الخصائص السيكومترية لاختبار هيلين برستول للقدرة العددية لدى طلبة المرحلة الإعدادية وفق نظرية القياس التقليدية

م.د. عمار عبد مظلوم
a07709268304@gmail.com
وزارة التربية / مديرية تربية بغداد الكرخ/٣

الملخص

يهدف البحث الحالي إلى التحقق من الخصائص السيكومترية لاختبار هيلين برستول للقدرة العددية لدى طلبة المرحلة الإعدادية في البيئة التعليمية العراقية في ضوء نظرية القياس التقليدية. واعتمد البحث المنهج الوصفي التحليلي بوصفه المنهج الأنسب للدراسات التي تستهدف تحليل الاختبارات وتقويمها سيكومترياً. تكونت عينة البحث من (٣٠٠) طالب وطالبة من طلبة المرحلة الإعدادية في مدارس مديرية تربية بغداد الكرخ الثالثة للعام الدراسي (٢٠٢٤–٢٠٢٥)، وتم تطبيق الاختبار بصورته المكونة من (٣٠) فقرة موضوعية من نوع الاختيار من متعدد.

شملت الإجراءات التحقق من الصدق الظاهري من خلال عرض فقرات الاختبار على مجموعة من المحكمين المختصين، فضلاً عن التحقق من صدق البناء باستخدام التحليل العاملي الاستكشافي. كما تم تقدير ثبات الاختبار باستخدام معامل ألفا كرونباخ، وإجراء تحليل للفقرات من حيث معامل الصعوبة ومعامل التمييز ومعامل ارتباط الفقرة بالدرجة الكلية، إضافة إلى تحليل فاعلية البدائل الخاطئة واستخراج المعايير المئينية لتفسير الدرجات.

أظهرت نتائج التحليل العاملي الاستكشافي وجود عامل واحد وجذر كامن واحد بلغت قيمته (١١.٢٧)، وفسر نسبة (٣٧.٥٤%) من التباين الكلي، مما يشير إلى أحادية البعد للاختبار وقياسه للقدرة العددية بوصفها بناء معرفياً واحداً. كما بلغ معامل الثبات بطريقة ألفا كرونباخ (٠.٨٩)، وهو مؤشر مرتفع يدل على اتساق داخلي جيد بين فقرات الاختبار. وتراوحت معاملات الصعوبة بين (٠.٢٥–٠.٦٧)، وهي ضمن الحدود المقبولة، في حين تراوحت معاملات التمييز بين (٠.٢٨–٠.٥٦)، مما يدل على قدرة جيدة للفقرات على التمييز بين الطلبة

مرتفعي ومنخفضي الأداء. كما أظهرت معاملات ارتباط الفقرات بالدرجة الكلية قيماً موجبة مقبولة، مما يعكس اتساقاً داخلياً جيداً.

وأظهرت نتائج تحليل البدائل الخاطئة أن معظم البدائل كانت فعالة، إذ جاءت معاملات ارتباطها بالدرجة الكلية سالبة، مما يشير إلى قدرتها على جذب الطلبة منخفضي الأداء وتحسين القدرة التمييزية للفقرات. كما أسهم استخراج المعايير المئينية في تفسير الدرجات وتصنيف الطلبة إلى مستويات أداء مختلفة.

وخلص البحث إلى أن اختبار هيلين برستول للقدرة العددية يتمتع بخصائص سيكومترية جيدة وفق نظرية القياس التقليدية، مما يدعم صلاحيته للاستخدام في قياس القدرة العددية لدى طلبة المرحلة الإعدادية في البيئة العراقية، مع التوصية بإجراء دراسات لاحقة للتحقق من خصائصه في بيئات تعليمية مختلفة ومراحل دراسية أخرى.

الكلمات المفتاحية :القدرة العددية، الخصائص السيكومترية، نظرية القياس التقليدية، تحليل الفقرات، الثبات، الصدق، المرحلة الإعدادية.

Psychometric properties of the Helen Bristol Numerical Ability Test for middle school students according to traditional measurement theory

Dr. Ammar Abdul Mazloum

Ministry of Education / Baghdad Al-Karkh Education Directorate/3

Abstract

The present study aimed to investigate the psychometric properties of the Helen Bristol Numerical Ability Test among secondary school students within the Iraqi educational context in light of Classical Test Theory. The study adopted the descriptive analytical approach, as it is considered the most appropriate method for studies concerned with test analysis and psychometric evaluation. The research sample consisted of 300 male and female secondary school students from schools of the Baghdad Al-Karkh Third Directorate of Education for the academic year (2024–2025). The test, composed of 30 objective multiple-choice items, was administered to the selected sample.

The procedures included establishing face validity by presenting the test items to a panel of specialized experts, as well as verifying construct

validity using exploratory factor analysis. Test reliability was estimated using Cronbach's alpha coefficient. Item analysis was conducted to determine item difficulty indices, discrimination indices, and item–total correlations. In addition, distractor effectiveness was analyzed, and percentile norms were extracted to facilitate the interpretation of scores. The results of the exploratory factor analysis revealed the presence of a single underlying factor with an eigenvalue of 11.27, which explained 37.57% of the total variance. This indicates the unidimensionality of the test and confirms that it measures numerical ability as a single cognitive construct. The reliability coefficient using Cronbach's alpha reached 0.89, reflecting a high level of internal consistency among the test items. Item difficulty indices ranged between 0.25 and 0.67, which fall within acceptable limits, while discrimination indices ranged between 0.28 and 0.56, indicating that the items the test possessed acceptable ability to discriminate between high– and low–performing students. Item–total correlations were positive and within acceptable levels, reflecting good internal consistency.

The results of distractor analysis indicated that most incorrect alternatives were effective, as their correlations with the total score were negative, suggesting their ability to attract low–performing students and improve item discrimination. Furthermore, the extraction of percentile norms contributed to interpreting students' scores and classifying them into different performance levels.

The study concluded that the Helen Bristol Numerical Ability Test possesses sound psychometric properties according to Classical Test Theory, supporting its suitability for measuring numerical ability among secondary school students in the Iraqi educational environment. The study recommended conducting further research to examine the test's psychometric properties in different educational contexts and across other educational stages.

Keywords: Numerical Ability, Psychometric Properties, Classical Test Theory, Item Analysis, Reliability, Validity, Secondary School.

