

המבחן הפסיכומטרי החדש: בדיקת תוקף רבי-ממדית

דביר קלפר ונועה סקה

מחקר זה מציג תוצאות של ניתוחים שנועדו לאפיין את המבנה של המבחן הפסיכומטרי במתכונתו החדשה, שהונהגה בספטמבר 2012. כדי לבחון את המבנה הפנימי שלו בוצע ניתוח גורמים מאשש (confirmatory factor analysis) ברמת סוג פריט, בהתאם לחלוקה הפנימית של המבחן. ראשית נבדק מודל תאורטי פשוט לתיאור מבנה הגורמים, ולאור הממצאים הוצע מודל משופר. במסגרת ניתוח הגורמים המאשש הופעלו גם פרוצדורות סטטיסטיות שונות להערכת מהימנות ותוקף. התוצאות מראות התאמה טובה למודל הבנוי משלושה גורמים שהם תחומי המבחן הפסיכומטרי (חשיבה כמותית, חשיבה מילולית וידע באנגלית). המודל המשופר, שהצביע על קשרים נוספים בין הסולמות לבין הגורמים השונים, הראה התאמה טובה מעט יותר מזו של המודל הראשוני שהוצע.

מילות מפתח: מבחנים פסיכומטריים, תוקף מבחנים, ניתוח גורמים

זה שנים רבות נעשה שימוש במערכת החינוך במבחנים גורליים (high-stakes tests) סטנדרטיים – מבחנים שעל בסיסם מתקבלות החלטות גורליות בנוגע לנבחנים בהם, כגון קבלה או אי-קבלה למוסד להשכלה גבוהה (Amrein & Berliner, 2002). מבחנים

* ד"ר דביר קלפר, המרכז הארצי לבחינות ולהערכה, ירושלים.

דואר אלקטרוני: dvir@nite.org.il

ד"ר נועה סקה, בית הספר לחינוך, האוניברסיטה העברית בירושלים, והמרכז הארצי לבחינות ולהערכה, ירושלים.

דואר אלקטרוני: noa.saka@mail.huji.ac.il

אלו היו נפוצים בעיקר במערכת המיון של ההשכלה הגבוהה (כגון מבחני ה-SAT וה-ACT בארצות הברית והמבחן הפסיכומטרי בישראל), אך בעשורים האחרונים גבר השימוש בהם בעולם המערבי בכל תחומי החיים, ובעיקר לדירוג נבחנים ולהערכת יכולות.

מבחן גורלי מוגדר לרוב כמבחן בעל חשיבות גדולה, שתוצאותיו עשויות להשפיע במידה ניכרת על אדם או על גוף מסוים ולגרום לו רווח או הפסד (Zwick, 2007). המבחן נחשב גורלי יותר ככל שהתוצאות של ביצועי הנבחן מקושרות לסיכונים או לסיכויים רבים יותר. מבחנים אלו ממלאים מקום חשוב בתהליך הקבלה לאוניברסיטאות (כגון המבחן הפסיכומטרי בישראל), בהכשרה למקצוע (למשל מבחני לשכה לעריכת דין) ואף במיון לעבודה, לפיכך נודעת חשיבות רבה לבחינה מדוקדקת של מאפייניהם הפסיכומטריים, להבטחת איכותם וליכולתם לנבא בעיילות את התבחינים (קריטריונים) שהם נועדו לנבא.

מבחני המיון להשכלה הגבוהה משמשים כלי לחיזוי הצלחתו של המועמד בלימודים הגבוהים ולדירוג המועמדים, וכל מוסד אקדמי רשאי להחליט מה המשקל שהוא מייחס למבחן בתהליך המיון. יש חילוקי דעות באשר לצורך בקיומם, לתוקף ולמהימנות של המבחנים, למשקל הגבוה שניתן להם בתהליך הקבלה וללחץ המופעל על המועמדים להצליח בהם, ולכן הבטחת איכותם הפסיכומטרית של המבחנים ותיאור מאפייניהם הפסיכומטריים חשובה עוד יותר.

מחקרים קודמים רבים עסקו בסוגיות הנוגעות למהימנות ולתוקף המבחנים הן בארץ והן בעולם (לסקירה ראו Zwick, 2007). עם זאת, מיעוט ניכר של מחקרים עסקו בסוגיית המבנה הפנימי של מבחנים אלו ובשאלה מה הם הגורמים הפנימיים המרכיבים אותם. מחקרים שונים בחנו את המבנה הפנימי של פריטי המבחן או את סולמותיו כחלק מבדיקת תוקף רחבה יותר או מתוך השוואת התפקוד בין קבוצות שונות (למשל, Cahalan-Laitusis, Cook, & Aicher, 2004; Huynh & Barton, 2006; Steinberg, Cline, & Sawaki, 2011).

מספר קטן של מחקרים בחן את המבנה הפנימי של מבחנים גורליים למיון להשכלה הגבוהה כדי להשיב על השאלה אם מדובר במדד אחד של יכולת כללית או במדד המייצג תחומי דעת מובחנים (Koenig, Frey, & Detterman, 2008). ככלל, ממצאים של מחקרים קודמים היו חלוקים בסוגיה אם מבחנים מסוג זה בודקים למעשה גורם אחד של יכולת כללית או אם ניתן לזהות תת-תחומים וכישורים שאינם מודדים יכולת כללית בלבד ומושפעים גם מלמידה מסוימת לתחום. למשל, רוק ועמיתיו (Rock, Bennett, & Jirele, 1988) בחנו את המבנה הפנימי של מבחני ה-SAT וה-GRE, הבוחנים יכולות לימודיות לקראת כניסה לאוניברסיטאות ולמכללות בארצות הברית (להלן ארה"ב). הם מצאו כי מבחן ה-SAT הורכב משני גורמים שתאמו את שני תחומי התוכן שהיו מיוצגים באותה עת במבחן, המילולי והכמותי, ומבחן ה-GRE הורכב משלושה גורמי חשיבה – מילולית, כמותית ואנליטית. ואולם, מחקרים אחרים מצאו כי מבחני מיון מסוג זה

עומדים במתאם גבוה עם גורם אחד - גורם $1g$ - של יכולת כללית, הן עבור מבחן ה־SAT (Koenig et al., 2008) והן עבור מבחן ה־ACT (Frey & Detterman, 2004). לאור הרלוונטיות של השאלה גם בנוגע למבחן הפסיכומטרי, מטרתו של מחקר זה הייתה לבחון את המבנה הפנימי של המבחן ולבדוק אם אכן קיימים בו גורמים מובחנים, התואמים את התחומים האקדמיים השונים המיוצגים בו (מילולי, כמותי, ידע באנגלית). שאלה זו נקשרת לדיון הציבורי הנוגע לרלוונטיות של המבחן - האם נבדקים במבחן תחומי תוכן מוגדרים הנקשרים באופן תאורטי להצלחה אקדמית (למשל, כישורים מילוליים, יכולת הבעה בכתב, ידע באנגלית)? האם המבחן משקף יכולת כללית המתקשרת לתפקודים אקדמיים? האם כדאי להשתמש בציון הכללי, המורכב אולי מיכולות שונות ומובחנות זו מזו, ועל כן לא רלוונטיות במידה שווה למיון גורף לכל תחומי הלימוד, כפי שנעשה ברוב המוסדות (למשל, יכולות מילוליות לעומת כמותיות)? להבנת המבנה הפנימי של המבחן יכולות להיות השלכות בנוגע לשאלות אלו ולשאלות נוספות.

מערכת המיון לאוניברסיטאות בישראל

על מנת לקבל תעודת בגרות נדרש כל תלמיד תיכון להיבחן במבחני בגרות במתכונת ארצית, הניתנים על-ידי משרד החינוך פעמיים בשנה. התלמיד נבחן לרוב בסיום כיתה יא ובסיום כיתה יב בחלק גדול מן המקצועות המרכיבים את תעודת הבגרות. ציון המבחן מצורף לציון התיכון - הציון השנתי שהתלמיד מקבל מבית הספר בכל מקצוע שלמד וניגש בו לבחינת הבגרות, ושני הציונים מרכיבים יחד את ציון הבגרות במקצוע. צירוף הציונים בכל מקצועות הבגרות מהווה את ציון תעודת הבגרות של התלמיד.

