

סקירת ספרות בנושא:

מבחנים להערכת יכולת כתיבה

דנה גנאור-שטרן

פברואר 1996



מרכז ארצי לבחינות ולהערכה
NATIONAL INSTITUTE FOR TESTING & EVALUATION
מיסודן של האוניברסיטאות בישראל
גבעת רם, שער נווה שאנן ת"ד 26015 ירושלים 91260 טל' 02-759555

תוכן העניינים

3	מבוא
4	הכתיבה מן ההיבט הקוגניטיבי
6	הכתיבה מן ההיבט הפונקציונאלי
7	מבנה הדוח
8	פרק א': מתכונת מבחן הכתיבה
8	מבחני כתיבה ישירים מול מבחני כתיבה עקיפים
9	פיתוח מבחני כתיבה ישירים
20	פרק ב': הערכה וציינון
20	שיטות הציינון
23	הקריטריונים לבחירה בין שיטות הציינון השונות
23	מימדי ההערכה
25	ממצאים אמפיריים לגבי מימדי ההערכה ושיטות הציינון
31	שיטות הציינון המקובלות במבחני כתיבה שונים בארץ ובעולם
35	פרק ג': מהימנות
35	גורמים הקשורים למטלת הכתיבה
37	גורמים הקשורים לנבחן
37	גורמים הקשורים לציינון הבחינה
43	חישוב מהימנות בהתבסס על מספר גורמי רעש
45	כיוול
46	פרק ד': תוקף
46	תוקף תוכן
46	תוקף כלפי קריטריון
55	תוקף מבנה
59	סיכום
61	מקורות

מטרת דוח זה היא לסקור את הספרות העוסקת במבחנים להערכת יכולת כתיבה. הדוח דן במתכונת מבחני הכתיבה, בשיטות ההערכה והציון הקיימות ובמחקרים באשר למהימנות ולתוקף של מבחנים מסוג זה.

למותר לציין כי יכולת הכתיבה או ההתבטאות בכתב חשובה ביותר בחיים האישיים, המקצועיים והאקדמיים. אף שההתבטאות בכתב נעשית בתדירות נמוכה יותר מההתבטאות בעל פה, יש לה חשיבות מיוחדת. שפה כתובה היא על פי הגדרתה מתועדת, לעומת התבטאות בעל פה הנשמרת רק בזיכרונם של בני אדם. מכיוון ששפה כתובה נשמרת לאורך זמן, היא נחשבת למחייבת יותר משפת דיבור, ועקב כך יש לה פונקציות רשמיות רבות: חוקים ונהלים מועלים על הכתב, התחייבויות וחוזים בין אנשים מנוסחים בכתב וכו'. כמעט כל הסכמה המושגת בין אנשים המעוניינים בשמירתה, מועלית על הכתב. נוסף על כך, מקובל גם לחשוב כי פנייה אל אדם או אל גורם רשמי בבקשה כלשהי תהיה אפקטיבית יותר אם תנוסח בכתב. בהקשר של לימודים אקדמיים מיומנות הכתיבה כמו גם מיומנות הקריאה הן חשובות ביותר. באמצעות הקריאה רוכש הסטודנט ידע חדש, בעוד שבדרך כלל הכתיבה משמשת אותו לסיכום החומר ולפיתוח רעיונות חדשים. נוסף על כך, לפחות בחלק מן הפקולטות מקבלים הסטודנטים במהלך לימודיהם מספר רב של מטלות כתיבה, במסגרת של עבודות ומבחנים, המהוות את הבסיס להערכה של הישגי הסטודנט. באמצעות מטלות הכתיבה המרצים יכולים להעריך את מידת הידע וההבנה של הסטודנט את החומר הנלמד ואת יכולתו לפתח רעיונות משלו. למרות חשיבותה של מיומנות הכתיבה בלימודים אקדמיים, בחינות המיון למסגרת זו שמו דגש רב יותר על בחינת יכולת הקריאה מאשר על הערכת כושר הכתיבה. הסיבה לכך היא כי את כושר הקריאה קל וטבעי יותר למדוד באמצעות מבחנים רבי ברירה, אשר במשך זמן רב נחשבו לשיטת הבחינה היעילה ביותר, ואולם שיטה זו אינה מתאימה למדידת יכולת כתיבה. כיום, עם המגמה של הרחבת השימוש בהערכות ביצוע המבוססות בדרך כלל על מבחנים פתוחים, ועם ההכרה כי מיומנות הכתיבה - על אף חשיבותה הרבה - אינה נחלת הכלל, יש עניין הולך וגובר בפיתוח מבחני כתיבה.

בפיתוח מבחני כתיבה יש לקחת בחשבון את המאפיינים היחודיים של תהליך הכתיבה. בגלל חשיבותה אין היצירה הכתובה נשארת בדרך כלל בצורתה הגולמית, אלא היא עוברת תהליכי עריכה ותיקון. בשפה הכתובה יש הקפדה על ניסוחים מדויקים יותר מבשפה מדוברת, היא עמוסה ומרוכזת יותר וכוללת פחות תזרות מבשפת דיבור, שכן הקורא יכול תמיד לחזור ולעיין שוב בטקסט. לכן אין לראות בשפה הכתובה גרסה כתובה של שפת הדיבור בלבד, אלא שפה בעלת אפיונים יחודיים משלה, הדורשת מן הכותב מיומנויות נוספות על אלה הנדרשות מן הדובר. בדרך כלל שפה מדוברת מתרחשת בסיטואציה חברתית ספציפית הכוללת אינטראקציה בין שני בני אדם לפחות. יצירה כתובה, לעומת זאת, נוצרת בנסיבות שבהן אין קשר ישיר בין הכותב לקורא. לעתים, הנמען הידוע לכותב, אינו נמצא בסמיכות לו, ובמקרים אחרים, הכותב מכון את יצירתו לקהל יעד רחב אשר אינו מוכר לו היכרות אישית. מאחר שהדיבור מתרחש במסגרת מפגש

בין אנשים הוא אינו ערוץ התקשורת היחיד, אלא נלוות לו הבעות פנים ותנועות גוף, אשר גם הן עוזרות להעברת המסר ולפענוחו. אם המאזין לא הבין את המסר, הוא יכול לבקש מן הדובר לחזור על דבריו או להבהירם. תקשורת כתובה אינה מתרחשת במסגרת של מפגש חי בין אנשים ולכן לא קיימים גורמים חוץ-טקסטואליים בין הכותב לקורא אשר יעזרו בפענוח הטקסט. על הכותב לנסח את היצירה הכתובה בצורה שתכיל בתוכה את כל המידע שרצה הכותב להעביר, ותוכל להיות מובנת בזכות עצמה. הכותב מייצר את הטקסט בעצמו, באופן מבודד, ללא שיתוף פעולה של הקורא או תגובה מידית שלו. אף על פי שתהליך הכתיבה נעשה ללא קשר ישיר עם הקורא, הרי שעל הכותב להתייחס לייצוג של קהל היעד אשר קיים במוחו; עליו להתאים את כתיבתו לקהל מסוים ולקחת בחשבון את הידע העומד ברשות קהל זה, את דעותיו ותגובות אפשריות שלו.

הכתיבה מן ההיבט הקוגניטיבי

ראשית המחקר על כושרי השפה הכתובה התמקד בתהליכים קוגניטיביים של היחיד. מודלים אשר עסקו בתהליכים קוגניטיביים בכתיבה הבחינו בין יצירת הרעיונות לבין יצירת הטקסט (De Beaugrande, 1980; Collins & Gentner, 1980; Hayes & Flower, 1980).

Collins & Gentner (1980) טוענים כי תהליך יצירת הטקסט כפוף לאילוצים רבים יותר ומוגבל יותר על ידי חוקים מתהליך יצירת הרעיונות. הטקסט הכתוב ברמת המילה, המשפט או הפסקה, צריך למלא אחר כללי הדקדוק והתחביר. ביצירת הרעיונות יש לאדם יותר חופש פעולה, אולם גם כאן הוא מוגבל על ידי הדרישה כי הרעיונות יהיו סבירים, רלוונטיים, ועקביים זה עם זה. לדעתם, הרבה מהקושי בכתיבה נובע מהאילוצים הרבים שפועלים בתהליך ניסוח הרעיונות לכדי טקסט כתוב.

De Beaugrande (1980) חילק את תהליך הכתיבה לארבעה שלבים: שלב התכנון שבו הכותב מתמקד במטרת הטקסט, בקהל היעד, ובסוג הכתיבה שבה יכתוב; שלב יצירת הרעיונות; שלב הפיתוח והניסוח המפורט של הרעיונות; ושלב הניסוח של הרעיונות במילים.

Hayes & Flower (1980) הציעו מודל תיאורי מפורט יותר של תהליך הכתיבה, המתבסס בעיקרו על ניתוח פרוטוקול. לפי מודל זה עולם הכותב מחולק לשלושה חלקים: סביבת המשימה, זיכרון לטווח ארוך ותהליך הכתיבה. סביבת המשימה והזיכרון לטווח ארוך הם ההקשר שבו מתרחש תהליך הכתיבה. סביבת המשימה כוללת גורמים חיצוניים לכותב אשר משפיעים על הביצוע במשימה: הוראות הבוחן, פרטי המשימה עצמה, תיאור של הנושא ושל הקהל המיועד. משלב מסוים בתהליך הכתיבה גם הטקסט שהופק הוא חלק מסביבת המשימה. בזיכרון לטווח ארוך של הכותב יש אינפורמציה תוכנית לגבי נושאי הכתיבה כמו גם ידע על אסטרטגיות כתיבה.

Hayes & Flower (1980) חילקו את תהליך הכתיבה לשלושה תת-תהליכים: תכנון (planning), תרגום (translating) וביקורת (reviewing). תהליכים אלה לא מתרחשים בהכרח בזה אחר זה, סדרם משתנה אצל כותבים שונים ובכתיבת מטלות כתיבה שונות, ונוסף על כך תיתכן חזרה על חלק מהתהליכים.

א. תכנון

בשלב התכנון נעשה שימוש במידע מסביבת המשימה ומהזיכרון לטווח ארוך כדי להציב מטרות וליצור תכנית אשר תנחה את יצירת הטקסט. התכנית יכולה להישלף מהזיכרון לטווח ארוך או להיות מיוצרת בזמן אמיתי.

התכנון מורכב גם הוא משלושה מרכיבים :

ייצור (generating) - מבוסס על שליפת אינפורמציה רלוונטית למשימת הכתיבה מהזיכרון לטווח ארוך, שליפה המונעת בעיקר על ידי נושא החיבור וקהל היעד. התוצר של תהליך זה יכול להיות מילים, חלקי משפטים או משפטים שלמים.

ארגון - בשלב זה נעשית הבחירה של החומרים השימושיים ביותר שעלו בתהליך הייצור וארגונם לכדי תכנית כתיבה, אשר יכולה להיות מנוסחת במונחי זמן ("קודם אדון בנושא א ואחר כך בנושא ב") או באופן היררכי ("תחת נושא א אדון בנקודות 1,2,3"). תהליך הארגון נעשה ברמות הכללה שונות : בהתחלה הוא נעשה ברמת נושא אחד או שני נושאים, אחר כך הוא מתרחב למספר גדול יותר של נושאים ולחלוקה לפרקים.

הצבת מטרות - חלק מהחומרים אשר נשלפים בתהליך הייצור נוגעים לכתיבה עצמה ולאופן שבו היא צריכה להיעשות, למשל, הכותב יכול לומר לעצמו "יש לכתוב זאת בפשטות". תהליך הצבת המטרות כולל זיהוי של קריטריונים להערכה שיעשה בהם שימוש בשלב העריכה.

ב. תרגום

בשלב זה מתורגמת האינפורמציה שנשלפה מן הזיכרון למשפטים קבילים מבחינה לשונית. ההנחה היא שבזיכרון האינפורמציה מאוחסנת בצורה של היגדים לוגיים (propositions) ולא בצורה של מבנים לשוניים.

ג. ביקורת

תהליך הביקורת מורכב מקריאה ומעריכה. מטרת שלב זה היא לשפר את איכות הטקסט שנוצר בתהליך התרגום על ידי מציאת ליקויים בכתוב ותיקונם. הליקויים יכולים להיות ברמה הלשונית או במידה שבה הטקסט משיג את מטרות הכותב. שלב הביקורת מתחיל באופן אוטומטי, ועם התחלתו הוא מפסיק תהליכים של תכנון או תרגום המתרחשים באותו זמן.

בהקשר של מבחני כתיבה סביבת הכתיבה מורכבת הן מהידע שבזכרוננו של הכותב והן מכל חומר אשר יסופק לו על ידי הבוחן. כדי להעשיר סביבה זו חלק ממבחני הכתיבה כוללים קטעי קריאה או תרשימים עליהם יתבסס הכותב. מבחני כיתה מן הגישה הקלאסית המתרחשים בזמן קצוב בוחנים את התוצר של הכתיבה, כלומר את החיבור, ואינם מתייחסים לתהליך, אולם סביר להניח שבנסיבות אלה לא כל התהליכים המוזכרים מתקיימים; נראה ששלב הביקורת במסגרת של בחינה יהיה מצומצם למדי בגלל קוצר הזמן.

הכתיבה מן ההיבט הפונקציונאלי

כיום נפוצה הגישה הסוציולינגוויסטית הרואה את הקריאה והכתיבה כפעולות לשוניות בעלות פונקציה קומוניקטיבית אשר מעוגנות במציאות חברתית מוגדרת. אף על פי שתהליך הכתיבה נעשה בידי אדם אחד בלבד, הוא למעשה חלק מסיטואציה חברתית: הכותב מייעד את דבריו לקהל מוגדר אשר אתו הוא נמצא באינטראקציה למטרה מסוימת. מאחר שכך הרי משימות במבחן כתיבה צריכות להיות לקוחות אף הן מסיטואציה חברתית מוגדרת (Bloom & Green, 1984). דוגמה למשימה כזו היא: "כתוב חיבור שבו אתה פונה למנהל בית הספר ומסביר לו את דעתך בנושא חובת לבישת תלבושת אחידה". להבדיל ממשימת הכתיבה הבאה התלושה מהקשר חברתי: "מהי עמדתך בעניין חובת לבישת תלבושת אחידה?".

בכל אקט של תקשורת יש חשיבות הן לנותן המסר והן למקבלו, והמשמעות של הטקסט אינה מצויה בטקסט עצמו אלא היא נבנית על ידי הקורא. הכותב אינו יכול להעביר במישרין את כוונותיו, אלא בכתיבתו הוא כולל רמזים בלבד באשר לכוונותיו (Takala, 1982). הקורא משתמש ברמזים אלה כדי לבנות מחדש את המשמעות או את המסר של הטקסט. הקריאה אם כן אינה תהליך פסיבי, אלא יש בה שיחזור או בנייה מחדש המושפעים הן מתהליכי bottom-up המונעים על ידי הנאמר בטקסט, והן מתהליכי top-down המונחים על ידי דעותיו וניסיונו של הקורא. המשמעות של הטקסט אינה קיימת בטקסט עצמו אלא באינטראקציה בין הקורא לבין הטקסט, והטקסט מאפשר יותר מפרשנות לגיטימית אחת. לפי גישה זו עיבוד טקסט הוא אינטראקציה קומוניקטיבית שבה מתבצע משא ומתן בין הקורא לכותב בדבר משמעות הטקסט. גם קריאה וגם כתיבה הם תהליכי יצירת משמעות (Sarig & Folman, 1988).

גישות האוריינות המודרניות טוענות עוד כי יש לראות בקריאה ובכתיבה חלקים של אותה המיומנות. (De Beaugrande 1980) טוען כי האופרציות העומדות בבסיס הכתיבה והקריאה הן סימטריות, אולם מתרחשות בסדר הפוך. הכותב צריך לתרגם את כוונותיו לתוכן מסוים ולנסח תוכן זה בטקסט מילולי כתוב. הקורא יוצא מהטקסט המילולי הכתוב, מסיק ממנו בדבר התוכן הנאמר, וממנו מקיש על כוונת הכותב. למרות הסימטריות של התהליכים אין להניח כי הקורא בהכרח מגיע לרעיונות ולכוונות המקוריות של הכותב.

הצדקה נוספת לטשטוש ההפרדה בין קריאה לכתיבה נובעת מהעובדה ששתי הפעולות נעשות בדרך כלל יחד במסגרת אותה המשימה. לעיתים תכופות נכתבים חיבורים על סמך קריאה במקורות, אולם גם כאשר קריאה כזו אינה הכרחית, הרי שבמהלך תהליך הכתיבה האדם קורא את הטקסט שכתב וקריאה זו מנחה אותו בשלבי העריכה והתיקון של הטקסט.

לסיכום, טוענות (Sarig & Folman 1988) כי הקשר ההדוק בין תהליכי הכתיבה והקריאה נובע משלושה גורמים: א. בתהליכי הקריאה והכתיבה יש אינטראקציה בה מתבצע משא ומתן על משמעות הטקסט. ב. בשני תהליכים אלו קורות פעולות קוגניטיביות מקבילות המתרחשות בסדר הפוך. ג. בהקשר האקדמי שתי הפעולות מתבצעות ברוב המקרים ביחד, ובכל משימה בה נדרשת כתיבה נדרשת גם קריאה. המשתמע מכך לדעתן הוא שאין ללמד קריאה וכתיבה בנפרד ואין

לבחון אותן בנפרד (Sarig & Folman, 1988). מומלץ, אם כך, שמבחן כתיבה האמור לבחון יכולת כתיבה באופן אותנטי יהיה בנוי משימות כתיבה המבוססות על קטעי קריאה.

מבנה הדוח

דו"ח זה יתמקד בסקירת הספרות הדנה בהערכת יכולת כתיבה, כאשר הדגש יהיה על מבחני כתיבה רחבי היקף, שמטרתם העיקרית היא דירוג הנבחנים על פי כושרם ואשר הם אינם חלק מתהליך הלימוד של הכתיבה. בפרק א' תידון מתכונת מבחני הכתיבה: יוצג הויכוח בין תומכי הגישה הישירה למדידת יכולת כתיבה הסבורים כי בחינה כזו צריכה לכלול משימות כתיבה לבין תומכי הגישה העקיפה הסבורים כי יש לכלול שאלות רבות ברירה. בהמשך הדוח הדגש יהיה בעיקר על דיון במבחני כתיבה ישירים. המשכו של פרק א' יעסוק בסוגי משימות הכתיבה, מאפייניהם, מספר המשימות שיש לתת במבחן, מתן בחירה בין משימות, מקבילות משימות הכתיבה והזמן המוקצב להן. פרק ב' יעסוק בנושא המורכב של הערכת החיבורים. הוא יכלול סקירה של שיטות הציון העיקריות תוך אזכור יתרונותיהן וחסרונותיהן, והצגת נתוני מחקרים אשר הישוו בין שיטות ציון שונות. בפרק ג' יידון נושא מהימנות מבחני כתיבה, עד כמה יציבים הציונים המתבססים על משימות כתיבה שונות ועל הערכות של בודקים שונים. בפרק ד' יידון נושא התוקף לסוגיו: תוקף תוכן, תוקף ניבוי ותוקף מבנה, ויוצגו תוצאות מחקרים אשר בדקו את התוקף של מבחני כתיבה שונים. הסיכום יכלול את המסקנות העולות מסקירת הספרות.

פרק א': מתכונת מבחן הכתיבה

פרק זה יעסוק בסוגיית הרכבת מבחן הכתיבה. בהתחלה יוזכרו שני הסוגים העיקריים של מבחני הכתיבה (ישירים ועקיפים) והיתרונות והחסרונות של כל אחד מהם. בהמשך הדיון יתמקד במבחני כתיבה ישירים בלבד, ויעסוק בחלוקה של משימות כתיבה לסוגים לפי ז'אנר הכתיבה הנדרש ולפי מידת ההתבססות של משימת הכתיבה על ידע. תידון ההשפעה של ז'אנר הכתיבה ומרכיב הידע על הערכת הכתיבה וההשלכות של השפעות אלה על הרכבת מבחן הכתיבה. המשך הפרק ידון בניסוח משימת הכתיבה, בסוגיית מקבילות משימות כתיבה שונות, ובשאלת הזמן המוקצב למשימת הכתיבה.

מבחני כתיבה ישירים מול מבחני כתיבה עקיפים

מבחנים להערכת יכולת כתיבה נחלקים לשני סוגים: מבחני כתיבה ישירים, הכוללים משימת כתיבה אחת או יותר, ומבחנים בלתי ישירים או עקיפים, הכוללים פריטים רבי ברירה. במבחני כתיבה ישירים מתבקשים הנבחנים לכתוב חיבור אחד או יותר בדרך כלל בזמן קצוב ועל נושא/ים שנקבעו מראש. בגלל אורך הזמן הנדרש לכתיבת חיבור מבחני כתיבה ישירים כוללים בדרך כלל מספר קטן של משימות כתיבה (פחות משלוש). פריטים רבי ברירה דורשים לפתרונם זמן קצר בהרבה ולכן מבחני כתיבה עקיפים כוללים מספר גדול יותר של פריטים. להלן דוגמאות לפריטים ממבחני כתיבה עקיפים (שלושת הראשונים מופיעים במבחן הכתיבה הכלול ב-SAT):

- א. Usage) שימוש - הפריט כולל משפט ועל הנבחן לאתר את החלק השגוי במשפט. הטעות יכולה להיות דקדוקית או תחבירית.
- ב. תיקון משפט - מוצג משפט ובצמוד לו מוצעות ארבע אפשרויות לתקן חלק ממנו. על הנבחן לבחור את האפשרות שיוצרת את המשפט הנכון ביותר.
- ג. תיקון בהקשר - מופיעה פסקה של טקסט המוצגת כטיוטה ראשונה, ואחריה מספר שאלות. בחלק מהשאלות הנבחן מתבקש לבחור את המשפט שמנסח מחדש בצורה הטובה ביותר חלק מהפסקה. חלק מהפריטים שואלים על שיטות הארגון או הכתיבה של הפסקה.
- ד. ארגון - מופיעים ארבעה משפטים אשר היו פסקה אחת קודם, ועל הנבחן לזהות מהו סדר ההופעה הנכון של המשפטים. ניתן לתת מספר פסקאות כאלה ועל הנבחן לזהות את הסדר של הפסקאות, נוסף על סידור המשפטים בכל פסקה.

תומכי הגישה הישירה טוענים כי גישתם תקפה בהגדרה: היא באה למדוד יכולת כתיבה על ידי ביצוע במשימת כתיבה. מבחן ישיר הוא היחיד שיכול לבדוק את שלב היצירה שהוא כה מרכזי בתהליך הכתיבה.

מתנגדיה של הגישה הישירה מעלים את הנקודות הבאות:

- א. התוקף של מבחן הכתיבה שהוזכר לעיל הוא רק לכאורה, יש הבדל רב בין כתיבה בזמן מבחן לכתיבה בסיטואציה טבעית. המבחן בודק יכולת לייצר טיוטה מהירה תחת לחץ זמן על נושא

- שבדרך כלל אינו מעניין את הנבחן. בתנאים מגבילים שכאלה גם לא נשאר לנבחן זמן לעריכה ולתיקונים, שלבים שהם אינטגרליים בתהליך הכתיבה האקדמית (Cooper, 1984).
- ב. תהליך הצינון של חיבורים מתבסס על הערכות סובייקטיביות של בודקים, ולכן הוא מכניס רעש רב למדידה.
- ג. מבחן כתיבה ישיר כולל מספר קטן של פריטים ולכן מהימנותו צפויה להיות נמוכה. קיימים ז'אנרים שונים של כתיבה ומבחן שכולל רק משימת כתיבה אחת או שתיים יכול להעריך יכולת כתיבה רק בז'אנרים שכלל.
- ד. אופי המבחן גורם לכך שקשה לבדוק את הקושי של משימות הכתיבה ולכיל בין נוסחים.

תומכי הגישה העקיפה טוענים שהיא מהווה פיתרון לרבים מחסרונות השיטה הישירה: הבדיקה היא אובייקטיבית, המבחן כולל מספר גדול של פריטים, ניתן להשוות את קושי הפריטים ולכיל בין נוסחים, מבחן עקיף מתמקד בכשרים הבסיסיים הנדרשים בכל סוגי הכתיבה ולכן אינו מוגבל רק לז'אנרים מסוימים.

החיסרון העיקרי של הגישה העקיפה הוא שהבדיקה של יכולת הכתיבה נעשית רק דרך זיהוי ולא באמצעות הפקה, כלומר אין במבחן כזה כתיבה של ממש. מבחן עקיף יכול לבדוק רק טווח מוגבל של כושרי כתיבה, כמו זיהוי של טעויות דקדוק ותחביר, והוא מתמקד רק בחלקים קטנים יותר של הטקסט כמו משפט או פסקה. לעומתו מבחן כתיבה ישיר בודק יכולות גבוהות, כלומר כשרים הקשורים לרמה הכוללת והגבוהה ביותר של כתיבת הטקסט (Cooper, 1984).

הוויכוח בין תומכי הגישה הישירה לבין תומכי הגישה העקיפה למדידת יכולת כתיבה לא הסתיים, והוא יידון גם בהמשך הדוח.