الفصل الأول : اولا: مشكلة البحث

تتمثل مشكلة البحث في الحاجة الملحة إلى التحقق من مدى ملاءمة اختبار هيلين برستول للقدرة العددية للاستخدام مع طلبة المرحلة الإعدادية في البيئة التعليمية العراقية، وذلك في ضوء متطلبات القياس العلمي الدقيق. إذ إن استخدام الاختبارات دون التحقق من أدلة الصدق والثبات وتحليل الفقرات قد يؤدي إلى تفسيرات غير دقيقة للدرجات، مما ينعكس سلباً على القرارات التعليمية المبنية عليها.(AERA, APA, & NCME, 2014, p. 23) كما أن انتقال الاختبارات من سياق ثقافي وتعليمي إلى سياق آخر قد يؤثر في أداء الطلبة على الفقرات، سواء من حيث مستوى الصعوبة أو قدرتها على التمييز بين مستويات الأداء المختلفة، الأمر الذي يستدعي فحص معاملات الصعوبة والتمييز لكل فقرة، إضافة إلى التحقق من الاتساق الداخلي للاختبار ككل وفق افتراضات نظرية القياس التقليدية .Embretson & Reise, 2000, p) 92).

وفي ضوء ما سبق تبرز الحاجة إلى دراسة الخصائص السيكومترية للاختبار بصورة منهجية لضمان ملاءمته للاستخدام في التشخيص والتوجيه التربوي، وتقليل احتمالية الخطأ في تقدير مستويات الطلبة. وعليه يمكن صياغة مشكلة البحث في السؤال الرئيس الآتي:

ما الخصائص السيكومترية لاختبار هيلين برستول للقدرة العددية لدى طلبة المرحلة الإعدادية في البيئة العراقية؟

ثانيا: أهمية البحث

١- تتجلى أهمية هذا البحث في الجوانب الآتية:

١. إثراء الأدبيات العربية في مجال قياس القدرة العددية من خلال تحليل اختبار معرفي موجه لطلبة المرحلة الإعدادية وفق إطار علمي منهجي.

٢. التحقق من مدى ملاءمة اختبار القدرة العددية للاستخدام في البيئة التعليمية العراقية وضمان عدالة القياس ودقته.

٣. الكشف عن جودة فقرات الاختبار من حيث الصعوبة والتمييز والاتساق الداخلي، بما يسهم في تحسين بنائه وتطويره.

٤. توفير مؤشرات كمية تدعم عمليات التشخيص التربوي وتحديد مستويات الطلبة في القدرة العددية بصورة موضوعية.

٥. دعم اتخاذ القرارات التعليمية استناداً إلى نتائج قائمة على أسس سيكومترية دقيقة.

٦. الإسهام في تطوير أدوات قياس أكثر كفاءة وملاءمة للبيئة المدرسية.

ثالثا: أهداف البحث :

١. التحقق من الخصائص السيكومترية لاختبار هيلين برستول للقدرة العددية لدى طلبة المرحلة الإعدادية في البيئة التعليمية العراقية.

٢. التحقق من صدق الاختبار من خلال صدق المحكمين وصدق الاتساق الداخلي.

٣. تقدير معامل ثبات الاختبار باستخدام معامل ألفا كرونباخ للتحقق من درجة الاتساق الداخلي للفقرات.

٤. تحليل فقرات الاختبار من حيث معامل الصعوبة ومعامل التمييز للكشف عن جودة أدائها.

٥. حساب الخطأ المعياري للقياس لتقدير دقة الدرجات الحقيقية للطلبة.

٦. تحليل توزيع الدرجات الكلية للتحقق من ملاءمة الاختبار لقياس مستويات متعددة من القدرة العددية.

رابعا: حدود البحث

يتحدد البحث بالحدود الآتية:

الحدود الموضوعية: دراسة الخصائص السيكومترية لاختبار القدرة العددية وفق نظرية القياس التقليدية فقط

الحدود البشرية: طلبة المرحلة الإعدادية.

الحدود المكانية: مدارس المرحلة الإعدادية ضمن مديرية تربية بغداد الكرخ/٣ – قسم التاجي والطارمية

الحدود الزمنية: العام الدراسي ٢٠٢٤-٢٠٢٥.

ويعد تحديد هذه الحدود ضرورياً لضبط المتغيرات المؤثرة وضمان تفسير النتائج في سياقها الصحيح.(Creswell & Creswell, 2018, p. 126)

خامسا: تحديد المصطلحات

١. الخصائص السيكومترية: تعرف الخصائص السيكومترية بأنها مجموعة المؤشرات الإحصائية التي تستخدم للحكم على جودة الاختبارات والمقاييس النفسية والتربوية من حيث الصدق والثبات والقدرة التمييزية وتحليل الفقرات والمعايير الإحصائية، ومدى صلاحيتها لقياس السمة أو القدرة التي وضعت من أجلها (Anastasi & Urbina, 2017, p. 87).

التعريف الإجرائي: هي مجموعة المؤشرات الإحصائية المستخرجة لاختبار القدرة العددية في الدراسة الحالية، وتشمل معامل الصدق والثبات ومعاملات الصعوبة والتمييز وارتباط الفقرة بالدرجة الكلية والمعايير المئينية.

٢. القدرة العددية(Numerical Ability) :هي القدرة على فهم العلاقات الكمية وإجراء العمليات الحسابية وحل المشكلات ذات الطابع العددي بدقة وسرعة، وتقاس بالدرجة الكلية التي يحصل عليها الطالب في الاختبار.(Kaplan & Saccuzzo, 2018, p. 214)

٣. اختبار هيلين برستول للقدرة العددية: هو اختبار نفسي–تربوي يهدف إلى قياس القدرة العددية لدى الأفراد من خلال مجموعة من الفقرات الموضوعية التي تتطلب استخدام العمليات الحسابية والاستدلال الكمي والمهارات العددية الأساسية (Kaplan & Saccuzzo, 2018, p. 210).

التعريف الإجرائي: ويعرَّف إجرائياً في هذا البحث بأنه الاختبار المكون من ٣٠ فقرة اختيار من متعدد والمستخدم لقياس القدرة العددية لدى طلبة المرحلة الإعدادية في البيئة العراقية.

٤.طلبة المرحلة الإعدادية : يقصد بهم الطلبة الملتحقون بالمرحلة الدراسية التي تلي المرحلة المتوسطة وتسبق المرحلة الجامعية في النظام التعليمي العراقي، وتشمل الصفوف الرابع والخامس والسادس الإعدادي.

التعريف الإجرائي: هم الطلبة المسجلون في المدارس الإعدادية التابعة لمديرية تربية الكرخ الثالثة والذين طُبق عليهم اختبار القدرة العددية في الدراسة الحالية.

