביים בלבד.

איור 2. איך מבצעים ניסוי מבוקר.



החיסרון העיקרי של ניסוי מבוקר עם שתי קבוצות הוא שמערך כזה אינו מאפשר הערכת אפקטיביות של רכיבים ספציפיים של ההתערבות. לדוגמה: אם כל הסטודנטים בקבוצת ההתערבות קיבלו שני סוגים של תת-התערבות, ואם נקבל תוצאה המעידה על השפעה של הפרויקט (חיובית או שלילית), לא נוכל לדעת אם ההשפעה נובעת משתי ההתערבויות גם יחד או שההשפעה נובעת רק מהתערבות אחת מתוך השתיים. מערך AB הוא מערך דומה המאפשר השוואה בין שני סוגי התערבות שונים. במקרים שבהם יש למעלה משתי התערבויות המערך יהיה מורכב אף יותר.

### מערך AB

במערך AB אין קבוצת ביקורת וזה ההבדל היחיד בינו לבין ניסוי מבוקר. ההשוואה במערך AB היא בין קבוצה A המקבלת סוג התערבות מסוים ובין קבוצה B המקבלת סוג התערבות אחר. כמו בניסוי מבוקר - ההקצאה לקבוצות מבוססת על הגרלה אקראית.

החיסרון העיקרי של מערך AB בא לידי ביטוי במקרים שבהם לא מתקבל הבדל בין שתי קבוצות ההתערבות השונות. במקרה כזה, אין דרך לדעת אם העדר ההבדל נובע מכך ששני סוגי ההתערבות היו אפקטיביים באותה מידה או ששניהם לא היו אפקטיביים כלל. כמו כן, אם שני סוגי ההתערבות היו אפקטיביים באותה מידה, לא נדע במערך AB מהי עוצמת האפקטיביות. לכן, אם מתאפשר, מוטב להוסיף קבוצת ביקורת כך שמערך ה-AB יהיה למעשה ניסוי מבוקר ובו שלוש קבוצות (או יותר), ובכך תיפתר בעיה זו. אולם, מערך הכולל קבוצות רבות דורש אוכלוסיית יעד גדולה יותר כדי שאפשר יהיה לזהות אפקטים מתונים.

### גודל אפקט מול גודל אוכלוסיית היעד

במחקרי הערכה יש קשר הדוק בין גודל האפקט הצפוי לבין גודלה הרצוי של אוכלוסיית היעד. ככל שהאפקט צפוי להיות קטן יותר, על מספר הפרטים באוכלוסיית היעד להיות גדול יותר כדי לאפשר זיהוי של האפקט. לדוגמה: אם אוכלוסיית היעד תכלול כאלף סטודנטים בכל סמסטר וכאלפי סטודנטים כעבור שנה, אפשר יהיה לזהות אפקטים בסדר גודל של כ-10 נקודות אחוז (או יותר) לטובת קבוצת ההתערבות אחרי סמסטר אחד, ושל כ-6 נקודות אחוז (או יותר) אחרי שני סמסטרים גדולים (סמסטר א וסמסטר ב). לכן, אם אנו בודקים את ההשפעה של פרויקט המיועד להגדיל את שיעור ההצלחה בקורסים של סטודנטים, ושיעור ההצלחה יעמוד על 56% בקבוצת ההתערבות לעומת 50% בקבוצת הביקורת, המחקר לא יוכל לזהות אפקט כזה כמובהק מבחינה סטטיסטית אחרי סמסטר אחד (כשיהיו אלף סטודנטים). אולם, אם אחרי שני סמסטרים יהיה אפקט זהה של 6 נקודות אחוז, אזי הוא יהיה מובהק מבחינה סטטיסטית. אם ההבדל הוא מובהק, אפשר לומר בביטחון גבוה למדי, שההבדל בתוצאה אינו מקרי אלא נוצר הודות להתערבות. לכן, כדי לאפשר זיהוי אפקטים מתונים לאחר שנת לימודים (או שני סמסטרים גדולים) רצוי שיהיו כ-500 (או יותר) סטודנטים בכל קבוצה בכל סמסטר של הפעלת הפרויקט. מובן שהוספת קבוצות התערבות למערך (כגון ביקורת + A + B ...) דורשת מספר גדול יותר של סטודנטים.