פיתוח מבחני כתיבה ישירים

חלק מהקושי בפיתוח מבחן כתיבה נובע מהעובדה כי מיומנות הכתיבה או יכולת הכתיבה היא רחבה במיוחד. קשה להגדיר את גבולותיה, מאחר שהשימוש בה נעשה במספר גדול של הקשרים ומשימות. התוצרים של יכולת הכתיבה רבים ומגוונים כל כך שקשה להסכים על קריטריונים להערכה (Quellmalz, 1986). אפשר לאפיין משימות כתיבה במספר רב של מימדים, אולם דוח זה יתמקד רק בשניים מהם: ז'אנר הכתיבה ותחום הידע. ז'אנר הכתיבה מתייחס לסוגי הכתיבה השונים אשר נוצרו להשגת מטרות שונות (כמו למשל תיאור לעומת שכנוע). תחום הידע מתייחס לנושא בו עוסקת הכתיבה.

ז'אנרים של כתיבה

קיימות דרכים רבות לסווג את החומר הכתוב. הסיווג יכול להתבסס על נושא הקטע, על מטרות הטקסט, על סגנון הכתיבה, על קהל היעד וכו'. הסיווגים הללו אינם בלתי תלויים שכן מטרות הטקסט מכתיבה פעמים רבות את קהל היעד ואת סגנון הכתיבה. בהקשר של הערכת יכולת כתיבה נפוצה החלוקה על פי ז'אנרים.

Breland et al. (1987) מבחינים בין שלושה ז'אנרים של כתיבה: 1. כתיבה תיאורית נרטיבית - המביעה ניסיון אישי וידע אוטוביוגרפי, ויש בה מעט התייחסות לעולם ולקהל החיצוני. 2. כתיבה הסברית אקספוזיטורית - הכוללת העברת מידע וניתוח של תופעות בעולם תוך התייחסות לקורא. 3. כתיבה שכנועית - אשר מטרתה העיקרית היא לשכנע את הקורא בצדקת טיעוני הכותב. פרויקט המשוב האמריקאי (NAEP - The National Assessment of Educational Progress) מבחין בין משימות כתיבה משלושת הז'אנרים הבאים: כתיבה אינפורמטיבית, כתיבה שכנועית וכתיבה נרטיבית המערבת דמיון. ניתוח מפורט יותר של משימות כתיבה בכלל ומשימות הכתיבה הדרושות באוניברסיטה בפרט נעשה בידי חוקרים האחראים על מבחן TOEFL, בחינת הידע באנגלית של הסטודנטים הזרים בארצות הברית (Bridgeman & Carlson, 1983). התוצאה הייתה מיון ל-10 סוגי משימות כתיבה:

- א. חיבור אישי - המאופיין בסובייקטיביות רבה.
- ב. תיאור סדרתי או כרונולוגי של אובייקטים או מאורעות.
- ג. תיאור מרחבי או פונקציונאלי - דומה לסוג ב', אולם מסתמך על מיומנויות ארגון מופשטות יותר.
- ד. השוואה והנגדה - דורש מן הנבחן לקשר בין כמה נושאים או אספקטים ללא נקיטת עמדה.
- ה. השוואה, הנגדה ונקיטת עמדה.
- ו. אקסטרפולציה - דורש נוסף על מה שנדרש במשימה ה' גם אלמנט של יצירתיות או דמיון.
- ז. כתיבה שכנועית המכוונת לקהל ספציפי.
- ח. תיאור ופרשנות של גרף או טבלה - דורש הבנה של אינפורמציה המוצגת באופן גרפי.
- ט. סיכום קטע - דורש מן הנבחן לקרוא קטע, להבין, להוציא ממנו את הרעיונות המרכזיים ולסכם אותו במילים שלו.
- י. סיכום קטע, ניתוח והערכה של תוכנו - נוסף על הנדרש במשימה ט' הנבחן יכול להשוות את הנאמר בקטע לרעיונות אחרים, להביע את דעתו או לבצע אקסטרפולציה.

נראה כי מבחן כתיבה המיועד למיון של מועמדים לאוניברסיטאות צריך לבחון יכולת כתיבה בז'אנר המקובל באוניברסיטאות, כלומר כתיבה אקדמית. כדי להגדיר מהו ז'אנר זה שאלו Bridgeman and Carlson (1983) מרצים מהם המאפיינים והדרישות של הכתיבה הנדרשת מן הסטודנטים במחלקות שונות ובאוניברסיטאות שונות. המחקר התמקד בדרישות מסטודנטים לתארים מתקדמים במחלקות: מנהל עסקים, הנדסה אזרחית, הנדסת חשמל, פסיכולוגיה, כימיה ומדעי המחשב. את הסטודנטים לתואר ראשון ייצגו סטודנטים מהחוג לאנגלית מאחר שנמצא כי רוב הכתיבה נעשית בקורסים לאנגלית. במחקר יש נתונים מ-190 מחלקות ב-36 אוניברסיטאות. מאחר שהמחקר נעשה בידי הצוות האחראי על מבחן TOEFL הוא אסף נתונים רק ממוסדות שבהם מספר גדול של סטודנטים זרים. המחקר התמקד בדרישות הכתיבה שמעמידים החוגים השונים לפני הסטודנטים, דרישות אשר אינן נקבעות בלעדית לאוכלוסיית

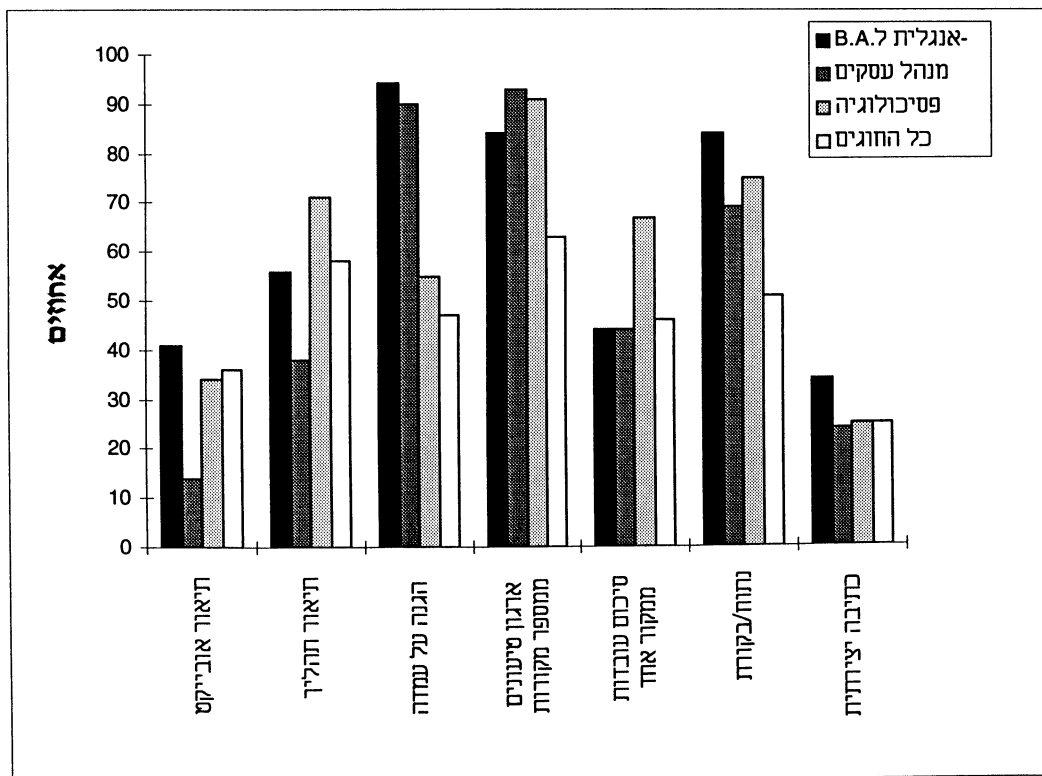
הסטודנטים הזרים, ולכן אפשר ללמוד מתוצאות מחקר זה על דרישות הכתיבה שמעמידים מוסדות אקדמיים לפני כלל הסטודנטים. להלן מוצגות ההערכות של מרצים מחוגים לאנגלית לתואר ראשון, מנהל עסקים, פסיכולוגיה לתואר מתקדם וממוצע כל החוגים (אשר פרט לני"ל כולל גם הנדסות, כימיה ומדעי המחשב).

סוגי המשימות האקדמיות הדורשות כתיבה בחוגים השונים

סוגי המשימות האקדמיות הדורשות כתיבה השכיחות בכל החוגים היו: סיכום קצר של מאמר, עבודות מחקר קצרות וארוכות, וכן מבחנים הכוללים כתיבה. דוחות מעבדה או ניסוי ניתנים בעיקר בחוגי ההנדסה השונים, בכימיה ובמדעי המחשב. כתיבה אקספוזיטורית/ביקורתית נדרשת בעיקר בלימודי אנגלית לתואר ראשון. ניתוחי מקרה (case studies) ניתנים בעיקר במנהל עסקים.

חשיבות מיומנויות כתיבה שונות להצלחה בשנה הראשונה ללימודים

החוקרים הגדירו שבע מיומנויות: תיאור אובייקט, תיאור תהליך, הגנה על עמדה, ארגון טיעונים ממספר מקורות, סיכום עובדות ממקור אחד, ניתוח והערכה וכתיבה יצירתית. המרצים התבקשו לדרג לגבי כל מיומנות את חשיבותה להצלחה בשנה הראשונה ללימודים על סולם בין 1 (חשיבות נמוכה) לבין 5 (חשיבות גבוהה). תרשים מספר 1 מייצג את אחוז המחלקות אשר דירגו מיומנות זו בציון 4 ומעלה. מן התרשים עולה כי כתיבה יצירתית כמו גם תיאור אובייקט אינם נתפסים כחשובים ברוב המחלקות. ואילו הגנה על עמדה, ארגון טיעונים ממספר מקורות וניתוח וביקורת דורגו כחשובים ביותר בעיקר בקרב מרצים לאנגלית לתואר ראשון, מנהל עסקים ופסיכולוגיה.



תרשים מספר 1: חשיבות מיומנויות כתיבה שונות - אחוז המחלקות אשר דירגו כל מיומנות בציון 4 ומעלה, מתוך (Bridgeman & Carlson 1983)

ניתוח המשימות על פני מימדים

המרצים התבקשו לדרג כל אחת מ- 10 משימות הכתיבה המופיעות בעמוד 10, כמשימה טובה, משימה סבירה, ומשימה לא מתקבלת למטרות של השמה של סטודנטים זרים. החוקרים בחנו את דפוסי ההערכות של המשימות השונות, והסיקו מכך על מידת הדמיון שבין המשימות. בהמשך נותחו הנתונים באמצעות תוכנית לסילוס רב-מימדי (SINDSCAL) והניתוח העלה שני מימדים ראשיים:

1. מעורבות אישית - המידה בה נדרש הכותב לבטא את עמדתו האישית במשימת הכתיבה. לדוגמה, "סכם קטע" דורש מעורבות מועטת כזו, ואילו "השווה ונקוט עמדה" דורש מעורבות רבה.
2. מורכבות - המידה שבה משימת הכתיבה מבוססת על תהליכי חשיבה וניתוח מורכבים. למשל, משימת תיאור דורשת חשיבה וניתוח מועטים, ואילו משימות כמו השוואה או אקסטרפולציה דורשות חשיבה וניתוח ברמות גבוהות.

דגמים של כתיבה עיונית

רוזנר (1985) מציינת ארבעה דגמים של כתיבה עיונית. הדגם הראשון הוא דגם של "כתיבה פתוחה". הנבחן נדרש לכתוב חיבור על נושא מוגדר, אולם ללא חומר עזר וללא עיגון בהקשר חברתי כלשהו. לדגם כתיבה זה מספר חסרונות: הוא אינו אותנטי, הוא תלוי במידה רבה בידע בתחום הדעת ובאינטליגנציה, הוא אינו מוגדר מספיק ולכן יוצר בעיות בשלב ההערכה. הדגם השני הוא החיבור המובנה - זוהי כתיבה על נושא כלשהו המעוגנת בסיטואציה חברתית והמכוונת לנמען מסוים. דגם זה הוא אותנטי יותר, ממוקד יותר ולכן מאפשר הערכה מהימנה יותר. אולם הוא עדיין מבוסס במידה רבה על ידע ועל אינטליגנציה והוא דורש מן הנבחן לבצע סימולציה מסויימת של הסיטואציה החברתית המוגדרת במטלה. הדגם השלישי הוא דגם של מאמר תגובה. הנבחן מקבל קטע קריאה ועליו לכתוב חיבור המתייחס לקטע זה. הדגם הרביעי דורש מן הנבחן לבנות מאמר חדש על סמך שילוב של רעיונות מתוך מספר מקורות הניתנים לו על ידי הבוחן. יתרונם העיקרי של שני דגמי הכתיבה האחרונים הוא באותנטיות שלהם, ובכך שהם אינם נשענים במידה רבה על ידע בתחום הדעת. חסרונם: הנבחן נדרש להפגין מידה מעטה יחסית של יצירתיות ודמיון. לטענת רוזנר (1985, 1990) שני הדגמים האחרונים הדורשים שילוב של הבנה והבעה הם המומלצים לבחינה הבודקת יכולת כתיבה עיונית בגלל שלהם מידת הדמיון הרבה ביותר למשימות הנדרשות מן הכותב העיוני.

נושא החיבור ומרכיב הידע

מבחן כתיבה בדרך כלל מכוון את הנבחן לכתובה בנושא מסוים, ולכן במידה רבה הוא מודד לא רק כושר כתיבה אלא גם התמצאות בעולם תוכן מסוים. מרכיב הידע מתבטא במשימת הכתיבה בשני אופנים: בעולם התוכן שבו יעסוק החיבור ובמידה שבה כתיבת החיבור תלויה בידע קודם. אפשר להגדיר ארבע רמות מבחינת רמת הידע הנדרשת על ידי מבחן הכתיבה:

1. מטלת כתיבה הבודקת יכולת הבעה בכתב של ידע ספציפי בעולם תוכן מסוים, למשל, "השווה את הגופים המחוקקים של מדינת ישראל לעומת אלה של ארצות הברית". משימה כזו מתאימה בייחוד למבחני הישג, וזו המטלה השכיחה ביותר במבחני כיתה בבית הספר ובאוניברסיטה. שאלות מסוג זה הן חלק ממבחן SAT stage II שהוא מבחן הישגים של ה-SAT (מבחן מיון לקולג'ים בארצות הברית).
2. משימת כתיבה המתייחסת לתחום מסוים באופן כללי, ומאפשרת לנבחן לבחור מהו הידע שישתמש בו לכתיבת החיבור. לדוגמה, "חברה המנסה למקסם הן את החופש והן את השוויון של אזרחיה מנסה לעשות את הבלתי אפשרי. צעדים הנקוטים כדי להבטיח חירות בוודאי ייצרו חוסר שוויון, וצעדים הנקוטים כדי לעודד שוויון יהיו על חשבון החירות. באיזו מידה אתה מסכים או מתנגד לעמדת הדובר שיש מתח בין חירות לשוויון? חזק את עמדתך בדוגמאות ספציפיות מהידע שלך בהיסטוריה או מאירועים עכשוויים". במשימה זו יש הסתמכות על ידע בתחום מסוים, אבל הניסוח הכללי של השאלה אינו מגדיר באופן ספציפי את הידע הנדרש, אלא מאפשר לכותב להשתמש בכל ידע שיש לו בתחום. לכן, יש כאן בחינה של ידע, אולם באופן רופף יותר מאשר בסעיף 1. זו השיטה שכנראה יאמץ ה-ETS בחיבור מבחן הכתיבה שיצורף ל-GRE: הוא יכלול שאלות במדעי הרוח, החברה והטבע, ויאפשר לנבחן לכתוב חיבור על נושא אחד מתוך שלושה לפי בחירתו.
3. משימת כתיבה דומה לזו המוזכרת בסעיף 2, אולם כללית יותר, שאינה מתמקדת בתחום מסוים ומתירה לנבחן לבחור הן את הידע הספציפי והן את התחום הכללי שעליו יכתוב. למשל, "כמה שינויים או המצאות שתוכננו כ"שיפורים" התגלו כבעלי השפעות בלתי צפויות או שליליות. חשוב ובחר שינוי אחד כזה, למשל במוצר, במכונה, בתהליך, במדיניות, או בארגון. תאר בקיצור את המצב לפני השינוי, הסבר את השיפור המתוכנן ואת ההפסדים שנבעו מן השינוי". יתרונו של ניסוח כללי כל כך של מטלת הכתיבה הוא בהקטנת ההשפעה של הידע על הביצוע, אולם המחיר יכול להיות במרחב הרב שהוא משאיר לפרשנות. ייתכן כי יהיו הבדלים רבים באופן הבנת השאלה, ולכן ייווצרו בעצם נוסחים שונים של אותה הבחינה.
4. ביסוס משימת הכתיבה על מידע נתון - בראשיתו של מבחן כזה ניתן לנבחן מידע בקטע קריאה, תרשים או טבלה ועל הנבחן לכתוב חיבור המתייחס לאינפורמציה שניתנה לו. כאשר ניתן לנבחן קטע קריאה השאלות יכולות לבקש מן הנבחן לסכם את הנאמר בטקסט, להביע את עמדתו באשר לנאמר בטקסט, ליישם את הנאמר על תחומים אחרים וכו'. לחילופין, אפשר לכלול כמה קטעים קצרים המתייחסים לאותו הנושא, והנבחן יתבקש בחיבור לעשות אינטגרציה בין הרעיונות המובאים בקטעים השונים, ולהעריך אותם. יתרונו של שיטה זו הוא בצמצום ההשפעה של ידע בעולם תוכן מסוים על הציון שיקבל הנבחן בכתיבה. נוסף על כך, היא קשורה במישרין למטלות אשר הסטודנט נדרש לבצע באוניברסיטה - לקרוא טקסט, לסכמו ולהעריך אותו. אחת הבעיות בשיטה זו היא שהכוונה היא שהנבחן יוכל להתבסס רק

על מידע הניתן בקטע הקריאה, אולם הדבר ניתן למימוש בעיקר בתחומים כמו מדעי הרוח והחברה. גם אנשים ללא כל רקע בתחום יוכלו להבין קטעים העוסקים בנושאים אלה. קשה הרבה יותר לבסס חיבור על קטע קריאה במדעי הטבע אשר יובן היטב על ידי אנשים חסרי ידע בתחום. מאפיין חשוב של שיטה זו הוא היעדר הבחנה ברורה בין מדידת הבנת הנקרא למדידת יכולת הכתיבה, ובכך היא מבטאת את עקרונות הגישה האוריינית. בבחינה אשר מתעדת לבחון יכולת כתיבה בלבד הדבר יוצר מידה של חוסר צדק. הבנת הקטע היא תנאי מקדים הכרחי להצלחה בכתיבה החיבור. מי שאינו מבין את הקטע, לא ניתנת לו האפשרות להפגין את יכולת הכתיבה שלו. במבחן הבגרות בהבעה יש ניסיון להתמודד עם בעיה זו באופן הבא: ניתנות מספר שאלות הבודקות את הבנת הנקרא, בהמשך ניתן נושא לחיבור, ונאמר לנבחנים כי הם יכולים, אך אינם חייבים, להתבסס על החומר בקטע הקריאה. מתן אינפורמציה כרקע לחיבור בצורה של גרפים או טבלאות אף מכניסה להערכת יכולת הכתיבה גם את הכשרים הכמותיים.

לא נמצאו מחקרים אשר בדקו את ההשפעה של מידת הידע הנדרש לכתיבת החיבור על הביצוע בו, ונראה כי הבחירה בסוג המבחן תלויה במידה רבה במטרתו; אם הכוונה היא למיין מועמדים לאוניברסיטאות על פי יכולתם בלבד ולא על סמך ידע בתחומים מסוימים אזי נראה כי מבחני כתיבה מסוג 3 או 4 הם המתאימים ביותר. לעומת זאת אם הכוונה היא למיין מועמדים לתארים מתקדמים הרי יש הצדקה לדרוש מן הנבחן להפגין גם ידע בתחומים ספציפיים (כנדרש במבחנים מסוג 1 או 2).

השפעת ז'אנר הכתיבה ונושא החיבור על הביצוע במבחני כתיבה

מטלות הכתיבה שמבצע הסטודנט באוניברסיטה מגוונות מבחינת ז'אנר הכתיבה הנדרש בהן, והן מבטאות כמה מיומנויות כתיבה. השאלה היא עד כמה אפשר להסיק מביצוע במשימת כתיבה מז'אנר כתיבה אחד על ביצוע במשימת כתיבה מז'אנר אחר. מחקרים אשר בדקו סוגיה זו העלו כי קיימים הבדלים רבים ביכולת הכתיבה של הנבחנים כפי שהיא מוערכת על סמך משימות כתיבה בז'אנרים שונים ובנושאים שונים (Boodoo & Garlinghouse, 1983; Breland et al, 1987; Moss, Cole and Khampalikit, 1987; Quellmalz, Capell & Chou, 1982). מחקרים אלה ידונו בהרחבה בפרק הדין במהימנות מבחני כתיבה. בפרק זה יידונו ההשלכות של ממצאים אלה על הרכבת הבחינה.

השלכות על פיתוח מבחני כתיבה

מחקרים הראו כי ביצועי הנבחנים אינם יציבים כאשר הם כותבים חיבורים על נושאים שונים ובז'אנרים שונים של כתיבה, לכן יש יכולת הכללה מוגבלת מביצוע במטלת כתיבה אחת. בסעיף זה יוצעו שתי דרכים להתמודד עם בעיה זו: מתן בחירה בין משימות כתיבה ומתן מספר משימות כתיבה.

מתן בחירה

אפשר להתמודד עם השפעת מרכיב הידע ונושא הכתיבה על הביצוע במבחן הכתיבה באמצעות מתן בחירה: להציג לפני הנבחנים מספר משימות כתיבה מתחומים שונים תוך הנחיה לבצע משימת כתיבה אחת מתוך המוצעות. ההנחה היא שהנבחנים יבחרו לכתוב על הנושא שבו הם בקיאים ומרגישים כי יש להם מה לומר לגביו.

יתרונות אפשריים של בחירה:

בחירה נותנת לנבחן הרגשת שליטה ותחושה כי קיימים התנאים שיאפשרו לו להפגין את יכולתו האופטימלית.
מתן בחירה יכול לנטרל השפעה של ידע מוקדם, כי הוא מקטין את ההסתברות שנבחן יידרש לכתוב על נושא לא מוכר לו.

חסרונות אפשריים:

העובדה שאנשים כותבים חיבורים על נושאים שונים יוצרת בעצם כמה נוסחי בחינה, והשאלה היא כיצד אפשר לוודא כי נוסחי בחינה אלו הם ברמת קושי שווה. לא ברור כיצד להשוות רמת קושי של משימות כתיבה בנושאים שונים.
אין לנו ידע בדבר השיקולים המנחים את הנבחן בבחור את הנושא שעליו יכתוב, ולכן אין לנו ביטחון כי הנבחנים יבחרו דווקא בנושא שבו הם בקיאים ביותר (White, 1985).

מחקר של ה-ETS חקר את סוגיית מתן בחירה בין נושאים במבחן הכתיבה באמצעות ניסוי (Powers, Fowles, Farnum, & Gerritz, 1992). בשלב הראשון הוצגו ל-491 סטודנטים לתואר ראשון 20 נושאים לחיבורים והם נתבקשו לדרג את העדפתם על סולם בין 1 ל-10. בשלב הבא הוצגו להם שני זוגות של נושאים הלקוחים מתוך ה-20 ובכל פעם על כל סטודנט היה לבחור נושא אחד מבין השניים ולכתוב עליו חיבור. בסך הכול היו במאגר השאלות 10 זוגות של נושאים: 3 הוגדרו חוויות אישיות, ו-7 האחרים הוגדרו נושאים כלליים. כל סטודנט כתב שני חיבורים, כאשר כל חיבור נבדק בידי זוג בוחנים בשיטה הוליסטית (המבוססת על התרשמות כללית).
כאשר נבדקו דירוגי ההעדפות של הסטודנטים נמצא כי כל נושא זכה לטווח רחב של ציוני העדפה, כלומר הוא נבחר בידי אחוז מסוים של האנשים כמועדף ביותר ובידי אחרים כמועדף פחות. למרות זאת היו נושאים שהיו מועדפים יותר מאחרים והם:

1. תאר את הקורס המעניין ביותר שלמדת.
2. תאר לעצמך שאתה יכול לבלות יום אחד עם כל אדם שתבחר מכל תקופה. את מי תבחר ואיך תבלה את היום?
3. האם אתה תומך בתלבושת אחידה? דון בעמדתך.
4. מהי עמדתך לגבי מי שמציעים לקיים שנת לימודים של 12 חודש בארצות הברית? הנושאים המועדפים פחות היו:

1. באיזו מידה אתה מסכים או מתנגד לעמדה שהבעיה של הצעירים האמריקניים היא שהם אינם רואים את העבודה כעניין המרכזי של הקיום שלהם.

2. מה אתה צופה שיהיה אחד הכוחות או התנועות הדומיננטיים בשנות ה-90? דון באלמנט זה ובאפקט שאתה צופה שיהיה לו על ארצות הברית.

3. תאר מקרה שבו אתה או אדם אחר שאתה מכיר היה צריך לעשות מעשה הקרבה, והסבר כיצד פעולה זו תרמה להתבגרותך.