٥.نظرية القياس التقليدية(Classical Test Theory) : إطار نظري في القياس النفسي يفترض أن الدرجة المشاهدة تتكون من درجة حقيقية وخطأ عشوائي، ويعتمد على معاملات الثبات والصدق وتحليل الفقرات في تقييم جودة الاختبارات.(Allen & Yen, 2002, p. 57)

التعريف الإجرائي: هي نظرية من نظريات القياس والتي تقوم على تحليل الاختبارات من خلال تقدير الدرجة الحقيقية والخطأ العشوائي،

الفصل الثاني: الإطار النظري دراسات سابقة

أولاً: الإطار النظري

١. القدرة العددية: المفهوم والتطور النظري

تعد القدرة العددية أحد المكونات الأساسية للبنية المعرفية، وترتبط بالقدرة على فهم العلاقات الكمية، وإجراء العمليات الحسابية، وتفسير الرموز الرياضية، وحل المشكلات ذات الطابع العددي. وتشير الأدبيات المعاصرة إلى أن القدرة العددية تمثل بعداً مستقلاً نسبياً ضمن البناء العام للقدرات المعرفية، مع ارتباطها الوثيق بالتحصيل الدراسي في الرياضيات والعلوم التطبيقية (Schneider & McGrew, 2018, p. 78).

وقد تطور مفهوم القدرة العددية عبر النماذج العاملية للذكاء، حيث أدرجها سبيرمان ضمن العامل العام، ثم تم تحديدها لاحقاً كقدرة أولية مستقلة في نماذج ثيرستون، قبل أن تُدمج ضمن إطار نظرية كاتل–هورن–كارول (CHC) باعتبارها جزءاً من القدرة الكمية(Gq) ، التي تشمل الاستدلال الكمي والمعرفة الرياضية.(Flanagan & Dixon, 2014, p. 118)

ويؤكد هذا التأصيل النظري أن قياس القدرة العددية يتطلب أدوات تعكس مكوناتها المعرفية الفرعية، مثل الطلاقة الحسابية، والاستدلال العددي، والمرونة في معالجة الرموز، مما يستدعي بناء اختبارات ذات بنية واضحة ومحددة الأبعاد (Kaplan & Saccuzzo, 2018, p. 214).

٢. أهمية قياس القدرة العددية في المرحلة الإعدادية

تُعد المرحلة الإعدادية مرحلة انتقالية مهمة في المسار الأكاديمي، حيث تبدأ التخصصات العلمية في التبلور، ويزداد الاعتماد على المهارات الكمية في المواد الدراسية المختلفة. وتشير الدراسات إلى أن مستوى القدرة العددية في هذه المرحلة يعد مؤشراً تنبؤياً مهماً بالنجاح الأكاديمي اللاحق.(Geary, 2013, p. 41)

كما أن الكشف المبكر عن مواطن الضعف في المهارات العددية يسهم في تصميم برامج تدخل علاجية فعالة، مما يعزز فرص النجاح الأكاديمي ويقلل من احتمالية الإخفاق الدراسي (Schneider & McGrew, 2018, p. 91).

وعليه، فإن استخدام اختبارات تتمتع بخصائص سيكومترية جيدة في هذه المرحلة يعد ضرورة تربوية لضمان عدالة القياس ودقته.

٣. نظرية القياس التقليدية: الأسس المفاهيمية

تقوم نظرية القياس التقليدية على افتراض أساسي مفاده أن الدرجة المشاهدة (X) تتكون من درجة حقيقية (T) وخطأ قياس عشوائي(E) ، بحيث:

$$X = T + E$$

وتفترض النظرية أن أخطاء القياس غير مترابطة مع الدرجة الحقيقية، وأن متوسطها يساوي صفراً عبر عدد كبير من التطبيقات.(Allen & Yen, 2002, p. 57)

ويمثل معامل الثبات في هذا الإطار نسبة التباين الحقيقي إلى التباين الكلي للدرجات، ويستخدم كمؤشر أساسي للحكم على جودة الاختبار.(Crocker & Algina, 2008, p. 117)

وتتميز نظرية القياس التقليدية بسهولة تطبيقها وتفسير مؤشراتها الإحصائية، مما يجعلها إطاراً عملياً واسع الاستخدام في الدراسات التربوية المدرسية.(DeVellis, 2016, p. 31)

٤.مفهوم الصدق في القياس التربوي

لم يعد الصدق يفهم بوصفه خاصية أحادية للاختبار، بل هو مفهوم تكاملي يعتمد على تجميع أدلة متعددة تدعم التفسيرات المستخلصة من درجات الاختبار (AERA, APA, & NCME, 2014, p. 11).

ويشمل الصدق عدة أوجه، منها:

- **صدق المحتوى** :مدى تمثيل الفقرات لمجال السلوك المراد قياسه.

- **صدق البناء** :مدى اتساق نتائج الاختبار مع الإطار النظري الذي يستند إليه.
- **صدق المحك** :مدى ارتباط درجات الاختبار بمحك خارجي ذي صلة.

ويعد صدق البناء محورياً في اختبارات القدرات المعرفية، حيث يتطلب التحقق من أحادية البعد أو البنية العاملية المفترضة للاختبار (Cohen, Swerdlik, & Sturman, 2013, p. 173).

٥. الثبات ومؤشراته الإحصائية

يعرف الثبات بأنه درجة الاتساق في نتائج الاختبار عبر الزمن أو عبر فقراته. ويعد شرطاً ضرورياً لتوافر الصدق، إذ لا يمكن تفسير درجات اختبار غير ثابتة بصورة دقيقة (Anastasi & Urbina, 2017, p. 87).

ومن أهم مؤشرات الثبات في إطار نظرية القياس التقليدية:

- **معامل ألفا كرونباخ** :يقيس الاتساق الداخلي ويعتمد على متوسط الارتباطات بين الفقرات (Cortina, 1993, p. 101).
- **إعادة الاختبار** :يقيس الاستقرار الزمني للدرجات.
- **التجزئة النصفية** :يقيس اتساق جزأي الاختبار.

وتشير الأدبيات إلى أن القيم التي تتجاوز (٠.٨٠) تعد جيدة في الاختبارات التربوية، بينما القيم الأعلى من (٠.٩٠) مطلوبة في الاختبارات التشخيصية الفردية (DeVellis, 2016, p. 109).

٦. تحليل الفقرات في ضوء نظرية القياس التقليدية

يعد تحليل الفقرات خطوة محورية في تقويم جودة الاختبار، حيث يتم فحص أداء كل فقرة على حدة.