ברוב המקרים המוסדות להשכלה גבוהה משקללים את ציון תעודת הבגרות ואת ציון המבחן הפסיכומטרי (לרוב במשקלות שווים) ומחשבים ציון חדש - ציון סכום. המועמד מתקבל למוסד האקדמי לפי ציון זה ולפי חתכים שונים של ציוני הקבלה בחוגים הלימוד השונים. אם כן, הציון הפסיכומטרי הוא בדרך כלל רכיב אחד משניים במערך המיון לאוניברסיטאות.

המבחן הפסיכומטרי בישראל

המבחן הפסיכומטרי הוא מבחן סטנדרטי המשמש כלי למיון מועמדים לתואר הראשון ולחיוזי סיכויי הצלחתם בלימודים במוסדות להשכלה גבוהה. המבחן מאפשר לדרג את

כל המועמדים בסולם הערכה אחיד, ולעומת אמצעי מיון אחרים הוא מושפע פחות מן הרקע של כל מועמד או ממשתנים סובייקטיביים אחרים. מחקרים רבים מראים כי כושר החיזוי של המבחן הפסיכומטרי טוב. כלומר, נבחנים שקיבלו ציונים גבוהים במבחן מצליחים לרוב בלימודי התואר הראשון יותר מנבחנים שקיבלו ציונים נמוכים. כמו כן, נמצא כי מכל הצירופים השונים של אמצעי המיון הקיימים, הצירוף של המבחן הפסיכומטרי ובחינות הבגרות הוא בעל כושר החיזוי הטוב ביותר (ראו למשל, בייט־מרום ועמיתיה, 1998; קנת־כהן, אורן, טורוול וכהן, 2013). עם זאת, המבחן הפסיכומטרי אינו כלי מיון מושלם. לרוב הוא אמנם מנבא היטב הצלחה בלימודים, אך ייתכן מיעוט מקרים שבהם מועמדים יצליחו בלימודיהם על אף הישגיהם הנמוכים בו ולהפך. לעתים מועלות טענות נגד קיומו של המבחן, הנוגעות לתוקפו הנראה, למידת תרומתו מעבר לבחינות הבגרות ולתחומי לימוד שונים ולאפליה אפשרית של נבחנים ממעמד חברתי־כלכלי נמוך ושל נבחנים מקבוצות תרבותיות שונות. כמו כן, המבחן אינו מודד במישרין גורמים כגון מוטיבציה, יצירתיות או התמדה, היכולים בהחלט להיות קשורים להצלחה בלימודים (בייט־מרום ועמיתיה, 1998). עם זאת, יש לציין כי מקצת התכונות האלה נבדקות בעקיפין הן במבחן הפסיכומטרי והן במבחני הבגרות.

לעומת זאת, המבחן הפסיכומטרי מעניק הזדמנות שנייה לנבחנים בעלי פוטנציאל גבוה שמסיכות שונות לא ניתנה להם הזדמנות לבטא את מלוא יכולתם בלימודים בבית הספר התיכון. כמו כן, בשל תרגומו לכמה שפות הוא מאפשר להחיל סולם מיון אחיד גם על מועמדים ששפת אמם אינה עברית או שאינם בעלי תעודת בגרות ישראלית. לנוכח מאפיינים אלו של המבחן, מחקר שוטף בנוגע לתוקפו ובנוגע למאפייניו הפסיכומטריים הוא הכרחי על מנת לתמוך בשימוש שנעשה בו במערכת ההשכלה הגבוהה. שני מחקרים בלבד בחנו את המבנה הפנימי של המבחן הפסיכומטרי, ושניהם עסקו במתכונות ישנות של המבחן. בודסקו (1985) בדק את מבנה המבחן הפסיכומטרי במתכונתו הראשונה, כאשר כלל חמישה תחומים: צורות, מתמטיקה, הבנה מילולית, אנגלית וידע כללי. במחקרו נמצאו שני גורמים: גורם של ידע, שכלל את תחומי הידע הכללי והאנגלית, וגורם של חשיבה כמותית, שכלל את התחום הכמותי והצורות, כאשר התחום המילולי נטען על שניהם. מחקר נוסף על אותה גרסה נעשה בידי בלר (Beller, 1990) וכלל שלוש שיטות של ניתוח מבנה: מודל עצים חיבורי – ADDTREE (Sattath & Tversky, 1977), ניתוח המרחב הקטן ביותר – SSA (Smallest Space Analysis), גם במחקר זה (Guttman, 1968) וניתוח גורמים מגשש (exploratory factor analysis). גם במחקר זה נמצאו שני גורמים מרכזיים: ידע ופתרון בעיות. מאז לא נעשו בארץ מחקרים שבחנו את המבנה הפנימי של המבחן, אף שנערכו בו כמה שינויים. באוקטובר 1990 שונתה גרסת המבחן, ומאז הוא כולל שלושה פרקים בלבד – חשיבה מילולית, חשיבה כמותית ואנגלית, ולאחרונה שונתה מתכונתו שוב, כפי שיפורט בהמשך.

המבחן הפסיכומטרי במתכונתו החדשה

לאחרונה נערכו שינויים של ממש במבנה המבחן הפסיכומטרי. מספטמבר 2012 הוא כולל חלק נוסף – מטלת כתיבה – שהוא 25% מן הציון בחלק המילולי, ובפרקים הסגורים חלו שינויים אחדים. בפרקי החשיבה המילולית הושמטו פריטים של אוצר המילים, הושמטו פריטים המבוססים על התאמת שורשים, והוזק החלק של הבנת הנקרא על-ידי הוספת פריטים המבוססים על הבנת הנקרא בטקסטים קצרים של משפט עד שניים על חשבון פריטים של היגיון פורמלי (הסקת מסקנות המבוססת על טיעונים ונתונים). בפרקי החשיבה הכמותית הושמטו פריטים המבוססים על השוואה כמותית. שינויים אלו נערכו על מנת לתמוך בתוקף הנראה של המבחן, לצמצם מטלות שאינן תורמות תרומה של ממש לניבוי לעומת מטלות אחרות, או מטלות שאינן ממטרות המבחן (למשל, שינון אוצר מילים). כמו כן, הוצע שינוי אפשרי בפרקטיקת חישוב הציונים, ולפיו לכל נבחן יחושב גם ציון בדגש מילולי וגם ציון בדגש כמותי, נוסף על הציון הפסיכומטרי הכללי. לאור שינויים אלו התעורר הצורך לחקור מחדש את המבנה הפנימי של המבחן במתכונתו החדשה כחלק מפעולות התיקוף שלו.

מטרת המחקר

מטרת המחקר הייתה לאפיין את המבנה של המבחן הפסיכומטרי במתכונתו החדשה, הנהוגה מספטמבר 2012, ולבחון היבטים שונים של תוקפו הפנימי. מעבר לחקירה מחודשת של מבנה המבחן הפסיכומטרי התמקד אפוא המחקר בבחינת השפעתם של השינויים, ובייחוד בהשפעתה של הוספת מטלת הכתיבה על המבנה הפנימי של היכולות הנמדדות במבחן.

שיטת המחקר

שיטות שונות לבדיקת מבנה פנימי

קיימות שיטות שונות לבדיקת מבנה פנימי של פריטי מבחנים וסולמותיהם. רבות מן השיטות מבוססות על ניתוח מטריצת מתאמים בין המשתנים ועל הרכבת מדדים כלליים מצירופיהם בעזרת שיטות שונות, כגון ניתוח אשכולות או ניתוח גורמים מגשש (Beller, 1990). בשנים האחרונות פותחו מודלים חדשים המאפשרים לבדוק השערות אפריוריות על מבנה פנימי של סולמות וגורמים בעזרת ניתוח גורמים מאשש (confirmatory factor analysis) (Bentler, 1990, 1992).