בהמשך נבדק הקשר בין העדפת נושאים לבין בחירה לכתוב עליהם, ונמצא קשר חיובי; ציון העדפה גבוה יותר היה מלווה בדרך כלל באחוז גבוה של נבדקים שבחרו לכתוב על נושא זה. נמצא כי הבדלים קטנים בציון ההעדפה היתרגמו להבדלים גדולים במידה שבה נבחנים בחרו את הנושא מתוך זוג הנושאים. למשל בזוג מסוים של נושאים היה הבדל של 0.9 בציוני ההעדפה שניתנו על סולם בן 10 דרגות, אולם הנושא המועדף נבחר פי 4 יותר מהנושא האחר ($N=413$).

הקשר בין העדפת נושאים לבין רמת הכתיבה

לכל נבחן השוו את הציון שקיבל על החיבור בנושא לו נתן דירוג העדפה גבוה יותר לעומת הציון שקיבל על החיבור בנושא לו נתן ציון העדפה נמוך יותר. הפער הממוצע בין ביצוע על הנושא היותר מועדף לפחות מועדף היה קטן מאוד, 0.11 על סולם של 6 דרגות לזכות החיבור על הנושא המועדף. סטיית התקן של הפערים הייתה 0.93. 36% מהנבחנים קיבלו ציון גבוה יותר על הנושא שהעדיפו, 33% קיבלו ציון גבוה יותר על הנושא האחר, ואצל 31% לא היה הבדל. לסיכום, לא נמצא קשר מובהק בין ציונים על החיבורים לבין מידת ההעדפה; כלומר לא נמצא כי נבחנים הצליחו יותר בחיבורים על נושאים אותם העדיפו ($N=413$).

קורלציות עם משתנים נוספים:

המתאם בין ציוני החיבורים בנושאים המועדפים ובין ציוני החיבורים בנושאים הלא מועדפים היה 0.52 ($p < 0.1$). חושבו מתאמים בין הציונים על החיבור בנושא המועדף יותר והציונים על החיבור בנושא המועדף פחות לבין המשתנים הבאים: ממוצע הציונים בקולג', מספר הקורסים שלמד הנבחן אשר דרשו כתיבה, וכן ציוני SAT מילולי וכמותי, והם מובאים בטבלה הבאה:

הקריטריון	N	מתאם עם ציון על הנושא הפחות מועדף	מתאם עם ציון על הנושא היותר מועדף
ממוצע ציוני תואר ראשון	411	0.27	0.32
מספר הקורסים אשר כללו משימות כתיבה שלקח הסטודנט במהלך לימודיו	373	0.28	0.24
SAT מילולי	312	0.47	0.51
SAT כמותי	312	0.30	0.29

טבלה מספר 1: מתאמים בין ציוני הכתיבה לבין מדדים אקדמיים, (Powers et al, 1992)

לא נמצא הבדל מובהק בין המתאמים של ציונים על החיבור בנושא המועדף עם הקריטריונים השונים לבין המתאמים של ציונים על החיבור בנושא הפחות מועדף עם אותם הקריטריונים. מחברי הדוח הסיקו מהיעדר הבדל זה כי מתן בחירה לא יפגע בתוקף המבחן, ולכן לדעתם יש לאפשר בחירה.

White (1985) מתנגד למתן בחירה, מאחר שלטענתו רמת הקושי של הנושאים השונים לעולם אינה שווה. לעתים קרובות, לדעתו, השאלות הקשות ביותר הן גם המעניינות ביותר, והן מושכות בעיקר את הסטודנטים הטובים. בשל כך סטודנטים טובים אשר היו מקבלים ציון טוב מאוד על נושא א', בחרו את נושא ב' בגלל העניין שבו, וקיבלו בו ציון גרוע. סטודנטים גרועים, לעומת זאת, יבחרו בשאלות הקלות יותר ויקבלו ציונים טובים. סטודנטים גרועים, אשר אינם יודעים להעריך את קושי השאלות, עלולים לבחור בנושא קשה באופן מקרי ועקב כך לקבל ציונים נמוכים יותר. ההצדקה של מתן בחירה בנושאי החיבור נשענת על ההנחה שסטודנטים בוחרים בנושא אשר מתאים להם ביותר, אולם לדעת White אין עדות לכך שהסטודנטים בוחרים בנושא שעליו יכתבו באופן הטוב ביותר. White (1985) מציע לבנות אלמנט של בחירה בתוך השאלה עצמה כמו למשל בפריט הבא: "תאר אובייקט מסוים שאתה מעריך, והסבר למה אתה מעריך אותו". בשאלה זו הנבחן יכול לבחור מהו האובייקט שעליו יכתוב.

כיצד נוהגים במבחני כתיבה אמריקניים המשמשים למיון או להשמה? מבחן ה-GRE מציע נושאים לשלושה חיבורים: במדעי הרוח, במדעי החברה ובמדעי הטבע, ועל הנבחן לכתוב חיבור אחד בנושא לפי בחירתו. המבחן של אוניברסיטת קליפורניה, לעומת זאת, אינו מציע בחירה, וכמוהו גם מבחן ה-SAT.

מתן מספר משימות כתיבה

דרך נוספת להתמודד עם השפעת הנושא וז'אנר הכתיבה על איכות הכתיבה היא לתת כמה משימות כתיבה מעולמות תוכן שונים ומז'אנרים שונים של כתיבה. מבחן כזה צפוי להיות מהימן יותר, ולתת ייצוג מקיף יותר של יכולת הכתיבה של הנבחן. החסרונות העיקריים של שיטה זו הם בעיקרם פרקטיים - הדבר יאריך את משך הבחינה ואת תהליך הצינון וייקר אותם. כפשרה אפשר לתת כמה מטלות כתיבה קצרות, אולם ספק אם כתיבת חיבור קצר דומה לכתיבת עבודות ארוכות בזמן ממושך. ייתכן כי רק חיבור ארוך יותר אשר בו אפשר לפתח דיון מעמיק יותר ייתן אומדן מהימן ליכולת הכתיבה של הנבחן. באשר לקריטריון, ולמטלות הכתיבה העומדות לפני הסטודנט באוניברסיטה נראה כי הסטודנט נדרש למטלות משני הסוגים: במבחני כיתה החיבורים הם בדרך כלל קצרים יחסית, ואילו במבחני בית ובעבודות הדרישה היא לכתיבה ארוכה יותר, הנשענת בדרך כלל על ביבליוגרפיה.

ניסוח משימת הכתיבה

לאחר שמוסכם מהו ז'אנר הכתיבה שאותו מעוניינים לבחון במבחן הכתיבה, ומהי רמת הידע הנדרשת, יש לנסח את משימת הכתיבה הספציפית. בניסוח משימת הכתיבה יש לשים לב למאפיינים הבאים:

א. בהירות - על הכותרת להיות מנוסחת בבירור, כדי שהנבחנים יבינו מה נדרש מהם מבחינת נושא החיבור, סגנון הכתיבה ואורך החיבור הרצוי (White, 1985).

- ב. הבניה - כדאי להבנות את משימת הכתיבה על ידי ציון מטרת החיבור (למשל תיאור, ניתוח וכו') ותיאור הקורא המשוער (פיקהולץ, 1984).
- ג. ספציפיות - בנקודה זו קיימים חילוקי דיעות בספרות. Miller and Crocker (1990) טוענים כי הנושא צריך להיות ספציפי ומוגדר. נושא פתוח או עמום המאפשר להבינו בכמה דרכים יגרום לשונות רבה בנושאים שבהם יתמקדו הכותבים בחיבוריהם, דבר שיקשה על הציינון. White (1985), לעומתם, המליץ על ניסוח פתוח של השאלה כדי לצמצם את ההישענות על ידע ספציפי.
- ד. עניין - על הנושא להיות מעניין ומעורר מחשבה כדי שיתאפשר דיון רציני ומעמיק, דבר שייצור שונות גדולה בין החיבורים. כמו כן נושא מעניין יעודד את הסטודנטים להשקיע מחשבה בכתיבה, ויקל על עבודת המעריכים.
- ה. פתיחות - Fowles and Fiero (1990) מוסיפים כי כדאי שמטלת הכתיבה תעלה נושא אשר בו נקודות מבט אחדות, ושאפשר לנקוט לגביו יותר מעמדה אחת. לא מומלץ לנסח מטלות כתיבה בנושאים שבהם רק עמדה אחת ניתנת להצדקה באופן רציני, או שהכותב יחשוד שיש תשובה מסוימת שהיא יותר מקובלת חברתית מאחרות. כמו כן כדאי להימנע מנושאים מעוררי מחלוקת בחברה או מנושאים שהם טעונים מדי מבחינה רגשית.
- ו. הטיה - יש להימנע מנושאים אשר נותנים יתרון לקבוצת נבחנים בעלי ידע או ניסיון מסוים.
- לאחר ניסוח משימות הכתיבה יש לבדוק אותן באופן ניסיוני על ידי העברתן למדגם של נבחנים. בשלב זה עולה השאלה של מהן משימות הכתיבה הטובות. Carlson, Bridgeman, Camp & Waanders (1985) הציעו את הקריטריונים הבאים להערכה של משימות הכתיבה עצמן:
1. האם הכותבים הבינו את הנושא?
 2. האם הכותבים ענו לנושא באופן ישיר, והאם ביצעו את ההוראות?
 3. האם החיבורים היו די מגוונים, והאם הם מראים כי הנושא היה רחב או צר מדי?
 4. האם נראה כי הכותבים הספיקו לסיים את חיבוריהם כראוי בזמן המוקצב?
 5. האם יש עדויות לקשיים בהבנת קטע הקריאה או התרשים אליהם היה צריך להתייחס בחיבור?

מקבילות משימות הכתיבה

משימות כתיבה אשר ניתנות לנבחנים במועדים שונים צריכות להיות מקבילות זו לזו ככל האפשר. לכן, אם מטלות הכתיבה מבוססות על ציטטות או על קטעי קריאה יש לדאוג כי הם יהיו שווים באורכם וברמת הקושי שלהם (Koenig, 1995). בתהליך פיתוח מבחן כתיבה למועמדים לבית ספר לרפואה (Koenig, 1995) אף הגביל את הכשרים הקוגניטיביים הנמדדים כדי לאפשר מקבילות בין משימות הכתיבה. הוחלט כי הכושר החשוב להם ביותר הוא ביצוע סינתזה, ועל הנבחנים להיות מסוגלים להנגיד שתי עמדות ולהציע דרך לפשר ביניהן. לכן, הוחלט על פורמט קבוע של משימות כתיבה אשר יכלול אזכור של שתי דעות נוגדות בעניין כלשהו. לצורך השגת

מקבילות עבר (Koenig (1995 ממדריך ציינון אשר חובר באופן ספציפי לכל משימת כתיבה לשימוש במדריך ציינון המתאים לכל המשימות, מאחר שציון הבנוי על מדריך המחובר למשימה אחת בלבד קשה להשוואה עם ציון המבוסס על מדריך אחר.

הזמן המוקצב

הזמן המוקצב לכתיבת החיבורים שונה במבחני הכתיבה השונים. במבחן הכתיבה של אוניברסיטת קליפורניה הזמן המוקצב לכתיבת חיבור אחד הוא שעתיים. הוא כולל גם את זמן הכתיבה וגם את זמן הקריאה של הקטע עליו מתבסס החיבור. קטע הקריאה הוא באורך שני עמודים ברווח יחיד. על הנבחנים לקרוא את הקטע ולכתוב חיבור שנושאו קשור לתוכן הקטע.

מבחן הכתיבה של ה-GRE מקציב לכתיבת חיבור שעה אחת. במבחן אין קטע קריאה. ניתנים 3 עמודים לכתיבת החיבור. בבחינות הבגרות בהבעה הזמן המוקצב לכתיבת חיבור אחד הוא בין שעתיים לשלוש שעות (תלוי בסוג הבחינה).

(Breland et al. (1987 הקציבו במחקרם 45 דקות לכתיבת חיבור תיאורי או הסברי. במבחן Test of Written English (TWE הניתן לאנשים אשר אינם דוברי אנגלית כשפת אם ואשר נלווה למבחן ה-TOEFL מוקצבת חצי שעה לכתיבת החיבור. מחקר שנעשה ב-ETS בדק את ההשפעה שתהיה להארכה של הזמן מ-30 ל-45 דקות (Hale, 1992), ואלה ממצאיו:

מהימנות בין מעריכים - לא נמצא הבדל במהימנויות בין חיבורים אשר נכתבו בהקצבות זמן שונות. הדבר נמדד על ידי השוואת המתאמים בין הערכות של בודקים שונים והשוואת הפערים בהערכות בודקים שונים.

מהימנות בין משימות כתיבה שונות - השוו את המתאמים בין שני חיבורים על נושאים שונים, האחד נכתב בזמן מוקצב של 30 דקות והאחר בזמן מוקצב של 45 דקות, למתאמים בין שני חיבורים על נושאים שונים שנכתבו שניהם ב-30 או ב-45 דקות, והמהווים מדד למהימנות מבחן מקביל. נמצא כי המתאם בין חיבורים אשר נכתבו ב-30 דקות לבין אלה אשר נכתבו ב-45 דקות אינו נמוך יותר ממהימנות מבחן מקביל. כלומר, אנשים דורגו באופן דומה בין אם ניתנו להם 30 דקות ובין אם ניתנו להם 45 דקות לכתיבת החיבור. כמו כן לא היה הבדל במהימנויות מבחן מקביל בין חיבורים אשר נכתבו ב-30 דקות לעומת חיבורים אשר נכתבו ב-45 דקות.

רמת ביצוע - כאשר הוקצו לנבחנים 45 דקות לכתיבת החיבור רמת הביצוע שלהם הייתה גבוהה יותר בכשליש סטיית תקן. לכן נראה היה שהמבחן הארוך יותר נתן לנבחנים הזדמנות לבטא בצורה טובה יותר את יכולתם. אף שהייתה עלייה בביצוע של כל הנבחנים, לא היה שינוי בדירוג שלהם האחד ביחס לשני. כמו כן העלייה בביצוע לא הושפעה מרמת הכתיבה של הנבחן; מתוספת הזמן נהנו גם תלמידים בעלי רמת כתיבה גבוהה וגם תלמידים בעלי רמת כתיבה נמוכה. המסקנה של החוקרים הייתה כי די בהקצבת זמן של 30 דקות לכתיבת החיבור. חשוב לציין כי מחקר זה בדק יכולת כתיבה בשפה זרה ולא בשפת אם.

פרק ב': הערכה וצינון

פרק זה יתמקד בנושא הערכת החיבורים וצינונם. בתחילה יוצגו שיטות הצינון הקיימות, תוך פירוט היתרונות והחסרונות של כל שיטה, ויוצגו הקריטריונים לבחירה בין שיטות הצינון. בהמשך הפרק יוזכרו מימדים שונים של החיבור שיכולים לשמש בסיס להערכה. החלק האחרון של הפרק יכלול ממצאים אמפיריים על עמדות מרצים לגבי שיטות ההערכה, על הקשר בין שיטות הצינון הקיימות, וכן סקירה קצרה של שיטות ההערכה המקובלות במבחני כתיבה שונים בארץ ובעולם.

שיטות הצינון

צינון הוליסטי

שיטת צינון זו יוצאת מהגישה התאורטית הרואה את הכתיבה כיכולת אחת, אשר אי אפשר או לא רצוי לעשות רדוקציה שלה או למצות אותה בדרך של התמקדות באספקטים בדידים שלה. לפיכך, ההערכה של כל חיבור ניתנת על ידי ציון אחד בלבד. למעשה יש שיטות שונות של צינון הוליסטי; יש הטוענות כי הציון צריך להתבסס על התרשמות כללית ראשונית של הקורא, ואילו אחרות מכוונות את הקורא לכמה אספקטים שעל הציון לשקף, והציון הוא מעין שקלול אינטואיטיבי של איכות החיבור במימדים הרלוונטיים. הצינון ההוליסטי הוא יחסי בטבעו, אין מודל של החיבור המושלם אשר כל החיבורים מושווים אליו. בתחילת שלב הצינון הצוות האחראי על הערכת הבחינה קורא מדגם גדול של חיבורים ודן באיכותם עד שמגיע להסכמה בדבר הדירוג של החיבורים הללו. בשלב זה נקבע הבסיס לצינון, וחשוב להגיע להסכמה רחבה ככל האפשר. באמצעות דירוג זה נבנה סולם של ציונים המתבסס על איכות החיבורים בנושא המסוים ובאוכלוסיה המסוימת והוא משמש בסיס למדריך הצינון. השכיחים ביותר הם הסולמות בעלי 5-7 דרגות. בצינון הוליסטי המבוסס על התרשמות כללית, המדריך לצינון בנוי רק ממספר חיבורים המייצגים את הרמות השונות על פני סולם הציונים. בצינון הוליסטי אשר מפרט את התכונות החשובות, המדריך מציין את התכונות הנדרשות מחיבור בכל דרגה על סולם הציונים. תפקידו של המדריך הוא להציג את הסטנדרטים לצינון באופן מפורש וכמו כן לעודד את השימוש בכל הטווח של הסולם בעזרת מתן דוגמאות למקרים קיצוניים. המדריך לצינון נכתב בכל פעם מחדש לפי נושא החיבור והאוכלוסיה הרלוונטיים. אין להשתמש במדריך שחובר לחיבור לאוכלוסיה שונה מאחר שזה הופך את הצינון-לצינון קריטריון, או במדריך שחובר לנושא אחר, מכיוון שזה מניח שהנושאים הם ברמת קושי שווה (White, 1985). שיטת הצינון ההוליסטית היא הנפוצה ביותר בפרוייקטים רחבי היקף של הערכת יכולת כתיבה.

יתרונות:

- לשיטה זו יתרונות פרקטיים חשובים: היא קלה יחסית למציין, ההערכה נעשית לאחר קריאה אחת בלבד של החיבור, ולכן דורשת זמן מועט.

חסרונות:

- ציינון הוליסטי המכוון את המעריך למספר מימדים דורש בעצם סיכום סובייקטיבי של הערכות על מימדים אלו. במקרה זה ייתכן רעש הנובע מהבדל במשקלות אשר נותנים מעריכים שונים למימדים השונים, ואשר נותן אותו מעריך בנקודות זמן שונות (נבו ועמיתיו, 1989).
- קל להגיע להסכמה בדבר חיבורים אשר קיבלו ציון גבוה ביותר (הם טובים בכל המימדים) ובדבר חיבורים שקיבלו ציון נמוך ביותר (הם גרועים בכל המימדים), אולם לא תמיד קל להגיע להסכמה בדבר הציון ההוליסטי שיש לתת לחיבור אשר הוא טוב מאוד במימד מסוים וגרוע לאחר. ייתכנו הרבה צירופים של חיבורים ברמות שונות על מימדים שונים, ומדריך הציינון אינו מורה במפורש כיצד לבצע את השקלול, דבר שיכול להיות מקור להרבה רעש במדידה.
- יש קושי להשוות בין ציונים של אותם אנשים על חיבורים בנושאים שונים ובין ציונים של אנשים שונים על אותו נושא.

ציינון על פי תכונה ראשונית

ציינון על פי תכונה ראשונית מעוגן בתאוריה הרואה את הכתיבה כאקט של קומוניקציה, וככזה הוא בא למלא פונקצייה מסוימת. לדוגמה, חיבור מסוים מטרתו לשכנע, אחר מטרתו להעביר מידע. הציון שמקבל החיבור מבוסס על המידה שבה הוא השיג את מטרתו. בדומה לציינון הוליסטי גם כאן יש ציון על מדד אחד, אולם בהבדל מן הציינון ההוליסטי, יש כאן הגדרה מדויקת ומוגדרת של מימד זה. ההערכה אינה מורכבת מאינטגרציה של אספקטים שונים, אלא מתמקדת במימד המרכזי, וכל חיבור מצוין רק על מימד זה. הציינון אינו מתמקד במאפיינים החשובים לכל סוגי הכתיבה, כמו למשל שטף או בהירות, אלא רק במאפיינים הנחוצים להשגת המטרה הספציפית של משימת הכתיבה.

יתרונות:

- שיטת ציינון זו לוקחת בחשבון את ההבדלים בין מטלות הכתיבה השונות.
- שיטה זו אינה דורשת שקלול סובייקטיבי של ציונים על מדדים שונים ליצירת ציון כולל.

חסרונות:

- שיטת ציינון זו נותנת משקל רב למשימת הכתיבה הספציפית ואינה רואה אותה כמייצגת את כל משימות הכתיבה. בשל כך יכולת ההכללה מביצוע במשימה אחת הוא מוגבל וכן קשה להשוות בין משימות כתיבה שונות.

ציינון אנליטי

הציינון האנליטי נשען על גישה הטוענת כי יש להתייחס אל הכתיבה כאל משימה הבנויה ממרכיבים שונים. בהתאם, הערכת הכתיבה צריכה להיות בנויה מהערכות של התכונות השונות בנפרד (למשל, רמת ארגון, סגנון). ציינון אנליטי מבוסס על זיהוי המימדים החשובים בכתיבה,

על אופרציונליזציה שלהם ועל מתן משקלות יחסיים למימדים השונים. לאחר קריאת החיבור המעריך נותן ציון על כל מימד; בדרך כלל כדי לציין את כל המימדים על המעריך לקרוא כל חיבור מספר פעמים. המימדים יכולים להיות מצוינים באופן כמותי או איכותי. ציון כמותי יכול לכלול ספירת טעויות כתיב, ספירת טעויות דקדוק, מספר המילים במשפט וכדומה והוא מכונה בדרך כלל ציון אטומיסטי. ציון איכותי כולל הערכה סובייקטיבית יותר של איכות התוכן או הסגנון והוא יכול להיות דיכוטומי או רציף. לבסוף מחושב הציון הכולל בעזרת שקלול סטטיסטי של הציונים על המימדים השונים. שיטה זו מניחה כי אפשר להעריך את המימדים הנפרדים באופן בלתי תלוי, וכי הציונים של המימדים השונים אינם משקפים התרשמות כללית או אפקט הילה.

בדרך כלל משתמשים באותם המימדים להערכה של משימות כתיבה שונות, אך אפשר גם לקבוע את המימדים הרלוונטיים לפי ז'אנר הכתיבה הנדרש. בניסוח המימדים האנליטיים יש לקחת בחשבון שתי דרישות מנוגדות (Witt, 1995). מצד אחד כדי להגיע למהימנות גבוהה בין שופטים יש לנסח את המימדים באופן מדויק וספציפי ומן הצד השני יש לנסח אותם באופן כללי ופתוח כדי שיתאימו למשימות כתיבה שונות. לטענת Witt (1995) אם משתמשים בשיטת ציון אנליטית יש לבנות את משימות הכתיבה תוך התחשבות בשיטה. יש לנסח משימות כתיבה אשר יפיקו שונות בחיבורים על פני כל המימדים ואשר יעמידו דרישות ברמת קושי דומה בפני הנבחנים. קיימות משימות כתיבה אשר "מתארגנות" בעצמן ואשר הארגון שלהן הוא קל יחסית (כמו משימת השוואה), ולכן לא צפויה בהן שונות במימד הארגון. משימות אחרות, כמו שכנוע, אינן מכוונות את הכותב לצורת ארגון מסוימת, ולכן צפויה בהן שונות רבה יותר בציונים על מימד זה.

יתרונות:

- יש קריטריונים ברורים לצייון, ודבר זה גורם לתהליך הערכה ממוקד וקפדני יותר (Quellmalz, 1986). הפירוק של הערכת החיבור למרכיבים עוזר לצמצם השפעה של גורמים לא רלוונטיים.
- ערך דיאגנוסטי - אפשר לתת לנבחנים משוב מפורט על הנקודות החזקות והחלשות בכתיבה שלהם.
- צירוף של אינפורמציה ממדדים שונים נעשה באופן מדויק ועיקבי.
- face validity - ציון כזה נראה אובייקטיבי והוגן יותר מציון הוליסטי. לנבחנים קל יותר להבין את המשמעות של ציון כזה.
- אם המימדים אינם תלויים בנושא החיבור הספציפי אפשר להשוות בין ציונים של חיבורים בנושאים שונים.

חסרונות:

- ציון אנליטי דורש זמן רב יחסית מאחר שיש צורך לקרוא כל חיבור מספר פעמים.

- השימוש באותם המימדים לצינון כל משימות הכתיבה מתעלם מההבדלים בין סוגי הכתיבה השונים (Cooper, 1977).

צירוף של ציינון אנליטי וצינון הוליסטי

שיטה זו מנסה ליהנות מהיתרונות של השיטה ההוליסטית והשיטה האנליטית על ידי שילוב של שתייהן. לאחר הקריאה הראשונה של החיבור ייתן המעריך ציון הוליסטי, ובהמשך הוא יתבקש לציין את החיבור גם על מדדים נפרדים. חשוב ביותר כי הציון ההוליסטי יינתן לפני הציונים הנפרדים, אחרת, יש סכנה שהמערכים, מתוך שאיפה לעקביות, יעשו מעין ממוצע משוקלל של הציונים הנפרדים, ובכך תאבד התרומה הייחודית של הציון ההוליסטי. מסיבה זו יש לאסור גם תיקון בדיעבד של הציון ההוליסטי. לשיטה זו עלות גבוהה ויש בה טעם רק אם נראה כי המימדים האנליטיים אינם במתאם גבוה בינם לבין עצמם ואינם במתאם גבוה עם הציון ההוליסטי.