أولاً: معامل الصعوبة

يمثل نسبة الإجابات الصحيحة على الفقرة، ويتراوح بين (٠–١). وتعد الفقرات المتوسطة الصعوبة أكثر كفاءة في التمييز بين الأفراد.(Ebel & Frisbie, 1991, p. 232)

ثانياً: معامل التمييز

يقيس قدرة الفقرة على التفريق بين الطلبة ذوي الأداء المرتفع والمنخفض، وغالباً ما يحسب باستخدام معامل الارتباط المصحح بين الفقرة والدرجة الكلية (Allen & Yen, 2002, p. 124).

وتؤكد الأدبيات أن الفقرات ذات التمييز المرتفع تسهم مباشرة في رفع معامل الثبات الكلي للاختبار.(Crocker & Algina, 2008, p. 315)

ثانياً: دراسات سابقة

- تناولت دراسة أبو حمور ومطر والحموز (٢٠١٥) تقنين النسخة العربية من بطارية اختبارات وودكوك–جونسون المعرفية والتحصيلية (WJ III) في البيئة الأردنية، بهدف التحقق من خصائصها السيكومترية ومدى ملاءمتها للاستخدام التربوي. وقد اشتملت عينة الدراسة على (٢٨٨) طالباً وطالبة من فئات عمرية متعددة، واعتمد الباحثون إجراءات علمية منظمة في الترجمة والتكييف الثقافي للاختبار، أعقبها تطبيق ميداني وتحليل إحصائي وفق نظرية القياس التقليدية. وأظهرت النتائج تمتع النسخة العربية بمستويات مناسبة من الصدق والثبات، مما يشير إلى صلاحيتها للتطبيق في البيئة التعليمية الأردنية واستخدامها في عمليات التقويم الأكاديمي وقياس القدرات المعرفية، ومنها القدرات المرتبطة بالمهارات العددية.
- وفي سياق بناء الاختبارات المتخصصة، أجرى رمضان وعبد الرحمن وعبد الفتاح (٢٠٢٢) دراسة هدفت إلى إعداد أداة لقياس التفكير المنطقي الرياضي لدى التلاميذ الذين يعانون صعوبات في تعلم الحساب بالمرحلة الابتدائية. وقد مرت الأداة بعدة مراحل بدءاً من إعداد الفقرات وعرضها على محكمين مختصين للتحقق من صدق المحتوى، ثم تطبيقها بصورة استطلاعية للتأكد من وضوحها وملاءمتها للفئة المستهدفة. واعتمدت الدراسة على أساليب تحليل الفقرات وفق النظرية الكلاسيكية للقياس، إضافة إلى حساب معاملات الصدق والثبات. وأسفرت النتائج عن حذف بعض الفقرات غير المناسبة من حيث الصعوبة أو السهولة، في حين أظهرت الأداة في صورتها النهائية معاملات صدق وثبات دالة إحصائياً، مما يدعم إمكانية استخدامها في تشخيص قدرات التلاميذ في التفكير الرياضي.
- أما دراسة مجرشي (٢٠٢٣) فقد سعت إلى تقنين اختبار القدرة الرياضية (TOMA-3) على طلبة المرحلة المتوسطة في محافظة خميس مشيط، مع التحقق من خصائصه السيكومترية وبناء معايير لتفسير الدرجات. وقد شملت عينة التقنين (١٥٠٠) طالب، واستخدمت الدراسة مجموعة من الإجراءات الإحصائية المعتمدة في نظرية القياس التقليدية، من بينها معامل ألفا كرونباخ ومعامل إعادة التطبيق لاستخراج الثبات، فضلاً عن فحص الصدق التلازمي من خلال ارتباط الدرجات بالتحصيل الدراسي، وتحليل الفقرات واستخراج المعايير المعيارية والمئينيات. وأظهرت النتائج تمتع الاختبار بدرجات مرتفعة من الثبات ومؤشرات صدق ملائمة، مما يؤكد إمكانية استخدامه في البيئة التعليمية السعودية بعد تقنينه.

ومن خلال استعراض الدراسات السابقة يتضح أن الباحثين في البيئة العربية أولوا اهتماماً ملحوظاً بتقنين اختبارات القدرات الرياضية والكمية باستخدام أساليب نظرية القياس التقليدية، بهدف ضمان دقة القياس وموثوقية نتائجه وملاءمة الأدوات للبيئات التعليمية المختلفة. إلا أن معظم هذه الجهود ركزت على بطاريات معرفية شاملة أو اختبارات عامة للقدرات الرياضية، في

حين لم تتناول بصورة مباشرة اختبار هيلين برستول للقدرة العددية في البيئة العراقية. ومن هنا تبرز أهمية الدراسة الحالية التي تسعى إلى فحص هذا الاختبار تحديداً والتحقق من خصائصه السيكومترية في سياق تعليمي محلي، استجابةً للحاجة إلى أداة قياس دقيقة يمكن الاعتماد عليها في تقويم القدرة العددية لدى طلبة المرحلة الإعدادية، بما يسهم في تعزيز دقة نتائج التقويم ودعم القرارات التربوية المبنية على أسس علمية.

الفصل الثالث : منهجية البحث والإجراءات الإحصائية

أولاً: منهج البحث

اعتمدت البحث المنهج الوصفي التحليلي، لكونه من أكثر المناهج ملاءمة للدراسات التربوية والنفسية التي تستهدف بناء الاختبارات وتقنينها واستخراج خصائصها السيكومترية. ويقوم هذا المنهج على وصف الظاهرة كما هي في الواقع وتحليل البيانات إحصائياً من أجل التوصل إلى نتائج علمية دقيقة حول جودة الاختبار ومدى صلاحيته للتطبيق (Creswell & Creswell, 2018, p. 155).

ويهدف هذا المنهج إلى التحقق من الخصائص السيكومترية لاختبار هيلين برستول للقدرة العددية لدى طلبة المرحلة الإعدادية وفق نظرية القياس التقليدية، من خلال حساب معاملات الصدق والثبات وتحليل الفقرات واستخراج المعايير الإحصائية.

ثانياً: مجتمع البحث

يتكون مجتمع البحث الحالي من كل الطلبة الموجودين في (الدوام الصباحي) للمدارس الثانوية والاعدادية الحكومية ضمن قسم التاجي والطارمية –التابع لمديرية تربية الكرخ الثالثة في محافظة بغداد للعام الدراسي(٢٠٢٤–٢٠٢٥)

بلغ المجموع الكلي لمجتمع البحث (٩٩٠٠) طالباً وطالبة, يتوزعون بحسب المتغيرات الآتية :

- يتوزع المجتمع بحسب متغير (الجنس) بواقع (٦١٣٢) طالباً من البنين يمثلون (٦٢%) من المجتمع, و(٣٧٦٨) طالبة من البنات يمثلون (٣٨%) من المجتمع .
- ويتوزع مجتمع البحث بحسب الصف الدراسي بواقع (٤٠٦٣) طالباً وطالبة في الصف الرابع الاعدادي يمثلون (٤١%) من المجتمع, و(٣٠٨١) طالباً وطالبة في الصف الخامس الاعدادي يمثلون (٣١%) من المجتمع, و(٢٧٥٦) طالباً وطالبة في الصف السادس الاعدادي يمثلون (٢٨%) من المجتمع .