מחקרים קודמים שנעשו על מבנה המבחן הפסיכומטרי השתמשו בשיטות מגששות לתיאור המבנה הפנימי (בודסקו, 1985; Beller, 1990). לעומת זאת, במחקר זה

השתמשנו בשיטת ניתוח הגורמים המאשש כדי לבחון את תוקפו של המבנה האמפירי של המבחן, כלומר כדי לבחון אם ציוני הנבחנים במבחן יתקבצו לפי שלושה גורמים על-פי המבנה התאורטי של שלושה תחומי תוכן מובחנים. שיטה זו מאפשרת להפעיל גם פרוצדורות סטטיסטיות נוספות על מנת לבדוק את התאמת המבנה התאורטי המוצע למבנה האמפירי המתקבל מן הנתונים. הפרוצדורות שהופעלו, בין השאר, היו בדיקת מתאמים בין חלקי המבחן, הערכת מהימנות, בדיקת תוקף מתכנס ותוקף מבחין ואומדן לשיעור השונות המוסברת על-פי כל גורם מתוך שונות המדידה. חלק ממדדים אלו הושו לאלו המתקבלים ללא המודל המאשש, ובנוגע לחלקם בוצעו תיקונים סטטיסטיים על מנת לקבל התאמה טובה יותר לנתונים, כפי שיפורט בהמשך.

בשלב הראשון נבדקו ציוני הנבחנים לפי מודל סטנדרטי, שכלל שלושה גורמים משוערים – חשיבה מילולית, חשיבה כמותית וידע באנגלית – כדי לבדוק את התאמתו של המודל לנתונים. בשלב שני שופר המודל בהתאם למדדי טיב ההתאמה, ולבסוף הוצע מודל לא סטנדרטי מתוקן, שתאם את הנתונים בדיוק רב יותר והצביע על קשרים חשובים נוספים בין חלק מתת-התחומים הנבדקים במבחן.

המבחן במחקר זה והנבחנים

לצורך ניתוח הגורמים בחרנו את המבחן הפסיכומטרי שהתקיים במועד דצמבר 2012 בשפה העברית. בנוסף זה נבחנו 3,704 אנשים, 38% מהם גברים ו-62% נשים, רובם צעירים ($M = 21.8$, $SD = 2.39$, $Md = 21.8$). כ-90% מן הנבחנים במבחן זו היו ילידי ישראל. כ-30.8% מהם דיווחו כי מצבם החברתי-כלכלי הוא הרבה או הרבה מאוד מעל הממוצע, 41.8% דיווחו כי מצבם מעט מעל הממוצע, 16.5% דיווחו כי מצבם מעט מתחת לממוצע, 10.4% דיווחו כי מצבם הרבה או הרבה מאוד מתחת לממוצע. נתונים אלו מייצגים באופן הולם את אוכלוסיית הנבחנים במועדי דצמבר ב-10 השנים האחרונות.

הליך המחקר

הנתונים נלקחו מתוך מאגר הנתונים של המרכז הארצי לבחינות ולהערכה, שבו נשמרים נתוני הנבחנים. הנתונים, בלא פרטים המזהים את הנבחנים, הועברו לקובץ נפרד לצורך עיבודם. המחקר אושר על-ידי המרכז הארצי לבחינות ולהערכה, ובעריכתו, באיסוף הנתונים ובניתוחם נשמרו כל הכללים האתיים.

כלים: מבנה המבחן ורכיביו

במבחן הפסיכומטרי תשעה פרקים, וכל פרק שייך לאחד משלושה תחומים: חשיבה כמותית, חשיבה מילולית או אנגלית. הפרק הראשון במבחן הוא מטלת כתיבה בתחום

החשיבה המילולית. שמונת הפרקים האחרים מורכבים משאלות בִּרְהָה מרובה, שבהן על הנבחן לבחור את התשובה הנכונה מתוך ארבע אפשרויות. פרקים אלו יכוננו בהמשך "פרקי בִּרְהָה", והם אינם מופיעים במבחן בסדר קבוע. בראש כל פרק מצוינים הזמן המוקצב לפתורנו ומספר השאלות בו.

בפרקי הבררה של כל אחד משלושת התחומים מופיעות שאלות מכמה סוגים. שאלות מאותו סוג מקובצות יחד ומסודרות ברמת קושי עולה, למעט שאלות בהבנת הנקרא (בפרקי החשיבה המילולית ובפרקי האנגלית), המסודרות לפי סדר הופעת הנושאים בקטע הקריאה. בשלושת התחומים נבדקות יכולות הדרושות באופן תאורטי להצלחה בלימודים אקדמיים.

בתחום החשיבה המילולית נבדקים כישורים מילוליים הבאים לידי ביטוי בלימודים אקדמיים: עושר לשוני, כישורים לוגיים, יכולת להבין קטעי קריאה מורכבים ולנתחם, יכולת לחשוב בבהירות ובשיטתיות ויכולת לגבש רעיון ולהביעו בכתב באופן מבוסס ומדויק.

בחלק זה שלושה פרקים: פרק אחד של מטלת כתיבה ושני פרקים של 23 שאלות בררה כל אחד, ובסך הכול 46 שאלות בררה ומטלת כתיבה אחת.

בתחום החשיבה הכמותית נבדקות היכולת להשתמש במספרים ובמונחים מתמטיים כדי לפתור בעיות כמותיות והיכולת לנתח נתונים המוצגים בצורות שונות, כגון טבלאות או תרשימים. בחלק זה שני פרקים של 20 שאלות כל אחד, ובסך הכול 40 שאלות. בתחום האנגלית נבדקת השליטה בשפה האנגלית הבאה לידי ביטוי, בין השאר, באוצר המילים וביכולת לקרוא קטעים ברמה אקדמית ולהבינם. בחלק זה שני פרקים של 22 שאלות כל אחד, ובסך הכול 44 שאלות.

בסך הכול המבחן הפסיכומטרי כולל 130 שאלות בררה ומטלת כתיבה אחת. הבטחת איכותן של השאלות: לפני הכללת השאלות בפרק המשמש לקביעת הציון יש לבחון את איכותן ולהבטיח שהן הוגנות ומבחינות בין נבחנים שיכולתם גבוהה לבין נבחנים שיכולתם נמוכה. לפיכך, שני פרקים מתוך התשעה מורכבים משאלות המצויות בתהליך של בדיקת איכות. פרקים אלו אינם תפעוליים, כלומר אינם משמשים לחישוב הציון. שאלות שיצליחו לעמוד בבדיקת האיכות באמות מידה סטטיסטיות ואחרות עשויות להופיע בעתיד בפרקים המשמשים בקביעת הציון, והשאלות האחרות, שלא יעמדו באמות המידה הנדרשות, לא ייכללו בעתיד בבחינות. כל הפרקים המשמשים בחישוב הציונים מורכבים משאלות שנבדקו בבדיקת איכות כזאת. לאור זאת, בניתוח הגורמים שנעשה לצורך המחקר נותחו רק ששת פרקי הבררה התפעוליים.

המשתנים

משתני המחקר הוגדרו לפי הפרקים ותת־הפרקים במבחן, וערכיהם מייצגים את ציוני הנבחנים המחושבים לכל פרק או תת־פרק. על־פי החלוקה הפנימית המופיעה בחוברת

המבחן, כל אחד מן התחומים במבחן מחולק לשניים או לשלושה פרקים. כאמור, בכל אחד מתחומי התוכן במבחן שניים מן הפרקים הם פרקי בררה תפעוליים. לשלושת הפרקים של תחום החשיבה המילולית נוסף פרק נפרד (פרק 1) של מטלת כתיבה. נוסף על כך, הפרק של ההבנה וההסקה בתחום החשיבה המילולית נחלק עוד לשלושה תת־פרקים: השלמת משפטים, היגיון ופריט יחיד של הבנת הנקרא (פריט יחיד המבוסס על קטע קריאה קצר). בסך הכול התקבלו 11 סוגי פריטים (תת־פרקים), שהוגדרו כמשתני המחקר, ועליהם בוצע ניתוח הגורמים. ערכיהם של 11 המשתנים חושבו כמדדים על בסיס התשובות לכל הפריטים שנכללו בכל תת־פרק במבחן, בהתאם לחישוב המקובל במרכז הארצי לבחינות ולהערכה. החלוקה הפנימית של המבחן מוצגת בתרשים 1.

