جامعة سوهاج

كلية التربية

المجلة التربوية

استخدام نظرية إمكانية التعميم في تقدير ثبات مقياس الحكمة الاختبارية لدى طلاب المرحلة الثانوية في ضوء التصميم المتقاطع ثنائي الوجه

إعداد

د/ رمضان السيد فرحات*

مدرس علم النفس التعليمي والإحصاء التربوي

كلية التربية بنين– جامعة الأزهر بالقاهرة

د/ أشرف سعد جاد الله*

مدرس علم النفس التعليمي والإحصاء التربوي

كلية التربية بنين– جامعة الأزهر بالقاهرة

تاريخ استلام البحث : ١ يونيو ٢٠٢٤م – تاريخ قبول النشر: ١٨ يونيو ٢٠٢٤م

DOI:

* البريد الإلكتروني: Ramadanfrahat2254.el@azhar.edu.eg

* البريد الإلكتروني: AshrafGadallah.8@azhar.edu.eg

مستخلص البحث:

هدف البحث إلى الكشف عن مصادر التباين الأكثر تأثيرًا على ثبات درجات مقياس الحكمة الاختبارية, والكشف عن أثر زيادة كل مصدر من مصادر التباين على ثبات المقياس, والوقوف على تأثير حجم العينة على قيم معاملات إمكانية التعميم النسبية والمطلقة للمقياس، وبلغ عدد المشاركين في البحث (٣٠٠) مشاركًا من طلاب وطالبات المرحلة الثانوية، تراوحت أعمارهم الزمنية ما بين (١٤–٢٠) عام, بمتوسط حسابي (١٦,٨٢٢), وانحراف معياري (١,٢٣٤), وتمثلت أداة البحث في مقياس الحكمة الاختبارية (إعداد الباحثين)، وتم استخدام الأساليب الإحصائية المتمثلة في تحليل دراسات التعميم والقرار لاستخراج معاملي إمكانية التعميم النسبي والمطلق وذلك باستخدام برنامج (EduG) V.6.1–e في ضوء التصميم المتقاطع ثنائي الوجه (الطلاب× الأبعاد× فترات التطبيق)، وتوصلت نتائج البحث إلى تمتع مقياس الحكمة الاختبارية بمعاملي إمكانية تعميم نسبي ومطلق جيدين جاءا على الترتيب (٠,٩٠– ٠,٨١), وأن أكبر مكون لتباين الخطأ في القياس النسبي يرجع إلى تفاعل (الطلاب × فترات التطبيق) بنسبة بلغت (٥٢,١٪) من إجمالي نسب التباين الكلي, وأكبر مكون لتباين الخطأ في القياس المطلق يرجع إلى (عدد الأبعاد) بنسبة بلغت (٥٠,٣٪) من إجمالي نسب التباين الكلي, وأن زيادة عدد الأبعاد له تأثير ضئيل في ارتفاع قيم معاملات إمكانية التعميم النسبية والمطلقة, وأن زيادة عدد فترات التطبيق يؤثر بشكل ملحوظ في ارتفاع قيم معاملات إمكانية التعميم النسبية والمطلقة, وأن قيم معاملي إمكانية التعميم النسبي والمطلق ترتفع بشكل دال إحصائيًا بزيادة حجم العينة, وقد تم مناقشة النتائج وتفسيرها في ضوء الإطار النظري والدراسات والبحوث السابقة, وقُدمت مجموعة من التوصيات والبحوث المقترحة.

الكلمات المفتاحية: نظرية إمكانية التعميم، الحكمة الاختبارية، طلاب المرحلة الثانوية, التصميم المتقاطع ثنائي الوجه.

Utilizing Generalizability Theory in Estimating the Reliability of the Test-Wiseness Scale among High School Students in the Context of a Two-Facet Crossed Design

Dr. Ramadan El Sayed Frahat *
Lecturer in Educational Psychology and Statistics
Faculty of Education (Boys), Al-Azhar University, Cairo

Dr.Ashraf Saad Gad Allah*
Lecturer in Educational Psychology and Statistics
Faculty of Education (Boys), Al-Azhar University, Cairo

ABSTRACT

This research aimed to identify the most influential sources of variance affecting the reliability of the test-wiseness scale scores, examine the impact of each variance source on the scale's reliability, and determine the effect of sample size on the values of both relative and absolute generalizability coefficients of the scale. The study involved 300 participants, both male and female high school students, aged between 14 and 20 years, with a mean age of 16.822 years and a standard deviation of 1.234 years. The research instrument used was the test-wiseness scale, developed by the authors. Statistical methods, specifically generalizability and decision studies, were employed to extract the relative and absolute generalizability coefficients using the EduG software version 6.1-e, based on a two-facet crossed design (Persons × Dimensions × Occasions). The results indicated that the test-wiseness scale demonstrated good relative and absolute generalizability coefficients of 0.90 and 0.81, respectively. The largest component of error variance in relative measurement was attributed to the interaction of (Persons × Occasions), which accounted for 52.1% of the total variance, while in absolute measurement, the largest error variance component was due to the number of dimensions, accounting for 50.3% of the total variance. It was found that increasing the number of dimensions had a negligible effect on the values of both relative and absolute generalizability coefficients. Conversely, increasing the number of occasions significantly affected the values of these coefficients. Additionally, the values of both relative and absolute generalizability coefficients significantly increased with the sample size. The results were discussed and interpreted within the framework of the theoretical background and previous studies, and several recommendations and proposed research ideas were provided.

Keywords: Generalizability Theory, Test-Wiseness, High School Students, Two-Facet Crossed Design.

* Email: Ramadanfrahat2254.el@azhar.edu.eg
* Email: AshrafGadallah.8@azhar.edu.eg

مقدمة:

تعتمد صحة القرارات النفسية والتربوية والاجتماعية المستخدمة في تشخيص قدرات الأفراد وتصنيفهم في مجموعات أو توجيههم لبرامج علاجية مناسبة على جودة الأدوات المستخدمة في عمليات القياس والتقويم, لذا أصبح اتصاف هذه الأدوات بالصدق والثبات والدقة أحد المعايير المستخدمة في الحكم على مدى صلاحية هذه الأدوات للاستخدام في عمليات التشخيص والتصنيف والمفاضلة بين الأفراد, فاستخدام أدوات قياس تشوبها أخطاء القياس والتقدير يترتب عليه اتخاذ قرارات غير دقيقة أو متحيزة.

لذا تتضح أهمية نظريات القياس النفسي في أنها تساعد في تحديد مصادر خطأ القياس ومستوى التأثير الذي تحدثه هذه المصادر, كما أنها تمثل الإطار الذي يتم من خلاله الكشف عن العوامل المؤثرة في الدرجة التي يحصل عليها الفرد, بالإضافة إلى أنها تمثل الأساس العلمي الذي ترتكز عليه إجراءات بناء أدوات القياس وتفسير الدرجات الناتجة عن تطبيق هذه الأدوات (سليمان, ٢٠٠٩, ٦٧)*.

وقد سيطرت نظرية القياس الكلاسيكية Classical test theory (CTT) وما يرتبط بها من نماذج وأساليب إحصائية تتعلق بالمفردات الاختبارية على منهجيات القياس النفسي طوال القرن الماضي, واستندت في مبادئها على عدد قليل من الافتراضات الخاصة بطبيعة البيانات الاختبارية, وقدمت إطارًا مرجعيًا لبناء وتقييم الاختبارات والمقاييس النفسية والتربوية, وتفسير درجاتها (علام, ٢٠٠٥, ٤٨).

ونظرًا لاعتمادها على مجموعة قليلة من الافتراضات؛ فإن ذلك أدي إلى ظهور كثير من الأخطاء في القياس كحصول الفرد على نسبتين مختلفتين في الذكاء مثلًا إذا تم اختباره باختبارين مختلفين يقيسان الذكاء أو موقفين مختلفين, وذلك لاختلاف درجة الخطأ المعياري في كل حالة (Domino& Domino, 2006, 34).

وتقدم نظرية القياس الكلاسيكية تقنيات فعالة في دراسة الثبات, إلا أنها ليست بالمرونة الكافية التي تتناسب مع جميع مشكلات الثبات التي يمكن أن تظهر في تطبيق الاختبارات والمقاييس النفسية, كأن يتم تقدير درجات الأفراد بواسطة أكثر من مصحح, أو استخدام صيغ متبادلة غير متكافئة تمامًا لنفس الاختبار, ففي هذه الحالات قد يتم الحصول

* يلتزم الباحثان في توثيقهما للمراجع بالنسخة السابعة لقواعد الجمعية الأمريكية لعلم النفس (APA).