הקריטריונים לבחירה בין שיטות הצינון השונות

- מטרת המבחן - אם אחת ממטרות המבחן היא להעביר לנבחנים משוב על ביצועיהם הרי שיש לבחור בשיטת ציינון אנליטית, שכן היא אינפורמטיבית יותר מהשיטות האחרות.
- משימת הכתיבה - ייתכן כי יש לציין משימות כתיבה שונות בדרכים שונות. נראה כי כתיבה ספרותית, למשל, קשה ולא רצוי לפרק למרכיביה ואילו משימות כתיבה מובנות יותר, כמו סיכום קטע, ניתנות יותר בקלות להערכה בכמה מימדים נפרדים.
- מהימנות בין שופטים - מאחר והסכמה בין מעריכים שונים לגבי אותו החיבור היא דרישה חשובה ביותר, יש לבחור בשיטת ציינון שתביא למהימנות הגבוהה ביותר בין שופטים.
- תוקף - אפשר להסתכל על סוגים שונים של עדויות לתוקף:
 - תוקף ניבוי - איזה ציונים ינבאו טוב יותר את הקריטריון שנקבע.
 - תוקף תוכן - איזו שיטה היא הערכה נכונה יותר של כושר הכתיבה.
- מובחנות המדדים - אם קיים מתאם גבוה בין המימדים השונים בצינון האנליטי הרי ייתכן כי בפועל הציון מבטא התרשמות כוללת, ואז עדיף להשתמש בשיטת ציינון הוליסטית.
- עלות של מערכת הצינון - מתבטאת בעיקר במשך הזמן שאורך תהליך ההערכה.

מימדי ההערכה

נוסף על השאלה של סוג הצינון המועדף, יש להגיע להסכמה בדבר המימדים שעליהם תתמקד ההערכה בין אם תהיה הוליסטית בין אנליטית. ברשימה הבאה מובאים מימדי הערכה אפשריים אשר נאספו ממחקרים בנושא.

רמת התוכן

הבנת הנושא

הבנת חומר העזר עליו מתבססת המשימה (אם קיים)

התייחסות לכל מרכיבי המשימה

התייחסות רחבה ומעמיקה לתוכן

מספר הרעיונות ואיכותם

מקוריות

רמת הטיעונים

רמת התימוכין

רלוונטיות הדוגמאות

מבנה

ארגון - סדר כתיבת הרעיונות

פתיחה וסיום

חלוקה לפסקאות התואמת את התוכן

קוהרנטיות - לכידות של כל הרעיונות בטקסט וקשר ברור שלהם זה לזה ואל משימת הכתיבה

פוקוס - שמירה על דיון ממוקד

סגנון כתיבה

התאמת הסגנון לסוג הכתיבה הנדרש ולקהל המיועד

בהירות

דיוק

שטף הכתיבה

דקדוק ותחביר

שימוש במשפטים מורכבים

בנייה נכונה של משפטים

שמירה על חוקי הלשון

אוצר מילים

שפה עשירה

שימוש באוצר מילים המתאים להבעה בכתב

איות נכון

שימוש מדויק במילים

בחירת המימדים הרלוונטיים להערכת החיבורים תלויה במטרת הבחינה. אם מטרתה היא לתת משוב על הוראת ההבעה בכתב הרי שיש להעריך את החיבורים על סמך המימדים הנלמדים. אם הכוונה היא לבדוק יכולת כתיבה בתחום מקצועי בלבד, כמו משפטים, הרי ששימוש בשפה עשירה

אינו מימד חשוב. אפשר להשתמש באותם המימדים לציון מטלות כתיבה שונות, או לבחור במימדים שונים לפי אופי משימת הכתיבה. למשל, מימד ההתייחסות לקהל היעד חשוב במשימת שכנוע יותר מאשר במשימת סיכום. אפשר גם לשנות את המשקלות של המימדים השונים לפי אופי משימת הכתיבה.

אחת הנקודות הבעיות בהערכת יכולת הכתיבה היא הקושי להבחין בין יכולת כתיבה לבין יכולת חשיבה המתבטאת על ידי תוכן הנכתב. מחד גיסא, המבחן נועד לבחון יכולת כתיבה ולא אינטליגנציה או מקוריות, ומאידך, אם מתעלמים באופן מוחלט מן התוכן הרי שאנו נותרים רק עם תמונה חלקית ביותר של יכולת הכתיבה של הנבחן, בעיקר על יכולתו הלשונית או הדקדוקית.

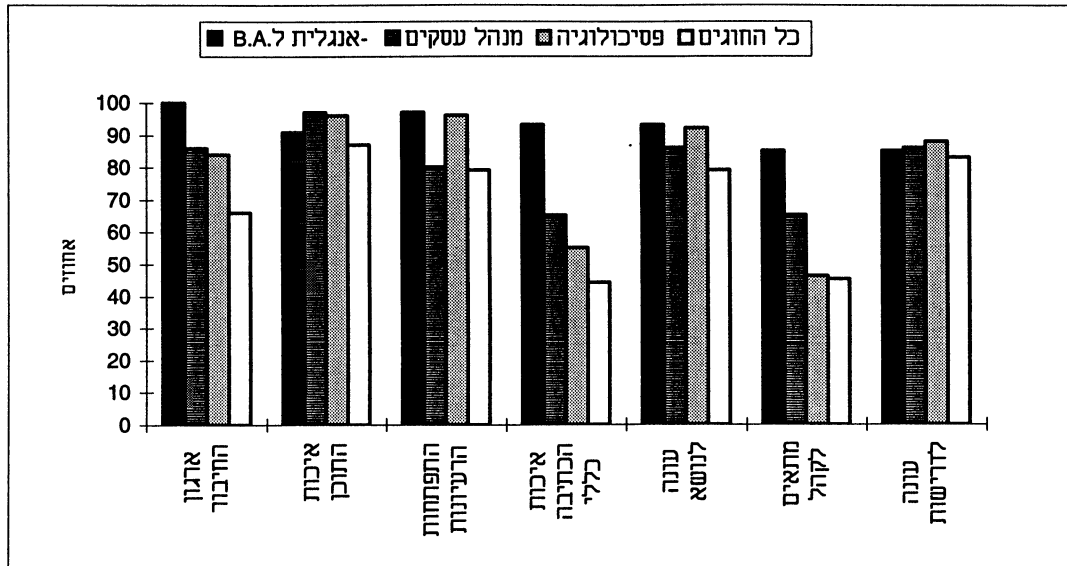
ממצאים אמפיריים לגבי מימדי ההערכה ושיטות הציון

מחקר על עמדות מרצים בדבר מימדי ההערכה וסוג הציון המבוקשים

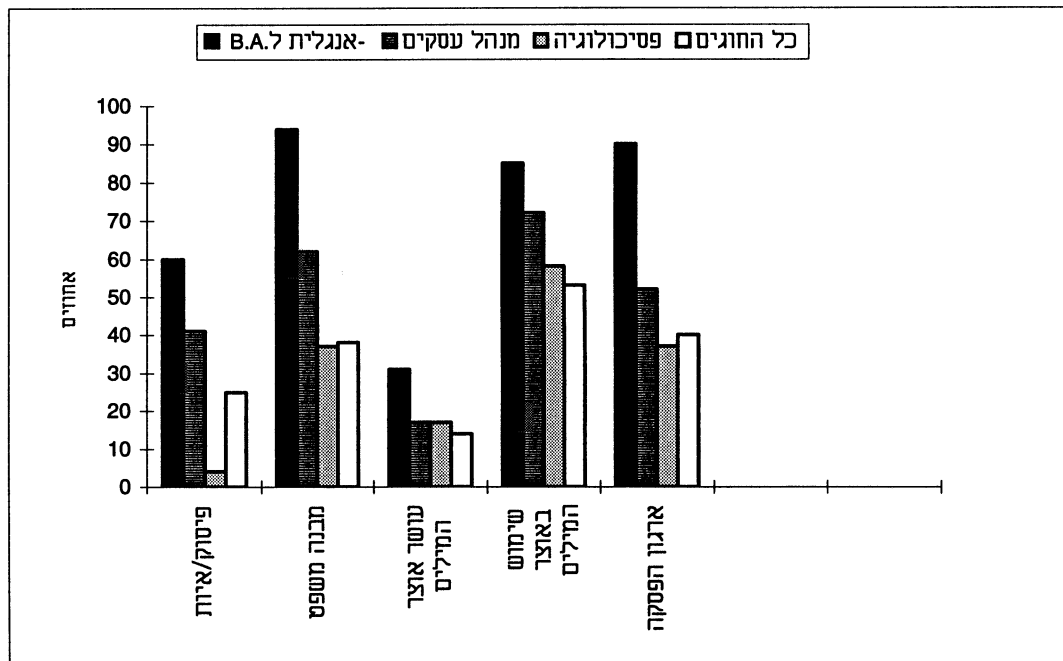
במחקר שנערך בידי הצוות האחראי על מבחן TOEFL נשאלו נציגים של מספר מחלקות באוניברסיטאות שונות על דעתם בעניין המימדים שעל פיהם יש להעריך כתיבה, ומהי שיטת הציון העדיפה בעיניהם (Bridgeman & Carlson, 1983). המחקר, שתואר כבר בפרק שדן במתכונת מבחן הכתיבה, כלל נתונים מן המחלקות הבאות: לימודים מתקדמים בהנדסת חשמל, הנדסה אזרחית, מחשבים, כימיה, מנהל עסקים ופסיכולוגיה. תלמידי החוג לאנגלית לתואר ראשון נבחרו כמייצגים את הדרישות לתואר ראשון. כאן מוצגים נתונים רק מן החוגים במדעי הרוח והחברה (פסיכולוגיה, מנהל עסקים, ואנגלית לתואר ראשון), ונתוני כל החוגים.

חשיבות המימדים השונים להערכת כתיבה

למרצים הוצגה רשימת מימדים והם נתבקשו לדרג על סולם בין 1 ל-5 באיזו מידה הציונים במטלות כתיבה בקורסים לסטודנטים מתחילים מושפעים מכל אחד מהמימדים הללו. התרשימים הבאים מציגים את אחוז המחלקות בכל תחום אשר דירג כל מימד בציון 4 ומעלה, ואשר מעיד לכן כי מיוחסת למימד זה חשיבות רבה. המימדים חולקו לשני סוגים: תכונות מקומיות הכוללות פיסוק/איות, מבנה המשפט, עושר אוצר המילים, שימוש באוצר מילים, ארגון כל פסקה, ותכונות גלובאליות הכוללות את ארגון החיבור כולו, איכות התוכן, התפתחות הרעיונות, איכות הכתיבה באופן כללי, המידה בה החיבור עונה לנושא, המידה בה החיבור מתאים לקהל והמידה בה החיבור עונה לדרישות המשימה. דירוגי התכונות המקומיות והתכונות הגלובאליות מוצגים בנפרד בתרשימים א2 ו-ב2.



תרשים מספר 2 א: חשיבות התכונות הגלובאליות להערכת משימות כתיבה - אחוז המחלקות אשר דירגו כל תכונה בציון 4 ומעלה, מתוך (Bridgeman & Carlson, 1983).



תרשים מספר 2 ב: חשיבות התכונות המקומיות להערכת משימות כתיבה - אחוז המחלקות אשר דירגו כל תכונה בציון 4 ומעלה, מתוך (Bridgeman & Carlson, 1983).

בדרך כלל נמצא כי התכונות הגלובאליות הוערכו כחשובות יותר מהתכונות המקומיות. אבל חשיבותן של התכונות הייתה שונה בחוגים השונים. בחוג לאנגלית לתואר ראשון גם מבנה המשפט, ארגון הפסקה ואוצר מילים היו חשובים במיוחד.

בהמשך התבקשו המרצים לבחור מהי לדעתם התכונה החשובה ביותר ומהי התכונה הפחות חשובה בהערכת חיבור. היה קונצנזוס בין המחלקות כי איכות התוכן היא תכונה חשובה ביותר, ולאחריה נבחרו המידה בה החיבור עונה לדרישות המשימה, התפתחות הרעיונות והמידה שבה החיבור עונה לנושא. התכונה הפחות חשובה: רוב המחלקות הסכימו כי עושר אוצר המילים אינו חשוב וכך גם איות ופיסוק.

סוג הצינון המועדף

המרצים נשאלו בדבר העדפתם בין שיטת צינור הוליסטית שתתבסס על התרשמות כללית לבין שיטת צינור אנליטית, שתהיה מורכבת מ-3 ציונים נפרדים המודדים תוכן (איכות הרעיונות), טעויות דקדוקיות מכניות, וארגון וקוהרנטיות בכתובה. בכל המחלקות הייתה העדפה לשיטת צינור אנליטית.

מחקרים המשווים בין שיטות הצינון השונות

כאמור, אחד השיקולים בבחירה בין שיטות צינור שונות הוא נתוני המהימנות והתוקף שלהם. כדי להשוות בין מקדמי תוקף ומהימנות של שיטות צינור שונות יש להשוות את הציונים של אותם החיבורים אשר צוינו בשיטות צינור שונות. דרישה זו יוצרת בעיה מתודולוגית מאחר שאם אותם השופטים מעריכים את החיבורים מספר פעמים בשיטות שונות, קשה להניח שההערכות הן בלתי תלויות. מאידך, אם נעזרים בקבוצות שונות של שופטים לביצוע הערכות מסוגים שונים, הרי הבדל בין התוצאות יכול לנבוע לא רק מהבדל בשיטות הצינון אלא גם מהבדל בין קבוצות המעריכים.

במחקרם של Veal and Hudson (1983) נערכה השוואה בין ארבעה אופני צינור של מבחני כתיבה ישירים: צינור אנליטי, צינור הוליסטי, צינור על פי תכונה ראשונית וצינור אטומיסטי המבוסס על ספירת טעויות כתיב ודקדוק. נוסף על כך הושוו מבחני כתיבה ישירים למבחנים עקיפים שבהם שאלות סגורות. הנבדקים היו 2400 תלמידים בכיתות י' בבתי ספר שונים. כל בית ספר קיבל שתי מטלות: חיבור לצינור אנליטי, הוליסטי או אטומיסטי וכן מבחן רב ברירה או חיבור נוסף לצינור על פי תכונה ראשונית. קבוצות שונות של בוחנים ביצעו את ההערכות השונות. מהימנות בין שופטים של הצינור ההוליסטי הייתה בטווח שבין 0.69 ל-0.76.

מהימנות בין שופטים אשר בדקו את התכונה הראשונית הייתה 0.72.

המימדים של שיטת הצינור האנליטית היו תוכן, סגנון וטעויות כתיב ודקדוק. מהימנות בין שופטים בצינור המימדים הנפרדים הייתה בטווח שבין 0.76 ל-0.82. מהימנות בין שופטים של הציון האנליטי הכולל הייתה 0.90. כלומר, שיטת הצינור האנליטית במחקר זה הייתה המהימנה ביותר.

המתאם בין הציון האנליטי הכולל לציון ההוליסטי היה 0.64, ובין הציון ההוליסטי לכל אחד מן המימדים הנפרדים היה בין 0.57 ל-0.62. בהתחשב במהימנות בין שופטים של הצינור ההוליסטי אפשר לראות כי המתאם בין הציון ההוליסטי לאנליטי היה גבוה ביותר. המתאם בין הציון ההוליסטי לבין הציון על פי תכונה ראשונית היה נמוך במיוחד (0.19). המתאם בין הציון ההוליסטי לציון אטומיסטי המבוסס על ספירת טעויות היה -0.25 (מתאם שלילי זה נובע מכך

שציון אטומיסטי גבוה פירושו מספר גבוה של טעויות). נראה, אם כן, כי ציון שנעשה על פי תכונה ראשונית בדק אספקטים שונים מאלה שנבדקו על ידי הציון ההוליסטי והאנליטי. במחקר זה נעשתה גם הערכת עלות ונמצא כי:

ציון הוליסטי דורש פחות זמן וכסף ולכן הוא החסכוני ביותר. ציון אנליטי הוא יקר יותר במונחי זמן, הוא במתאם גבוה עם המדדים ההוליסטיים ויתרונם שהוא מספק לנבחנים מידע מפורט יותר. ציון אטומיסטי וציון של תכונה ראשונית היו במתאם נמוך עם המדדים ההוליסטיים והאנליטיים ועלותם הייתה גבוהה יחסית, לכן הם מומלצים רק למחקר.

מחקר של Moss, Cole and Khampalikit (1982) השווה בין מבחן רב ברירה לבין 2 מטלות כתיבה אשר כל אחת נבדקה בשתי שיטות ציון: הוליסטית ואטומיסטית (ספירת טעויות). הנבדקים היו 84 תלמידים מכיתה ד', 45 מכיתה ז' ו-98 מכיתה י'. מטלות הכתיבה היו: כתיבת מכתב להשגת מטרה מסוימת (אשר הייתה שונה לכל קבוצת גיל) וכן כתיבת דוח קצר לשיעור במדעים. המבחן רב הברירה והציון האטומיסטי בדקו את ארבעת המימדים הבאים: כתיב, שימוש באותיות ראשיות (Capital letters), פיסוק והתבטאות (ידע בדקדוק, בבחירת מילים ובארגון). בציון ההוליסטי הודגשו הארגון, הבהירות והנכונות הדקדוקית. 4 בוחנים בדקו כל חיבור ונתנו ציון בין 1 ל-4. המתאמים בין המדדים השונים מובאים בטבלה מספר 2 (המתאמים שבתוך הסוגריים מתוקנים בעבור חוסר מהימנות המעריכים).

זוג המדדים	כיתה ד' - N= 84	כיתה ז' - N= 45	כיתה י' - N= 98
מבחן רב ברירה וציון אטומיסטי	0.42 (0.68)	0.27 (0.60)	0.60 (0.76)
ציון אטומיסטי וציון הוליסטי	0.21 (0.48)	0.24 (0.62)	0.42 (0.72)
מבחן רב ברירה וציון הוליסטי	0.12 (0.20)	0.39 (0.67)	0.47 (0.75)

טבלה מספר 2: מתאמים בין מדדי הכתיבה השונים, מתוך Moss, Cole & Khampalikit (1982)

המתאם הגבוה ביותר נמצא בין ציונים במבחן רב ברירה לבין הציונים האטומיסטיים. נתון זה אינו מפתיע מאחר שהם בדקו את אותם המימדים. אחריו בא המתאם בין הציון ההוליסטי לבין הציון האנליטי והנמוך ביותר היה המתאם בין הציון ההוליסטי לציון על מבחן רב ברירה. כפי שאפשר לראות בטבלה היו הבדלים במקדמי המתאם בין הגילאים השונים: בכיתה י' נמצאו מתאמים גבוהים יותר ממה שנמצאו בשתי הכיתות האחרות. נראה שבמחקר זה הציונים הושפעו ממימדי ההערכה יותר משהושפעו מסוג המבחן (מבחן פתוח מול מבחן רב ברירה).

Carlson, Bridgeman, Camp, & Waanders (1985) בדקו את יכולת הכתיבה של 172 סטודנטים (רובם זרים ומיעוטם אמריקנים) באמצעות ארבע משימות כתיבה, שתיים מהן מסוג "תאר ופרש גרף" והשתיים האחרות מסוג "השווה והנגד". כל החיבורים צוינו בשיטה הוליסטית ונוסף על כך מחציתם (אחד מכל סוג) צוינו גם בשיטה אנליטית על שני מימדים (מדד של מכניקה ודקדוק - "רמת המשפט" ומדד של ארגון וקוהרנטיות - "רמת הטקסט"). טבלה מספר 3 מציגה את המתאמים בין הערכות ברמת הטקסט לבין הערכות ברמת המשפט בתוך כל נושא ומעבר לנושאים (המתאמים שבתוך הסוגריים מתוקנים בעבור חוסר מהימנות המעריכים).

נושא 2 - רמת המשפט	נושא 1 - רמת המשפט	נושא 2 - רמת הטקסט	
0.72 (0.88)	0.87 (1.06)	0.66 (0.81)	נושא 1 - רמת הטקסט
0.86 (1.02)	0.63 (0.75)		נושא 2 - רמת הטקסט
0.73 (0.87)			נושא 1 - רמת המשפט

טבלה מספר 3: מתאמים בין ציונים על מימדים שונים של החיבור, מתוך Carlson, Bridgeman, Camp, & Waanders (1985)

מהטבלה עולה כי הערכות של רמת הטקסט הן במתאם גבוה למדי עם הערכות ברמת המשפט. מעבר לכל הנבדקים ומעבר לנושאים השונים נמצא מתאם של 0.90 בין ציונים המבוססים על רמת הטקסט לבין ציונים המבוססים על רמת המשפט. ממצאי מחקר זה מראים כי בעצם הערכות של רמת הטקסט והערכות של רמת המשפט אינן בודקות דברים נפרדים, וכי המתאמים בין ציונים על מימדים שונים של אותם החיבורים גבוהים מהמתאמים בין ציונים שניתנו לחיבורים שונים על אותו מימד. ייתכן שזה נובע מהעובדה שאותם המעריכים נתנו את שני סוגי הציונים לאותו החיבור בזה אחר זה. נוסף על כך נבדק הקשר בין הציונים ההוליסטיים לציונים על המימדים הנפרדים ונמצא מתאם של 0.87 בין הציונים ההוליסטיים לציונים על רמת המשפט, ומתאם של 0.88 בין הציונים ההוליסטיים לציונים על רמת הטקסט. אם כך, המחקר מראה מתאמים גבוהים מאוד בין הערכות על מימדים שונים, ובשל כך הסיקו החוקרים כי שלושת סוגי הציונים: הציונים ההוליסטיים, הציונים על רמת הטקסט והציונים על רמת המשפט מודדים בעצם את אותו הדבר, ולכן אפשר להסתפק בציונים ההוליסטיים שהם בעלי העלות הנמוכה ביותר.

כהן, איתן ובן-צור (1986) בדקו את המתאמים בין ארבעת המימדים להערכת החיבורים במבחן יע"ל, הבודק ידע בעברית של מועמדים לאוניברסיטאות אשר אינם דוברי עברית כשפת אם. ארבעת מימדי הציון הם: תוכן, ארגון, עושר הלשון ותקינות הלשון, והמתאמים ביניהם מוצגים בטבלה מספר 4.

	תוכן	ארגון	עושר הלשון	תקינות הלשון
ארגון	0.81			
עושר הלשון	0.82	0.69		
תקינות הלשון	0.64	0.65	0.78	
ציון כולל	0.91	0.88	0.92	0.87

טבלה מספר 4: מתאמים בין ציונים על ארבעת מימדי הציונות של חיבורי יע"ל, מתוך כהן, איתן, בן-צור (1986)

מן הטבלה עולה כי קיים מתאם גבוה למדי בין ארבעת המימדים, עם זאת נראה כי מדד תקינות הלשון הוא חריג במקצת. ייתכן כי מדד זה בודק יותר את רמת המשפט מאשר את רמת הטקסט ואם כך הוא הדבר הרי שבמבחן יע"ל יש מתאמים נמוכים יותר בין המדד של רמת המשפט ומדד של רמת הטקסט מאשר אלה שנמצאו ב- TOEFL.

חשוב לציין כי שני המחקרים האחרונים, שמצאו מתאם גבוה בין המימדים השונים בציון אנליטי, התבססו על חיבורים אשר בחנו כתיבה בשפה שאינה שפת אם.

במסגרת NAEP (National Assessment of Educational Progress), מחקר רחב היקף להערכת יכולת הכתיבה של תלמידי כיתות ד', ח', ו-י"א בארצות הברית, נתבקשו התלמידים לבצע משימות כתיבה משלושה ז'אנרים: כתיבה אינפורמטיבית, כתיבה שכנועית וכתיבה נרטיבית (Applebee, Langer, Mullis, & Junkins, 1990). כל משימת כתיבה צוינה בשלוש שיטות: ציון על פי תכונה ראשונית (המידה שבה הצליח התלמיד בהשגת מטרת החיבור); ציון הוליסטי (הערכת הרעיונות, הארגון, המכניקה והתחביר); ציון אטומיסטי שבדק את השליטה בשימוש בכללי השפה האנגלית הכתובה (על ידי בדיקת איות, פיסוק ודקדוק).

המתאמים בין הציונים שניתנו בדרך של תכונה ראשונית ובין הציונים שניתנו בדרך של ציון הוליסטי היו בטווח שבין 0.38 ל-0.66.

Miller and Crocker (1990) מזכירים מחקרים נוספים אשר בדקו את הקשר בין ציון אנליטי לציון הוליסטי: Prater and Padia (1983) דיווחו על מתאמים בסדר גודל של 0.75 ל-0.76 בין ציון הוליסטי לציון אנליטי בכיתות ד' ו-ו'. Olson and Swadener (1984) מצאו מתאם מרובה של 0.79 בין ציונים הוליסטיים לבין 4 מדדים אנליטיים (ארגון, סגנון, מכניקה ולוגיקה). לסיכום, ברוב המחקרים שהוזכרו נמצא מתאם גבוה למדי בין ציונים הוליסטיים לאנליטיים, ומתאם נמוך יחסית ביניהם לבין ציונים על פי תכונה ראשונית.