הציון הפסיכומטרי

אנגלית	חשיבה מילולית	חשיבה כמותית
--------	---------------	--------------

הבנת הנקרא	ניסוח מחדש	השלמת משפטים	הבנת הנקרא	הבנה והסקה	אנלוגיות	מטלה כתיבה	הסקה מתרשים	שאלות ובעיות
------------	------------	--------------	------------	------------	----------	------------	-------------	--------------

הבנת הנקרא בדיד	היגיון	השלמת משפטים
-----------------	--------	--------------

תרשים 1: מבנה המבחן הפסיכומטרי

ממצאים

בלוח 1 מוצגת סטטיסטיקה תיאורית של 11 משתני המחקר (סוגי הפריטים) – ממוצעי הציונים, סטיות תקן, מקדם מהימנות בשיטת קרונבך (Cronbach, 1951), מספר השאלות בכל משתנה ורמת הקושי.
 מטלת הכתיבה כוללת פריט יחיד, ולפיכך מהימנותה נקבעה לפי אומדן ($\alpha = .50$) שהתבסס על אומדני ביצוע של נבחנים חוזרים בבחינות שנערכו בארה"ב (Brelmad, 2004). (Kubota, Nickerson, Trapani, & Walker, 2004).

לוח 1: מאפיינים של משתני המחקר: ממוצע, סטיית תקן, מקדם מהימנות, מספר השאלות ורמת הקושי

משתנה	מספר שאלות	רמת קושי	M	SD	α
שאלות ובעיות	32	0.65	20.70	5.73	.84
הסקה מתרשים	8	0.59	4.68	1.84	.55
מטלת כתיבה	24	0.60	14.35	4.08	.50
אנלוגיות	12	0.65	7.81	2.28	.60
השלמת משפטים בעברית	6	0.60	3.60	1.44	.46
היגיון	10	0.69	6.86	1.84	.55
הבנת הנקרא בדיד	6	0.70	4.21	1.43	.50
הבנת הנקרא	12	0.62	7.41	2.91	.77
השלמת משפטים באנגלית	16	0.75	12.06	3.66	.85
ניסוח מחדש באנגלית	8	0.71	5.66	1.90	.67
הבנת הנקרא באנגלית	20	0.66	13.29	4.22	.82

קשרים בין חלקי המבחן

בלוח 2 מוצגים מתאמי פירסון בין המשתנים המייצגים את תת-הפרקים הכוללים במבחן, ובלוח 3 מוצגים המתאמים בין שלושת תחומי המבחן והציון הכולל במבחן הפסיכומטרי.

לוח 3: מתאמי פירסון בין תחומי המבחן

ציון פסיכומטרי כולל	אנגלית	מילולי	כמותי	
			--	כמותי
		--	0.67	מילולי
	--	0.65	0.55	אנגלית
--	0.87	0.89	0.83	ציון פסיכומטרי כולל

המתאמים בין ציוני התחומים השונים לבין עצמם נמצאים בטווח הבינוני-גבוה, וממצא זה משקף את הקשרים המשמעותיים ביניהם. כמו כן, המתאמים בין כל אחד מן התחומים לבין הציון הכללי הם גבוהים, באופן המשקף את תרומתו של כל אחד מן התחומים לציון הכללי המחושב. מובן שיש להביא בחשבון את העובדה כי הציון הפסיכומטרי הכללי מורכב בחלקו גם מציוני התחומים, והדבר תורם למתאמים הגבוהים.

מודל תאורטי תלת-גורמי סטנדרטי

המודל המוצע לבדיקה הוא מודל תלת-גורמי התואם את החלוקה הפנימית של המבחן לתחומים ולפרקים, כפי שהוצג בתרשים 1. כלומר, המודל כולל שלושה גורמים המייצגים משתנים חבויים (latent exogenous variables) לפי תחומי החשיבה הנמדדים במבחן – חשיבה כמותית, חשיבה מילולית ואנגלית – המחולקים ל-11 מדדים גלויים (manifest endogenous indicators) של סוגי הפריטים (תת-פרקים) הנמדדים במבחן ומוסברים כל אחד על-פי התחום שהוא משתייך אליו.

מודל תאורטי תלת-גורמי סטנדרטי: מדדים לטיב ההתאמה

בניתוח גורמים מאשש נהוג להשתמש בכמה מבחנים סטטיסטיים כדי לקבוע באיזו מידה המודל מתאים לנתונים (Shur, 2006). האמירה כי למודל יש התאמה טובה לנתונים מצביעה על היותו קביל, כלומר מתקבל על הדעת (Schermelleh-Engel, Moosbrugger, & Müller, 2003) ומייצג באופן סביר את הנתונים, אך היא אינה קובעת בהכרח שהמודל נכון. סיכום של מדדי טיב ההתאמה ושל משמעותם מצוי אצל קליין (Kline, 2010) ואצל האטשר (Hatcher, 1994).

ערכיהם של ארבעה מחמשת המדדים לטיב ההתאמה של המודל התאורטי הסטנדרטי מראים התאמה טובה מאוד לנתונים, $RMSEA = .0551$, $GFI = .9763$, $NFI = .9756$, $NNFI = .9695$, $CFI = .9773$. מדד χ^2 הוא היחיד שאינו מצביע על התאמה של המודל לנתונים בשל היותו מובהק סטטיסטית, $\chi^2(41) = 502$, $pv < .0001$. נציין כי במדגמים גדולים כמו המדגם במחקר זה, מדד χ^2 עשוי להיות מובהק סטטיסטית גם אם המודל מראה התאמה טובה (James, Mulaik, & Brett, 1982), ולכן מקובל להשתמש בערך χ^2

מחולק בדרגות החופש שלו. עם זאת, גם ערך זה שהתקבל בניתוח, $\chi^2(41) / 41 = 12.24$, אינו מראה התאמה טובה בשל היותו גבוה מן הרף המקובל של 2 (ראו, למשל, Ullman, 2001). אם כן, בהמשך נציע מודל לא סטנדרטי, ונראה שערך זה ישתפר במידה ניכרת.