على نتائج مضللة عند حساب معاملات الثبات والخطأ المعياري, كما أنه يصعب تطبيق نظرية القياس الكلاسيكية لاستخراج جميع مكونات الخطأ؛ حيث إنها توفر مكون خطأ واحد فقط غير متمايز (Crocker& Algina, 2008, 157).

فالدرجة التي يحصل عليها الفرد لا تمثل المستوى الفعلي أو الحقيقي لقدرته أو مهارته أو سلوكه المعتاد، ولمعرفة الدرجة الحقيقية فإن الأمر يتطلب معرفة الاتساق في أدائه وذلك من خلال عمليات القياس المتعددة، سواء باستخدام نفس الاختبار أو استخدام اختبارات مكافئة للاختبار، ويرجع اختلاف الدرجات التي يحصل عليها الفرد على نفس الاختبار وفي المرات المتعددة لعوامل متعددة مثل: الجهد الذي يبذله في المواقف الاختبارية المتعددة، وتركيزه وانتباهه للمهمة المراد منه القيام بها، واختلاف الأوقات التي يؤخذ فيها الاختبار، والنمو الطبيعي، والتغيرات في صحته، والتدريب والتعليم، والتغيرات في الشخصية، والاختلاف في المفردات الاختبارية من حيث صعوبتها وسهولتها ووضوحها وغموضها، وأثر الصدفة، وتؤثر مصادر الاختلاف على القرار الذي قد يتخذ في شأنه (الطريري، ١٩٩٧، ٤٦-٤٧).

لذا ظهرت نظرية إمكانية التعميم Generalizability Theory لتحسين إجراءات القياس للحصول على بيانات ثابتة كبديل للنظرية الكلاسيكية التي يتم فيها تقدير مصدر واحد لأخطاء القياس, حيث تسمح نظرية إمكانية التعميم بالتقدير المتزامن لمصادر التباين المتعددة والتفاعلات بينها, وتختلف تقديرات الثبات وفق نظرية إمكانية التعميم بشكل ملحوظ عن تقديرات الثبات وفق نظرية القياس الكلاسيكية خاصة إذا تفاعلت أخطاء القياس المرتبطة بالأوجه المختلفة لتصميم البحث مع بعضها البعض Heitman et al., 2009, (48).

وتقدم نظرية إمكانية التعميم إطارًا مفاهيميًا واسعًا للعلوم الاجتماعية والتربوية والنفسية وبناءً شاملًا لمصادر القياس المتعددة باستخدام تحليل التباين (ANOVA), وتعد نظرية إحصائية قوية يمكنها التعامل مع مصادر متعددة للخطأ باعتبارها امتدادًا لكل من نظرية القياس الكلاسيكية وتحليل التباين (Teker et al., 2015, 635).

ويعتمد جوهر نظرية إمكانية التعميم على فكرة أن الأداة تتضمن عينة صغيرة من مفردات ومهام كثيرة كان من الممكن تضمينها بدلًا من ذلك، والهدف النهائي من إجراء الاختبار هو الحصول على معلومات كافية من هذه العينة المحدودة لتعميمها عبر النطاق

الأشمل، ويعتمد نجاح التعميم من عينة محدودة إلى النطاق الأوسع على طبيعة ومقدار الخطأ الذي يتم مواجهته بسبب أخذ عينة صغيرة من النطاق المستهدف، وتوفر نظرية إمكانية التعميم طريقة لتحديد مقادير خطأ القياس التي تجعل التعميم من العينة إلى النطاق فيه خطورة، وتقدم تقديرًا لاتساق هذا التعميم من خلال فحص مدى اختلاف الدرجات بسبب الفرد ومصادر خطأ القياس المختلفة (Li et al., 2015, 1323).

فإذا كانت نظرية القياس الكلاسيكية تتعامل مع مصدر واحد فقط من مصادر التباين وهو الأفراد المختبرين؛ فإن نظرية إمكانية التعميم تقسم مصادر تباين الثبات إلى مصادر متعددة بناءً على عدد الأوجه أو الأبعاد المستخدمة في تصميم الدراسة, فكل وجه يمثل خاصية من خصائص جمع البيانات والتي من المتوقع أن تؤدي إلى تباين في الدرجة الكلية الملاحظة (Albano, 2018, 181).

ويمكن أن تكون التصميمات الخاصة بدراسات نظرية إمكانية التعميم متقاطعة أو متداخلة، ففي التصميمات المتقاطعة كليًا، يكون لموضوعات القياس درجات على جميع الأوجه والشروط، وتكون التصميمات متداخلة عندما يكون لموضوعات القياس درجات على بعض شروط الأوجه، وعادة ما يكون الأفراد هم موضوع القياس، ولا يعتبرون مصدرًا للخطأ، لأن اختلاف درجاتهم حقيقي ومنتظم بينما يهدف القياس إلى إيجاد تباين الخطأ بإجراء قياسات في مناسبات عديدة وأوجه مختلفة، والتي قد تكون مصادر محتملة لأخطاء القياس (Yelboga & Tavsancil, 2010, 1849).

وفي العديد من المواقف يجب تصنيف شروط القياس إلى بعدين أو أكثر، فعندما يستجيب المفحوص على عدة أسئلة مقالية، ويتم تصحيحها من خلال عدة مقدرين للدرجات، فإن مجموعة الأسئلة المقالية ومجموعة المقدرين يؤلفان أوجه القياس، ويسمى التصميم في هذه الحالة بالتصميم ثنائي الوجه (Crocker& Algina, 2008, 177).

ويمكن أن يهتم الباحثون بستة أوجه في نظرية إمكانية التعميم وهي المقدر، والعبارة، والوقت، والسياق، والطريقة، والبعد، وهذه الأوجه لها نظائرها أو ما يقابلها في نظرية القياس الكلاسيكية، فالتعميم من خلال المقدر يقابله في نظرية القياس الكلاسيكية ثبات المحكم، ويناظر تعميم الوقت ثبات إعادة التطبيق، ويماثل تعميم السياق الصدق المرتبط بالمحك، ويقابل تعميم العبارة ثبات الاتساق الداخلي أو صدق البناء، ويناظر تعميم الطريقة

الصدق التقاربي، وأخيرًا يماثل تعميم البعد الصدق التمايزي (Cone, 1977, Gresham & Carey, 1988 كما ورد في: أمين، ٢٠١٦، ١٣٩).

ونظرًا لأن التقويم يشكل جزءًا لا يتجزأ من عمليات التعليم والتعلم؛ فإن كيفية تعامل المتقدمين للاختبار مع المهام التي يتم تكليفهم بها, والتعرف على الاستراتيجيات التي يستخدمونها أثناء إجراء الاختبار يعد أمرًا مثيرًا للاهتمام من قِبل كل من المعلمين والمتعلمين وخبراء القياس على حد سواء, لذا تشكل استراتيجيات إجراء الاختبار وجوانبها المتعددة أحد الموضوعات ذات الصلة التي تركز عليها العديد من البحوث والدراسات (Bumbalkova, 2021, 647).

حيث يشير علماء القياس والتقويم التربوي إلى وجود عوامل متعددة تؤثر في تقدير درجات التقييم لنتائج تحصيل الطلاب, منها: قدرات الطلاب, ونوع الاختبار, وظروف التطبيق, وقد ركزت الكثير من البحوث والدراسات على عامل آخر يؤثر بشكل فعال في تباين درجات الطلاب وهو ما يسمى بالحكمة الاختبارية Test-Wiseness والذي حظي باهتمام المتخصصين في المجال التربوي بشكل عام ومجال القياس النفسي بشكل خاص (العبيدي, ٢٠٢٣, ٦٢٣).

وتعد الحكمة الاختبارية من العوامل الشخصية التي تؤثر إيجابيًا في درجات الطلاب, فهي تعمل على تهيئتهم نفسيًا, وتساعدهم في تنظيم دراستهم, وتعطي لهم خبرة في التعرف على أنواع الأسئلة وكيفية التعامل معها, واتباع التعليمات, وتقسيم زمن الاختبار على عدد الأسئلة وأنواعها وفقًا لمستويات صعوبتها (سعادة وعيسى, ٢٠٢٢, ٢٢١).