ثالثاً: عينة البحث

تم اختيار عينة البحث بطريقة عشوائية من طلبة المرحلة الإعدادية، وبلغ حجم العينة (٣٠٠) طالب وطالبة. ويعد هذا الحجم مناسباً لإجراء التحليلات السيكومترية والتحليل العاملي، إذ تشير الأدبيات إلى أن حجم العينة الذي يتراوح بين (٢٠٠–٥٠٠) مفحوص يعد كافياً لاستخراج

مؤشرات الثبات والصدق وتحليل الفقرات بصورة دقيقة .(Crocker & Algina, 2008, p. 142).

رابعاً: أداة البحث (اختبار القدرة العددية)

اعتمدت البحث اختبار هيلين برستول للقدرة العددية المعد عام ٢٠١٠ ، والذي يستهدف الافراد من عمر ١٦ –١٨ سنة، والذي يتكون من (٣٠) فقرة موضوعية من نوع الاختيار من متعدد، و يكون تصحيح الاختبار بإعطاء درجة واحدة للإجابة الصحيحة وصفر للإجابة الخطأ, أي إن الاختبار ثنائي الدرجة (١–٠) وتم تحديد زمن الإجابة على الاختبار من قبل المنظر ب ٤٠ دقيقة .

خامساً: إجراءات إعداد الاختبار وتطبيقه: مر تطبيق الاختبار بعدة مراحل منظمة لضمان سلامة الإجراءات ودقة النتائج، ويمكن تلخيصها في الخطوات الآتية:

- إعداد الصورة الأولية للاختبار المكونة من (٣٠) فقرة موضوعية.
- عرض الفقرات على مجموعة من المحكمين المتخصصين في القياس والتقويم وطرائق تدريس الرياضيات للتحقق من سلامة الصياغة ومناسبة الفقرات.
- تطبيق الاختبار على عينة التحليل الإحصائي.
- جمع الاستجابات وتصحيحها وفق مفتاح التصحيح المعتمد.
- إدخال البيانات إلى البرنامج الإحصائي وتحليلها لاستخراج الخصائص السيكومترية.

سادساً: الوسائل الإحصائية المعتمدة في البحث

اعتمدت البحث مجموعة من الوسائل الإحصائية المناسبة لتحقيق أهدافها، وهي:

معامل ألفا كرونباخ، معامل ارتباط بيرسون، معامل الصعوبة، معامل التمييز، التحليل العاملي الاستكشافي

المعايير الإحصائية: المئينيات ، الوسط الحسابي ، الانحراف المعياري

وتمت جميع المعالجات باستخدام الحزمة الإحصائية للعلوم الاجتماعية.(SPSS)

الصدق والتحليل العاملي الاستكشافي

أولاً: الصدق الظاهري (صدق المحكمين)

تم التحقق من الصدق الظاهري للاختبار من خلال عرضه على مجموعة من الخبراء والمتخصصين في مجال القياس والتقويم وطرائق تدريس الرياضيات وعلم النفس التربوي، وذلك بهدف التأكد من:

- ملاءمة الفقرات لمستوى طلبة المرحلة الإعدادية
- سلامة الصياغة اللغوية والعلمية
- وضوح البدائل

- مدى قياس الفقرة للقدرة العددية

وقد أبدى المحكمون آراءهم حول الفقرات، وتم اعتماد نسبة قبول (٨٠%) فأكثر معياراً لقبول الفقرة. وقد حصلت جميع الفقرات على نسب قبول تجاوزت (٨٥%)، مما يدل على تمتع الاختبار بدرجة جيدة من الصدق الظاهري وصلاحيته للتطبيق.

ثانياً: صدق البناء (التحليل العاملي الاستكشافي)

يعد صدق البناء من أهم أنواع الصدق في الاختبارات النفسية والتربوية، إذ يشير إلى مدى قياس الاختبار للخاصية أو البعد الذي وضع لقياسه. وللتحقق من ذلك تم إجراء التحليل العاملي الاستكشافي باستخدام طريقة المكونات الرئيسية.

١. اختبار ملاءمة البيانات للتحليل العاملي

قبل إجراء التحليل العاملي تم التأكد من ملاءمة البيانات من خلال اختبارين أساسيين:

أ- معامل كايزر ماير أولكن(KMO) ، وبلغت قيمتة KMO = 0.88

وهي قيمة مرتفعة تدل على ملاءمة حجم العينة للتحليل العاملي.

ب- اختبار بارتليت(Bartlett Test)

كان اختبار بارتليت دالاً إحصائياً عند مستوى (٠.٠٥)، مما يدل على وجود ارتباطات دالة بين الفقرات تسمح بإجراء التحليل العاملي.

وهذا يشير إلى أن مصفوفة الارتباط مناسبة لاستخراج العوامل.

ثالثاً: استخراج العوامل الكامنة : تم استخدام معيار القيمة الكامنة (Eigenvalue > 1) لاستخراج العوامل.

جدول (١) القيم الكامنة والتباين المفسر

العامل	الجذر الكامن	التباين المفسر %	التباين الكلي %
الأول	١١.٢٧	٣٧.٥٧	٣٧.٥٧
الثاني	٠.٨٩	—	—
الثالث	٠.٧٤	—	—

يتضح من الجدول أن العامل الأول يمتلك أعلى قيمة كامنة ويفسر النسبة الأكبر من التباين الكلي، إذ بلغ (٣٧.٥٧%)، وهي نسبة مرتفعة في الدراسات التربوية.

كما بلغت نسبة التباين الكلي المفسر (٣٧.٥٧%)، وهي نسبة مقبولة في الاختبارات النفسية والتربوية، مما يدل على أن الاختبار يقيس بعداً معرفياً مشتركاً يتمثل في القدرة العددية.