רוזנר (בדפוס) הישוותה בין הערכה הוליסטית לאנליטית של משימת כתיבה אחת אשר נכתבה על ידי 40 תלמידי כיתות ז'. המעריכים היו מורים ללשון וכל אחד מהם העריך כל חיבור בשתי השיטות. מהימנות בתוך מעריכים הייתה גבוהה יותר בשיטת ההערכה האנליטית (טווח שבין

0.53 ל- 0.82), מאשר בשיטת ההערכה ההוליסטית (טווח שבין 0.22 ל- 0.76). המתאמים שבין הציונים ההוליסטיים לציונים האנליטיים היו בין 0.40 ל- 0.72, והמתאם הגבוה ביותר היה בין הציון ההוליסטי לציון על מימד הלשון, כנראה בגלל שהמעריכים היו מורים ללשון. המתאם בין הציון ההוליסטי לממוצע הציונים האנליטיים היה 0.66. כמו כן המחקר הראה כי הערכה הוליסטית נתנה ציונים גבוהים יותר מאשר הערכה אנליטית.

שיטות הציון המקובלות במבחני כתיבה שונים בארץ ובעולם

בחינות הבגרות בהבעה בכתב הנערכות בישראל:

כל חיבור מצויין בידי שני מעריכים באופן אנליטי, ומימדי ההערכה על פי משקלותיהם הם: תוכן והבנה (40 נקודות),

מבנה - לכידות וקישוריות (20 נקודות),

מבע - עושר הלשון והסגנון (20 נקודות),

תקינות הלשון (20 נקודות).

מפתח ההערכה לחיבורים:

1. תוכן והבנה - סך הכול 40 נקודות:

א. הבנת הנושא - ברמה המילולית וברמת המשמעות לפי ניסוחו - מעבר למרכיבי הבודדים - 16 נקודות

ב. ידע - הצגת עובדות ותופעות רלוונטיות; הדגמות הן מתוך הטקסט הצמוד והן מחוצה לו - 8 נקודות.

ג. תהליכי חשיבה - במאמר העיוני: עקביות, היגיון, שיטתיות, הבחנה בין עיקר לטפל, השוואה, עימות, ניתוח, הנמקה, רעיונות, מקוריות, עומק. בכתיבה היצירתית: מקוריות, דמיון, עומק, כנות, עושר - 16 נקודות.

2. מבנה (לכידות וקישוריות) - עריכה לפי מבנה, לפי דרך הפיתוח שהתלמיד בחר בה (מאמר עיוני, תיאור אישי, רשמים וכו').

לדוגמה במאמר העיוני:

א. פתיחה מתאימה לנושא (כגון הגדרת מושגים); הבהרת הנושא או הצגת הבעיה; הגדרת סוג הכתיבה והנמקתו

ב. גוף החיבור: הרעיונות העיקריים והפסקאות המיוחדות להם; הקישור בין הרעיונות והמעברים מפסקה לפסקה

ג. הסיום: הסיכום, תמצית הטיעונים והמסקנה המבוססת על הטיעונים הללו.

3. מבע (עושר הלשון והסגנון) -

א. התאמת הלשון לסוג המבע, למטרותיו ולאופיו (מאמר עיוני, פולמוס, רשימה סאטירית,

תיאור קליל של רשמים, כתיבה יצירתית); הבחנה בין "סלנג" לבין הבעה בכתב (10

נקודות).

ב. אוצר מילים מגוון ; שימוש גמיש באלטרנטיביות תחביריות ; סגנון בהיר ומדויק (10 נקודות).

4. תקינות הלשון -

א. תחביר ופיסוק ב. כתיב ג. צירופי לשון וצורות דקדוק (15 נקודות).

ד. צורה חיצונית ; כתב יד ; שוליים ; חלוקה לפסקאות (5 נקודות).

מבחן ידע בעברית (יע"ל) הניתן על ידי המרכז הארצי לבחינות ולהערכה :

מבחן זה ניתן למועמדים לאוניברסיטאות אשר אינם דוברי עברית כשפת אם, והוא מתלווה לבחינה הפסיכומטרית. הוא כולל 2 פרקים של פריטים רבי ברירה ומשימת כתיבה אחת. כל חיבור מוערך בידי 2 מעריכים על סולם בין 1 ל- 7 על פני ארבעה מימדים בעלי משקל שווה. המימדים הם :

תוכן ההבעה (רעיונות וקשר למשימת הכתיבה כפי שהוגדרה)

ארגון ההבעה (מבנה וארגון הכתיבה)

עושר ההבעה (שימוש באוצר מילים מגוון, ביטויים וכו')

דיוק הלשון (טעויות בתחביר, כתיב, כתב ופיסוק)

סקר של משרד החינוך ואוניברסיטת תל אביב על הבעה בכתב בבית הספר היסודי (נבו, 1989) :

בשנת 1989 נערך סקר על הוראת ההבעה בכתב בכיתות ו'. הוא כלל נתונים מ- 98 כיתות ב- 57 בתי ספר. כל נבחן כתב שני חיבורים אשר צוינו בשיטה הוליסטית תוך התבססות על ארבעה מימדים : תוכן, מבנה, לשון (עושר, תקינות דקדוקית ותחבירית) ומכניקה (שימוש נכון בכללי פיסוק, היעדר שגיאות כתיב, כתב יד קריא, סדר וניקיון).

GRE הניתן על ידי ה- Educational Testing Service :

מבחן זה הינו מבחן כניסה ללימודים מתקדמים באוניברסיטאות בארצות הברית, המורכב מפרקים הכוללים פריטים רבי ברירה בשלושה תחומים : מילולי, כמותי ואנליטי. יש כוונה לצרף למבחן זה גם מבחן כתיבה ישיר, הנמצא בשלב נסיוני בלבד כיום. החיבורים במבחן הכתיבה הניסיוני מצוינים בשיטה הוליסטית על סולם מ- 1 ל- 6. בהוראות לנבחן נאמר כי הבודקים יעריכו את האיכות הכוללת של החיבור, ובכלל זה את רמת הארגון, רמת הפיתוח וההצגה של הרעיונות בנושא, השימוש במידע תגמך והשליטה בחוקי האנגלית. במדריך למציין מוצגים תיאורים של חיבורים ברמות השונות. מתיאור החיבורים עולה כי המדדים החשובים הם : הבנה של הנושא שהוצג, הצגת העמדה בצורה אפקטיבית עם דוגמאות מתאימות, שמירה על דיון ממוקד בנושא, ביטוי הרעיונות בבהירות ובדיוק, שימוש בשפה שוטפת ואפקטיבית, ובכלל זה עם מבני משפט שונים ואוצר מילים המתאים לנושא.

University of California - The Universitywide Subject A Examination - מבחן זה הוא מבחן פנימי של רשת האוניברסיטאות של מדינת קליפורניה. מבחן Subject A הוא חובה לכל הנרשמים לתואר ראשון, והוא משמש להשמה לקורסי אנגלית. הוא כולל משימת כתיבה המתבססת על קטע קריאה. החיבור מוערך באופן הוליסטי - בידי שני בודקים - על סולם בין 1 ל-6. ממדריך הציון עולה כי המימדים החשובים הם: פיתוח הרעיונות, סגנון הכתיבה, המידה בה החיבור עונה לנושא, המידה בה הטיעון משכנע ומחוזק בדוגמאות, שימוש נכון במילים, בנייה נכונה של משפטים, שמירה על חוקי האנגלית (Gadda, 1992).

Basic Skills Assessment הניתן על ידי ה-Educational Testing Service תכנית הערכה זו נועדה לאתר תלמידים מכיתות ח' עד י"ב אשר אינם עומדים בדרישות המינימום בתחומי הקריאה והכתיבה. מבחן הכתיבה כולל 45 דקות של פריטים רבי ברירה וכן 4 מטלות כתיבה:

1. מילוי טופס
 2. כתיבת מכתב עסקי
 3. העברת אינפורמציה בצורה מדויקת
 4. כתיבה יצירתית
- Fowles (1978) מציעה להתאים את שיטת הציון למטלת הכתיבה; מטלות מילוי טופס והעברת אינפורמציה לציון באופן אנליטי, כתיבה יצירתית וכתיבת מכתב עסקי לציון באופן הוליסטי.

: The IEA study of written composition

מחקר זה נועד להשוות את ההישגים של תלמידים במספר מקצועות במדינות שונות. יכולת ההבעה בכתב של תלמידים מ-14 מדינות נבחנה על ידי משימות כתיבה מז'אנרים שונים. במחקר זה בחרו להשתמש בשיטת ציון אנליטית כי היא מאפשרת להשוות בין ביצועים של משימות שונות.

המימדים שבהם בחרו להשתמש:

1. איכות התוכן והיקפו
2. ארגון התוכן והצגתו
3. סגנון
4. דקדוק ותחביר
5. איות

בשלב מאוחר יותר הוחלט להוסיף גם מדד של רושם כללי, ההערכה על מדד זה תיעשה בתחילה ואחריה יעברו המעריכים למדדים האנליטיים. לא יותר תיקון בדיעבד של מדד הרושם הכללי (Gorman, Purves, & Degenhart, 1988).

לסיכום, מבחני כתיבה רחבי ההיקף הנערכים כיום מוערכים רובם בדרך הוליסטית, המבוססת על הגדרה של המימדים החשובים בהערכה. בחינה של שיטות הציינון המקובלות מראה הן שיטת הציינון ההוליסטית והן שיטת הציינון האנליטית מתמקדות בשלושה מימדים עיקריים: מימד תוכן, מימד ארגון ומימד לשוני. המימד הלשוני מתחלק בדרך כלל לסגנון ועושר הלשון למול תקינות הלשון.

פרק ג': מהימנות

בפרק זה תידון שאלת מהימנות מבחני הכתיבה. מהימנות הבחינה משקפת את המידה שבה הציונים אומדים נכונה את היכולת האמיתית של הנבחנים (אנסטאזי, 1990). בכל בחינה יש מרכיב של טעות הנובע מכך שהציון בה אינו רק תוצאה של היכולת האמיתית אלא הוא מושפע גם מגורמים מקריים שאינם תלויים ביכולת זו. בהקשר של מבחני כתיבה ישירים גורמים אלה יכולים להיות קשורים למטלת הכתיבה, לנבחן ולציון הבחינה. פרק זה מחולק לסעיפים, כאשר כל סעיף מתמקד בגורם רעש אחר שעלול לפגוע במהימנות הבחינה, והוא סוקר מחקרים אשר בדקו באופן אמפירי השפעות אלו. בדיקה זו נעשתה באחת משתי השיטות הבאות: א. באמצעות חישוב מתאם בין ציונים שניתנו על ידי מעריכים שונים או ציונים שנקבעו על סמך פריטים שונים. ב. באמצעות תורת ההכללה, אשר יוצאת מן ההגדרה של תורת המבחנים הקלאסית לפיה הציון הנצפה של אדם במבחן מורכב מציון אמיתי וממרכיב טעותי, וממשיכה בניסיון להפריד בין הגורמים התורמים למרכיב הטעותי. נעשה שימוש רב בשיטה זו לבדיקת מהימנות מבחני כתיבה שכן הציונים במבחני כתיבה המוערכים על ידי בודקים מושפעים ממספר גורמי רעש (פריטים, מעריכים, והאינטראקציות ביניהם). פירוק השונות בציונים למקורותיה נעשה על ידי ניתוח שונות. השונות הכללית מפורקת לשונות המוסברת על ידי הנבחנים, על ידי המעריכים, על ידי הפריטים וגורמים אפשריים אחרים והאינטראקציות בין הגורמים הללו. השונות המוסברת על ידי הנבחנים היא המקבילה של תורת ההכללה לשונות הציונים האמיתיים של הנבחנים, והשונות המוסברת על ידי כל הגורמים האחרים והאינטראקציות ביניהם לגורם הנבחנים משקפת שונות טעותית. ככל שגורם הנבחנים מסביר חלק גדול יותר בשונות הציונים הנצפים כך יכולת ההכללה מציוני המבחן גבוהה יותר. (Brennan, 1992).

גורמים הקשורים למטלת הכתיבה

נושא החיבור - כמו יכולת הבנת הנקרא גם יכולת הכתיבה היא תחום-על אשר אינו ניתן לבדיקה באופן בלתי תלוי בעולם תוכן מסוים. חיבור, כמו קטע הבנת הנקרא, נכתב על נושא ספציפי בתוך עולם תוכן כלשהו. צפוי כי אדם יצליח יותר במשימות הבנת הנקרא וכתיבה אם המשימות יעסקו בנושא בו הוא בקי מאשר אם הם יעסקו בנושא אותו הוא אינו מכיר. שאלת יציבות הציונים במשימות כתיבה בנושאים שונים חשובה במיוחד במבחן כתיבה מאחר שבדרך כלל במבחן כזה מספר קטן מאוד של פריטים, והוא אינו מכסה טווח רחב של נושאים. מספר מחקרים בדקו את השפעת הנושא על הערכת יכולת הכתיבה.

מחקרם של Breland et al. (1987) כלל נתונים על יכולת הכתיבה של 267 סטודנטים משש אוניברסיטאות. כל סטודנט נתבקש לכתוב שישה חיבורים בשלושה ז'אנרים של כתיבה (סיפורי, אקספוזיטורי ושכנועי). בניתוח שונות נמצא כי האינטראקציה בין נושא הכתיבה לנבחנים (22%) הסבירה אחוז גדול יותר מן השונות מאשר האינטראקציה בין ז'אנר הכתיבה לנבחנים (10%).

האינטראקציה בין נושא החיבור לנבחנים חזרה על עצמה גם כאשר הניתוח נעשה בתוך כל ז'אנר כתיבה בנפרד. יש לציין כי החיבורים לא התבססו על ידע בתחום מסוים אלא על ידע כללי בלבד. מחקרים נוספים שבדקו כתיבה המתבססת על ידע ספציפי הגיעו למסקנות דומות. במחקר של Boodoo and Garlinghouse (1983) 34 סטודנטים לחינוך נתבקשו לכתוב שלושה חיבורים שבהם הם מביעים את מידת הסכמתם עם תאוריות פסיכולוגיות שלמדו. צייון החיבורים התמקד במידה שבה משכנעים הכותבים בטענותיהם והוא נעשה בידי שלושה מעריכים על סולם שבין 1 ל-5. הציונים נותחו בניתוח שונות, והגורמים היו נבחנים, שאלות ומעריכים. נמצא כי האינטראקציה בין הנבחנים לשאלות הסבירה חלק גדול מהשונות (הגורם השני בגודלו). דבר זה מראה שחלק מהנבחנים הצליחו יותר בשאלות מסוימות ואחרים הצליחו יותר באחרות, וייתכן שזה נובע מרמת ההיכרות שלהם עם התאוריות שעליהם נתבקשו להגיב. הנתונים נותחו גם על ידי ניתוח גורמים אשר מצא שלושה גורמי תוכן המייצגים את שלוש השאלות. מחקר זה הראה כי גורם התוכן של השאלות קובע במידה רבה את יכולת הנבחנים לענות. אולם יש לזכור כי החיבורים במחקר זה התבססו על ידע ספציפי. המשמעות של ממצאים אלה היא שאין עקביות גבוהה בדירוג יכולת הכתיבה של הנבחנים כאשר זו נבדקת על סמך חיבורים בנושאים שונים, ופירוש הדבר הוא מהימנות נמוכה בין משימות שונות. ייתכן כי הסיבה לכך נעוצה בכך שיכולת הכתיבה מושפעת מהמידה בה הנושא קרוב, מוכר, ורלוונטי לנבחנים.

ז'אנר הכתיבה - מבחני כתיבה הכוללים מספר משימות כתיבה בדרך כלל כוללים משימות מז'אנרים שונים. מספר מחקרים בדקו את מידת העקביות בדירוג הנבחנים המתבסס על משימות מז'אנרים שונים. ל-217 תלמידים אמריקניים מכיתות ד', ז', ו-י ניתנו שתי מטלות כתיבה משני ז'אנרים: כתיבת מכתב להשגת מטרה מסוימת (אשר השתנתה לפי גיל הנבחנים), המייצגת כתיבה שכנועית, וכתיבת דוח קצר לשיעור במדעים, המייצגת כתיבה אינפורמטיבית, אנליטית יותר. כל חיבור נבדק בשתי שיטות ציון. ניתוחי שונות על ציוני הנבחנים הראו כי בשתי שיטות הציון האינטראקציה בין הנבחן למטלה הסבירה חלק גדול למדי מהשונות בציונים (ברוב המוחלט של המקרים האינטראקציה היא הגורם המסביר את החלק הגדול ביותר מן השונות), (Moss, Cole and Khampalikit 1982). אפשר להעיר על מחקר זה כי ההבדל בין המטלות אינו רק בז'אנר הכתיבה אלא גם בעולם התוכן ואולי בדרגת הקושי.

מחקר נוסף בחן את יכולת הכתיבה של 200 תלמידי כיתות י"א ו-י"ב בארצות הברית על ידי מבחן רב ברירה וחיבורים מ-2 ז'אנרים (סיפורי ואקספוזיטורי) ומצא כי היה הבדל גדול בין הז'אנרים בעיקר בהערכת רושם כללי וארגון החיבור (Quellmalz, Capell & Chou, 1982). במחקר על הבעה בכתב בבית הספר היסודי שערכו דוד נבו ועמיתיו (1989) ניתנו מטלות בשלושה ז'אנרים של כתיבה: רגשי, מעשי ולימודי. המתאמים בין שלוש המטלות לאחר תיקון לחוסר מהימנות המעריכים היו בין 0.35 ל-0.45.

מתאמים גבוהים יותר בין משימות כתיבה בז'אנרים שונים נמצאו על ידי (Carlson et al., 1985). מחקר שעשה הצוות האחראי על מבחן TOEFL בדק ביצוע של סטודנטים אמריקניים וזרים על שתי משימות כתיבה מן הסוג של "נתח את האינפורמציה מתוך גרף או טבלה", ושתיים מן הסוג של "השווה והנגד" (Carlson et al., 1985). המתאמים בין חיבורים על נושאים שונים באותו ז'אנר כתיבה והמתאמים בין חיבורים שונים בז'אנרים שונים של כתיבה היו בטווח שבין 0.66 ל-0.73. לאחר תיקון לחוסר מהימנות המעריכים המתאמים היו אף גבוהים יותר והגיעו לטווח שבין 0.81-0.86.

לסיכום, נראה כי המסקנה העולה מהנתונים היא שרמת הכתיבה של נבחן משתנה כאשר הוא נדרש לכתוב בז'אנרים שונים. ייתכן כי הדבר נובע מהבדל בכללים המנחים ובסגנונות המקובלים בז'אנרים השונים. בגלל המהימנות הנמוכה של משימות כתיבה מז'אנרים שונים יש לכלול מספר משימות כאלה, ולנסות ולבחור את הז'אנרים כך שיהיו הרלוונטיים ביותר למטרת המבחן.

גורמים הקשורים לנבחן

גורמים כמו עייפות, מחלה או מוטיבציה יכולים להשפיע על הרמה שבה ביצע הנבחן את משימת הכתיבה. גורמי רעש אלה אינם ייחודיים למבחני כתיבה ישירים אלא קיימים גם במבחנים רבי ברירה. נשאלת השאלה אם השפעתם גדולה יותר במבחנים פתוחים כאשר נדרש מן הנבחן להפיק תשובה ולא רק לבחור בין תשובות. לא נמצאו מחקרים בספרות הנותנים תשובה לשאלה זו. בכל מקרה, אפשר לצמצם רעש הנגרם על ידי גורמים אלה באמצעות מתן מספר מבחנים במועדים שונים, דבר שמעורר קשיים פרקטיים וטכניים רבים.

גורמים הקשורים לצינון הבחינה

אחד ההבדלים החשובים בין מבחנים רבי ברירה למבחנים פתוחים הוא בתהליך הצינון; בעוד שמבחנים רבי ברירה מצויינים באופן מכני, מבחנים פתוחים מוערכים בידי אנשים. צינון הנעשה בידי מעריכים עלול להוסיף רעש למדידה. רעש זה יכול לנבוע מחוסר עקביות בעבודתו של המעריך היחיד ו/או מחוסר הסכמה בין מעריכים שונים.

א. גורמים הקשורים למעריך

בדומה לנבחן, גם עבודתו של המעריך יכולה להיות מושפעת מגורמים מצביים, כמו למשל מצב רוח או עייפות. (Townsend, Kek and Tuck 1989) בדקו את ההשפעה של מצב רוחו של המעריך על הציונים שנתן. הוצגו לפני מעריכים סרטים כדי להכניסם למצב רוח חיובי או שלילי, ואחר כך נתבקשו המעריכים לציין סדרה של חיבורים. נמצא אפקט קטן של מצב רוח רק על הצינון של החיבור הראשון בסדרה (מצב רוח שלילי הביא לצינון גבוה יותר), ולא לגבי שאר החיבורים בסדרה. כמו כן לא הייתה אינטראקציה בין מצב הרוח לבין איכותם של החיבורים כפי שהוערכו קודם לכן בידי מעריכים אחרים. מכך הסיקו החוקרים כי אין השפעה של מצב הרוח של המעריך על הערכותיו, בהסתייגויות הבאות: המדובר בחיבורים קצרים ובצינון הוליסטי.

המחקר בדק מצב רוח שנגרם מגירויים חיצוניים, ואולי אין להקיש ממנו על מצב רוח שמקורו פנימי.

ג. הבדלים בין מעריכים

במקרים רבים מדווחת מידת ההסכמה בין מספר שופטים המציינים את אותם החיבורים כמייצגת את המהימנות של מבחן הכתיבה כולו. אולם יש לזכור כי גורם המעריך הוא רק מקור אחד לרעש במדידה ולכן מדד של מהימנות בין מעריכים מספק בדרך כלל הערכת יתר למהימנות הבחינה כולה שמושפעת גם משונות בין משימות (Breland et al. 1987).

בספרות מוזכרים מספר מדדים להסכמה בין מעריכים: אחוז החיבורים אשר קיבלו הערכות זהות מידי 2 המעריכים, אחוז החיבורים אשר ההבדלים בהערכות של שני המעריכים היה גדול מערך מסוים, מקדם מתאם ממוצע בין כל שתי הערכות, נוסחת ספירמן-בראון, ומקדם ההכללה. בגלל השימוש במדדים שונים להערכת הסכמה קשה להשוות בין הנתונים המדווחים במבחני הכתיבה השונים (Cherry & Meyer, 1993).

Breland (1983) סקר את הספרות המחקרית בנושא מהימנות בין מעריכים במבחני כתיבה ישירים. בטבלה מספר 5 מוצגים סיכום נתוני מהימנות בין מעריכים בציון משימת כתיבה אחת מן המחקרים השונים. מהימנות מעריך אחד חושבה על ידי מתאם ממוצע בין הערכות של שני מעריכים. מהימנות שני מעריכים ומעלה חושבה על ידי תיקון ספירמן-בראון.

פרטי המחקר	שיטת הציון	הסולם	מעריך אחד	2 מעריכים	3 מעריכים	4 מעריכים	5 מעריכים
Akeju (1972)	לא מפורטת		0.72	0.84	0.88		
Breland (1983)	אטומיסטית	1-20	0.40	0.57			
	אנליטית	3-15	0.67	0.80			
	הוליסטית	1-4	0.54	0.70			
Coffman (1966)	הוליסטית	1-3	0.39	0.56	0.65	0.72	0.76
Coffman (1971)	הוליסטית	1-4		0.70			
Conry & Jeroski (1980)	הוליסטית	1-9			0.74		
Coward (1952)	הוליסטית	1-9	0.54	0.69			
	אנליטית		0.70	0.82			
(1982) ETS	הוליסטית	1-4		0.71			
Finlayson (1951)	הוליסטית	1-20	0.71	0.83	0.88	0.91	
Hackman & Johnson (1977)	הוליסטית	1-5	0.61				
Huddleston (1954)	אנליטית		0.60				
Michael et al. (1980)	הוליסטית	1-4	0.66	0.80			
Myers et al. (1966)	הוליסטית	1-4	0.41	0.58	0.67	0.73	
Powills et al. (1979)	הוליסטית	1-4			0.81		
Traxler & Anderson (1935)	אנליטית	1-10		0.94			
	אנליטית	1-10		0.84			
מהימנות חציונית			0.64	0.70	0.78		

טבלה מספר 5: מהימנות בין מעריכים במחקרים השונים על מבחני כתיבה, מתוך Breland (1983)

בהסתמך על נתונים אלה הצביע Breland (1983) על שני גורמים המשפיעים על מהימנות בין מעריכים: א. טווח הציונים - ככל שטווח הציונים גדול יותר המהימנות גבוהה יותר. ב. שיטת הציון - מתוך שלושה מחקרים אשר בהם החיבורים נבדקו בשיטות ציון שונות, נראה כי שיטת הציון האנליטית מהימנה יותר מאחרות.

המחקרים הבאים בדקו את סוגיית המהימנות בין מעריכים במבחני כתיבה מאז סקירת הספרות שערך Breland (1983). Veal & Hudson (1983) דיווחו על מהימנויות בין הערכות הוליסטיות של שופטים שונים בטווח שבין 0.69 ל-0.76. Moss, Cole and Khampalikit (1982) דיווחו על מהימנויות בין שופטים של מעל 0.85 הן להערכות הוליסטיות והן להערכות אנליטיות. כאשר

חישוב המהימנות מתבסס הן על שונות בין מעריכים והן על שונות בין מטלות הכתיבה המדדים יורדים ל- 0.41-0.50 בעבור הערכות הוליסטיות ול- 0.30-0.73 להערכות אנליטיות.