وتزداد قدرة الطلاب على الإفادة من مهارات الحكمة الاختبارية في ظل عدم وعي كثير من المعلمين بمباديء الحكمة الاختبارية, وعدم رغبتهم في التعرف على معايير صياغة مفردات الاختبارات, والتحقق من صدقها وثباتها, وتحديد معاملات صعوبة وتمييز هذه المفردات, بالإضافة إلى أن أداء بعض الطلاب قد يكون منخفضًا في الاختبارات المقننة بسبب عدم إلمامهم بمبادئ الحكمة الاختبارية, لذا تعد الحكمة الاختبارية عامل عام يؤثر بطرق متنوعة في أداء الطلاب في الاختبارات المختلفة (Hayati & Ghojogh, 2008, 170).

فقد يكون الأداء السيء لبعض الطلاب في الاختبارات ليس بسبب نقص معرفتهم بمحتوى الاختبار, ولكن بسبب نقص مكونات الحكمة الاختبارية لديهم أو بسبب استخدامهم

لاستراتيجيات غير فعالة, فالحكمة الاختبارية لها تأثيرات مباشرة وغير مباشرة على أداء الطلاب؛ حيث إنها تساعدهم بشكل مباشر في زيادة درجاتهم في الاختبارات من خلال الاستخدام الفعال لوقتهم وجهدهم وظروف ومتطلبات الاختبار, كما أنها تؤثر على عوامل أخرى مهمة للموقف الاختباري مثل تقليل قلق الاختبار وتحسين اتجاهات الطلاب نحو أداء الاختبارات (Dodeen,2015, 109).

لذا تعد الحكمة الاختبارية مصدرًا من مصادر تباين درجات الاختبار بين الطلاب في الاختبارات التي يتقدمون إليها, فالطلاب الذي يمتلكون مستوى مرتفع من الحكمة الاختبارية يحصلون على ميزة ذات قيمة مقارنة بزملائهم, كما أنها من الممكن أن تكون عنصرًا إيجابيًا في النمو الذاتي للطلاب, حيث إن معرفتهم بمبادئ الحكمة الاختبارية تُسهل استخدامهم لها في اختبارات العلوم التطبيقية والاجتماعية والفنون, كما أن هناك قصور في وجود مقاييس مقننة للحكمة الاختبارية في البيئة التعليمية والثقافية (Opesemowo at el., 2018, 467).

ويتضح مما سبق أهمية نظريات القياس النفسي بصفة عامة ونظرية إمكانية التعميم بصفة خاصة في التحقق من ثبات الاختبارات والمقاييس النفسية والتربوية، حيث إنها تعد امتدادًا لنظرية القياس الكلاسيكية وتختلف عنها في التعامل مع مصادر متعددة للخطأ في وقت واحد باستخدام أسلوب تحليل التباين، كما يتبين أهمية متغير الحكمة الاختبارية باعتباره من العوامل التي تساعد الطلاب في الإجابة عن أسئلة الاختبار بفاعلية، حيث إنها تساعدهم على الاستعداد للاختبار، والتعامل مع مختلف أنواع الأسئلة، والتعامل مع ورقة الإجابة، وتنظيم وقتهم بكفاءة، مما يتطلب إعداد مقياس للحكمة الاختبارية والتحقق من ثباته في ضوء نظرية إمكانية التعميم حتى يمكن استخدامه في التعرف على مستوى الحكمة الاختبارية لدى طلاب المرحلة الثانوية.

مشكلة البحث:

تُعد الاختبارات من أدوات القياس الأكثر شيوعًا في معظم الأنظمة التعليمية في العالم, سواء كانت اتجاهات الطلاب إيجابية نحوها أو سلبية، فهذه حقيقة ويجب التعامل معها, وقد أصبحت أهمية الاختبارات تتجاوز المدارس الأكاديمية, حيث يتم اتخاذ العديد من القرارات الحاسمة التي تؤثر على حياة الأفراد بناءً على اختبارات محددة, وبالنظر إلى هذه

القضايا، ينبغي بذل كل الجهود لمساعدة المتقدمين للاختبار بشكل عام والطلاب بشكل خاص على الأداء الجيد في الاختبارات التي يتقدمون إليها (Dodeen, 2015, 108).

وبالنظر إلى واقع العملية التعليمية بمصر حاليًا نجد أن الاختبارات التحصيلية تكاد تكون الأداة الوحيدة التي تستخدم لتحديد مستوى الطلاب بمختلف المراحل التعليمية, وعلى الأخص في نهاية المرحلة الثانوية العامة, حيث لا توجد أداة غيرها تستخدم في ذلك, لذا فإنه يجب على الباحثين التأكد من أن درجات الطلاب تعبر بشكل فعلي عن مستواهم المعرفي وذلك إما من خلال عزل المتغيرات التي تؤثر في درجاتهم التحصيلية, أو على الأقل تحديد مدى إسهام هذه المتغيرات في درجات الطلاب ليسهل بعد ذلك على متخذي القرار وضع تلك المتغيرات ومدى إسهامها بعين الاعتبار عند تصنيف الطلاب بناءً على درجاتهم, وتأتي الحكمة الاختبارية على رأس هذه المتغيرات, لذا فإن قياسها وتحديدها يعد ضرورة ملحة باعتبارها مصدر من مصادر تباين درجات الطلاب (محمد, ٢٠١٨, ٧٠-٧١).

وتبرز مشكلة البحث من خلال أهمية توافر مقياس للحكمة الاختبارية لدى طلاب المرحلة الثانوية يتمتع بمعاملات ثبات قابلة للتعميم في العديد من المواقف والظروف, وتحديد أكثر مصادر التباين تأثيرًا على ثبات درجات الأفراد على هذا المقياس.

فقد يتمتع بعض الطلاب بفهم جيد لموضوعات المادة الدراسية, وقدرة جيدة على القراءة, لكنهم لا يستطيعون إدارة وقت الاختبار بحكمة, كما أن هناك بعض الطلاب غير قادرين على تفسير تعليمات الاختبار, وتنقصهم مهارات الاستعداد الجيد لأدائه, كما أنهم قد يشعرون بمستويات قلق مرتفعة في قاعات الاختبارات, وربما يرجع ذلك إلى نقص امتلاكهم لمكونات الحكمة الاختبارية من أجل تحسين أدائهم وتحقيق أقصى قدر ممكن من أدائهم الاختباري(Mustapha, 2014, 42) .

وتتأثر درجات الطلاب في الاختبارات التحصيلية إلى حد كبير بمستوى الحكمة الاختبارية لديهم, حيث توجد شكاوي متكررة من بعض الطلاب تفيد بعدم حصولهم على درجات مرتفعة في الاختبارات تتناسب مع ما يبذلوه من جهد كبير في الاستعداد للاختبارات, بينما يحصل طلاب آخرون لديهم نفس مستوى القدرات المعرفية وبذلوا نفس الجهد وربما أقل في الاستعداد للاختبارات على درجات مرتفعة, وقد يرجع ذلك بدرجة كبيرة إلى اختلاف مستوى الحكمة الاختبارية لدى الطلاب, لذا يجب توجيه الاهتمام بتحديد أهم مكونات الحكمة

الاختبارية التي يمكن أن يستخدمها الطلاب, وقياس مدى استخدامهم لها أثناء الاختبارات (الزهراني, ٢٠١٥, ٢٢٢).

ولقياس مهارات الحكمة الاختبارية لدى الطلاب لابد من توافر أداة قياس مناسبة والتي يجب أن يتوافر فيها مجموعة من الخصائص وشروط الصلاحية للاستخدام, وذلك للحصول على معلومات وبيانات دقيقة تسهم في اتخاذ قرارات موضوعية ودقيقة (أبو الحسن, ٢٠٢٢, ٥٤).

ونظرًا لأن الحكمة الاختبارية مستقلة عن معرفة المختبر بموضوعات الاختبار؛ فيمكن اعتبارها أحد العوامل التي تساهم في تباين خطأ الدرجات الملاحظة, وبالتالي فإنه ما لم يكن جميع الطلاب يمتلكون نفس مستوى الحكمة الاختبارية؛ فإن الاختلافات في درجات الاختبار الناتجة عن مستويات الحكمة الاختبارية لدى الطلاب قد تؤثر في ثبات الاختبار, كما أنها قد تجعل درجات بعض الطلاب أقل من درجاتهم الحقيقية مما يشكل تهديدًا لصدق الاختبار وصلاحيته (Gbore & Osakuade, 2016, 35).

وتنبثق مشكلة البحث أيضًا من خلال ما تؤكده الأطر النظرية بضرورة استخدام نظرية إمكانية التعميم في تقدير ثبات الاختبارات والمقاييس النفسية مقارنة بالنظرية الكلاسيكية, كونها تأخذ في حساب الثبات جميع مصادر التباين المؤثرة فيه والتفاعل بين هذه المصادر.