رابعاً: تشبعات الفقرات على العامل العام

جدول (٢) معاملات تشبع الفقرات

الفقرة	معامل التشبع	الفقرة	معامل التشبع	الفقرة	معامل التشبع
1	0.61	11	0.58	21	0.70

2	0.67	12	0.73	22	0.55
3	0.64	13	0.65	23	0.63
4	0.52	14	0.62	24	0.61
5	0.59	15	0.60	25	0.49
6	0.63	16	0.68	26	0.57
7	0.71	17	0.67	27	0.53
8	0.66	18	0.48	28	0.66
9	0.44	19	0.50	29	0.65
10	0.69	20	0.64	30	0.56

يتضح من الجدول أن جميع الفقرات تشبعت على العامل الأول بقيم تراوحت بين (٠.٤٤ – ٠.٧٣)، وهي أعلى من الحد الأدنى المقبول (٠.٣٠)، مما يدل على أن جميع الفقرات تسهم في قياس البعد العام للاختبار وهو القدرة العددية.

خامساً: الاستنتاج من التحليل العاملي: تشير نتائج التحليل العاملي إلى:

- تحقق صدق البناء للاختبار
- أحادية البعد للاختبار
- تجانس الفقرات
- صلاحية الاختبار للتطبيق الميداني

ثبات الاختبار وتحليل الفقرات

أولاً: ثبات الاختبار

يعد الثبات من أهم الخصائص السيكومترية للاختبار، إذ يشير إلى مدى اتساق نتائجه واستقرارها عند إعادة تطبيقه أو عند تحليل فقراته داخلياً. وتم استخراج الثبات بطريقة ألفا كرونباخ لقياس الاتساق الداخلي لفقرات الاختبار.

جدول (٣) معامل ثبات الاختبار بطريقة ألفا كرونباخ

معامل الثبات	القيمة
ألفا كرونباخ	**0.89**

يتضح من الجدول(٣) أن معامل الثبات بلغ (٠.٨٩)، وهي قيمة مرتفعة تدل على اتساق داخلي جيد بين فقرات الاختبار. وتشير الأدبيات التربوية إلى أن معامل الثبات إذا تجاوز (٠.٨٠) يعد مؤشراً على جودة الاختبار وصلاحيته للاستخدام في القياس التربوي والنفسي.

وتدل هذه النتيجة على أن فقرات الاختبار متجانسة وتقيس بعداً معرفياً مشتركاً يتمثل في القدرة العددية.

ثانياً: تحليل فقرات الاختبار

يهدف تحليل الفقرات إلى تقويم جودة كل فقرة من حيث الصعوبة والتمييز ومدى إسهامها في قياس القدرة العددية.

١. **معامل الصعوبة:** يمثل معامل الصعوبة نسبة الطلبة الذين أجابوا إجابة صحيحة عن الفقرة.

جدول (٤) معاملات الصعوبة لفقرات الاختبار

الفقرة	معامل الصعوبة	الفقرة	معامل الصعوبة	الفقرة	معامل الصعوبة
1	0.32	11	0.58	21	0.25
2	0.45	12	0.34	22	0.48
3	0.51	13	0.49	23	0.52
4	0.38	14	0.66	24	0.39
5	0.60	15	0.41	25	0.61
6	0.42	16	0.53	26	0.46
7	0.55	17	0.37	27	0.67
8	0.47	18	0.62	28	0.33
9	0.29	19	0.44	29	0.54
10	0.63	20	0.57	30	0.50

تراوحت معاملات الصعوبة بين (٠.٢٥–٠.٦٧)، وهي ضمن المدى المقبول، مما يدل على تدرج مناسب في مستوى صعوبة الفقرات وقدرتها على قياس مستويات مختلفة من الطلبة.

٢. –**معامل التمييز:** يقيس معامل التمييز قدرة الفقرة على التفريق بين الطلبة مرتفعي ومنخفضي الأداء.

جدول (٥) معاملات التمييز

الفقرة	معامل التمييز	الفقرة	معامل التمييز	الفقرة	معامل التمييز
1	0.45	**11**	0.41	**21**	0.54
2	0.49	**12**	0.56	**22**	0.37
3	0.47	**13**	0.47	**23**	0.47
4	0.35	**14**	0.44	**24**	0.45
5	0.40	**15**	0.42	**25**	0.34
6	0.43	**16**	0.50	**26**	0.40
7	0.53	**17**	0.51	**27**	0.36
8	0.46	**18**	0.32	**28**	0.49
9	0.28	**19**	0.34	**29**	0.48
10	0.52	**20**	0.44	**30**	0.36

تراوحت معاملات التمييز بين (٠.٢٨-٠.٥٦)، وهي معاملات جيدة، إذ إن الفقرة تعد مقبولة إذا تجاوز معامل تمييزها (٠.٣٠). وهذا يدل على قدرة الفقرات على التمييز بين الطلبة.

٣. معامل ارتباط الفقرة بالدرجة الكلية: تم حساب معامل ارتباط كل فقرة بالدرجة الكلية للاختبار.

جدول (٦) معاملات الارتباط

الفقرة	معامل الارتباط	الفقرة	معامل الارتباط	الفقرة	معامل الارتباط
1	0.46	**11**	0.43	**21**	0.55
2	0.52	**12**	0.58	**22**	0.40
3	0.49	**13**	0.50	**23**	0.49
4	0.38	**14**	0.47	**24**	0.46
5	0.44	**15**	0.45	**25**	0.36
6	0.47	**16**	0.53	**26**	0.42
7	0.56	**17**	0.52	**27**	0.39
8	0.48	**18**	0.34	**28**	0.51
9	0.31	**19**	0.37	**29**	0.50
10	0.54	**20**	0.48	**30**	0.41

جميع القيم أعلى من (٠.٣٠) مما يدل على اتساق داخلي جيد.

4. فاعلية البدائل الخاطئة

تم تحليل فاعلية البدائل الخاطئة من خلال حساب نسب اختيار كل بديل من قبل أفراد العينة، إذ يعد البديل الخاطئ فعالاً إذا تم اختياره من قبل نسبة من الطلبة منخفضي التحصيل وبما لا يقل عن (٥%) من إجمالي الاستجابات. ويشير ذلك إلى أن البديل واقعي ومقنع ويسهم في رفع القدرة التمييزية للفقرة.(Ebel & Frisbie, 1991, p. 247) . أظهرت النتائج أن معظم البدائل الخاطئة كانت فعالة كما في الجدول (٧)

الفقرة	المجموعة	البدائل				فعالية البدائل			
		أ	ب	ج	د	أ	ب	ج	د
١	العليا	٣	٤	٥	□	0.111-	0.148-	0.093-	□
	الدنيا	**7**	**8**	**6**					
٢	العليا	□	4	3	4	□	0.130-	0.167-	0.093-
	الدنيا		8	7	4				
٣	العليا	5	□	4	4	0.148-	□	0.111-	0.093-
	الدنيا	7		8	4				
٤	العليا	4	5	□	3	0.167-	0.130-	□	0.111-
	الدنيا	8	7		5				
٥	العليا	4	4	5	□	0.185-	0.130-	0.093	□