### **האם מניעת התערבות היא מהלך מוסרי-אתי?**

טענה רווחת - של מפעילי תכניות התערבות וקובעי מדיניות - נגד ביצוע ניסוי מבוקר היא שאין זה אתי, או מוסרי או הוגן למנוע מחברי קבוצת הביקורת את קבלת ההתערבות שאמורה לעזור להם. טענה זו מוצדקת רק במקרים שהאפקטיביות של ההתערבות מוכחת או ידועה מראש. ללא הוכחת האפקטיביות כל אחת מן האפשרויות הבאות יכולה להיות נכונה: יתכן שלהתערבות יש השפעה חיובית, יתכן שאין לה השפעה כלל ויתכן שיש לה השפעה שלילית.

לאור זאת, הקצאה אקראית לקבלה או לאי-קבלה של ההתערבות עשויה להיות הוגנת יותר מהמצב הקיים, שהוא לעתים תוצר של הטיה לטובת סטודנטים בעלי מאפיינים מסוימים (כגון סטודנטים



אסרטיביים יותר, או בעלי מוטיבציה גבוהה יותר וכו'). למעשה, אפשר לטעון שההקצאה האקראית מאפשרת חלוקה הוגנת של משאב קבלת ההתערבות – יותר משיטות אחרות כמו שיטת "כל הקודם זוכה" (ראו דיון דומה בנושא [מהבנק העולמי](#)). זאת ועוד: התערבות חדשה - שאינה טובה - המועברת ללא קבוצת ביקורת, עשויה להימשך ולהזיק לסטודנטים ולמשאבי האוניברסיטה, מבלי דרך לדעת זאת.

### באלו מקרים מומלץ לערוך ניסוי מבוקר?

מומלץ לערוך ניסוי מבוקר במקרים הבאים:

- ✓ במקרים שבהם הדבר מתאפשר מבחינה לוגיסטית, טכנית ואתית.
- ✓ במקרים שבהם מדובר בפרויקטים המיועדים להיות רחבי היקף.
- ✓ בפרויקטים שלצורך יישומם המלא יידרשו משאבים רבים (כוח אדם, פיתוח, תקציב).
- ✓ בפרויקטים שבהם המטרה ברורה (למשל שיפור שיעורי הצלחה, הפחתת נשירה וכו').

אין מומלץ לערוך ניסוי מבוקר במקרים הבאים:

- ההתערבות מיועדת למספר קטן של משתתפים ואין כוונה להרחיב את הפעילות.
- הפרויקט אינו בשל להערכה ועדיין נמצא בשלבי פיתוח מבחינת הגדרת אוכלוסיית יעד ומודל פעולה מתאימים.
- תוצאות בדיקת האפקטיביות אינן יכולות לגרום לשינוי בפרויקט המוערך - כולל פרויקטים בעלי חשיבות ערכית/חברתית שאינה תלויה באפקטיביות (למשל פרויקטים המיועדים לאוכלוסיות מיעוט).

בשני המקרים האחרונים, נמליץ לערוך מחקר הערכה מעצבת שיוכל לסייע לשיפור יישום הפרויקט.