حيث تعد نظرية القياس الكلاسيكية من النظريات الواهنة والمحدودة في معالجتها لمفهومي ثبات وصدق الاختبارات نظرًا لأنها لا تميز بين أخطاء القياس, وإنما تعطي قيمة تقديرية كلية لمصادر أخطاء القياس المتعددة المتعلقة بصيغة الاختبار المستخدم بواسطة مختبر أو فاحص معين وفي ظروف معينة, وقد حاولت نظرية الاستجابة للمفردة Item response theory (IRT) التغلب على بعض أوجه القصور في نظرية القياس الكلاسيكية, وعلى الرغم من أن كلا النظريتين تعالج تباين أخطاء القياس معالجة مختلفة؛ إلا أن كليهما ينظر إلى مصدر هذه الأخطاء نظرة ضيقة ومحدودة، فالمعلومات أو البيانات المستمدة من أي إجراءات قياس تكون مشوبة بدرجة معينة من الخطأ، ولذلك تعتمد نظرية إمكانية التعميم على تحديد ظروف القياس التي تسهم في الخطأ، لكي يتسنى تعميم القياس عليها (علام, ٢٠٠٦, ٦٩٨– ٧٠٠).

وتلعب نظرية إمكانية التعميم دورًا فريدًا ولا غنى عنه في تصميم وتقييم إجراءات القياس، ولهذا فإن معايير الاختبارات التربوية والنفسية، التي تم تطويرها بشكل مشترك من قبل جمعية البحوث التربوية الأمريكية، والجمعية النفسية الأمريكية، والمجلس الوطني للقياس في التعليم، تؤكد على ضرورة الرجوع إلى نظرية إمكانية التعميم عند التحقق من صدق وثبات إجراءات الاختبار أو الملاحظة (Cardinet et al., 2010, 2).

ويجب أن يدرك المتخصصون مدى تفوق نظرية إمكانية التعميم على نظرية القياس الكلاسيكية عند حساب معامل الثبات نتيجة لتعاملها مع جميع مصادر تباين الخطأ في آن واحد، هذا بالإضافة إلى أن نظرية القياس الكلاسيكية تهمل التعامل مع التفاعل الذى قد يظهر بين المصادر المختلفة لأخطاء القياس (الثبيتي، ١٩٩٨، ١٣٠).

ولن يكون من المناسب الاعتماد على درجة واحدة أخذت من موقف واحد ووفق ظروف دون أخرى واعتمدت على وجه واحد (مصدر خطأ) دون بقية الأوجه كما في نظرية القياس الكلاسيكية، بل لابد من الأخذ في الاعتبار كل الأوجه التي يشتمل عليها موقف الاختبار أو بعضها في ضوء مبررات قوية وواضحة، وعليه فإن الدرجة الحقيقة في هذه الحالة ستكون مرتبطة بل ومعتمدة على نوعية النطاق أو الأوجه التي تؤخذ في الاعتبار كمنطلق من أجل الحكم على درجة الفرد (أمين، ٢٠١٦، ١٣٩).

وعلى الرغم من الاهتمام القوي بنظرية إمكانية التعميم داخل مجتمع القياس، والاعتراف بها على نطاق واسع كإطار شامل لقياس ثبات المقاييس، وفوائدها المحتملة، إلا أن تطبيقاتها لا تزال نادرة وتم استخدامها بشكل منخفض في كتابة تقارير نتائج مقاييس الفروق الفردية، وقياس وتقييم وتحسين الخصائص السيكوترية للمقاييس، وقد تكون الأسباب المحتملة لهذا الإهمال هي مفاهيم النظرية، وتجاهل الارتباط بينها وبين نظرية القياس الكلاسيكية، وصعوبة إيجاد البرامج وتشغيلها لإجراء التحليلات الإحصائية التي تتطلبها النظرية (Vispoel et al., 2018, 1).

كما تبرز مشكلة البحث أيضًا من خلال توصيات الدراسات والبحوث السابقة التي تؤكد ضرورة استخدام نظرية إمكانية التعميم في تقدير ثبات الاختبارات والمقاييس النفسية والتربوية، واختيار الأوجه المناسبة حسب هدف البحث وطبيعة المتغير المقاس ومكوناته، فيوصي Ogunka and Orluwene (2019, 1) باستخدام نظرية إمكانية التعميم في

تقدير الخصائص السيكومترية لإيجاد مفردات اختبارية ثابتة، ويوصي Orluwene and Memory (2020, 328) أيضًا باستخدام نظرية إمكانية التعميم في تحديد مدى ثبات المقاييس نظرًا لقدرتها على فصل مصادر الخطأ المتعددة.

ويشير Dorathy et al. (2021, 178) على أنه يجب استخدام نظرية إمكانية التعميم لتحديد مدى ثبات الأداة حيث إنها ستراجع مصادر متعددة للتباينات، كما يشير العرايضة والشريفين (٢٠٢٢أ، ٥٨٢) أهمية إجراء بعض الدراسات المستقبلية التي تثري أدبيات القياس والتقويم باستخدام نظرية إمكانية التعميم كالكشف عن مدى مساهمة مصادر تباين الخطأ لأدوات قياس مختلفة كالاستبانات واختبارات الاستعداد.

ويبين الغرايبة وبني أحمد (٢٠٢٣، ٩٩) أن استخدام نظرية القياس الكلاسيكية أصبح غير كافي في تقدير ثبات الدرجات وذلك لأنها لا تأخذ بعين الاعتبارات جميع مصادر الأخطاء (المصححين وفترات التقييم وطرق التقييم والمهام وشكل الاختبار)، ولذلك أصبحت الحاجة مهمة إلى إجراء بحوث تسلط الضوء على نظرية إمكانية التعميم في تقدير ثبات الاختبارات.

ويشير Huang (2009, 887) إلى أنه يمكن استخدام بعض الأساليب لخفض مكونات التباين: كدمج أكبر عدد من الأوجه عندما يكون ذلك ممكنًا, واستخدام تصميمات متقاطعة بدلًا من تصميمات متداخلة, ودمج الفترة كوجه من أوجه القياس.

ويرى Jackson et al. (2022, 54) أن ثبات المقاييس يعتمد على عوامل متعددة منها المقدرين، والمفردات، والأبعاد، وجميع التفاعلات الممكنة بين هذه العناصر، وخطأ البواقي الناتج عن تأثيرات أخرى غير معروفة.

كما يوضح طباع ودعيدش (٢٠٢٣، ١٤–١٥) أن درجات الطلاب يمكن أن تتأثر بالفترة التي أنجزت فيها التقييمات، ولكن استثنت الفترة من مصادر التباين ولم تلق اهتمامات بحثية مقارنة بمصادر التباين الأخرى كالمهمة والمقدر، ويعد مصدر تباين الفترة وتفاعله مع الأبعاد الأخرى خاصة مصدر رئيس للتباين لا يمكن تجاهله أثناء تقدير الثبات.

ويتضح مما سبق، أهمية استخدام نظرية إمكانية التعميم في التحقق من ثبات مقياس الحكمة الاختبارية، ولذا سعى البحث الحالي إلى تحديد بعض أوجه مصادر الخطأ

(عدد الأبعاد/ مكونات الحكمة الاختبارية، وفترات التطبيق) بهدف حساب تبايناتها المختلفة وتقدير معاملي الثبات النسبي والمطلق للمقياس في ضوء النظرية.

وبناءً على ذلك يمكن صياغة مشكلة البحث الحالي في التساؤلات التالية:

١- ما مصادر التباين (الطلاب, وعدد الأبعاد, وفترات التطبيق) الأكثر تأثيرًا على ثبات درجات مقياس الحكمة الاختبارية لدى طلاب المرحلة الثانوية؟

٢- ما أثر زيادة كل مصدر من مصادر التباين (عدد الأبعاد, وفترات التطبيق) على ثبات مقياس الحكمة الاختبارية لدى طلاب المرحلة الثانوية؟

٣- ما أثر اختلاف حجم العينة (١٠٠- ٢٠٠-٣٠٠) على قيم معاملات إمكانية التعميم النسبية والمطلقة لمقياس الحكمة الاختبارية لدى طلاب المرحلة الثانوية؟

هدف البحث:

يهدف البحث الحالي إلى:

١- الكشف عن مصادر التباين (الطلاب, وعدد الأبعاد, وفترات التطبيق) الأكثر تأثيرًا على ثبات درجات مقياس الحكمة الاختبارية لدى طلاب المرحلة الثانوية.