המודל האמפירי

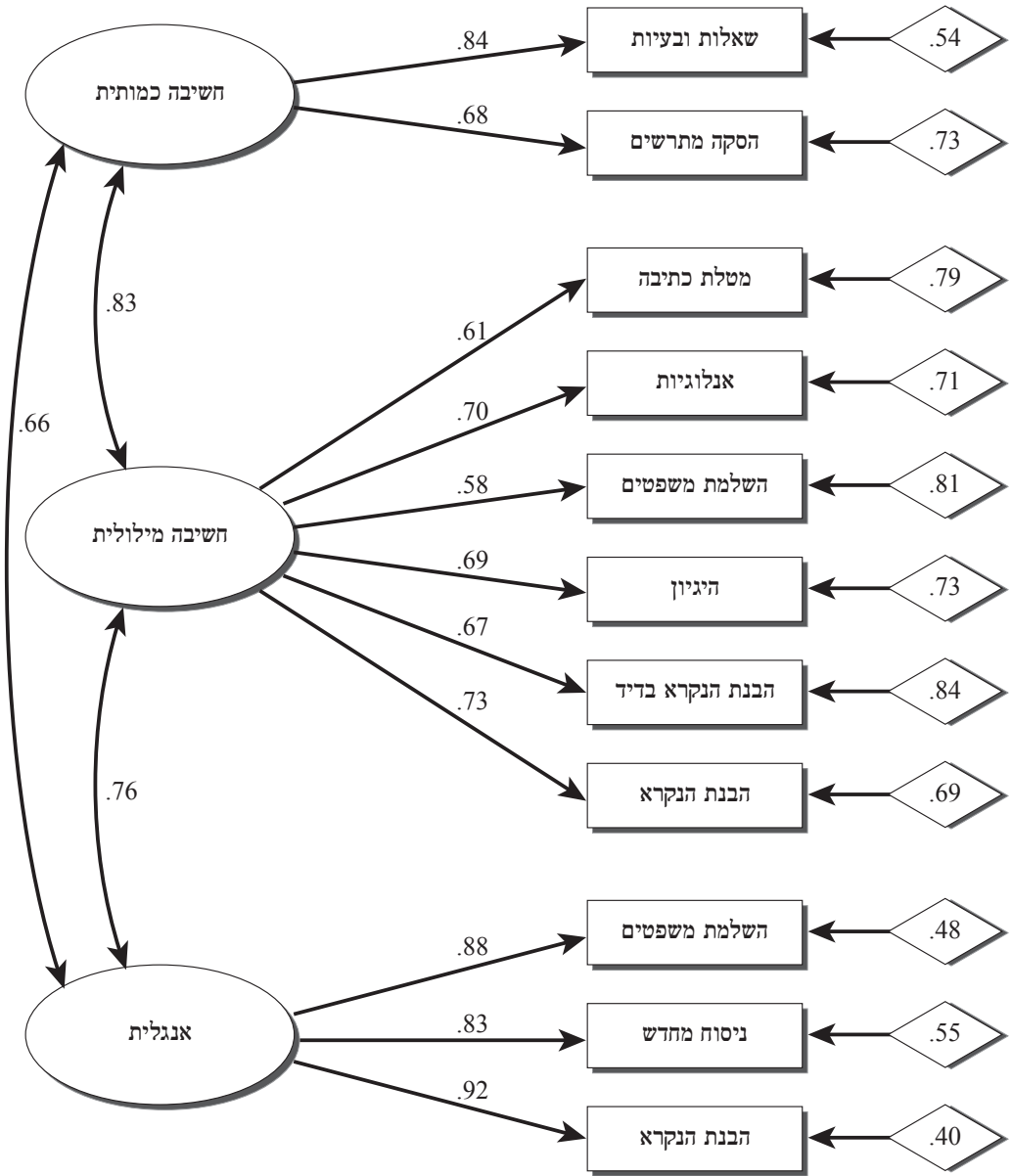
תרשים 2 מציג את המודל האמפירי שהתקבל מניתוח גורמים של ציוני הנבחנים, לרבות טעינות מתוקנת של כל אחד מן המדדים על הגורם שהוא שויך אליו באופן תאורטי. לתרשים זה הוספנו את ערכי הטעינות של טעויות המדידה ואת המתאמים בין שלושת הגורמים. כמקובל בדיווח על ניתוח גורמים מאשש, מלבנים מסמלים את המשתנים הגלויים – מדדים המחושבים מכמה פריטים המרכיבים פרק או תת־פרק, אליפסות מסמלות את המשתנים החבויים (הגורמים), ומעוינים מסמלים את טעויות המדידה.

המודל האמפירי: בדיקת המהימנות

מהימנות אינדיקטור: מהימנות אינדיקטור של משתנה היא ריבוע המתאם בין המשתנה החבוי (הגורם) לבין המדד המחושב מערכי פריטיו, והיא מציינת את שיעור השונות במדד המוסברת על־ידי הגורם שהוא אמור למדוד (Long, 1983).

מהימנות מורכבת בניתוח גורמים מאשש עבור כל משתנה חבוי (גורם) במודל היא אנלוגית למקדם המהימנות לפי קרונברך, והיא משקפת את העקיבות הפנימית של המדד המודד את הגורם.

אומדן לשונות המחולצת: מדד אומדן השונות המחולצת מעריך את שיעור השונות המוסברת על־ידי גורם מסוים מתוך סך כל שונות המדידה (Fornell & Larcker, 1981).



תרשים 2: המודל התאורטי המוצע: תוצאות ניתוח גורמים מאשש

בלוח 4 מוצגים שלושת מדדי המהימנות לכל משתנה ולכל תחום:

לוח 4: המודל האמפירי: טעינות מתוקננת, מהימנות אינדיקטור, מהימנות תחום ואומדן לשונות מחולצת

אומדן לשונות מחולצת	מהימנות תחום	מהימנות אינדיקטור*	טעינות מתוקננת	משתנה	תחום
.58	.74	.71	.84	שאלות ובעיות	חשיבה כמותית
		.46	.68	הסקה מתרשים	
.44	.83	.37	.61	מטלת כתיבה	חשיבה מילולית
		.50	.70	אנלוגיות	
		.34	.58	השלמת משפטים בעברית	
		.47	.69	היגיון	
		.45	.67	הבנת הנקרא בדיד	
		.53	.73	הבנת הנקרא	
.77	.91	.77	.88	השלמת משפטים באנגלית	אנגלית
		.69	.83	ניסוח מחדש באנגלית	
		.84	.92	הבנת הנקרא באנגלית	

* מהימנות אינדיקטור מחושבת כריבוע הטעינות המתוקננת של המדד על הגורם, כלומר ריבוע המתאם בין הגורם לבין המדד המחושב מערכי פריטיו.

פורנל ולרקר (Fornell & Larcker, 1981) קבעו סף של 0.5 בשונות המחולצת כערך סביר (כך שיהיה גדול יותר מן השונות הנובעת מטעות). מלוח 4 עולה כי תחום החשיבה הכמותית והידע באנגלית עומדים בסף זה, ואילו תחום החשיבה המילולית אינו עומד בו.

המודל האמפירי: בדיקת התוקף

תוקף מתכנס (convergent validity) ותוקף מבהין (discriminant validity) קשורים בדרך כלל בניתוח בשיטת "רב-תכונה רב-שיטה" (Multi-Trait Multi-Method) MTMM, שכמה מבנים (תכונות) בה מוערכים ביותר משיטה אחת (Campbell & Fiske, 1959). יש הטוענים כי שיטת MTMM מציגה הוכחות חזקות יותר לתוקף המתכנס והמבהין

מן השיטה שהצענו כאן (למשל, Schmitt & Netemeyer, Johnston, & Burton, 1990; Stults, 1986; Widaman, 1985). למרות זאת, הפרוצדורה המוצעת כאן שופכת אור על המבנה הפנימי של המבחן, והיא שימושית כאשר לא ניתן להשתמש בשיטת ה-MTMM. תוקף מתכנס: תוקף מתכנס מודגם כאשר כלים שונים המשמשים כדי למדוד מבנה מסוים, מראים תוצאות מתואמות ביניהם. במחקר זה הוערך התוקף המתכנס על-פי מובהקות הטעינות על הגורמים. טעינות מובהקות סטטיסטית של כל אחד מן המדדים המודדים מבנה אחד היא אינדיקציה תומכת לתוקף המתכנס של אותם מדדים (Anderson & Gerbing, 1988). כל הטעינויות במחקר זה נמצאו מובהקות סטטיסטית, וממצאים אלו תומכים, כפי שנאמר, בתוקף המתכנס של המבנה המוצע. תוקף מבחין: תוקף מבחין מודגם כאשר כלים שונים המשמשים למדידת גורמים שונים מראים מתאמים חלשים יחסית ביניהם. מבחן שמראה תוקף מבחין מוכיח שהוא אינו מודד מבנה שהוא לא תוכנן למדוד.