(Breland et al 1987) חישוב מתאמים בין מעריכים בעבור כל מטלה בנפרד, ובהמשך הפעילו את נוסחת ספירמן-בראון כדי להעריך כמה תגדל המהימנות אם יוספו מעריכים נוספים. ממוצע המתאמים בין זוגות השופטים היה בטווח 0.51 - 0.65 (והוא נחשב למדד של מהימנות השופט היחיד), כאשר באמצעות נוסחת ספירמן-בראון נעשה תיקון ל- 2 מעריכים הטווח היה 0.68 - 0.79, וכאשר נעשה תיקון ל- 3 מעריכים הטווח היה 0.76 - 0.84.

(Applebee et al 1990) סקרו את תוצאות מחקר NAEP, מחקר רחב היקף להערכת יכולת הכתיבה של תלמידי כיתות ד', ח', ו-י"א בארצות הברית. במחקר השתתפו 18,000 תלמידים אשר נבחנו ב- 1984 וב- 1988. מהימנות ההערכות על פי תכונה ראשונית חושבה על סמך 20% של החיבורים אשר צוינו בידי שני מעריכים, ונתוני המהימנות מוצגים בטבלה מספר 6.

כיתה	טווח אחוז הסכמה מלאה	ממוצע אחוז הסכמה	טווח מקדמי מהימנויות	ממוצע מקדמי מהימנויות
ד'	88 - 97	92.4	0.88 - 0.99	0.94
ח'	75 - 94	82.9	0.64 - 0.92	0.77
י"א	85 - 93	90.5	0.86 - 0.93	0.90

טבלה מספר 6: מהימנות בין מעריכים ב-NAEP, מתוך (Applebee et al 1990)

חיבורים שנכתבו לחלק ממשימות הכתיבה נבדקו גם באופן הוליסטי. מהימנות בין שופטים חושבה על סמך 20% של החיבורים אשר צוינו בידי שני מעריכים והיא הייתה בטווח שבין 0.64 ל- 0.83.

(Applebee, Langer, Mullis, Latham, & Gentile 1994) דיווחו על תוצאות סקר NAEP שהתקיים ב- 1992. משימות הכתיבה הוערכו על סולם בין 1 ל- 6 בדרך של תכונה ראשונית, בהתאם לזיאנר שנדרש במשימה. מהימנות בין שופטים נאמדה על סמך אחוז המקרים בהם שני מעריכי החיבורים היו בהסכמה מלאה לגבי הציון. בחיבורים שנכתבו על ידי תלמידי כיתה ד' הייתה הסכמה מלאה ב- 84% מהחיבורים, בחיבורים מתלמידי כיתה ח' הייתה הסכמה מלאה ב- 80% מהמקרים, ובחיבורים מתלמידי כיתה י"ב הייתה הסכמה מלאה ב- 79% מהחיבורים.

מחקר על הבעה בכתב בבית הספר היסודי שנערך על ידי נבו ועמיתיו (1989) כלל משימות כתיבה בשלושה ז'אנרים (רגשי, מעשי ולימודי). נכללו במחקר 2,600 תלמידים אשר כל אחד מהם כתב שני חיבורים. החיבורים הוערכו בידי 12 בודקות באופן הוליסטי. מהימנות בין שופטים חושבה על מדגם של 15 חיבורים מכל מטלה אשר נבדקו בידי כל המעריכים. מקדמי המהימנות חושבו באמצעות מתאמים ובאמצעות מקדמי הכללה. ממוצע מהימנויות בין שופטים למטלה היו בין 0.62 ל- 0.90 כשחושבו באמצעות מתאמים, ובין 0.81 ל- 0.94 כשחושבו על פי מקדמי הכללה. המהימנויות הממוצעות לז'אנרים השונים על פי מתאמים היו 0.73, 0.84, ו- 0.75 ועל פי מקדמי הכללה 0.89, 0.92 ו- 0.84 לכתיבה רגשית, מעשית ולימודית בהתאמה.

מהימנותם של מבחני כתיבה ישירים לנבחנים בשפה שאינה שפת אימם נבדקה בארץ ובארצות הברית. בארץ נבדקה מהימנות החיבור הכלול במבחן יע"ל. מהימנות החיבור מבוססת על חישוב של מתאם בין שתי ההערכות הניתנות לכל חיבור. המתאם הממוצע בין המעריכים לכלל הנבחנים בשנים 1986 - 1989 היה 0.73. כמו כן חושבה גם מהימנות בין שופטים במבחן יע"ל במועד אפריל 1989 רק בעבור זוגות מעריכים שהעריכו בצוותא 30 חיבורים לפחות (סה"כ 47 זוגות). ממוצע המתאמים היה 0.72 (סטיית תקן של 0.11), וטווח המתאמים היה 0.43-0.85 (אללוף ויונה, 1990). שנתיים לאחר מכן דיווחו יונה ואללוף (1992) כי המתאם הממוצע בין 44 המעריכים של כלל החיבורים (14,885) בשנים 1991 - 1990 היה 0.81. כמו כן חושבה גם מהימנות בין שופטים רק בעבור זוגות מעריכים שהעריכו בצוותא 30 חיבורים לפחות (סה"כ 54 זוגות). ממוצע המתאמים היה 0.80 (סטיית תקן של 0.08) וטווח המתאמים היה 0.54-0.93. לאחר הפעלת נוסחת ספירמן בראון אומדן המהימנות הגיע ל-0.89. בבדיקה מאוחרת יותר דיווחו יונה ואללוף (1994) כי המתאם הממוצע בין 37 המעריכים של 18,338 החיבורים בשנים 1993 - 1992 היה 0.84. בעבור מועד יולי 1993 חושבה מהימנות בין שופטים לזוגות מעריכים שהעריכו בצוותא 30 חיבורים לפחות (סה"כ 33 זוגות). ממוצע המתאמים היה 0.84 (סטיית תקן של 0.10) וטווח המתאמים היה 0.56-0.97. לאחר הפעלת נוסחת ספירמן בראון אומדן המהימנות הגיע ל-0.91.

בארצות הברית נבדקה מהימנותו של מבחן הכתיבה המצורף ל-TOEFL. Carlson et al. (1985) דיווחו על ממוצע המתאמים בין 46 מעריכים בטווח שבין 0.66 ל-0.74, ולאחר שהופעלה נוסחת ספירמן-בראון המתאמים הגיעו ל-0.80 עד 0.85.

ג. הבדלים בין קבוצות מעריכים

אם מטרת המבחן היא לבא יכולת כתיבה כפי שתבטא בלימודים האקדמיים בחוגים השונים הרי שיש לבדוק אם יש הסכמה בין מרצים ללשון ולכתיבה לבין מרצים במחלקות השונות בדבר הערכת איכותה של הכתיבה. מחקר שנעשה על ידי הצוות האחראי על מבחן TOEFL בדק את הקשר בין הערכה הוליסטית של כתיבה כפי שנעשתה בידי מורים לאנגלית ומורים לאנגלית כשפה זרה לבין הערכה של מרצים מן החוגים מדעי החברה וכלכלה (Carlson et al., 1985). המחקר הראה מידה גבוהה של הסכמה: מתאם של 0.86 בין הערכות מורים לאנגלית לבין הערכות מרצים ממדעי החברה, ומתאם של 0.92 בין הערכות מורים לאנגלית לבין הערכות מרצים מהפקולטה להנדסה. מידת ההסכמה בין מרצים מפקולטות חברה והנדסה הייתה 0.92. מכך אפשר להסיק כי יש הסכמה בין מרצים ממוסדות אקדמיים על הערכה של רמת הכתיבה הנדרשת לפחות מסטודנטים זרים.

ד. מאפיינים חיצוניים של החיבור

עד היום רוב המבחנים הפתוחים ובכללם מבחני הכתיבה נכתבים בכתב יד. מחקרים בדקו אם יש השפעה שיטתית של צורת כתב היד על ההערכה הניתנת. הדבר נעשה על ידי העתקה של

חיבורים ברמות שונות בכתבי יד שונים והערכתם. נמצא כי חיבורים הכתובים בכתב יד יפה יותר קיבלו ציונים גבוהים יותר מחיבורים הכתובים בכתב שהוערך כלא יפה (Hughes, Keeling and Tuck, 1983). אפקט זה כמובן אינו רצוי, ויש להעיר את תשומת לב המעריכים על הטיה אפשרית זו על מנת לנסות ולהקטינה.

בעתיד, עם התרחבות השימוש במחשבים בבתי הספר יש להיערך לאפשרות של כתיבת החיבורים באמצעות מחשב. לכתיבת חיבור במחשב יתרונות ברורים - קל לחזור לטקסט לאחר שנכתב כדי לתקן טעויות או לערוך אותו מחדש, כמו כן אפשר להשתמש בקלות בפונקציות הבודקות איות ודקדוק. אולם מאחר שלא כל הנבחנים מכירים את המחשב ומעבד התמלילים, ומתמצאים בשימוש בהם, נראה כי במבחן כתיבה יש לתת לנבחנים יכולת לבחור בין כתיבה באמצעות נייר ועיפרון לבין כתיבה באמצעות מחשב. הדבר ייצור הבדל צורני גדול בין שני סוגי החיבורים. השאלה היא אם הדבר ישפיע על הערכת איכות הכתיבה. 2 מחקרים בדקו שאלה זו.

Powers, Fowles, Farnum, and Ramsey (1992) נתנו ל-500 נבחנים לכתוב שני חיבורים - אחד בכתב יד והאחר במחשב. החיבורים צוינו בשיטה הוליסטית בידי שני מעריכים על סולם בין 1 ל-6. מדגם של 32 חיבורים נבחר כדי לייצג את כל טווח הביצוע בכל סוג מבחן. חיבורים אשר נכתבו בכתב יד הודפסו במחשב, וחיבורים אשר הודפסו במחשב הועתקו בכתב יד. השוואה של הציונים שניתנו לחיבורים בכתב יד לבין הציונים שניתנו לאותם החיבורים אשר הוקלדו במחשב הראתה כי ציוני החיבורים המוקלדים במחשב היו נמוכים יותר. השוואה בין ציונים שניתנו לחיבורים אשר במקור הוקלדו במחשב לבין הציונים שניתנו לאותם חיבורים שהועתקו בכתב יד, הראתה כי הציונים לחיבורים בכתב יד היו גבוהים במעט מהציונים לחיבורים המקבילים להם אשר הודפסו. נמצאה, אם כך, מגמה עקבית של מתן ציונים נמוכים יותר לחיבורים מודפסים. הגורמים הבאים יכולים אולי להסביר מגמה זו:

- א. אורך החיבורים: חיבורים המודפסים ברווח יחיד נראים קצרים יותר מחיבורים בכתב יד.
- ב. בחיבורים מודפסים טעויות הכתיב והדקדוק בולטות מאוד, ואילו בחיבורים הכתובים ביד, כתב לא מסודר יכול לטשטש את הטעויות.
- ג. בחיבורים בכתב יד נראים התיקונים והשיפוצים של הכותבים, ואילו בחיבורים מודפסים אין עדות לכך. ייתכן שמעריכים תגמלו חיבורים כאשר ראו שנעשו הרבה ניסיונות לשפרם.
- ד. ייתכן כי מעריכים נוטים להזדהות עם כותבי חיבורים בכתב יד. חלק מהמעריכים אמרו כי כאשר אינם מבינים את כתב היד הם נותנים לנבחנים ליהנות מהספק.
- ה. מאחר שרוב החומר המקצועי שאדם קורא הוא מודפס, ייתכן שמראש יש ציפייה שחומר מודפס יהיה ברמה גבוהה יותר מחומר הכתוב בכתב יד, ולכן כאשר ציפייה זו מתבדה, יש הפחתה בציון.

בהמשך המחקר נוספו מעריכים חדשים אשר תודרכו לנסות ולהתעלם מהאפקט של אופן הכתיבה ומאורך החיבור, והחיבורים המודפסים הודפסו ברווח כפול כדי לצמצם את האפקט של האורך. התוצאות הראו כי היתרון בציונים לחיבורים בכתב יד ירד, אולם עדיין נשאר מובהק.

מחקר של חוקרים מן ה-ACT מצא תוצאות הפוכות. Wolfe, Bolton, Feltovich, & Welch (1993) נתנו למחצית מן הנבחנים על פי בחירתם לכתוב את החיבור בכתב יד, ולמחצית השנייה להדפיס את החיבור במחשב. החיבורים צוינו בשיטה הוליסטית, ונמצא כי חיבורים שנכתבו במחשב קיבלו ציונים גבוהים יותר מחיבורים שנכתבו ביד. הבעיה העיקרית של מחקר זה היא שלא הייתה הקצאה אקראית של הנבחנים; סביר להניח שהייתה סלקציה עצמית - הנבחנים הטובים, מכירים את המחשב ומתמצאים בו בחרו לכתוב את חיבוריהם וקיבלו ציונים גבוהים יותר, ומסיבה זו קשה לבטוח בתוצאות. נקודה מעניינת שהעלה מחקר זה נגעה לתהליכי הבדיקה של חיבורים מודפסים לעומת חיבורים הכתובים בכתב יד. מחקר זה מצא כי כאשר המעריכים קוראים חיבור בכתב יד הם קוראים קטע קטן יותר של החיבור בכל פעם, ועוצרים לעתים תכופות כדי להעיר הערות. לעומת זאת, בקריאת חיבורים מודפסים המעריכים עוצרים פחות ומעירים את הערותיהם בסוף הקריאה.

ה. גורמים קונטקסטואליים

בדרך כלל מעריכים מציינים סדרה של חיבורים של נבחנים שונים ברצף, והשאלה היא אם יש השפעה של החיבורים הקודמים על ההערכה שיקבל החיבור הבא אחריהם. Daly & Dickson- Markman (1982) השוו את הציון שקיבל חיבור מסוים כאשר צוינו לאחר ארבעה חיבורים טובים, לאחר ארבעה חיבורים גרועים, לבדו, או אחרי ארבעה חיבורים אשר נבחרו באופן אקראי וקיבלו את התוצאות הבאות:

- א. כאשר החיבור נקרא אחרי קבוצה של חיבורים טובים, הציון הניתן אינו נמוך יותר מהציון הניתן כאשר החיבור נקרא לבדו.
- ב. אפקט חיובי של סדרה - כאשר החיבור נקרא אחרי אוסף רנדומאלי של חיבורים הוא קיבל ציון גבוה יותר מהציון שקיבל כאשר הוא נקרא לבדו.
- ג. אפקט חיובי של קונטרסט - כאשר החיבור נקרא אחרי קבוצה של חיבורים גרועים הוא קיבל ציון גבוה יותר מאשר כשהוא בא לבדו.

המסקנה שעולה מן הממצאים היא שיש לערבב את החיבורים לפני הגשתם לציון ולהגישם למעריכים בסדר אקראי.

חישוב מהימנות בהתבסס על מספר גורמי רעש

בסקירת הספרות של Breland (1983) מוצגים נתונים באשר למהימנות הציונים במבחן הכתיבה המבוססים על מספר משימות כתיבה ומספר מעריכים. המתאם החציוני מעבר לכל המחקרים בין 2 משימות כתיבה שנבדקו בידי 2 מעריכים היה 0.66. המתאם בין 3 משימות כתיבה אשר נבדקו בידי 2 מעריכים היה 0.70. לאחר תיקון לפי נוסחת ספירמן-בראון התקבלה מהימנות של 0.69 עבור 2 משימות ו-0.76 עבור 3 משימות.

Breland et al (1987) בדקו את מהימנות מבחן הכתיבה בהתבסס על תיאוריית ההכללה, ולקחו בחשבון את הגורמים הבאים כמקורות אפשריים לרעש: נושאי כתיבה, ז'אנרים של כתיבה, מעריכים, ואינטראקציות בין הגורמים הללו לגורם הנבחנים. על סמך הנתונים האמפיריים שברשותם הם חישבו בכמה תגדל המהימנות עם כל תוספת של חיבור בז'אנר כתיבה כלשהו, בנושא כלשהו ובמספר המעריכים. הנתונים מופיעים בטבלה מספר 7.

# ז'אנרים	# נושאים בכל ז'אנר	# כולל של חיבורים	מעריך אחד	2 מעריכים	3 מעריכים	4 מעריכים	אינסוף מעריכים
1	1	1	0.42	0.53	0.58	0.63	0.72
	2	2	0.59	0.70	0.74	0.77	0.84
	3	3	0.69	0.77	0.81	0.84	0.89
2	1	2	0.57	0.68	0.72	0.76	0.83
	2	4	0.73	0.81	0.84	0.86	0.90
	3	6	0.80	0.86	0.88	0.90	0.93
3	1	3	0.66	0.75	0.79	0.82	0.87
	2	6	0.80	0.86	0.88	0.90	0.93
	3	9	0.85	0.90	0.92	0.93	0.95

טבלה מספר 7: הערכת המהימנות כפונקציה של מספר הז'אנרים, מספר הנושאים ומספר המעריכים, מתוך Breland et al. (1987)

מן הטבלה עולה התרומה הגדולה של הגדלת מספר החיבורים וסוגיהם למהימנות, תרומה שהיא גדולה יותר מזו הנובעת מהגדלת מספר המעריכים לכל משימה. למשל, בהתייחס למבחן הכולל משימת כתיבה אחת, עלייה במספר המעריכים מ-2 ל-3 תביא להעלאת המהימנות מ-0.53 ל-0.58, ואילו תוספת של משימת כתיבה בנושא אחר אך מאותו ז'אנר ללא הגדלת מספר המעריכים תעלה את המהימנות מ-0.53 ל-0.70.

Cope (1995) מראה אף הוא כי הגדלת מספר משימות הכתיבה תתרום יותר למהימנותו של מבחן כתיבה ישיר מאשר הגדלת מספר המעריכים.

מן המחקרים הללו עולה כי יש תועלת רבה יותר בהגדלת מספר המשימות במבחן כתיבה מאשר בהגדלת מספר המעריכים.

Koenig (1995) בתיאורו את פיתוח מבחן כתיבה למועמדים לבית ספר לרפואה מדווח על שלושה מדדי מהימנות: 1. מתאם בין ביצוע של שתי משימות כתיבה ותיקונו על ידי נוסחת ספירמן-בראון - על פי חישוב זה התקבלו מקדמי מהימנות בטווח שבין 0.77 ל-0.80. 2. מהימנות מבחן חוזר - הייתה בטווח שבין 0.61 ל-0.68. 3. מקדמי ההכללה - היו בטווח שבין 0.76 ל-0.86.

כיוול

כיוול בא לתקן הבדלים שיטתיים בין מדדים שונים הבאים למדוד את אותה התכונה. הבדלים כאלה יכולים לנבוע, למשל, ממבחנים ברמות קושי שונות. מטרת הכיוול היא להעמיד על סולם אחד ציונים המחושבים על סמך מדדים שונים. במבחן כתיבה קיימים שני מקורות להבדלים שיטתיים בין מדדים: הנוסחים השונים של הבחינה והמעריכים השונים הבודקים את החיבורים.

כיוול בין נוסחים:

ניתן לכייל בין נוסחים רק אם מניחים כי הנוסחים השונים מודדים את אותה התכונה. הנחה זו סבירה כאשר המבחן דוגם באופן מייצג את כל האספקטים של תכונה זו. מבחני כתיבה ישירים כוללים בדרך כלל מספר קטן מאוד של משימות כתיבה וקשה להניח כי הם דוגמים את יכולת הכתיבה באופן מלא. בנוסף, מחקרים הראו כי האינטראקציה בין המשימה לנבחנים הסבירה חלק ניכר מן השונות בציונים (Cope, 1995). כתוצאה מכך, יש קושי רב לכייל בין נוסחים שונים של מבחני הכתיבה. בנוסף טוען Cope (1995) כי הקושי בכיוול מבחני הכתיבה מחמיר עוד עקב השימוש בסולם ציונים צר (בדרך כלל בן 5 או 6 דרגות). כדי לסייע בכיוול בין נוסחים שונים של מבחן הכתיבה מציעים מספר חוקרים להשתמש בשאלות רבות ברירה הבודקות מיומנויות כתיבה (Breland, 1983 ; Cooper, 1984).

כיוול בין מעריכים:

באשר למבחנים פתוחים אשר נבדקים בידי מעריכים שונים, יש להעמיד את הציונים שניתנו בידי המעריכים השונים על סולם אחד. נמצא כי למרות ההדרכה הרבה בסדנאות הציון עדיין ייתכנו הבדלים שיטתיים בין המעריכים השונים. המעריכים נבדלים זה מזה במידה בה הם מחמירים או מקלים. בספרות הוצעו שיטות שונות לכיוול בין מעריכים.

Braun (1988) מציע שיטה סטטיסטית לכיוול ציונים של המעריכים השונים הנשענת על אינפורמציה המופקת מניסוי שמתבצע במהלך ההערכה של המבחנים התפעוליים. Cohen (1993) מציע שיטה לכיוול בין מעריכים המבוססת על מציאת סולם גלובאלי אשר אליו יכוילו כל המעריכים. שיטה זו בנויה על חיבורים המשותפים לזוגות המציינים. Engelhard (1992) מציע שיטה המבוססת על המודל הרב-שטחתי של Rasch ואשר מאפשרת לכייל את מבחן הכתיבה גם על הבדלים ברמת הקושי של משימות הכתיבה וגם על ההבדלים בין המעריכים.

פרק ד': תוקף

לפי ה-American Standards for Educational and Psychological Testing שנוסחו על ידי ה-Psychological Association תוקף המבחן מתייחס למידה שבה מוצדקות המסקנות שמסיקים מהמבחן או הפרשנות הניתנת לציוני המבחן. כלומר, התוקף אינו תכונה של המבחן אלא הוא מתייחס למסקנות המוסקות ממנו או לפרשנות הניתנת לציון שלו. עדויות לתוקף יש לחפש במקורות שונים: תוקף תוכן, תוקף כלפי קריטריון ותוקף מבנה. לא די בהוכחות לסוג אחד של תוקף כדי להצדיק את השימוש במבחן. פרק זה יכלול סקירה של מחקרים שבדקו את שלושת סוגי העדויות לתוקף של מבחני כתיבה.

תוקף תוכן

תוקף תוכן מתייחס למידה שבה אפשר להסיק מן המבחן על תחום ההתנהגות שאותו מבקשים ללמוד. לצורך תיקוף תוכן של מבחן כתיבה יש לנתח בשיטתיות את כושר הכתיבה, להגיע להגדרה ממצה של כל מרכיביו, ולבנות את המבחן כך שייצג את כל ההיבטים במשקלות המבטאים את חשיבותם היחסית. חשוב לתת ייצוג לכל המרכיבים הרלוונטיים ולהימנע מלייצג מרכיבים שאינם רלוונטיים.

תוקף תוכן בדרך כלל אינו מוערך באופן כמותי אלא הוא בנוי על הערכה סובייקטיבית של מומחים בתחומים הרלוונטיים. על פניו נראה כי למבחן לבדיקת יכולת כתיבה הדורש מן הנבחן לכתוב חיבור (מבחנים ישירים) יש תוקף תוכן גבוה יותר מלמבחן הכולל שאלות רבות ברירה (מבחנים עקיפים), שאינן דורשות מן הנבחן לכתוב.

מחקרים על מבחני כתיבה ישירים הראו כי יש שונות ביכולת הכתיבה של אותו אדם כפי שהוא מוערך בעזרת משימות כתיבה שונות, לכן, כדי שמבחן כתיבה יהיה תקף מבחינה תוכנית יש לכלול מספר רב של משימות כתיבה או - אם הדבר אינו אפשרי - להתמקד רק בסוג הכתיבה שאותו מעוניינים למדוד, ולהגביל את המסקנות מן המבחן רק לסוג כתיבה זה. תוקף תוכן מוערך לא רק על ידי אפיונים של משימת הכתיבה אלא גם על ידי תהליך הציינון. אם תהליך הציינון מתמקד רק באספקטים של דקדוק וסגנון הרי אין להסיק מהתוצאות של מבחן זה על אספקטים אחרים של יכולת הכתיבה של הנבחנים, כמו ארגון ותוכן.

תוקף כלפי קריטריון

תוקף כלפי קריטריון בודק את ביצועיו של הנבחן במבחן כנגד קריטריון, כלומר כנגד מדד ישיר ובלתי-תלוי של מה שהמבחן נועד למדוד (אנסטאזי, 1990). כאשר הקריטריון הוא ביצוע של הנבחנים בעתיד מכונה סוג תוקף זה תוקף ניבוי, וכאשר הקריטריון הוא ביצוע של הנבחנים במבחן אחר בהווה מכונה סוג תוקף זה תוקף מקביל. אחת השאלות העולות בהקשר של תוקף כלפי קריטריון היא מהו הקריטריון, או מהו הדבר אותו אנו מנסים לנבא. המחקרים שבדקו את יכולת הניבוי של מבחני כתיבה השתמשו בקריטריונים הבאים: ציונים בקורסים לכתיבה או

הערכת המרצה את יכולת הכתיבה, ציונים בקורסים בתחומים שונים הכוללים מטלות כתיבה וציונים בתום שנה א' בקולג'. קריטריונים אלה נבדלים במידת הקירבה שלהם לחזאי. קריטריון כמו ציונים או הערכות מקורסים לכתיבה הוא קרוב יחסית ליכולת הנמדדת על ידי החזאי, בעוד שציונים בתום שנה א' מושפעים מגורמים רבים אחרים פרט ליכולת הכתיבה. יכולת הניבוי של ציונים אלה תלויה באופן שבו נקבעים ציוני שנה א'; היא צפויה להיות גבוהה יותר בחוגים הדורשים משימות כתיבה רבות (כמו במדעי הרוח) ונמוכה יותר בחוגים שבהם הסטודנטים אינם נדרשים להתבטא בכתב לעתים תכופות (כמו במדעים המדויקים).