٢- الكشف عن أثر زيادة كل مصدر من مصادر التباين (عدد الأبعاد, وفترات التطبيق) على ثبات مقياس الحكمة الاختبارية لدى طلاب المرحلة الثانوية.

٣- الوقوف على تأثير حجم العينة على قيم معاملات إمكانية التعميم النسبية والمطلقة لمقياس الحكمة الاختبارية لدى طلاب المرحلة الثانوية.

أهمية البحث:

يمكن أن تتضح أهمية البحث الحالي في النقاط التالية:

١- تسليط الضوء على نظرية إمكانية التعميم وإسهاماتها السيكومترية في مجال تقدير ثبات الاختبارات والمقاييس والنفسية.

٢- توجيه نظر الباحثين في مجال القياس النفسي إلى استخدام نظرية إمكانية التعميم في تقدير ثبات أدوات القياس النفسي والتربوي بما ينعكس إيجابيًا على جودة بناء هذه الأدوات والتحقق من خصائصها السيكومترية.

٣- التحقق من ثبات مقياس الحكمة الاختبارية وفق نظرية إمكانية التعميم يراعي في حسابه تنوع مصادر تباين الخطأ, يمكن الاستفادة منه في تقييم مستوى الحكمة الاختبارية لدى

طلاب المرحلة الثانوية, في ظل ندرة استخدام هذه النظرية في استخراج معاملات ثبات الاختبارات والمقاييس النفسية.

٤- تقديم دلالات إحصائية تتعلق بأثر زيادة عدد أبعاد مقياس الحكمة الاختبارية وفترات التطبيق من خلال دراسات القرار, والتي يتم من خلالها تزويد الباحثين بأفضل شروط القياس المناسبة للحصول على معاملات تعميم جيدة لمقياس الحكمة الاختبارية لدى طلاب المرحلة الثانوية.

التعريف الإجرائي لمصطلحات البحث:

١- نظرية إمكانية التعميم Generalizability Theory

إحدى نظريات القياس التي تهتم بتقدير ثبات الاختبارات والمقاييس النفسية والتربوية من خلال تحليل مصادر التباين أو الأوجه المتعددة لأخطاء القياس والتفاعلات بينها في وقت واحد لتقدير مدى إسهامها في خطأ القياس الكلي وتحديد الأوجه التي لها التأثير الأكبر على الثبات الذي يعبر عنه بمعاملي إمكانية التعميم النسبي والمطلق (الباحثان).

٢- الحكمة الاختبارية Test-Wiseness

قدرة الطالب على التعامل بفاعلية مع الموقف الاختباري من خلال الاستعداد لأداء الاختبار بما يضمن تهيئته نفسيًا وعقليًا للحصول على درجات مرتفعة, والإدارة الجيدة لوقت الاختبار بما يساعده على تنظيم وقت الإجابة مع الاحتفاظ بالدقة, والتعامل المنظم مع ورقتي الأسئلة والإجابة بما يمكنه من فهم متطلبات كل سؤال وتنظيم عرض الإجابة, واستنتاجه للمؤشرات المستترة داخل أسئلة الاختبار بما يساعده على التوصل إلى الإجابة الصحيحة, وتحديد أوجه القوة والقصور في إجابته سواء من خلال الإجابات التي قدمها لكل سؤال أو من خلال تقييمه لأدائه الاختباري بشكل عام بعد الانتهاء من الاختبار (الباحثان).

٣- التصميم المتقاطع ثنائي الوجه (الطلاب× الأبعاد× فترات التطبيق)

Two-Facet crossed design (persons x Dimensions x Occasions)

أحد تصميمات نظرية إمكانية التعميم والذي يتطلب أن يستجيب جميع الطلاب المشاركين في البحث على جميع شروط أوجه مصادر الخطأ المتمثلة في الأبعاد "من البعد الأول إلى السابع", وفترات التطبيق "الأولى, والثانية" (الباحثان).

حدود البحث:

تتمثل حدود البحث الحالي في مقياس الحكمة الاختبارية المعد في البحث الحالي, ونظرية إمكانية التعميم المستخدمة في التحقق من ثبات المقياس في ضوء التصميم ثنائي الوجه المتقاطع كليًا (الطلاب, وعدد الأبعاد, وفترات التطبيق), والبرامج الإحصائية المستخدمة في البحث الحالي, وطلاب المرحلة الثانوية الذين اشتقت منهم عينة البحث في الفصل الدراسي الثاني من العام الدراسي (٢٠٢٣/ ٢٠٢٤م).

الإطار النظري للبحث:

المحور الأول: نظرية إمكانية التعميم:

أولًا: نشأة وتطور نظرية إمكانية التعميم:

نوقشت المبادئ الأساسية لنظرية إمكانية التعميم لأول مرة في المقالات التي نشرها Cronbach, Rajaratnam, Gleser ١٩٦٣ و١٩٦٥، وفي الواقع، بدأ استخدام تحليل التباين في دراسات الثبات قبل عمل Cronbach وزملائه، حيث ناقش Burt ١٩٣٦، وHoyt ١٩٤١، Jackson, Ferguson ١٩٤١ استخدام تحليل التباين في التنبؤ بالثبات، ثم أتبع ذلك الإسهامات التي قدمها Alexander ١٩٤٧، و Ebel ١٩٥١، و Finlayson١٩٥١، وLoveland ١٩٥٢، وBurt ١٩٥٥، وتم تقديم المعالجة الأولية والمحددة للنظرية من قبل Cronbach, Gleser, Nanda, Rajaratnam ١٩٧٢في كتاب ثبات القياسات السلوكية، ونشر Brennan ١٩٨٣ كتاب عناصر نظرية إمكانية التعميم، وقدم Shavelson, Webb ١٩٩١ دراسة مختصرة تصف أساسيات النظرية، وقدم Brennan ٢٠٠١ معالجة موسعة حديثة للنظرية، وتتناول اتساق أو ثبات تفسيرات الدرجات من خلال معرفة وتقدير حجم خطأ القياس من مصادر متعددة (Teker et al., 2015, 636؛ Li et al., 2015, 1322-1323 ؛ Brennan, 2003, 1).

ويُعد Cronbach مؤسس نظرية إمكانية التعميم، من خلال الأدبيات التي شارك في تأليفها والتي تظل حتى اليوم مراجع أساسية للباحثين الراغبين في دراسة منهجية واستخدامات النظرية، وتكمن أصالة منهجيتها في إدخالها تغييرًا جذريًا في منظور نظرية القياس وممارساته قائمًا على تحليل التباين، الذي يتمثل هدفه الأساسي في تقسيم التباين

الكلي للبيانات إلى عدد من المصادر التفسيرية المحتملة، وعلى الرغم من هذا التغيير العميق في المنظور، فإن النظرية لا تتعارض مع نتائج وإسهامات نظرية القياس الكلاسيكية (Cardinet et al., 2010, 1-2).

وقد قام Brennan& Crick بتصميم برنامج حاسوبي يسمى تحليل التعميم لنظام التباين (GENOVA) ١٩٨٣، ومع ذلك، نظرًا لأن النظرية والبرنامج المعد لها ظهرا معقدين للغاية بالنسبة للمستخدمين في باديء الأمر، فقد ظلت الدراسات المتعلقة بالنظرية محدودة حتى نشر Shavelson Webb ١٩٩١ كتابهما مقدمة في نظرية إمكانية التعميم، والذي جعل أساسيات النظرية أكثر قابلية للفهم والتنظير والتطبيق، وبالتالي تزايد انتشار النظرية، فنشر Brennan ٢٠٠١ كتاب نظرية إمكانية التعميم وقام بتطوير برنامج mGENOVA للتحليلات متعددة المتغيرات، وبرنامج urGENOVA للتنبؤ بمكونات التباين في التأثيرات العشوائية في التصميمات المتوازنة وغير المتوازنة (Teker et al., 2015, 636).