במחקר זה נעשה שימוש במבחן "הפרשי χ^2 " כדי לבדוק את התוקף המבחין בין שלושת הגורמים המוצעים (Anderson & Gerbing, 1988; Bagozzi & Phillips, 1982). במבחן זה משווים את ערך ה- χ^2 של המודל המקורי למודלים חלופיים, שבכל אחד מהם מאלצים את המתאם בין שני גורמים פנימיים (בכל פעם שניים אחרים) להיות 1. אם ערך ה- χ^2 של המודלים החלופיים גבוה יותר בצורה מובהקת מזה של המודל המקורי, הדבר מעיד על כך שהגורמים הפנימיים אינם זהים (כלומר, המתאם ביניהם אינו 1), ולכן מתקבלת תמיכה בתוקף המבחין. בעמודה השנייה בלוח 5 מוצג ערך ה- χ^2 המקורי וכן ערכי ה- χ^2 של שלושה מודלים חלופיים, והמתאם בין זוג אחד של תחומים נקבע בכל אחד מהם להיות 1. בעמודה השלישית מוצגים ההפרשים בין מדדי χ^2 עבור שלושה מודלים חלופיים בהשוואה לערך מבחן χ^2 במודל המקורי לצורך בדיקת התוקף המבחין.

לוח 5: מודל לבדיקת תוקף מבחין

$\Delta\chi^2(\Delta df)$	$\chi^2(df)$	המודל הנבדק
	502*** (41)	המודל המקורי
273*** (1)	775*** (42)	1 = (מילולי, כמותי) r
746*** (1)	1248*** (42)	1 = (אנגלית, כמותי) r
1838*** (1)	2340*** (42)	1 = (אנגלית, מילולי) r

*** p < .001

כל ההפרשים המוצגים בלוח 5 מובהקים סטטיסטית, ולפיכך הם תומכים בתוקף המבחין של שלושת תחומי התוכן במבחן.

מודל מותאם לא סטנדרטי

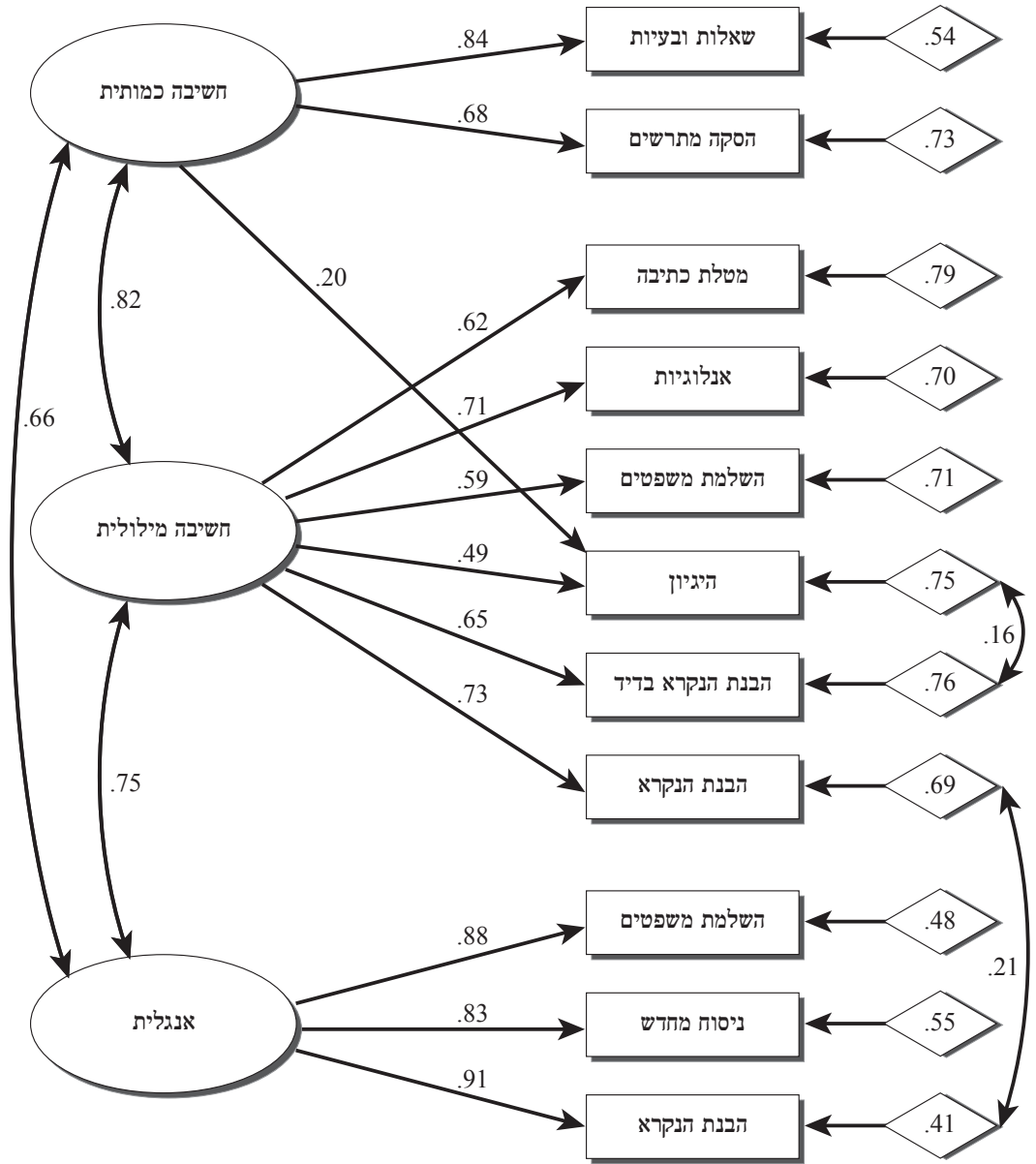
עקב אי-התאמות מסוימות לנתונים שנמצאו במודל המקורי הסטנדרטי שנבדק, נציע מודל לא סטנדרטי שהוא גרסה של המודל הסטנדרטי שהוצע. המודל המקורי תוקן על-פי המידע שהתקבל משני מדדי התיקון שהתוכנה הסטטיסטית מציעה: מבחן ואלד (Wald test), המסייע לזהות נתיבים וקשרים שניתן לגרוע מן המודל, ובעיקר מבחן כופל לג'ראנג' (Lagrange multiplier test), המזהה נתיבים וקשרים בין משתנים, שייתכן שהם אמורים להתווסף למודל.

מדדי התיקון האלה הציעו קשרים נוספים שלא נמצאו במודל הסטנדרטי המקורי. המודל הלא סטנדרטי שונה אפוא מן המודל הסטנדרטי בשני היבטים כלליים: (1) מדד המודד משתנה יכול להיות מושפע מיותר מגורם אחד; (2) תיתכן שונות משותפת בין הטעויות. בפרט, המדדים הראו כי המשתנה "היגיון", המשוך במקור לתחום החשיבה המילולית, מושפע למעשה משני תחומים: הן מן החשיבה המילולית והן מן החשיבה הכמותית. נוסף על כך, מדדי התיקון הראו קשר בין המשתנים "הבנת הנקרא" ו"הבנת הנקרא באנגלית" לבין המשתנים "היגיון" ו"הבנת הנקרא פריט בדיד". כפי שנעשה בנוגע למודל הסטנדרטי, כדי לבחון את המבנה הפנימי של המודל המוצע הלא סטנדרטי ואת מערך הקשרים בו נעשה ניתוח גורמים מאשש, וכן נבדקו עבורו מדדים לטיב ההתאמה.

מודל מותאם לא סטנדרטי: טעינות המשתנים על הגורמים

בתרשים 3 מוצג המודל האמפירי שהתקבל בניתוח הגורמים המאשש. בתרשים מצוינים ערכי הטעינות של הטעויות והמתאמים בין הגורמים לבין הטעויות. השינויים לעומת המודל הסטנדרטי המופיע בתרשים 2 מסומנים בחצים עבים.

כאשר משווים את תרשים 3 לתרשים 2, ניתן לראות שהטעינויות והמתאמים דומים יחסית, והטעינות של משתנה ההיגיון על תחום החשיבה המילולית במודל הסטנדרטי ($L = .69$) נטענת במודל הלא סטנדרטי על שני תחומים: על חשיבה כמותית ($L = .20$) ועל חשיבה מילולית ($L = .49$).