## מקורות מומלצים לקריאה

[Behavioral Insights in Higher Education: KCLxBIT Project Report 2015-2017 published by The Behavioural Insights Team and King's College London.](#)

Connolly, P., Keenan, C. and Urbanska K. (2018) The trials of evidence-based practice in education: a systematic review of randomised controlled trials in education research 1980–2016, *Educational Research*, 60:3, 276-291, DOI: [10.1080/00131881.2018.1493353](#)

[Feldman, D.H. and Webber. D.A. \(2019\). Not all studies are created equal: A reader's guide. Inside Higher Education, June 25, 2019.](#)

Herodotou, C, Heiser, C. and Rienties, B. (2017) Implementing randomised control trials in open and distance learning: a feasibility study, *Open Learning: The Journal of Open, Distance and e-Learning*, 32:2, 147-62, DOI: [10.1080/02680513.2017.1316188](#)

[Leigh, Andrew, 2018, \*Randomistas. How Radical Researchers Changed our World.\* New Haven: Yale University Press.](#)

[Mckenzie, D. Ethical Validity Response #2: Is random assignment really that unacceptable or uncommon? World Bank Blogs.](#)

[Ravallion M. Should the randomistas \(continue to\) rule? \*Centre for Global Development Working Paper\*, 492; 2018.](#)

[Wikipedia Entry: Randomized Controlled Trial](#)

What Works Clearinghouse Evidence Snapshot: [Linked Learning Communities.](#)

What Works Clearinghouse Infographic: [How the WWC Rates A Study](#)

What Works Clearinghouse Report: [When Schools Stay Open Late: The National Evaluation of the 21<sup>st</sup> Century Community Learning Centers Program.](#)

[רשות החדשנות \(2019\). חדשנות בישראל: תמונת מצב 2018-2019](#)

## נספח – דוגמאות להמחשה

### 1. תכנית ASAP של המכללה הציבורית של העיר ניו-יורק

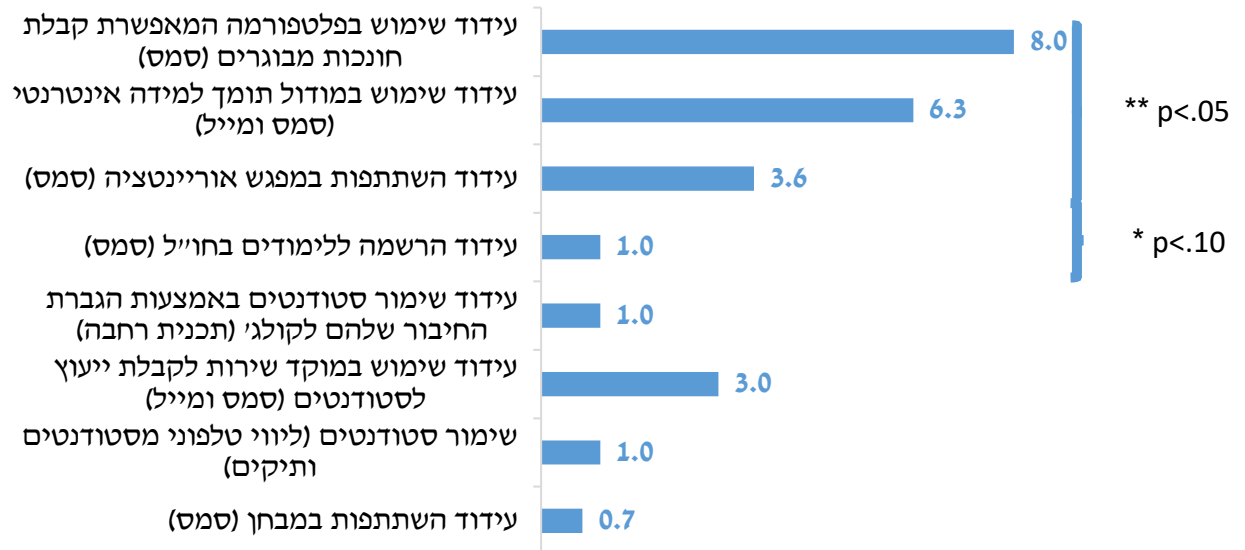
תכנית מקיפה של מכללות קהילתיות (community colleges) בעיר ניו-יורק נחשבת כאחת מהמצליחות ביותר בתחום מניעת נשירה מההשכלה הגבוהה. רק סטודנטים בהגדרה של full time, והחייבים לעבור קורס מכינה אקדמית יכולים להשתתף. חלקם הוקצו לקבוצת ההתערבות וחלקם הוקצו לקבוצת הביקורת שלמדה במתכונת רגילה. התכנית מספקת תמיכה רבה למשתתפיה ובכלל זה:

- ✓ תמיכה כספית: מלגות, ספרי לימוד ונסיעות
  - ✓ ייעוץ מקיף ואישי: מפגשים אישיים, חונכות, הכוון תעסוקתי
  - ✓ תהליך הרשמה מובנה: סדנה טרם ההרשמה והרשמה מוקדמת
  - ✓ שיטת co-remediation: למידה במכינה אקדמית (לדוגמה במתמטיקה) במקביל ללימודים בקורסים אקדמיים רגילים הדורשים מיומנויות מתמטיות, לעומת סטודנטים אחרים הלומדים במכינה לפני הקורסים האקדמיים הרגילים
- התערבות מקיפה זו הכפילה את שיעור ההתמדה של משתתפיה (לעומת קבוצת הביקורת), הן במדינת ניו-יורק והן בתכנית דומה במדינת אייווה.

### 2. הנאדג'ים לטובת הסטודנטים בקינגס קולג' לונדון

בקינגס קולג' בלונדון הוקמה יחידת What Works (מה עובד) במטרה לבדוק את אפקטיביות השירותים השונים הניתנים באוניברסיטה שמטרתם לשפר את השירותים ולהגביר את ההתמדה של סטודנטים בעלי סיכון לנשירה. ביוזמה חדשנית זו הם בדקו סדרה רחבה של התערבויות. מהתערבויות קטנות או "נאדג'ים" (כגון שליחת סמס לעידוד השתתפות במפגשי אוריינטציה) ועד לתכניות רחבות ושפתניות (כגון תכנית לעידוד שימור סטודנטים). נתון מעניין הוא שחצי מההתערבויות שנבדקו נמצאו אפקטיביות וחצי נמצאו לא אפקטיביות (איור 4). כל ההתערבויות נבדקו במערך ניסוי מבוקר, דבר שאיפשר לקינגס קולג' להרחיב את התערבויות שנמצאו כאפקטיביות ויעילות ולהסיר או לשנות את אלה שנמצאו לא אפקטיביות.

איור 4. ההשפעות של התערבויות שונות של קינגס קולג' בלונדון (בנקודות אחוז)



### 3. מיילים המיועדים לסייע בהתמודדות עם חסמים התנהגותיים של סטודנטים

המכללה הציבורית של מדינת ניו-יורק איתרה סטודנטים "בסיכון" (בעלי סיכויי נשירה גבוהים). אלו שהוקצו לקבוצת ההתערבות קיבלו הודעות באמצעות הדוא"ל - המיועדות למזער סדרה של התנהגויות לא רצויות כגון: אי-נוכחות במפגשי למידה, הכנה לא מספקת למבחנים ועבודות וחוסר ניצול שירותים של שיעורי עזר ללא תשלום.

הניסוחים בהודעות נבחרו על סמך שיטות מבוססות במדעי ההתנהגות:

- ✓ פישוט ופרסונליזציה של מיילים המזהירים את הסטודנטים שהם בסיכון להיכשל בקורס
- ✓ הדגשת "דדליינים": תזכורת שבועית שמזכירות את הדדליין לביצוע המשימה
- ✓ הכוונה: מייל המעודד את השימוש בשיעורי עזר חינוכיים לפני מבחנים
- ✓ תמיכה: מייל המדגיש שהסטודנט יכול להתמודד עם בעיות שעומדות לפניו
- ✓ הדגשת הפסד: למשל, "אתה משלם כ-280 ש"ח לכל מפגש למידה"
- ✓ השפעה חברתית: "סטודנטים שמצליחים מקדישים 20 שעות שבועיות בלמידה"

חבילה זו של התערבויות הביאה לירידה של 10% בנשירה מקורסים וגידול של 27% בקבלת חונכות בקבוצת ההתערבות.