وعلى الرغم من أن البرامج التي طورها Brennan استخدمت بشكل متكرر من قبل الباحثين، إلا أن تعقيدها تسبب في حدوث مشكلات خاصة للمستخدمين لأول مرة، وأدى عدم وجود برنامج حاسوبي سهل الاستخدام إلى تقييد الاستخدام الواسع النطاق للنظرية لفترة طويلة، وبحثًا عن حل لهذه المشكلة، قام Duquesne بتطوير برنامج يسمى ETUDGEN في أوائل الثمانينات، وعلى الرغم من أن البرنامج يلبي الاحتياجات الأكاديمية الأساسية للباحثين في هذا المجال؛ إلا أنه ظل محدودًا فيما يتعلق بإجراء بعض التطبيقات، وقدم Mushquash, O'Connor بعد ذلك بعض الصيغ المكتوبة لإجراء تحليلات إمكانية التعميم من خلال برامج SAS, SPSS, وMATLAB في مقال ٢٠٠٦، وتم تطوير برنامج EduG بواسطة Cardinet ٢٠٠٦ كبرنامج متعدد الاستخدامات وسهل الاستخدام لإجراء تحليلات إمكانية التعميم (Cardinet et al., 2010, 9 ؛ Teker et al., 2015, 636-637).

ومع مرور أكثر من (٤٠) عامًا منذ أن نشر Cronbach, Gleser Nanda, Rajaratnam أطروحتهم الأساسية ١٩٧٢حول نظرية إمكانية التعميم، أدى عملهم إلى توسيع وجهات النظر حول النظرية بشكل كبير من خلال توفير إطار شامل لتقدير ثبات الدرجات مع الإشارة إلى مصادر متعددة لخطأ القياس (Vispoel et al., 2018, 1).

ومنذ نشأتها قدمت نظرية إمكانية التعميم إطارًا لوضع تصور وتقييم وتحسين ثبات الدرجات الناتجة عن المقاييس الموضوعية والذاتية في العديد من التخصصات، فعند إجراء بحث في قاعدة بيانات PsycNet باستخدام الكلمات الرئيسة نظرية إمكانية التعميم، ظهرت (٢٤٧٤) نتيجة، وظهر نصفها تقريبًا (١٢٢٦) في الأدبيات البحثية منذ عام ٢٠١٢ نظرًا لمميزاتها المتعددة حيث تتضمن تعريفات واضحة للمجالات التي يتم تعميم النتائج عليها، ومؤشرات تعكس مدى إمكانية تعميم النتائج على تلك المجالات، وتقدير مدى إمكانية تغيير معامل إمكانية التعميم والاعتمادية للدرجات عند تغيير إجراءات القياس (Vispoel et al., 2023, 545).

ويتضح مما سبق أن عالم علم النفس الشهير Lee Cronbach وزملاؤه Gleser, Nanda and Rajaratnam هم مؤسسي نظرية إمكانية التعميم، حيث ساهموا بشكل كبير في تقديم الأفكار والمفاهيم النظرية لها، ومبررات استخدامها في تقدير ثبات الاختبارات والمقاييس النفسية والتربوية، كما حاول العديد من العلماء تطوير الجوانب العملية للنظرية وتصميم البرامج الإحصائية المختلفة التي تساعد الباحثين على إجراء العمليات الإحصائية للبيانات بسهولة ودقة وسرعة عالية مثل: Brennan الذي صمم برنامجي mGENOVA، urGENOVA، وDuquesne حيث طور برنامج ETUDGEN ، وCardinet الذي قام بتصميم برنامج EduG الذي اعتمد عليه الباحثان في تحليل البيانات وحساب معاملات الثبات.

ثانيًا: الفروق بين نظرية إمكانية التعميم ونظرية القياس الكلاسيكية:

توجد عدة منطلقات لنظرية إمكانية التعميم تتمثل في: (أ) اعتبار درجة المفحوص على الاختبار على أنها نموذج لملاحظات شاملة، والشمولية عبارة عن وصف للأوضاع والظروف التي يكون عليها المفحوص أثناء الاختبار والتي من الممكن أن تؤثر على النتائج. (ب) الدرجة الشاملة هي الدرجة المتوقعة من بين مجموعة الدرجات المستنتجة من بين مجموعة من الملاحظات. (ج) توجد مجموعة من النطاقات يتم بناءً عليها تطوير الاختبار أو المقياس، ولذلك ستكون مهمة مطور المقياس تحديد النطاق المناسب لعملية القياس، مع الأخذ في الاعتبار الأبعاد المشكلة للنطاق. (د) تُوظف لتحقيق هدفين هما بناء أدوات القياس بغرض التأكد من أن نتائج الاختبار متكافئة عندما يتم استنباطها من مواقف اختبارية متعددة وذات ظروف مختلفة ومستويات أبعاد متنوعة ويطق عليها دراسات التعميم، واستخدام نتائج

الاختبار لاتخاذ قرارات في حق الأفراد ويطلق عليها دراسات القرار (الطريري، ١٩٩٧، ٤٩-٥١).

وتقدم نظرية إمكانية التعميم إطارًا فكريًا مرجعيًا لكيفية تقدير ثبات القياسات السلوكية التي يتم الحصول عليها بواسطة إجراءات أو أدوات القياس المختلفة، مثل: الاختبارات، وقوائم الملاحظة أو المراجعة، وموازين التقدير، ومقاييس الميول والاتجاهات والشخصية، ولا تعد هذه النظرية امتدادًا للنظرية الكلاسيكية فحسب، بل هي أسلوب جديد يساعد الباحث ومصمم المقياس في معرفة ما إذا كانت قياسات معينة تحقق الهدف الذى ينشده من القياس، مثل: تقويم مقرر دراسي، أو انتقاء الأفراد، أو تقويم بنائي أو تجميعي لبرنامج معين أو الحصول على أدلة تتعلق بصدق التكوين الفرضي بصفة عامة (علام، ٢٠٠٦، ٧٠٠).

حيث تهتم نظرية القياس الكلاسيكية بنوع واحد فقط من أخطاء القياس في كل مرة، ولا تأخذ في الاعتبار تأثيرات التفاعل المحتملة لتباين خطأ القياس، ولا تعترف بالتمييز بين الثبات الذي يتضمن قرارات مطلقة يتم اتخاذها في سياق درجات القطع مقابل الثبات الذي يتضمن قرارات تأخذ في اعتبارها الاستقرار في الوضع النسبي أو التصنيف، بينما تعتبر نظرية إمكانية التعميم الملاحظة السلوكية عينة واحدة فقط من نطاق غير محدود من الملاحظات المقبولة، وتهتم بمدى نجاح تعميمها على هذا النطاق (Yelboga & Tavsancil, 2010, 1848,1850).

وتعد نظرية إمكانية التعميم طريقة لتقدير دقة القياس في المواقف التي تخضع فيها القياسات لمصادر متعددة للخطأ، وتوفر وسيلة لتقدير ثبات القياسات التي تم إجراؤها بالفعل، وتستخدم المعلومات حول إسهامات الأخطاء لتحسين إجراءات القياس في التطبيقات المستقبلية، حيث إن جميع إجراءات القياس لا يمكن أن تكون دقيقة تمامًا، سواء في العلوم الاجتماعية والصحية، أو الطبيعية أيضًا، لأن السمات أو الشروط التي يتم تقديرها يصعب تعريفها بأي معنى مطلق، ولا يمكن عادة ملاحظتها بشكل مباشر (Cardinet et al., 2010, 1).

ويتمثل الهدف الرئيس لنظرية إمكانية التعميم في تعميم درجات أداة قياس محددة من مجموعة معينة إلى نطاق التعميم الذي يتكون من نطاق الملاحظات المقبولة ودراسات إمكانية التعميم، ونطاق دراسات القرار (Teker et al., 2015, 635-636).

وتفترض نظرية القياس الكلاسيكية أن الدرجة الملاحظة للفرد (X) هي مجموع الدرجة الحقيقية (T) ودرجة الخطأ (E)، وتمثل الدرجة الحقيقية الدرجة المتوقعة أو المتوسطة الملاحظة للفرد عبر عدد لا نهائي مفترض من عمليات إعادة القياس دون أي تأثيرات للاختلاف، ومعامل الثبات = تباين الدرجة الحقيقية ÷ (تباين الدرجة الحقيقية+ تباين الخطأ)، ويُستبدل مفهوم الدرجة الحقيقية في نظرية القياس الكلاسيكية بمفهوم الدرجة الشاملة في نظرية إمكانية التعميم، ومعاملات إمكانية التعميم في نظرية إمكانية التعميم تشبه معاملات الثبات في نظرية القياس الكلاسيكية، ومعامل إمكانية التعميم= تباين الدرجة الشاملة ÷ (تباين الدرجة الشاملة+ المصادر الفردية لتباين الخطأ) (Vispoel et al., 2018, 2).