מחקרים אשר בדקו את תוקף הניבוי באופן כמותי הסתכלו בדרך כלל על שני מדדים: א. המתאם בין ציוני מבחן הכתיבה לבין הקריטריון. ב. התרומה הייחודית של מבחן הכתיבה לניבוי הקריטריון נוסף על החזאים שכבר קיימים.

Breland (1983) סוקר ממצאי 3 מחקרים שבדקו את תוקף הניבוי של מבחני כתיבה בין השנים 1977 ל-1980. מן הנתונים עולה כי המתאם הגבוה ביותר (בסביבות 0.50) היה בין הציון על כתיבת חיבור בתחילת השנה לבין הציון על כתיבת חיבור מאוחר יותר במהלך השנה. המתאם בין ציוני החיבור לציוני אנגלית היה בטווח של 0.26 - 0.32, והמתאם בין ציוני החיבור לציוני שנה א' בקולג' היה בין 0.21 ל-0.25. כמו כן הוא גם מסכם נתונים אשר בדקו את התוקף השולי של מבחן הכתיבה. מחקר של Breland & Gaynor (1979) (מצוטט ב- Breland (1983)) בדק את התרומה השולית של מבחן הכתיבה לניבוי הציונים באנגלית בשנה א' בקולג' נוסף על השימוש בחזאים הבאים: דירוג בבית הספר התיכון, ציון באנגלית בבית הספר התיכון, הציון המילולי ב-SAT והציון במבחן ה- TSWE (Test of Standard Written English), שהוא מבחן כתיבה עקיף. נמצא כי התוקף השולי נע בין 0.02 ל-0.04. מחקר של Checkettes and Christensen (1974) (מצוטט ב- Breland (1983)) מצא כי מבחן החיבור הוסיף 0.06 לתוקף הניבוי של ציוני אנגלית בשנה א' בקולג' מעבר לתוקף של מבחן הכתיבה העקיף. מספר מחקרים בדקו את התוקף של המימדים הנפרדים בשיטת צינור אנליטית כלפי ציוני בית ספר תיכון או ציוני אנגלית בבית ספר תיכון. נמצא כי למדדי סגנון, תחביר ודקדוק תוקף גבוה יותר מאשר למדדים של תוכן וארגון, על אף שהאחרונים נחשבים למדדים של כשרים גבוהים יותר. ייתכן שהדבר נובע ממהימנות גבוהה יותר בהערכת המימדים הלשוניים לעומת המימדים התוכניים.

Breland et al. (1987) בדקו את יכולת הכתיבה של 267 סטודנטים בקולג' שכל אחד מהם כתב 6 חיבורים משלושה ז'אנרים: נרטיבי, אקספוזיטורי ושינוע. תוקף הניבוי של כל חיבור נבדק אל מול שלושה קריטריונים: הציונים בחיבורים האחרים במחקר, הערכת המרצה את יכולת הכתיבה של הסטודנט וציונים בקורסים לאנגלית בשנה א' בקולג'. בטבלה מספר 8 מוצגים מקדמי התוקף של כל ז'אנר כתיבה כלפי כל אחד משלושת הקריטריונים על פי מספר המעריכים.

זיאנר הכתיבה (2) חיבורים מכל זיאנר)	מספר מעריכים	מתאם עם כתיבה בחיבורים האחרים	מתאם עם הערכת מורים	מתאם עם ציונים בקורסים באנגלית בקולג' באנגלית
נרטיבי	1	0.52	0.40	0.37
נרטיבי	2	0.59	0.45	0.42
נרטיבי	3	0.63	0.47	0.44
אקספוזיטורי	1	0.57	0.38	0.34
אקספוזיטורי	2	0.64	0.42	0.39
אקספוזיטורי	3	0.67	0.44	0.41
שכנועי	1	0.53	0.48	0.45
שכנועי	2	0.60	0.53	0.49
שכנועי	3	0.62	0.56	0.51

טבלה מספר 8: מקדמי התוקף של החיבורים כלפי הקריטריונים השונים, מתוך Breland et al. (1987)

מן הטבלה עולות הנקודות הבאות:

- המתאמים בין הציונים על חיבור מזיאנר אחד לבין ציונים על חיבורים אחרים במחקר היו בטווח של 0.5 עד 0.6 כאשר הם נקראו בידי מעריך אחד בלבד; המתאמים עלו לטווח של 0.6 ל- 0.67 כאשר הם נקראו בידי 3 מעריכים.
 - המתאמים בין הציונים על החיבורים לבין הערכות המורים: בין 0.38 ל- 0.48 לאחר הערכה אחת, בין 0.44 ל- 0.56 לאחר שלוש הערכות.
 - המתאמים בין הציונים על החיבורים לבין ציוני אנגלית בקולג': 0.34 - 0.45 לאחר הערכה אחת, 0.36 עד 0.50 לאחר שתי הערכות, ו- 0.41 עד 0.51 לאחר שלוש הערכות.
- מבין שלושת הזיאנרים - לחיבורים מזיאנר שיכנועי התוקף הגבוה ביותר בניבוי הערכות מורים וציונים בקורסים באנגלית.
- אם כך, ביצוע על חיבור אחד ניבא באופן הטוב ביותר ביצוע בחיבורים האחרים באותו המחקר, אולם גם תוקף הניבוי כלפי הערכת מורים וציוני אנגלית בקולג' היה גבוה למדי.
- בנוסף, נבדקו גם המתאמים שבין ציוני החיבורים לבין ציונים במבחנים רבי-ברירה הבודקים כשרים קרובים: SAT מילולי, TSWE, ו-ECT. שני האחרונים הם מבחני כתיבה עקיפים.

ECT	TSWE	SAT מילולי	ציוני החיבורים
0.61-0.66	0.63-0.71	0.54-0.57	

טבלה מספר 9: מתאמים בין ציונים בחיבורים לבין ציונים במבחנים רבי ברירה, מתוך Breland et al. (1987)

לצורך השוואה הציג Breland et al. (1987) את המתאמים בין הציונים ב-SAT מילולי וב-TSWE ובין הערכות המורים והציונים באנגלית בשנה א' בקולג':

הערכות מורים	ציונים באנגלית בשנה א'	
0.46	0.44	SAT מילולי
0.48	0.41	TSWE

טבלה מספר 10: מתאמים בין ציונים במבחנים סגורים ובין הערכות המורים וציונים בשנה א' בקולג', מתוך Breland et al. (1987)

נראה, אם כן, כי החיבורים אינם נופלים בתוקף הניבוי שלהם ממבחני ה-SAT המילולי וממבחן ה-TSWE, על אף שמבחנים אלה כוללים מספר גדול בהרבה של פריטים. כאשר חושב התוקף משילוב של חיבור אחד וציונים במבחני TSWE ו-ECT אזי מקדמי התוקף לניבוי הערכת המורים היו בין 0.55 ל-0.60, ומקדמי התוקף לניבוי ציונים בקורסים לאנגלית היו בין 0.51 ל-0.60. שילוב המבחנים אלה הביא להקטנת התרומה של הגדלת מספר המעריכים ולהקטנת ההבדלים בין הז'אנרים השונים של החיבורים.

תוקף שולי - חושבה תרומתם של החיבורים לתוקף מעבר לניבוי המבוסס על TSWE ו-ECT. התרומה הממוצעת של חיבורים אשר הוערכו בידי שני מעריכים לניבוי הערכות המורים הייתה 0.068, ולניבוי ציוני הקורסים באנגלית - 0.071. מתוך שלושת סוגי החיבורים, חיבורים שכנועיים תרמו את התרומה המשמעותית ביותר לתוקף. החיבור האקספוזיטורי ניבא בצורה הטובה ביותר את הציונים של שאר החיבורים, ואילו החיבור השכנועי ניבא בצורה הטובה ביותר את הערכות המורים והציונים באנגלית בקולג'. יש לציין כי בעוד שאר החיבורים נכתבו על ידי הסטודנטים בכיתה, החיבור השכנועי נכתב בבית, הוגש למורים להערכה, ותוקן בהתאם. הציון הסופי ניתן על סמך הגירסה המתוקנת.

Bridgeman (1991) חקר את תוקף הניבוי של מבחני כתיבה ישירים ועקיפים כלפי ציוני שנה א' בקולג'. במחקרו היו שני מדגמים:

א. 5,430 סטודנטים אשר התקבלו לקולגים של המדינה בניו ג'רסי ואשר נבחנו ב-Basic Skills Assessment שכולל בתוכו 2 מבחנים להערכת יכולת כתיבה: מבחן רב ברירה שבו 35 פריטים ב-35 דקות, וכן מבחן של 20 דקות לכתיבת חיבור המוערך בשיטה הוליסטית בידי שני בודקים. מהימנות בין בודקי החיבורים הוערכה להיות 0.83. המתאם בין החיבור למבחן הכתיבה הסגור היה 0.44.

ב. 6,088 סטודנטים אשר התקבלו לקולגים סלקטיביים יותר ולכן אפשר לראותם כסטודנטים בעלי יכולת גבוהה יותר מהסטודנטים במדגם הראשון. הם נבחנו ב-English Composition Test שבו 70 שאלות רבות ברירה ב-40 דקות, ו-20 דקות של כתיבת חיבור המוערך בשיטה הוליסטית.

התוצאות נותחו בנפרד לשני המדגמים :

מדגם א' -

בטבלה מספר 11 מוצגים מקדמי התוקף של מבחני הכתיבה בשילוב עם ציוני SAT מילולי וכמותי, ודירוג בית ספר תיכון, כלפי ציוני שנה א' בקולג' כפי שהתקבלו בקולגיים השונים.

חזאי	קולג' 1	קולג' 2	קולג' 3	קולג' 4	ממוצע
מספר תלמידים	1499	884	1635	1413	
דירוג ביי"ס תיכון + SAT מילולי + SAT כמותי	0.404	0.436	0.418	0.525	0.445
דירוג בית ספר תיכון + SAT מילולי + SAT כמותי + ציון כתיבה כולל	0.420	0.448	0.422	0.533	0.454
דירוג בית ספר תיכון + SAT מילולי + SAT כמותי + מבחן כתיבה סגור	0.424	0.451	0.429	0.530	0.457
דירוג בית ספר תיכון + SAT מילולי + SAT כמותי + מבחן כתיבה סגור + חיבור	0.424	0.451	0.430	0.533	0.459
דירוג בית ספר תיכון + SAT כמותי + מבחן כתיבה סגור + חיבור	0.414	0.443	0.408	0.526	0.446

טבלה מספר 11: מקדמי התוקף של החזאים השונים כלפי ציוני שנה א' בקולג' במדגם א', מתוך *Bridgeman (1991)*

מן הטבלה עולה כי הציונים של מבחני כתיבה תרמו מעט (בין 0.01 ל- 0.02) לתוקף הניבוי מעבר לתרומה של דירוג בית ספר תיכון ומבחני SAT המילולי והכמותי. כאשר בוחנים בנפרד את התרומה של מבחן הכתיבה הישיר ומבחן הכתיבה העקיף לתוקף, נראה כי התרומה העיקרית היא מן המבחן העקיף. ייתכן כי הדבר נובע מכך שרבים מציוני שנה א' בקולג' מבוססים על מבחנים הכוללים פריטים רבי ברירה.

מדגם ב' -

מדגם זה כלל נבחנים מ- 21 קולגיים. בטבלה הבאה מוצגים המתאמים הממוצעים והחציוניים של שלושת החזאים: ציון SAT מילולי, ציון על החלק הסגור של ECT, וציון על החיבור של ECT עם ציוני שנה א' בקולג'.

N	מדד	SAT-V	ECT - חלק סגור	ECT - חיבור
6,088	ממוצע	0.29	0.30	0.17
	חציון	0.30	0.30	0.16
	טווח	0.26-0.31	0.28-0.33	0.15-0.20

טבלה מספר 12: מקדמי המתאם של כל חזאי עם ציוני שנה א' בקולג', במדגם ב' מתוך *Bridgeman (1991)*

למבחן החיבור של ECT התוקף הנמוך מתוך השלושה אולם יש לזכור כי הוא מורכב רק מפריט אחד ואילו שאר המבחנים בנויים ממספר רב של פריטים. מהימנות בין מערכי החיבורים הייתה 0.72 (מהימנות נמוכה יותר מהמהימנות במדגם א', אולי בגלל שבמדגם ב' אוכלוסיה יותר הומוגנית). בטבלה מספר 13 מוצגת התרומה השולית של מבחני הכתיבה לתוקף במדגם ב':

חזאי	טווח	ממוצע
דירוג ביי"ס תיכון + SAT מילולי + SAT כמותי	0.34-0.58	0.469
דירוג בית ספר תיכון + SAT מילולי + SAT כמותי + מבחן כתיבה סגור	0.34-0.58	0.482
דירוג בית ספר תיכון + SAT מילולי + SAT כמותי + מבחן כתיבה סגור + חיבור	0.34-0.58	0.486

טבלה מספר 13 : מקדמי התוקף של החזאים השונים עם ציוני שנה א' בקולג', במדגם ב' מתוך Bridgeman (1991).

התוספת השולית הממוצעת של מבחן הכתיבה הסגור מעבר לדירוג בית הספר התיכון ולציוני SAT כמותי ומילולי היא 0.013 והטווח בין 0 ועד 0.03.

התוספת השולית הממוצעת של מבחן הכתיבה הפתוח מעבר לדירוג בית הספר התיכון, לציוני SAT כמותי ומילולי ולציוני המבחן הסגור היא 0.004 והטווח בין -0.01 ועד 0.014.

מן הנתונים עולה כי התרומה השולית של מבחן הכתיבה הסגור קטנה למדי אך היא קיימת, ואילו המבחן הפתוח אינו תורם לניבוי.

החוקרים בדקו גם את התוספת לתוקף כאשר הקריטריון הוא רק קורסים באנגלית וכאשר הקריטריון הוא ציון ממוצע של כל הקורסים חוץ מקורסי המדע, ומצאו שבשני המקרים התרומה של החיבור לא הייתה משמעותית: -0.001, 0.004 בהתאמה.

לסיכום הסיקו החוקרים כי מבחן הכתיבה הסגור תרם תרומה קטנה לניבוי ציוני שנה א' בקולג' על פני ציוני SAT ודירוג בית הספר התיכון, ומבחן החיבור לא תרם כל תרומה נוספת לניבוי.

מגבלות המחקר: מתייחס רק לחיבור קצר אשר נכתב במגבלות זמן ואשר מוערך בשיטה הוליסטית. יתכן שאפשר לשפר את הניבוי על ידי מתן יותר משימות כתיבה מז'אנרים שונים.

ייתכן גם שמבחן הכתיבה יתרום יותר לתוקף אם הקריטריון לא יורכב מציונים בקורסים אשר ניתנו רק על סמך מבחנים סגורים.

מחקר של ה-ETS ניסה לבדוק מה יהיו ההשפעות של המתכונת החדשה של ה-SAT על תוקף הניבוי שלו (Hale et al. 1992). אחד השינויים העיקריים במתכונת החדשה הוא הכנסת מבחן כתיבה ל-SAT-II אשר כולל שאלות רבות-ברירה ומטלת כתיבה. מבחן זה מחליף את מבחן ה-Test of Standard Written English (TSWE) שהוא מבחן רב ברירה שהתלווה לכל בחינה של ה-SAT, וכן את מבחן ה-English Composition Test (ECT) אשר כלל משימת כתיבה רק פעם אחת בשנה.

המחקר נעשה על ידי מתן מבחן SAT במתכונת ישנה וחדשה לסטודנטים ב-19 קולגיים ברחבי ארצות הברית. הסטודנטים ביצעו את המבחנים במהלך שבוע האוריינטציה שלהם תוך ידיעה שהישגיהם לא ישפיעו עליהם בשום צורה, וייתכן שהיעדר מוטיבציה השפיע על ביצועיהם בבחינה. במסגרת הניסוי כל סטודנט ביצע שני חלקים, אשר כל אחד מהם ארך 30 דקות. מבחן ה-SAT במתכונתו הישנה כלל שני חלקים מילוליים ושני חלקים כמותיים. המבחן החדש כלל שני חלקים מילוליים חדשים, שני חלקים כמותיים חדשים ומבחן כתיבה הבנוי משאלות רבות ברירה ומחיבור, אשר צויין בשיטה הוליסטית בידי 2 מעריכים. ציון הכתיבה חושב על ידי שקלול של הציונים על הפריטים הסגורים (70%) ושל הציון על החיבור (30%). מהימנותו של מבחן הכתיבה הסגור הייתה 0.86, ויכולת ההבחנה של הפריטים, כפי שנמדדה על ידי מתאם biserial הממוצע הייתה 0.5. מהימנות בין שופטים של מבחן הכתיבה הפתוח הייתה 0.81.

קשר בין החזאים השונים: המתאם בין ציון הכתיבה הכולל (הבנוי משקלול של מבחן רב ברירה וחיבור) למבחן המילולי היה 0.70, ובינו לבין המבחן הכמותי היה 0.42. נבדק תוקף ניבוי כלפי ציוני שנה א' בקולג', והתוצאות מוצגות בטבלה מספר 14:

החזאים		SAT מתכונת ישנה	SAT מתכונת חדשה
מבחן יחיד			
	מילולי	0.30	0.35
	כמותי	0.28	0.35
	כתיבה (70% סגור/30% פתוח)		0.37
צירוף מבחנים (משקלות שווים)			
	מילולי וכמותי	0.33	0.41
	מילולי, כמותי וכתיבה		0.43
צירוף מבחנים (משקלות אופטימליים)			
	מילולי וכמותי	0.35	0.42
	מילולי, כמותי וכתיבה		0.47

טבלה מספר 14: תוקף ניבוי של החזאים השונים כלפי ציוני שנה א' בקולג', מתוך Hale et al., 1992

מן הטבלה עולה כי כחזאי יחיד למבחן הכתיבה תוקף דומה לתוקף של המבחן המילולי והכמותי. כאשר בוחנים את התרומה השולית של מבחן הכתיבה מעבר לציוני SAT מילולי וכמותי נראה כי מבחן הכתיבה הוסיף תרומה בינונית לתוקף כאשר המשקלות היו אופטימליים, ואילו במשקלות שווים הוא הוסיף תרומה קטנה בלבד. לטענת הכותבים סביר להניח כי התרומה האמיתית היא בין שני הערכים. אולם יש לזכור כי בקבלה לאוניברסיטה בארצות הברית משתמשים בחזאי נוסף - הצלחה בבית ספר תיכון - ויש לבדוק אם למבחן הכתיבה תרומה ייחודית לניבוי הקריטריון נוסף על דירוג בבית ספר התיכון וציוני SAT. התוצאות מוצגות בטבלה מספר 15.

חזאי	SAT מתכונת ישנה תוקף ניבוי	SAT מתכונת ישנה תוקף שולי	SAT מתכונת חדשה תוקף ניבוי	SAT מתכונת חדשה תוקף שולי
------	-------------------------------	------------------------------	-------------------------------	------------------------------

משקלות שווים של החזאים

דירוג בית ספר תיכון	0.45		0.43	
דירוג בית ספר תיכון וצירוף של SAT מילולי וכמותי	0.51	0.06	0.52	0.09
דירוג בית ספר תיכון, צירוף של SAT מילולי וכמותי, וציון כתיבה			0.52	0.09

משקלות אופטימליים של החזאים

דירוג בית ספר תיכון	0.45		0.43	
דירוג בית ספר תיכון וצירוף של SAT מילולי וכמותי	0.53	0.08	0.53	0.09
דירוג בית ספר תיכון, צירוף של SAT מילולי וכמותי, וציון כתיבה			0.55	0.12

טבלה מספר 15: תוקף ניבוי של צירוף החזאים השונים כלפי ציוני שנה א' בקולג', מתוך Hale et al. 1992

כאשר מסתכלים על תוקף שולי מעבר לדירוג בית ספר תיכון רואים כי לציוני SAT מילולי וכמותי תרומה ייחודית לתוקף, ואילו למבחן הכתיבה תרומה לתוקף רק כאשר קיבל משקל אופטימלי, ותרומה זו נעלמה במשקלות שווים. ייתכן כי הדבר נובע ממתאם גבוה בין מבחן הכתיבה לדירוג בית הספר התיכון.

מחקר נוסף שנעשה על אותו המדגם בדק את התוקף של המתכונת החדשה של ה-SAT כמבחן השמה לקורסים באנגלית (Bridgeman, Hale, Lewis, Pollack and Wang, 1992). מקדמי התוקף לניבוי ציונים בקורסים באנגלית מוצגים בטבלה מספר 16.

חזאי	מתאם עם ציונים בקורסים באנגלית
דירוג בבית ספר תיכון	0.30
ציוני SAT מילולי	0.19
מבחן כתיבה רב ברירה	0.29
מבחן חיבור	0.22
מבחן כתיבה רב ברירה וחיבור	0.32
צירוף של SAT מילולי ושני מבחני הכתיבה	0.29

טבלה מספר 16: תוקף ניבוי של החזאים השונים כלפי ציוני שנה א' בקורסים באנגלית, מתוך Bridgeman et al, (1992)

כאשר הקריטריון היה ציונים בקורסים באנגלית נראה כי כל אחד ממבחני הכתיבה ניבא טוב יותר מציון SAT מילולי, וצירוף של שני מבחני הכתיבה היה החזאי הטוב ביותר.

מחקר של (Bridgeman and Lewis, 1994) השווה ביצוע של 821 סטודנטים במבחנים פתוחים וסגורים למטרות השמה במספר מקצועות (היסטוריה אמריקנית, היסטוריה אירופאית, אנגלית וביולוגיה) ובדק את המתאם שלהם עם ציוניהם בסיום הקולג'. המבחן באנגלית הוא הרלוונטי ביותר לנושא מבחני הכתיבה - המבחן הפתוח כלל שלוש משימות כתיבה, ואילו המבחן הסגור כלל 65 שאלות רבות ברירה. כאשר הקריטריון היה ממוצע כל הציונים בתום הקולג', התוקף של המבחן הסגור לא היה שונה מן התוקף של החיבור (0.25 לעומת 0.26), והציון המשוקלל של שניהם ניבא את הקריטריון טוב יותר מכל אחד מהחזאים בנפרד (0.31). התוצאות היו דומות גם כאשר הקריטריון היה ציונים בקורסים לאנגלית בלבד.

רוב המחקרים השתמשו במדדים כמותיים להערכת התוקף פרט למחקרם של Powers, Fowles, and Willard (1994) שבדקו את התוקף המקביל של מבחני כתיבה באופן איכותי. התוקף נבדק באמצעות בדיקת הקשר בין ציונים במבחני כתיבה ישירים (בשיטה הוליסטית על סולם בין 1 ל-6) לבין הערכות שנתנו מרצים מאוניברסיטאות שונות על אותם החיבורים. המרצים נתבקשו להשיב על השאלה "באיזו מידה תתיה מרוצה מרמת הכתיבה של נבחנים שכתבו את החיבורים הללו כמועמדים לקורסים או לתכנית שלך? ההערכות שנעשו ללא ידיעת הציונים היו על סולם בן 5 דרגות: "מרוצה מאוד", "מרוצה למדי", "לא מרוצה ולא לא מרוצה", "לא מרוצה למדי", "מאד לא מרוצה". נמצא קשר חיובי בין הערכות המרצים לציונים שניתנו קודם לכן, אולם הוא לא היה ליניארי: ניתנו יותר הערכות של "מאד לא מרוצה" לציונים בקצה הנמוך של הסולם (2) מאשר הערכות של "מאד מרוצה" לציונים בקצה הגבוה של הסולם (5 או 6). בנוסף, הפער בין

ציונים 2 ל- 3 בקצה התחתון של סולם הציונים התרגם לפער גדול יותר בהערכות מאשר הפער בין ציונים 5 ל- 6 בקצה העליון של הסולם.

לסיכום, כאשר מסתכלים על מספר מחקרים אשר בדקו את נושא תוקף הניבוי של מבחני הכתיבה עולות הנקודות הבאות: ברוב המחקרים נמצא תוקף גבוה למדי למבחני כתיבה ישירים כחזאי יחיד בניבוי הצלחה בקורסים לאנגלית בפרט ובניבוי הצלחה בלימודים בכלל. במספר מקרים תוקף הניבוי של ציון על חיבור אחד בלבד היה דומה לזה של המבחן המילולי ב-SAT או של מבחני כתיבה עקיפים המורכבים ממספר רב של פריטים. תרומה שולית של מבחני הכתיבה הישירים, לעומת זאת, נמצאה רק בחלק מן המחקרים בספרות. התרומה הגדולה ביותר נמצאה במחקר של Breland et al. (1987). ייתכן כי הדבר נובע מהעובדה שזה המחקר היחיד אשר כלל מספר גדול יחסית של משימות כתיבה (6) משלושה ז'אנרים. בנוסף, המחקרים מצביעים על כושר ניבוי טוב למבחני כתיבה סגורים. חוקרים רבים הציעו כי החזאי הטוב ביותר הוא שילוב של משימת כתיבה ומבחן רב ברירה (Cooper, 1984). יתרון נוסף להכללת פריטים סגורים הוא שאפשר להשתמש בהם לכיול ציונים על חיבורים בנושאים שונים (Breland et al. 1987).