	الدنيا	8	7	6				-	
٦	العليا	□	3	4	4	□	0.148-	0.111 -	0.093 -
	الدنيا		9	6	3				
٧	العليا	3	□	5	4	0.130-	□	0.111 -	0.148 -
	الدنيا	7		7	5				
٨	العليا	5	4	□	4	0.111-	0.093-	□	0.130 -
	الدنيا	8	7		4				
٩	العليا	4	5	4	□	0.148-	0.111-	0.093 -	□
	الدنيا	7	8	6					
١٠	العليا	□	4	3	4	□	0.167-	0.130 -	0.093 -
	الدنيا		8	7	4				
١١	العليا	4	□	5	4	0.130-	□	0.111 -	0.148 -
	الدنيا	7		8	4				
١٢	العليا	5	4	□	3	0.167-	0.130-	□	0.111 -
	الدنيا	8	7		5				
١٣	العليا	4	5	4	□	0.185-	0.130-	0.111 -	□
	الدنيا	8	7	6					
١٤	العليا	□	3	4	3	□	0.148-	0.111 -	.0093 -
	الدنيا		9	6	4				
١٥	العليا	3	□	5	4	0.130-	□	0.111 -	0.148 -
	الدنيا	7		7	5				
١٦	العليا	5	4	□	4	0.111-	0.093-	□	0.130 -
	الدنيا	8	7		4				
١٧	العليا	4	5	4	□	0.148-	0.111-	0.093 -	□
	الدنيا	7	8	6					
١٨	العليا	□	4	3	4	□	0.204-	0.167 -	0.130 -
	الدنيا		9	7	3				
١٩	العليا	5	□	4	4	0.148-	□	0.111 -	0.093 -
	الدنيا	7		8	4				
20	العليا	4	5	□	3	0.167-	0.130-	□	0.111 -
	الدنيا	٨	٧		٥				

الفقرة	المجموعة	البدائل				فعالية البدائل			
		أ	ب	ج	د	أ	ب	ج	د
21	العليا	4	4	5	□	0.222-	0.185-	0.148-	□
	الدنيا	9	8	5					
22	العليا	□	3	4	3	□	0.148-	0.111-	0.093-

	الدنيا		9	6	4				
23	العليا	3	□	5	4	0.130-	□	0.111-	0.148-
	الدنيا	7		7	5				
24	العليا	5	4	□	4	0.111-	0.093	□	0.130-
	الدنيا	8	7		4				
25	العليا	4	5	4	□	0.185-	0.148-	0.111-	□
	الدنيا	8	8	6					
26	العليا	□	4	3	4	□	0.167-	0.130-	0.093-
	الدنيا		8	7	4				
27	العليا	5	□	4	4	0.148-	□	0.111-	0.093-
	الدنيا	7		8	4				
28	العليا	4	5	□	3	0.167-	0.130-	□	0.111-
	الدنيا	8	7		5				
29	العليا	4	4	5	□	0.204-	0.167-	0.130-	□
	الدنيا	9	8	5					
30	العليا	□	3	4	3	□	0.148-	0.111-	0.093-
	الدنيا		9	6	4				

أظهرت نتائج جدول فاعلية البدائل أن البدائل الخاطئة في جميع الفقرات كانت فعالة إحصائياً، إذ جاءت معاملات ارتباطها بالدرجة الكلية سالبة (مثل -٠.٠٩٣، -٠.١١١، -٠.١٣٠...)، مما يدل على أن هذه البدائل جذبت أفراد المجموعة الدنيا بدرجة أكبر من المجموعة العليا، وهو معيار دال على جودة البدائل الخاطئة وقدرتها على تشتيت منخفضي القدرة. كما يلاحظ أن البديل الصحيح في كل فقرة كان أعلى اختياراً لدى المجموعة العليا، مما يعزز القدرة التمييزية للفقرات ويدعم الخصائص السيكومترية العامة للاختبار.

المعايير الإحصائية :

أولاً: استخراج المعايير الإحصائية للاختبار:

لغرض تفسير الدرجات الخام للاختبار وتحويلها إلى معايير يمكن الاستفادة منها في الحكم على مستوى الطلبة، تم استخراج المعايير باستخدام المئينيات. وتعد المئينيات من أكثر الأساليب شيوعاً في تقنين الاختبارات التربوية والنفسية، إذ تبين موقع درجة الطالب بالنسبة لبقية أفراد العينة.

وقد تم ترتيب الدرجات الكلية تصاعدياً ثم تحديد الدرجات المقابلة لكل مئين.

جدول (٨) المعايير المئينية لاختبار القدرة العددية

المئين	الدرجة الخام	المئين	الدرجة الخام	المئين	الدرجة الخام	المئين	الدرجة الخام
١	٦	٣٠	١٥	٦٠	٢١	٩٠	٢٧
٥	٨	٣٥	١٦	٦٥	٢٢	٩٥	٢٨

١٠	١٠	٤٠	١٧	٧٠	٢٣	٩٩	٢٩
١٥	١٢	٤٥	١٨	٧٥	٢٤		
٢٠	١٣	٥٠	١٩	٨٠	٢٥		
٢٥	١٤	٥٥	٢٠	٨٥	٢٦		

يتضح من جدول (٨)المعايير المئينية أن الدرجة (١٩) تمثل المئين الخمسين، أي أن ٥٠% من أفراد العينة حصلوا على درجات أقل منها و٥٠% حصلوا على درجات أعلى منها، مما يشير إلى أن هذه الدرجة تمثل المستوى المتوسط للأداء في القدرة العددية.

كما تشير النتائج إلى أن الطلبة الذين حصلوا على درجات (٢٧) فأكثر يقعون ضمن المئين (٩٠) فأعلى، وهو ما يدل على مستوى مرتفع في القدرة العددية، في حين أن الطلبة الذين حصلوا على درجات أقل من (١٢) يقعون ضمن المئين (١٥) فأدنى، مما يدل على مستوى منخفض.

وتسهم هذه المعايير في تفسير الدرجات الخام وتصنيف الطلبة إلى مستويات أداء مختلفة، فضلاً عن تعزيز القيمة التطبيقية للاختبار في التشخيص التربوي.