وينتج الخطأ غير المتمايز في نظرية القياس الكلاسيكية عن أسباب عشوائية فقط، بينما يُقسم في نظرية إمكانية التعميم إلى خطأ منتظم وخطأ عشوائي، أي أن الدرجة الملاحظة تتكون من درجة حقيقية وخطأ منتظم وخطأ عشوائي، ويمثل الخطأ المنتظم تباين الأوجه الذي يمكن تقسيمه بشكل أكبر بناءً على عدد الأوجه المطلوبة في تصميم الدراسة، وتقدر المصادر المتعددة للأخطاء بشكل منفصل في تحليل واحد، مما يساعد الباحث على اتخاذ قرار والتحقق من مدى ثبات الدرجات لأنواع مختلفة من التفسيرات للوضع النسبي للمختبر والمستوى المطلق للأداء (Dorathy et al., 2021, 174).

وتركز معاملات الثبات في ضوء نظرية القياس الكلاسيكية على اتساق درجات الاختبار، فمثلًا يأخذ ثبات إعادة التطبيق توقيت الاختبار في الاعتبار فقط، ويهتم ثبات الصور المتكافئة بصور الاختبار فقط، ويُمثل الاتساق الداخلي للمفردات المصدر الوحيد للخطأ، ولذلك، فإن عدم كفاية تفسير العديد من مصادر الخطأ، وعدم القدرة على تحديد مصادر الخطأ المحتملة وفحصها في نفس الوقت، وعدم فحص تأثيرات التفاعل بين مصادر الخطأ المختلفة، والتعامل مع الخطأ على أنه عشوائي دون التمييز بين الخطأ المنتظم والخطأ العشوائي؛ أدى إلى تطوير نظرية إمكانية التعميم (Akindahunsi & Afolabi, 2021, 147-148).

وتتميز نظرية إمكانية التعميم عن نظرية القياس الكلاسيكية بـ (أ) التكامل من الناحية المفاهيمية والتقييم المتزامن لثبات إعادة الاختبار، والاتساق الداخلي، والصدق التقاربي، والثبات بين المقدرين. (ب) تقدير تأثير أوجه القياس والتفاعلات بينها. (ج)

تحسين تقييم الثبات مع مراعاة الاقتصاد في التكلفة والوقت، فيمكن الحصول على معلومات حول عدد المناسبات والمفاهيم والمقدرين اللازمين للوصول إلى نتيجة يمكن الاعتماد عليها. (د) توفر معلومات حول المستوى المطلق والنسبي. (هـ) تتضمن افتراضات أقل تقييدًا. (و) التعامل بسهولة مع التصميمات غير المتوازنة، (ز) معالجة خطأ القياس باعتباره موضوع متعدد الأوجه (Clayson et al., 2021, 175 ؛ Yin & Shavelson, 2008, 276).

وتم تأسيس نظرية إمكانية التعميم لمعالجة القيود المفروضة على نظرية القياس الكلاسيكية، حيث تستخدم نظرية إمكانية التعميم النمذجة الرياضية والمعايرة الإحصائية لتقسيم مصادر التباين المتنوعة بدقة، فتقوم بتقسيم التباين الكلي إلى مكونات متمايزة تمثل البنية المستهدفة مقابل الخطأ، وتُقدر بعد ذلك الأحجام النسبية لمكونات التباين باستخدام تقنيات تحليل التباين، وتسمح بتقييم فاعلية نتائج القياس للأغراض معيارية أو محكية المرجع، والتحقق من تغيرات الثبات في ظل ظروف الاختبار المتغيرة (Zhao et al., 2024, 2-3).

وفي ضوء ما سبق؛ يلخص الباحثان أهم الفروق بين نظرية إمكانية التعميم ونظرية القياس الكلاسيكية في جدول (١):

جدول (١)
مقارنة بين نظرية إمكانية التعميم ونظرية القياس الكلاسيكية

م	وجه المقارنة	نظرية إمكانية التعميم Generalizability theory (GT)	نظرية القياس الكلاسيكية Classical test theory (CTT)
١	قياس أخطاء القياس	تقيس مصادر أخطاء القياس المختلفة في وقت واحد	تقيس مصدر واحد فقط للقياس في المرة الواحدة دون الاهتمام بالمصادر الأخرى
٢	التفاعل بين أخطاء القياس	تهتم بحساب التفاعل بين أخطاء القياس	تُهمل حساب التفاعل بين أخطاء القياس
٣	أسباب أخطاء القياس	أخطاء منتظمة وعشوائية	أخطاء عشوائية
٤	الدرجة الملاحظة	مجموع الدرجة الشاملة ودرجات الأخطاء العشوائية والمنتظمة	مجموع الدرجة الحقيقية ودرجة الخطأ العشوائي
٥	كيفية حساب الدرجة الشاملة أو الحقيقية	تمثل الدرجة الشاملة متوسط درجات الفرد التي يحصل عليها في جميع شروط القياس الممكنة	تمثل الدرجة الحقيقية متوسط درجات الفرد التي يحصل عليها في شرط واحد من شروط القياس
٦	حساب الثبات	تركز على حساب ثبات القياسات عبر الظروف ومصادر أخطاء القياس المختلفة	تركز على حساب ثبات قياس واحد في ظروف معينة
٧	معامل الثبات	تميز بين معاملات الثبات التي تهتم بالقرارات النسبية (معامل إمكانية التعميم النسبي) والقرارات المطلقة (معامل إمكانية التعميم المطلق)	تميز بين معاملات الثبات بناءً على أحد مصادر الخطأ المؤثرة على القياس (الفترة الزمنية, تكافؤ الصور, تقسيم الاختبار, الاتساق الداخلي)
٨	كيفية حساب معامل الثبات	تستخدم تحليل التباين لتقدير إسهام مصادر الخطأ المختلفة في التباين الكلي	تستخدم معامل ألفا أو معامل ارتباط بيرسون لتقدير معامل الثبات
٩	التعميم	تؤكد على إمكانية تعميم القياسات عبر الظروف المختلفة في ضوء نوع الأوجه عشوائية أم ثابتة	تركز على ظروف قياس محددة دون الاهتمام بتعميم القياسات وفق الظروف المختلفة
١٠	اتخاذ القرار	تقيم مكونات التباين لاتخاذ قرارات بشأن إجراءات القياس المستقبلية	تتخذ قرارات بناءً على الدرجات الملاحظة دون الاهتمام بمصادر الخطأ أو الظروف المختلفة

ثالثًا: المفاهيم الأساسية المستخدمة في نظرية إمكانية التعميم:

١- النطاق الشامل Universe domain:

يُعبر النطاق الشامل عن ظروف أو شروط جمع الملاحظات أي المثيرات والمواقف التي يتم في ضوئها جمع الملاحظات أو الحصول على القياسات حتى يمكن تعميم هذه التقديرات أو القياسات من موقف إلى آخر ومن عينة من الأفراد إلى عينة أخرى، ويمكن أن تنتمى ملاحظات أو قياسات معينة إلى نطاقات شاملة مختلفة، ولذلك يجب أن يحدد الباحث

النطاق تحديدًا دقيقًا؛ لكي يتأكد من دقة تعميم الملاحظات أو القياسات على نطاق شامل معين (سليمان، ٢٠٠٩، ٨٨؛ علام، ٢٠٠٦، ٧٠٢).

ويشير النطاق الشامل للملاحظات المقبولة إلى المدى الكامل للشروط الممكنة للأوجه وكيفية تعريف تلك الأوجه ودمجها، أو جميع الشروط الممكنة للأوجه التي يريد الباحث التعميم عليها (Clayson et al., 2021, 175).

٢- الدرجة الشاملة Universe score:

تعرف الدرجة الشاملة بالقيمة المتوقعة لدرجات الفرد الملاحظة التي يحصل عليها في مختلف المواقف التي تنتمى إلى النطاق الشامل المطلوب، ويمكن تقدير تباينها باستخدام آلية تحليل التباين المستخدمة في نظرية إمكانية التعميم، وتُعادل الدرجة الحقيقية في نظرية القياس الكلاسيكية، وتعتمد على فكرة تعدد النطاقات الشاملة التي يود متخذ القرار تعميم قياساته عليها، فالدرجات التي تنتمى إلى نطاق شامل معين ربما تتباين في أكثر من جانب أو وجه، وتعد هذه الأوجه مصدرًا مهمًا من مصادر التباين التي يجب أن يحددها الباحث (سليمان، ٢٠٠٩، ٨٨؛ علام، ٢٠٠٦، ٧٠٢؛ Tavakol & Brennan, 2013, 221).