תוקף מבנה

תוקף מבנה מתייחס למידה שבה המבחן מודד תכונה או מבנה תיאורטיים. מבנים תיאורטיים בהגדרה אינם ניתנים לצפייה באופן ישיר, אלא הם מוסקים מעדויות התנהגותיות. מבנה מוגדר בדרך כלל בשתי רמות: א. הגדרה אופרציונאלית - תרגום המבנה התיאורטי למדדים אמפיריים. לרוב יש מספר דרכים למדוד אותו המבנה, וצפוי קשר חיובי חזק בין המדדים השונים של אותו המבנה (תוקף מתכנס). ב. בהתבסס על ההגדרה התיאורטית גזירת השערות לגבי קשרים בין מבנה זה למבנים אחרים ובדיקתן באופן אמפירי. השיטות הסטטיסטיות המקובלות לבדיקת תוקף מבנה הן ניתוח גורמים ושיטת Multitrait-Multimethod (Crocker & Algina, 1986). תוקף המבנה של מבחן כתיבה מתייחס למידה שבה מבחן זה אכן מודד יכולת כתיבה ואותה בלבד. עדויות לתוקף מבנה מתקבלות מבחינת הקשר בין מבחני כתיבה שונים. באופן ספציפי אפשר להסתכל על הקשר בין ביצוע במשימות כתיבה שונות, וכן על הקשר בין ביצוע במבחני כתיבה ישירים ועקיפים.

הקשר בין משימות כתיבה שונות

סקירה של מידת הקשר בין משימות הכתיבה השונות ניתנה בפרק הדין במהימנות הבחינה. סקירה זו העלתה כי המתאם בין ביצוע במשימות כתיבה בז'אנרים ובנושאים שונים אינו תמיד גבוה. כאן נזכיר ניתוחי גורמים שביצעו Breland et al (1987) על נתונים של 267 סטודנטים, אשר כל אחד מהם כתב 2 חיבורים בכל אחד משלושה ז'אנרים (נרטיבי, אקספוזיטורי ושיכנועי).

טעינות הגורמים ומדדי ה- goodness of fit הצביעו על קיומו של גורם כללי של כתיבה, שהסביר כ- 78% מן השונות. נראה כי קיים גם תת-גורם של נושא החיבור, ותת-גורם פחות מובהק של זיאנר החיבור. תת-הגורם של זיאנר החיבור נבע בעיקר מהבדלים בביצוע בין הכתיבה השיכנועית לשני הזיאנרים האחרים. ייתכן שהבדלים אלה נבעו מהעובדה כי החיבורים השיכנועיים היו מטלת בית, בעוד שאת החיבורים האחרים הסטודנטים כתבו בכיתה. ההתאמה הטובה ביותר לנתונים הושגה מניתוח גורמים המבוסס על מודל אשר לקח בחשבון גם את השונות שנבעה מנושא החיבור.

Quellmalz, Capell, and Chou (1982) נתנו ל- 200 תלמידי כיתות י"א ו-י"ב מספר מבחני כתיבה: שתי משימות כתיבה ארוכות בשני זיאנרים (נרטיבי ואקספוזיטורי), כתיבת חיבור קצר ושאלות רבות ברירה. מחצית התלמידים כתבו שני חיבורים מאותו זיאנר ומחציתם כתבו שני חיבורים משני זיאנרים שונים. החוקרים בדקו את הקשר בין הערכות על חיבורים בזיאנרים שונים תוך התייחסות למדדי ההערכה הנפרדים. החיבורים הוערכו על המימדים הבאים: בהירות, ארגון, תמיכה בטיעונים ודיוק לשוני. בניתוח גורמים LISREL עלו 2 גורמים המייצגים את הזיאנרים של הכתיבה. בנוסף נמצא הבדל ברמת היציבות של הציונים על המימדים השונים שקיבלו אותם אנשים אשר כתבו שני חיבורים בשני זיאנרים שונים. היציבות נמדדת באמצעות המתאמים שבין הציונים בכל מימד בשני החיבורים. המתאמים הגבוהים ביותר היו במימדי הדיוק הלשוני (0.545) והבהירות (0.327) והמתאמים הנמוכים ביותר היו במימדי הארגון (0.007) והתמיכה בטיעונים (0.087).

הממצא החוזר של מתאמים נמוכים בין ציונים שקיבלו אותם אנשים במשימות כתיבה שונות גרם לחוקרים לטעון כי יכולת הכתיבה היא מורכבת ורב-מימדית והיא מתבטאת באופן שונה במשימות כתיבה שונות (Cherry & Mayer, 1993). מחקרם של Quellmalz, Capell, and Chou (1982) שבדק את המתאם בין מימדי ההערכה האנליטיים הנפרדים שופך אור נוסף על תופעה זו. תוצאותיו מראות כי חלק מכושרי הכתיבה (דיוק לשוני ובהירות) הם יציבים על פני משימות כתיבה שונות, ואילו כשרים אחרים (ארגון ותמיכה בטיעונים) מתבטאים באופן שונה במשימות כתיבה שונות.

הקשר בין מבחני כתיבה ישירים לבין מבחני כתיבה עקיפים

Cooper (1984) מסכם מחקרים אשר בדקו את הקשר בין מבחני הכתיבה הישירים ובין מבחני הכתיבה העקיפים. התוצאות מוצגות בטבלה מספר 17.

המחקר	הנבדקים	N	מתאם בין מבחני כתיבה ישירים לעקיפים
Breland, Colon & Rogosa (1976)	סטודנטים בשנה א'	96	0.42
Breland & Gaynor (1979)	סטודנטים בשנה א'	2,200	0.58-0.63
Huntley, Schmeiser, & Stiggins (1979)	סטודנטים	50	0.43-0.67
Godshalk, Swineford & Cofman (1966)	תלמידי תיכון	646	0.46-0.75
Hogar and Mishler (1980)	תלמידי כיתות ג' ו-ח'	300	0.65-0.68
Moss, Cole & Khampalikit (1981)	תלמידי כיתות ד'	84	0.2-0.68
	תלמידי כיתות ז'	45	0.6-0.67
	תלמידי כיתות י'	98	0.75-0.76

טבלה מספר 17: תוצאות מחקרים שבדקו את הקשר בין מבחני כתיבה ישירים ועקיפים, מתוך Cooper (1984)

מן הטבלה עולה כי המתאמים בין מבחני כתיבה ישירים לבין מבחני כתיבה עקיפים הם גבוהים יחסית, במיוחד בהתחשב במהימנות הנמוכה של מבחני הכתיבה הישירים הנובעת מהבדלים בין מעריכים ומהבדלים בין משימות כתיבה שונות.

Quellmalz, Capell, and Chou (1982) בדקו גם הם את הקשר בין הערכות של יכולת כתיבה המבוססות על משימות כתיבה לעומת אלה המבוססות על שאלות רבות ברירה. הפריטים רבי הברירה נבנו כך שימדדו כשרים דומים לאלה הנמדדים במשימות הכתיבה. לצורך זה הם התבססו על קטעים בז'אנרים נרטיבי ואקספוזיטורי והם בדקו תכונות של בהירות, ארגון ותמיכה בטיעונים. בפריטים רבי ברירה בניגוד למשימות כתיבה לא היה אפקט של ז'אנר הקטע על הביצוע של הנבחנים.

החיבורים הארוכים והקצרים צוינו על ארבעה מימדים: מיקוד, ארגון, תמיכה לטיעונים ודיוק לשוני; המבחן העקיף צויין רק על שלושת המימדים הראשונים (ללא דיוק לשוני). בטבלה מספר 18 מוצגים המתאמים בין מימדי הכתיבה כפי שנבדקו על ידי מבחני הכתיבה השונים.

מדד	מתאם בין שני החיבורים-	מתאם בין החיבורים למבחן העקיף	מתאם בין החיבורים לחיבור הקצר
בהירות	0.42	0.26	0.23
		0.38	0.27
ארגון	0.25	0.14	0.18
		0.31	0.23
תמיכה לטיעונים	0.31	0.20	0.35
		0.17	0.29
דיוק לשוני	0.59		

טבלה מספר 18: מתאמים בין ציונים על מספר מימדי כתיבה על בסיס שלושה מבחני כתיבה, מתוך Quellmalz, Capell, & Chou (1982)

מהטבלה עולה כי הקשר בין החיבורים למבחן העקיף במחקר זה הוא חלש יחסית, בדרך כלל, הוא נמוך מהקשר בין החיבורים השונים. בניתוח גורמים שכלל את מימדי הציינון ואת סוגי הבחינה השונים התקבל מערך של 7 גורמים: 4 גורמי בחינה (חיבור נרטיבי, חיבור אקספוזיטורי, חיבור קצר ומבחן עקיף) ושלושה מדדי הערכה (מדד קוהרנטיות המשלב את המיקוד והארגון, מדד תמיכה בטיעונים ומדד דיוק לשוני). הקשר החלש בין מבחני כתיבה ישירים לעקיפים נובע כנראה מההבדל שבין המטלות הנדרשות מן הנבחן בכל סוג מבחן.

אם כך, נתונים ממבחנים שונים להערכת יכולת כתיבה מראים מידה נמוכה יחסית של התכנסות. הדבר יכול לנבוע מהיות יכולת הכתיבה תכונה רב מימדית, ממהימנות נמוכה יחסית בציינון או מצירוף של שני הדברים.

Messick (1989) מרחיב את הגדרת המושג תוקף וטוען כי בתהליך איסוף עדויות לבדיקת התוקף של מבחן יש לקחת בחשבון גם את ההשלכות החברתיות של השימוש בו. יש, לדעתו, לבדוק את התוצאות המתוכננות ואת תוצאות הלואי של המבחן, את ההשלכות הממשיות ואת ההשלכות הפוטנציאליות. המידה בה השלכות אלה נחשבות כחיוביות או רצויות תשפיע אף היא על הערכת התוקף של המבחן. נראה כי קיומו של מבחן כתיבה יביא להגברת ההוראה של מיומנות הכתיבה בבתי-הספר העל-יסודיים ובמסגרות הקדם-אקדמיות. תוצאה זו נתפסת בעיני רוב האנשים כרצויה.

סיכום

סקירת הספרות בנושא מבחני כתיבה העלתה את הנקודות הבאות:

- קיימים שני סוגים עיקריים של מבחני כתיבה: מבחנים ישירים הכוללים משימות כתיבה ומבחנים עקיפים הכוללים פריטים רבי ברירה. כיום גובר השימוש במבחני כתיבה ישירים. בגלל טבע המשימה הנדרשת מבחני כתיבה ישירים כוללים מספר קטן של פריטים.
- ז'אנר הכתיבה הינו אחד המאפיינים החשובים ביותר של משימות כתיבה. הז'אנר נקבע על ידי מטרת הכתיבה וממנו נובעים סגנון הכתיבה הרצוי, המבנה המקובל וכו'.
- כאשר יכולת הכתיבה נבחנת על ידי ביצוע במשימות כתיבה עולה הבעיה של היעדר עקביות בין ביצוע במשימות מז'אנרים שונים. יכולת ההכללה מביצוע במשימת כתיבה מז'אנר אחד אל ביצוע במשימת כתיבה מז'אנר אחר היא מצומצמת. לכן, יש לכלול במבחן מספר משימות מז'אנרים שונים של כתיבה, ולבחור את המשימות הרלוונטיות למטרת הבחינה. אם מטרת הבחינה להיות כלי מיון למועמדים לאוניברסיטה הרי שיש לכלול בה משימות כתיבה בז'אנרים הנדרשים באוניברסיטה.
- אחד הגורמים לשונות בביצוע על פני משימות כתיבה שונות הוא מידת הידע הנדרש לצורך מילוי המשימות. כדי להקטין את השפעת מרכיב הידע בביצוע במבחן הכתיבה מומלץ לכלול בו גריינים (קטעי קריאה טבלאות או תרשימים) אשר ישמשו בסיס לכתיבת חיבורים. למשימת כתיבה המתבססת על גריין יתרון נוסף; היא דומה יותר למשימות כתיבה המתבצעות באופן שכיח במסגרות מקצועיות ואקדמיות - כתיבה הנשענת על חומר עזר.
- בגלל מיעוטם של הפריטים במבחן הכתיבה וההבדל ברמת הביצוע של נבחנים במשימות כתיבה מז'אנרים שונים, יש בעייתיות רבה ועם זאת גם צורך רב ביצירת מבחנים מקבילים.
- אחת הסוגיות המורכבות ביותר במבחן הכתיבה היא נושא הציון. הדבר נובע מהעובדה שהציון נעשה בידי מעריכים, והקושי להגיע להסכמה בציון. קיימות שתי שיטות עיקריות לציון חיבורים - שיטה הוליסטית - שיפוט כולל של איכות החיבור ומתן ציון אחד ושיטה אנליטית - הערכת החיבור על פני מספר מימדים.
- שיטת הציון הרווחת ביותר במבחני כתיבה רחבי היקף היא השיטה ההוליסטית. היתרון העיקרי של שיטת ציון זו הוא עלותה הנמוכה.
- ברובם המכריע של מבחני הכתיבה כל חיבור נבדק בידי שני מעריכים, ובמקרה של חוסר הסכמה הוא מועבר לבדק שלישי.
- נתוני מהימנות בין שופטים המוזכרים בספרות נעים בטווח רחב למדי. קשה להשוות בין נתונים של מחקרים שונים המבוססים על מדדי מהימנות שונים, על משימות כתיבה שונות, ועל שיטות ציון שונות. אם זאת, עולה מן המחקרים כי שיטת הציון האנליטית מהימנה יותר מהשיטה ההוליסטית. בשתי השיטות נראה שניתן להגיע למהימנות בין שופטים גבוהה יחסית לאחר תהליך הדרכה קפדני.

- מחקרים הראו כי כדי להעלות את מהימנות מבחן הכתיבה הגדלת מספר משימות הכתיבה עדיפה על פני הגדלת מספר המעריכים.
- תוקף הניבוי של מבחני הכתיבה כלפי הצלחה באוניברסיטה הוא גבוה על אף המספר הקטן של פריטים הכלולים בהם, ובחלק מן המבחנים הוא משתווה לזה של מבחנים רבי ברירה, כמו החלק המילולי של SAT. כאשר בודקים את התוספת השולית לתוקף, התרומה של מבחן הכתיבה נעה בין תרומה אפסית לתרומה של 0.06.
- צפוי כי לקיומו של מבחן כתיבה יהיו השלכות חברתיות חיוביות של עידוד השקעת משאבים רבים בהוראת הכתיבה בבתי הספר התיכוניים ובמוסדות אקדמיים.
- בגלל המאפיינים המיוחדים של מבחן הכתיבה (היותו מורכב ממספר קטן של פריטים, העובדה שהציון בו מושפע הן משונות בין מעריכים והן משונות בין משימות) יש קושי לכייל בין נוסחים.
- כדי להגדיל את המהימנות ואת תוקף הניבוי וכדי לפתור את בעיית הכיול המליצו רוב המחקרים על מבחן כתיבה המורכב ממשימות כתיבה ומשאלות רבות ברירה הבודקות כושרי כתיבה.

מקורות

אללוף א', יונה י' (1990) בחינת ידע בעברית דוח ארבע-שנתי (1986-1989) דוח מספר 125. מרכז ארצי לבחינות ולהערכה.

אנסטאזי א' (1990) מבחנים פסיכולוגיים. מהדורה שישית. האוניברסיטה הפתוחה.

יונה י', אללוף א' (1992) בחינת ידע בעברית דוח דו-שנתי (1990-1991) דוח מספר 153. מרכז ארצי לבחינות ולהערכה.

יונה י', אללוף א' (1994) בחינת ידע בעברית דוח דו-שנתי (1992-1993) דוח מספר 186. מרכז ארצי לבחינות ולהערכה.

כהן י', איתן מ' ובן-צור א' (1986) מבחן ידע בעברית לדוברי שפות זרות. דוח מספר 49. מרכז ארצי לבחינות ולהערכה.

נבו ד', וולף י', גולדבלט א', לוי א', מרק נ' וקניאל ש' (1989) הבעה בכתב בבית הספר היסודי: תהליכים הישגים ועמדות. משרד החינוך והתרבות ואוניברסיטת תל אביב.

פיקהולץ ש' (1984) השפעת ההבנייה של מטלות כתיבה על ההישגים בהבעה של תלמידים בכיתות י"ב. עבודת גמר לקראת תואר "מוסמך במדעי הרוח". אוניברסיטת תל-אביב.

רוזנר ר' (סתיו תשנ"ה) דגמים שונים לכתיבה עיונית. במכללה, 6, 17-27.

רוזנר ר' (בדפוס) הערכה הוליסטית לעומת הערכה אנליטית במשימת כתיבה. בתוך הערכה חלופית: הלכה ומעשה, בעריכת א. לוי, משרד החינוך והתרבות, גף להכשרת עובדי הוראה, מכון מופת.

רוזנר ר' (1990) ארבעים שנות חיבור - סקירת בחינות בגרות בחיבור בשנים תש"ט - תשמ"ח. סוגיות בחינוך (4), 162-173.

Applebee, A.N., Langer, J.A., Mullis, I.V.S., & Jenkins, L.B. (1990) The Writing Report Card 1984-1988. U.S. department of Education.

Applebee, A.N., Langer, J.A., Mullis, I.V.S., Latham, A.S. & Gentile, C.A. (1994) NAEP 1992 Writing Report Card 1984-1988. U.S. department of Education.

- De Beaugrande, R. (1980). Text, Discourse, and Process. ABLEX. Norwood, New Jersey.
- Bloom, D., & Green J. (1984) Directions in the sociolinguistic study of reading. In P. D. Pearson (Ed.), Handbook of Reading Research. Longman: New York & London.
- Boodoo, G. M., & Garlinghouse, P. (1983). Use of the essay examination to investigate the writing skills of undergraduate education majors. Educational and Psychological Measurement, 43, 1005-1014.
- Braun, H. I. (1988). Understanding scoring reliability. Journal of Educational Statistics, 13(1), 1-18.
- Breland, H. M. (1983). The direct assessment of writing skill: A measurement review. College Board Report No. 83-6. College Entrance Examination Board, New York.
- Breland, H.M., Camp, R., Jones, R.J., Morris, M.M., & Rock, D.A. (1987) Assessing writing skill. Research monograph No. 11. College entrance examination board, New York.
- Brennan, R.L. (1992) Elements of Generalizability Theory. ACT publications, Iowa City, Iowa.
- Bridgeman, B. (1991) Essays and multiple-choice tests as predictors of college freshman GPA. Research Report 91-3. Educational Testing Service, Princeton, New Jersey.
- Bridgeman, B., & Carlson, S. (1983) Survey of academic writing tasks required of graduate and undergraduate foreign students. Research report 15. TOEFL. Educational Testing Service, Princeton, New Jersey.
- Bridgeman, B., Hale, G.A., Lewis, C., Pollack, J., & Wang, M. (1992). Placement validity of a prototype SAT with an essay. Research Report 92-28. Educational Testing Service, Princeton, New Jersey..

- Bridgeman, B., & Lewis, C. (1994). The relationship of essay and multiple choice scores with grades in college courses. Journal of Educational Measurement, 31(1), 37-50.
- Carlson, S.B., Bridgeman, B., Camp, R., & Waanders, J. (1985) Relationship of admission test scores to writing performance of native and nonnative speakers of English. Research Report 19 . TOEFL. Educational Testing Service, Princeton, New Jersey.
- Cherry, R.D., & Meyer, P.R. (1993) Reliability issues in holistic assessment. In M.M. Williamson & B.A. Huot (Eds.) Validating holistic scoring for writing assessment. Hampton Press, Inc. Cresskill, New Jersey.
- Cohen, Y. (1993). Linear Multiple Calibration: The method and its application in calibrating essay raters. National Institute for Testing and Evaluation , Research Report no. 174.
- Collins, A., & Gentner, D. (1980). A framework for a cognitive theory of writing. in L. Gregg and E. R. Steinberg (Eds.) Cognitive Processes in Writing. LEA, Hillsdale, New Jersey.
- Cooper, C. R. (1977). Holistic evaluation of writing in C.R. Cooper and L. Odell Evaluating writing. National Council of Teachers of English.
- Cooper, P. L. (1984). The assessment of writing ability: A review of research. GRE Board research Report. Educational Testing Service, Princeton, NJ.
- Cope, R. T. (1995). Cautionary observations on reliability and equating of forms in high stakes performance assessment: The problem of granularity. Paper presented at the annual meeting of the American Association for Educational Research. San Fransisco, California.
- Crocker, L., & Algina, J. (1986). Introduction to classical and modern test theory. Harcourt Brace Jovanovich, Orlando, Florida.
- Daly, J. A., & Dickson-Markman, F. (1982). Contrast Effects in Evaluating Essays. Journal of Educational Measurement 19(4) 309-316.

- Engelhard, G. (1992). The measurement of writing ability with a many-faceted Rasch model. Applied Measurement in Education, 5(3), 171-191.
- Fowles, M.E. (1978) Basic skills assessment - manual for scoring the writing sample. Educational Testing Service, Princeton, New Jersey.
- Fowles, M., & Fiero, D. (1990). Steps in Developing Constructed-response Items for Stage II . Educational Testing Service, Princeton, New Jersey.
- Gadda, J. (1992). The universitywide Subject A Examination book. University of California.
- Gorman, T.P, Purves, A.C., & Degenhart, R.E. (1988). The IEA study of written composition I: The international writing tasks and scoring scales. Pergamon Press.
- Hale, G. (1992) Effects of amount of time allowed on the test of written English. Research Reports. TOEFL. Educational Testing Service, Princeton, New Jersey..
- Hale, G.A., Bridgeman, B., Lewis, C., Pollack, J.M., & Wang, M. (1992) A Comparison of the predictive validity of the current SAT and an experimental prototype. Educational Testing Service, Princeton, New Jersey.
- Hayes, J. R., & Flower, L. S. (1980) Identifying the organization of writing processes in L. Gregg and E. R. Steinberg (Eds.) Cognitive Processes in Writing. LEA, Hillsdale, New Jersey.
- Hughes, D. C., Keeling, B., & Tuck, B. F. (1983). Effects of achievement expectations and handwriting quality on scoring essays. Journal of Educational Measurement 20(1) 65-70.
- Koenig, J. A. (April, 1995) Examination of the comparability of MCAT writing sample test forms. Paper presented at the annual meeting of the American Association for Educational Research. San Fransisco, California.
- Messick, S. (1989) Validity. In R.L. Linn (Ed.), Educational Measurement. New York: Macmillan

- Miller, M., & Crocker L. (1990). Validation methods for direct writing assessment. Applied Measurement in Education, 3(3), 285-296.
- Moss, P. A., Cole, N. S., & Khampalikit, C. (1982). Comparison of procedures to assess written language skills at Grades 4,7, and 10. Journal of Educational Measurement, 19(1), 37-47.
- Powers, D. E., Fowles, M. E., Farnum, M., & Gerritz, K. (1992). Giving a choice of topics on a test of basic writing skills: does it make any difference? Educational Testing Service, Princeton, New Jersey.
- Powers, D. E., Fowles, M. E., Farnum, M., & Ramsey, P. (1992). Will they think less of my handwritten essay if others word process theirs? Effects on essay scores of intermingling handwritten and word-processed essays. Educational Testing Service, Princeton, New Jersey.
- Powers, D. E., Fowles, M. E., & Willard, A. E. (1994). Direct assessment, direct validation? An Example from the assessment of writing. Educational Assessment, 2(1), 89-100.
- Quellmalz, E. S. (1986) Writing skills assessment. In R. A. Berk (Ed.) Performance Assessment (pp. 492-506). The Johns Hopkins University Press. Baltimore & London.
- Quellmalz, E. S., Capell, F.J., and Chou,C-P. (1982). Effects of discourse and response mode on the measurement of writing competence. Journal of Educational Measurement, 19(4), 241-258.
- Sarig, G., & Folman, S. (1988). Issues in testing academic literacy: An integrative approach. The Open University in Israel. Tel Aviv, Israel.
- Stubbs, M. (1980) Language and Literacy. Routledge & Kegan Paul. London, Boston and Henley.
- Takala, S. (1982) On the origins, communicative parameters and processes of writing. Evaluation in Education. 5, 209-230.

- Townsend, M. A. R., Kek, L.Y., & Tuck, B. (1989). The effect of mood on the reliability of essay assessment. British Journal of Educational Psychology 59(2), 232-240.
- Veal L. R., & Hudson, S. A. (1983). Direct and indirect measures for large-scale evaluation of writing. Research in the teaching of english, 17,290-296.
- White, E. (1985). Teaching and assessing writing. Jossey-Bass Publishers: San Francisco.
- Witt, E. A. (1995). Issues in constructing an analytical scoring scale for a writing assessment. Paper presented at the annual meeting of the American Association for Educational Research. San Fransisco, California.
- Wolfe, E., Bolton, S., Feltovich B., & Welch, C. (1993). A comparison of word-processed and handwritten essays from a standardized writing assessment. ACT Research Report Series.