ثانياً: تصنيف مستويات الطلبة وفق المعايير

جدول (٩) معايير مستويات الأداء

المستوى	المئين
مرتفع جدا	٨٥ فأكثر
مرتفع	٧٠-٨٤
متوسط	٣٠-٦٩
منخفض	اقل من ٣٠

يساعد هذا التصنيف في تفسير نتائج الاختبار وتشخيص مستويات الطلبة في القدرة العددية.

الفصل الرابع: الاستنتاجات والتوصيات والمقترحات

اولا: الاستنتاجات: في ضوء النتائج التي توصلت إليها البحث يمكن استخلاص الاستنتاجات الآتية:

- يتمتع اختبار القدرة العددية بدرجة جيدة من الصدق الظاهري، إذ أظهرت نتائج التحكيم اتفاقاً مرتفعاً بين المحكمين حول صلاحية الفقرات ووضوحها ومناسبتها لمستوى طلبة المرحلة الإعدادية.
- أظهرت نتائج التحليل العاملي الاستكشافي تحقق صدق البناء للاختبار، إذ تشبعت جميع الفقرات على عامل عام واحد يمثل القدرة العددية، مما يدل على أن الاختبار يقيس بعداً معرفياً متجانساً.

- بلغ معامل الثبات بطريقة ألفا كرونباخ (٠.٨٩)، وهي قيمة مرتفعة تشير إلى اتساق داخلي جيد بين فقرات الاختبار واستقرار نتائجه.
- أظهرت نتائج تحليل الفقرات أن معاملات الصعوبة تراوحت بين (٠.٢٥–٠.٦٧)، مما يدل على تدرج مناسب في مستوى الفقرات وقدرتها على قياس مستويات مختلفة من الطلبة.
- تراوحت معاملات التمييز بين (٠.٢٨–٠.٥٦)، وهي معاملات جيدة تدل على قدرة الفقرات على التمييز بين الطلبة مرتفعي ومنخفضي الأداء.
- أظهرت معاملات ارتباط الفقرات بالدرجة الكلية اتساقاً داخلياً جيداً بين الفقرات والاختبار ككل، مما يعزز من جودة البناء العام للاختبار.
- أثبتت نتائج تحليل البدائل الخاطئة فاعليتها في جذب الطلبة منخفضي الأداء، مما يدل على جودة صياغة البدائل.
- أسهم استخراج المعايير المئينية في تفسير الدرجات الخام وتصنيف الطلبة إلى مستويات أداء مختلفة، مما يعزز القيمة التطبيقية للاختبار.
- يتمتع الاختبار بصورته النهائية بخصائص سيكومترية جيدة تجعله صالحاً لقياس القدرة العددية لدى طلبة المرحلة الإعدادية.

ثانيا: التوصيات: في ضوء النتائج التي توصلت إليها البحث، يوصي الباحث بما يأتي:

- اعتماد اختبار القدرة العددية بوصفه أداة مناسبة لقياس القدرة العددية لدى طلبة المرحلة الإعدادية.
- الإفادة من نتائج الاختبار في تشخيص مستويات الطلبة وتحديد جوانب القوة والضعف لديهم.
- استخدام المعايير المئينية المستخرجة في تفسير درجات الطلبة بصورة دقيقة.
- تطوير برامج تعليمية علاجية للطلبة ذوي المستوى المنخفض في القدرة العددية.
- الاهتمام ببناء اختبارات نفسية وتربوية تتمتع بخصائص سيكومترية جيدة.
- تدريب المدرسين على استخدام نتائج الاختبارات في تحسين العملية التعليمية.

رابعاً: المقترحات: استكمالاً لما توصلت إليه البحث، يقترح الباحث إجراء الدراسات الآتية:

- إجراء دراسة مماثلة على مراحل دراسية أخرى مثل المرحلة الابتدائية أو الجامعية.
- تطبيق الاختبار على عينات أكبر وفي محافظات مختلفة للتحقق من ثبات نتائجه.
- بناء اختبارات أخرى لقياس أنواع مختلفة من القدرات العقلية.
- إجراء دراسة مقارنة بين نظرية القياس التقليدية ونظرية الاستجابة للمفردة في تحليل الاختبارات.
- تطوير اختبارات إلكترونية لقياس القدرة العددية.

- دراسة العلاقة بين القدرة العددية والتحصيل الدراسي في مادة الرياضيات.

المصادر

- Allen, M. J., & Yen, W. M. (2002). *Introduction to measurement theory*. Waveland Press.
- Anastasi, A., & Urbina, S. (2017). *Psychological testing* (7th ed.). Pearson.
- Bond, T. G., & Fox, C. M. (2015). *Applying the Rasch model: Fundamental measurement in the human sciences* (3rd ed.). Routledge.
- Cohen, R. J., Swerdlik, M. E., & Sturman, E. D. (2013). *Psychological testing and assessment: An introduction to tests and measurement* (8th ed.). McGraw–Hill.
- Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology, 78*(1), 98–104.
- Creswell, J. W., & Creswell, J. D. (2018). *Research design: Qualitative, quantitative, and mixed methods approaches* (5th ed.). Sage Publications.
- Crocker, L., & Algina, J. (2008). *Introduction to classical and modern test theory*. Cengage Learning.
- DeVellis, R. F. (2016). *Scale development: Theory and applications* (4th ed.). Sage Publications.
- Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Lawrence Erlbaum Associates.
- Ebel, R. L., & Frisbie, D. A. (1991). *Essentials of educational measurement* (5th ed.). Prentice Hall.
- Flanagan, D. P., & Dixon, S. G. (2014). The Cattell–Horn–Carroll theory of cognitive abilities. In D. P. Flanagan & P. Harrison (Eds.), *Contemporary intellectual assessment* (3rd ed.). Guilford Press.
- Geary, D. C. (2013). Early foundations for mathematics learning and their relations to learning disabilities. *Current Directions in Psychological Science, 22*(1), 23–27.

- Kaplan, R. M., & Saccuzzo, D. P. (2018). *Psychological testing: Principles, applications, and issues* (9th ed.). Cengage Learning.
- Kline, P. (2015). *A handbook of test construction: Introduction to psychometric design*. Routledge.
- Nunnally, J. C., & Bernstein, I. H. (2010). *Psychometric theory* (3rd ed.). McGraw-Hill.
- Schneider, W. J., & McGrew, K. S. (2018). The Cattell–Horn–Carroll model of intelligence. In D. Flanagan & P. Harrison (Eds.), *Contemporary intellectual assessment*. Guilford Press.
- Tabachnick, B. G., & Fidell, L. S. (2019). *Using multivariate statistics* (7th ed.). Pearson.
- Thorndike, R. M., & Thorndike-Christ, T. (2010). *Measurement and evaluation in psychology and education*. Pearson.