תרשים 3: המודל המתוקן הלא סטנדרטי: תוצאות ניתוח גורמים מאשש

מודל מותאם לא סטנדרטי: מדדי טיב ההתאמה
בלוח 6 מוצגים מדדי טיב ההתאמה עבור המודל הלא סטנדרטי המוצע לעומת המדדים שהתקבלו עבור המודל הסטנדרטי.

לוח 6: מדדי טיב ההתאמה עבור המודל הסטנדרטי והלא סטנדרטי

המודל הלא סטנדרטי	המודל הסטנדרטי	המדד
332***	502***	$\chi^2(df)$
(38)	(41)	
8.76	12.24	χ^2 / df
.0458	.0551	RMSEA
.9842	.9763	GFI
.9837 / .9790	.9753 / .9695	NNFI / NFI
.9855	.9773	CFI

*** $p < .001$

מלוח 6 עולה כי המודל הלא סטנדרטי אכן מתאים לנתונים יותר מן המודל הסטנדרטי – כל המדדים מראים התאמה טובה יותר לנתונים. יש לשים לב שגם במודל זה ערך χ^2 הוא מובהק סטטיסטית (תוצאה שאינה טובה), אך היחס בין ערך χ^2 לבין דרגות החופש שלו קטן מן הערך המקביל במודל התאורטי (ועדיין גבוה מרף ה-2).

דיון ומסקנות

מחקר זה בחן היבטים שונים של תוקף המבחן הפסיכומטרי במתכונתו החדשה, ובפרט את תוקף המבנה ואת התוקף המתכנס והמבחין, בעזרת ניתוח גורמים מאשש (Bentler, 1990). ככלל, הממצאים תמכו בתוקף המבנה של המבחן הפסיכומטרי במתכונתו החדשה והעלו שלושה גורמים אמפיריים, התואמים את תחומי התוכן הקיימים במבחן – חשיבה מילולית, חשיבה כמותית ואנגלית. ממצאים אלו תואמים גם ממצאי מחקרים שנערכו בעולם (Rock et al., 1988) עבור מבחנים גורליים בעלי מבנה דומה.

ניתוח הגורמים המאשש תמך בקיומו של מבנה פנימי תלת-גורמי, שבו אפשרנו לכל מדד להיטען אך ורק על הגורם שהוא משויך אליו תאורטית. עבור מודל זה, המודל הסטנדרטי, ערכיהם של ארבעה מתוך חמישה מדדים לטיב ההתאמה היו בתחום הקביל ואף למעלה מכך. כלומר, מודל תלת-גורמי סטנדרטי הוא קירוב טוב לתיאור המבנה

הגורמי של סולמות המבחן השונים. עם זאת, כאשר אפשרנו לחלק מסולמות המבחן להיטען גם על גורמים אחרים מאלו שהם שויכו אליהם מבחינה תאורטית, וכאשר התאפשרה תלות בין הגורמים (המבוטאת במודל במתאם בין טעויות המדידה), התקבלו מדדים טובים יותר של טיב התאמה. מודל זה פונה המודל הלא סטנדרטי, והוא הניב מדדי התאמה מצוינים.

באשר לטעינות של סולמות מסוימים על הגורמים השונים, מטלת הכתיבה נטענה, כמצופה, על גורם החשיבה המילולית, וממצא זה תומך בהכללתה בחישוב הציון המילולי, כפי שנעשה כיום. כמו כן, נמצא מתאם משמעותי בין הבנת הנקרא בשפה העברית לבין הבנת הנקרא בשפה האנגלית. ייתכן כי ממצא זה מרמז על קיומו של גורם לשוני כללי, המשפיע על התפקוד הן בשפת האם והן בשפה זרה. ואכן, מחקרים קודמים תומכים בקיומה של העברה (transfer) של כישורים משפת האם לשפה זרה, וממצאיהם מעידים כי הכישורים בשפת האם הם בין המנבאים הטובים ביותר לכישורי השפה הזרה (Kahn-Horwitz, Shimron, & Sparks, 2005; Sparks, Patton, Ganschow, & Humbach, 2008; Sparks, Patton, Ganschow, Humbach, & Javorsky, 2008). לבסוף, נמצא כי המדד "היגיון" טעון במידה מסוימת גם על גורם החשיבה הכמותית, אף שבמקור הוא משויך לחשיבה מילולית. ייתכן כי ממצא זה, שלא שוער מראש, מצביע על מערכת קשרים בין תפקודי השפה לתפקודי החשיבה הכמותית. יתר על כן, ייתכן כי חשיבה סילוגיזמית, כלומר היכולת לבצע היסק מנתונים קיימים, המיוצגת בפריטי ההיגיון הפורמלי, נשענת במידה מסוימת גם על מיומנויות הנדרשות לפתרון בעיות בתחום הכמותי, כגון יכולות אנליטיות שונות ויכולת חשיבה מופשטת. ואכן, הספרות מציעה שני מסלולים דיפרנציאליים להסבר התהליכים הקוגניטיביים העומדים בבסיסה של החשיבה ההיסקית – המסלול התחבירי, שהוא מסלול לשוני-שפתי בבסיסו, והמסלול החזותי-מרחבי, הדורש יכולת לחולל מניפולציה מרחבית, שהיא כושר קוגניטיבי המקושר באופן הדוק ליכולות כמותיות-מתמטיות. מחקרים מראים כי קיימות למעשה שתי מערכות נפרדות התורמות לחשיבה היסקית, ויש ביניהן קשר ברמת התפקוד של המוח (Goel, Buchel, Frith, & Dolan, 2000).

השלכות המחקר וכיוונים למחקר עתידי

מגבלות המחקר

למחקר זה כמה מגבלות. ראשית, נציין כי המחקר נערך על נוסח אחד של מבחן, שהתקיים במועד מסוים. אף שהמדגם גדול וייצוגי, יש מקום לבחון את שחזור הממצאים במדגמים נוספים, שישתמשו בנוסחים נוספים ובמועדי מבחן שונים. שחזור הממצאים במחקרים כאלה יספק תמיכה של ממש בתוקף המודל שהוצע. נוסף על כך, מדדי התוקף המתכנס והמבחן מבוססים על תוצאות ניתוח הגורמים בלבד, ואף על פי שהם תומכים בתוקף המבחן, הם בבחינת בדיקה ראשונית וחלקית בלבד.

השלכות ומסקנות

מאחר שממצאי המחקר תומכים במבנה התאורטי המשוער של המבחן הפסיכומטרי ובחלוקה התאורטית לשלושה תחומי תוכן שונים, הם מחזקים את הרציונל העומד בבסיס המבחן ואת אופן חישוב הציונים השונים בו. יתר על כן, נראה כי בהתאם לכוונת מפתחיו, המבחן אכן בודק כישורים שפתיים, כישורים כמותיים ורמת ידע באנגלית. עקב כך נראה גם כי הפרקטיקה החדשה לחישוב הציונים, שהוכנסה לשימוש בשנתיים האחרונות ומאפשרת לקבל ציון בדגש מילולי, ציון בדגש כמותי וציון כללי, תואמת את הממצאים האמפיריים ויכולה לשמש פרקטיקה חלופית תקפה לקבלת החלטות בדבר קבלה או דחייה בתחומים שונים.

זאת ועוד, ממצאי המחקר תומכים בתוקף הנראה של המבחן. הם מלמדים שהשינויים שחלו בו אינם פוגעים במבנים התאורטיים העומדים בבסיסו וניתן להמשיך להשתמש במתכונתו החדשה.

סיכום וכיוונים למחקר נוסף

המבחן במתכונתו החדשה מתקיים מאז דצמבר 2012. מחקרים בעתיד יידרשו לבחון את תוקף הניבוי של המבחן במתכונת זו, ובפרט אם חלו שינויים בתוקף הניבוי של המבחן עקב החלת השינויים במבנה ואת תפקידה של מטלת הכתיבה – מטלה בעלת אופי שונה מאוד משאר רכיבי המבחן.

כמו כן, יש לחזור, כפי שצוין, על המחקר עם מדגמים נוספים כדי לבחון את יציבות המודל. נוסף על כך, במחקר זה התבססה בדיקת התוקף המתכנס והמבחין על ניתוח הגורמים המאשש. מחקר בעתיד יוכל להשתמש בכדיקות תוקף נוספות המקובלות בספרות, דוגמת MTMM, לשם תמיכה נוספת בתוקף המבחין ובתוקף המתכנס של הגורמים שנמצאו במחקר זה.

לסיכום, המחקר בדק ממדים שונים של תוקף המבחן הפסיכומטרי במתכונתו החדשה, ובפרט את מבנה הגורמים, והוא מספק תמיכה ראשונית טובה לתוקף המבחן ותומך בהשערה שהשינויים שבוצעו בתוכנו ובמבנהו לא פגעו בתוקף המדידה. מחקרים בעתיד יבחנו את השאלה אם השינויים אכן שיפרו את תוקף הניבוי של המבחן ואת איכותו.

מקורות

- בודסקו, ד' (1985). ניתוח גורמים של מערכת מבחני הכניסה הכלל אוניברסיטאיים (דוח מס' 21). ירושלים: המרכז הארצי לבחינות ולהערכה.
- בייט-מרום, ר', בלר, מ', בראון, ה', ברנן, ר', גפני, נ', המבלטון, ר'. ... וצלגוב, י' (1998). מקומה של הבחינה הפסיכומטרית בבחירת תלמידים למערכת ההשכלה הגבוהה. ירושלים: המרכז הארצי לבחינות ולהערכה.
- קנת-כהן, ת', אורן, כ', טורוול, א' וכהן, י' (2013). תוקף הניבוי של מערכת המיון לחוגים למשפטים באוניברסיטאות (מחזוריים תשנ"ה-תשס"ט). ירושלים: המרכז הארצי לבחינות ולהערכה.
- Amrein, A. L., & Berliner, D. C. (2002). High-stakes testing & student learning. *Education Policy Analysis Archives*, 10, 18. Retrieved from <http://epaa.asu.edu/epaa/v10n18/>.
- Anderson, J. C., & Gerbing, D. W. (1988). Structural equation modeling in practice: A review and recommended two-step approach. *Psychological Bulletin*, 103, 411-423.
- Bagozzi, R. P., & Phillips, L. W. (1982). Representing and testing organizational theories: A holistic construal. *Administrative Science Quarterly*, 27, 459-489.
- Beller, M. (1990). Tree versus geometric representation of tests and items. *Applied Psychological Measurement*, 14(1), 13-28.
- Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, 107(2), 238-246.
- Bentler, P. M. (1992). On the fit of models to covariances and methodology to the Bulletin. *Psychological Bulletin*, 112(3), 400-404.
- Brelmad, H., Kubota, M., Nickerson, K., Trapani, C., & Walker, M. (2004). *New SAT writing prompt study: Analysis of group impact and reliability* (College Board Research Report No. 2004-1 ETS RR-04-03). New York, NY: College Entrance Examination Board.
- Cahalan-Laitusis, C., Cook, L. L., & Aicher, C. (2004, April). *Examining test items for students with disabilities by testing accommodation*. The annual meeting of the National Council on Measurement in Education, San Diego, CA.
- Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 56, 81-105.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of test. *Psychometrika*, 16, 297-334.

- Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research, 18*, 39-50.
- Frey, M. C., & Detterman, D. K. (2004). Scholastic assessment or g? The relationship between the scholastic assessment test and general cognitive ability. *Psychological Science, 15*, 373-378.
- Goel, V., Buchel, C., Frith, C., & Dolan, R. J. (2000). Dissociation of mechanisms underlying syllogistic reasoning. *Neuroimage, 12*(5), 504-514.
- Guttman, L. (1968). A general nonmetric technique for finding the smallest coordinate space for a configuration of points. *Psychometrika, 33*(4), 469-506.
- Hatcher, L. (1994). *A step by step approach to using SAS for factor analysis and structural equation modeling*. Cary, NC: SAS Institute.
- Hooper, D., Coughlan, J., & Mullen, M. R. (2008). Structural equation modeling: Guidelines for determining model fit. *Journal of Business Research Methods, 6*, 53-60.
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*(1), 1-55.
- Huynh, H., & Barton, K. E. (2006). Performance of students with disabilities under regular and oral administrations of a high-stakes reading examination. *Applied Measurement in Education, 19*(1), 21-39.
- James, L. R., Mulaik, S. A., & Brett, J. M. (1982). *Causal analysis*. Beverly Hills, CA: Sage.
- Kahn-Horwitz, J., Shimron, J., & Sparks, R. L. (2005). Predicting foreign language reading achievement in elementary school students. *Reading and Writing, 18*, 527-558.
- Kline, R. B. (2010). *Principles and practice of structural equation modeling* (3rd ed.). New York, NY: Guilford.
- Koenig, K. A., Frey, M. C., & Detterman, D. K. (2008). ACT and general cognitive ability. *Intelligence, 36*(2), 153-160.
- Long, J. S. (1983). *Confirmatory factor analysis: A preface to LISREL*. Sage University Paper Series on Quantitative Application in the Social Sciences, 07-033. Beverly Hills, CA: Sage.
- Marsh, H. W., Balla, J. R., & McDonald, R. P. (1998). Goodness-of-fit indexes in confirmatory factor analysis: The effect of sample size. *Psychological Bulletin, 103*, 391-410.

- Netemeyer, R. G., Johnston, M. W., & Burton, S. (1990). Analysis of role conflict and role ambiguity in a structural equations framework. *Journal of Applied Psychology, 75*, 148-157.
- Rock, D. A., Bennett, R. E., & Jirele, T. (1988). Factor structure of the graduate record examinations general test in handicapped and nonhandicapped groups. *Journal of Applied Psychology, 73*(3), 383-392.
- Sattath, S., & Tversky, A. (1977). Additive similarity trees. *Psychometrika, 42*(3), 319-345.
- Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research Online, 8*(2), 23-74.
- Schmitt, N., & Stults, D. M. (1986). Methodology review: Analysis of multitrait-multimethod matrices. *Applied Psychological Measurement, 10*, 1-22.
- Shur, D. D. (2006). *Exploratory or confirmatory factor analysis? Statistics and data analysis, 200-231*. The Proceedings of the 31st Annual SAS Users Group International Conference. Cary, NC: SAS Institute.
- Sparks, R. L., Artzer, M., Ganschow, L., Patton, J., Siebenhar, D., & Plageman, M. (1997). Prediction of foreign language proficiency. *Journal of Educational Psychology, 89*(3), 549-561.
- Sparks, R. L., Patton, J., Ganschow, L., & Humbach, N. (2008). Long-term relationships among early first language skills, second language aptitude, second language affect, and later second language proficiency. *Applied Psycholinguistics, 30*, 725-755.
- Sparks, R. L., Patton, J., Ganschow, L., Humbach, N., & Javorsky, J. (2008). Early first-language reading and spelling skills predict later second-language reading and spelling skills. *Journal of Educational Psychology, 100*(1), 162-174.
- Steinberg, J., Cline, F., & Sawaki, Y. (2011). *Examining the factor structure of a state standards-based science assessment for students with learning disabilities*. College Board Research Report (ETS RR-11-38). New York, NY: College Entrance Examination Board.
- Ullman, J. B. (2001). Structural equation modeling. In B. G. Tabachnick & L. S. Fidell (Eds.), *Using multivariate statistics* (4th ed., pp. 653-771). Needham Heights, MA: Allyn & Bacon.
- Widaman, K. F. (1985). Hierarchically nested covariance structure models for multitrait-multimethod data. *Applied Psychological Measurement, 9*, 1-26.
- Zwick, R. (2007). *College admission testing*. Arlington, VA: National Association for College Admission Counseling.