وتشير الدرجة الشاملة إلى القيمة المتوقعة لجميع الملاحظات عبر النطاق الشامل لفرد ما، أو متوسط الدرجات التي يمكن توقعها عبر جميع الاختلافات المحتملة في إجراء القياس مثل: مقدري الدرجات أو صور الاختبار أو المفردات المختلفة Li et al., 2015, (Akindahunsi & Afolabi, 2021, 148؛ 1323) .

٣- أوجه مصادر الخطأ Facet of error sources:

تحدد نظرية إمكانية التعميم مصادر خطأ القياس، وتفصلها، وتقدر كل منها، وتعتبر سمات موقف القياس التي لا يهتم بها متخذ القرار مثل: صور الاختبار، أو المفردات، أو المناسبات، أو المقدرين، مصادر محتملة للخطأ وتسمى بأوجه القياس، فعندما يجيب الطلاب عن (٥٠) مفردة لأحد الاختبارات في مناسبتين، تكون المفردات والمناسبات هما أوجه القياس (Dorathy et al., 2021, 173؛ Webb & Shavelson, 2005, 717).

حيث تسمى ظروف القياس في نظرية إمكانية التعميم بالوجه، فعندما يتم تقييم أداء عمال المصنع من قبل اثنين من المشرفين تحت أعباء العمل الثقيلة والمتوسطة والخفيفة، فإن هذا التصميم يتضمن وجهين هما المشرف وعبء العمل، وعندما يدرس أحد الباحثين الكتابة التعبيرية للأطفال، في كل مناسبة من أربع مناسبات، ويكتب كل طفل تعبير حول

موضوعين مختلفين، ويقوم ثلاثة مقدرين بتقييم جميع التعبيرات، فإن هذا التصميم يتضمن ثلاثة أوجه: المناسبة والموضوع والمقدر (Crocker & Aigina, 2008, 159).

وتسمى الأوجه أحيانًا بالأبعاد Dimensions في أدبيات التصميم التجريبي، وتشكل جميع الأوجه النطاق الشامل للملاحظات المقبولة، وتقع على الباحث مسؤولية تحديد عدد الأوجه التي ينبغي تضمينها في نطاق الملاحظات المقبولة، وتقديم الأدلة الداعمة حول أهمية كل وجه لاتساق ودقة إجراء القياس (Tavakol& Brennan, 2013, 221).

٤- الشروط Conditions:

يُعبر عن القياسات أو الملاحظات السلوكية بالشروط، فالاختبار التحصيلي الذي يحتوي على (٤٠) مفردة من نوع الاختيار من متعدد يحتوي على (٤٠) شرطًا لوجه المفردات، واختبار الأداء الذي يحتوي على (١٠) مهام يتم تقييمها من قبل اثنين من المقدرين له (١٠) شروط لوجه المهام وشرطين لوجه المقدرين (Sanders, 2005, 711).

وتعرف الشروط بالطرق المنهجية التي تختلف بها ظروف القياس، أو مستويات أوجه الخطأ، فعند تطبيق الاختبار في فترات زمنية مختلفة تكون الفترات هي الشروط كالفترة الزمنية الأولى والثانية، وعند وجود عدد من المصححين للمقياس يكون المصححين هم الشروط كالمصحح الأول والثاني (الحربي والحربي، ٢٠١٧، ٤٣٢؛ Clayson et al., 2021, 175).

٥- الأوجه العشوائية والأوجه الثابتة Random facet and fixed facet:

يمكن التعامل مع الأوجه على أنها ثابتة أو عشوائية، فإذا كان الوجه عشوائيًا فإن الباحث يعتبر أن الشروط هي عينة من عدد أكبر من الشروط ويريد التعميم على جميع هذه الشروط، وإذا كان الوجه ثابتًا فإن الباحث يريد التعميم على تلك الشروط التي تظهر في الدراسة فقط (Crocker & Aigina, 2008, 159 ؛Vispoel et al., 2017, 3).

فإذا كان الغرض هو التعميم بما يتجاوز الشروط الواردة في دراسة إمكانية التعميم المحددة، فيجب اعتبار الوجه عشوائيًا، أي أن جميع الملاحظات داخل الوجه قابلة للتبادل تمامًا وتمثل عينة عشوائية من نطاق الملاحظات المقبولة، وإذا كان الغرض هو التعميم فقط على تلك الظروف التي تمت ملاحظتها، فيجب اعتبار الوجه ثابتًا، أي أنه تم أخذ عينات من جميع الشروط المهمة لتعميم الوجه (Clayson et al., 2021, 176).

٦- التصميم المتقاطع والتصميم المتداخل Cross design and nested design:

تختلف التصميمات في عدد الأوجه (واحد أو أكثر) وطبيعتها (عشوائية أو ثابتة) وكيفية إدارة شروطها (متقاطعة أو متداخلة)، ويعني التصميم المتقاطع أنه يجب على الأفراد الاستجابة لجميع شروط أوجه مصادر الخطأ المحددة، ويعني التصميم المتداخل أنه يجب على الأفراد الاستجابة لمجموعة مختارة من شروط الأوجه (Sanders, 2005, 711).

وعندما يرغب الباحث في التحقق من مفردات الاختبار كمصدر واحد لخطأ القياس، فإنه يعد تصميم ذو وجه واحد له نوعين، فإذا تم تطبيق نفس عينة المفردات على مجموعة من الطلاب، فإن التصميم يكون متقاطعًا حيث يجيب جميع الطلاب (s) على جميع المفردات (i)، ويرمز له بالرمز i × s، وعندما يجيب كل طالب عن مجموعة مختلفة من المفردات، فسيكون التصميم متداخلًا ويرمز له بالرمز i:s، وترتبط هذه التصميمات بنطاق الملاحظات المقبولة، والغرض الرئيس منها جمع البيانات التي يمكن استخدامها لتقدير مكونات التباين، وتوفر مكونات التباين الخاصة بنطاق الملاحظات المقبولة تحليلًا لإجمالي التباين الملاحظ إلى الأجزاء المكونة له، وتعكس هذه الأجزاء المكونة الإسهام التفاضلي لمختلف الأوجه، أي أن مكون التباين الكبير نسبيًا المرتبط بوجه ما يشير إلى أن هذا الوجه له تأثير كبير نسبيًا على درجات الطلاب، فإذا تم تقدير مكون التباين الخاص بالفاحصين مثلًا على أنه مرتفع، فيتم استنتاج أنهم لم يتصرفوا بشكل متسق في تقييمهم للمتغير (Tavakol& Brennan, 2013, 221-222).

٧- دراسات إمكانية التعميم Generalizability studies (G) أو دراسة (G- (G) study) ودراسات القرار Decision studies أو دراسة (D) (D-study):

يتم في دراسة (G) تقدير مكونات تباين الدرجات المرتبطة بمصادر مختلفة باستخدام إجراءات تحليل التباين، بينما تأخذ دراسة (D) مكونات التباين المقدرة لتقييمها وتحسينها بين البدائل للقياس التالي، حيث تشكل الأساس لسلسلة من الدراسات لاتخاذ قرارات بشأن إجراءات القياس المستقبلية مثل: عدد المفردات التي يجب تضمينها في الاختبار أو عدد المقدرين الذين يجب عليهم تقييم الاستجابات (Sanders, 2005, 711-712 ؛ Li et al., 2015, 1325).

ويهتم الباحث في دراسة (G) بمدى تعميم عينة من القياسات على نطاق القياسات الأكثر شمولًا، مثل: الدراسات المعنية باستقرار الاستجابات بمرور الوقت، أو تكافؤ الدرجات على صورتين أو أكثر، أو العلاقة بين درجات المقاييس الفرعية أو المفردات على المقياس،

ويتم في دراسة (D) جمع البيانات لاتخاذ القرار، حيث توفر بيانات لوصف المختبرين أو تحديد المستوى، أو مقارنة المجموعات في تجربة ما، أو التحقق من العلاقة بين متغيرين أو أكثر، ولا يمكن تصنيف الدراسة بناءً على تصميمها، ولكن هدف الباحث هو العامل الحاسم، فإذا طبق أحد الباحثين اختبارًا تحصيليًا على (٤٠٠٠) طفلًا تم اختيارهم عشوائيًا من المدارس العامة والخاصة في نفس المنطقة، فإذا كان الغرض تحديد ثبات الاختبار أو أنه يمكن الاعتماد عليه بنفس القدر للطلاب في المدرستين؛ فإنها تعتبر دراسة (G)، أما إذا أراد الباحث مقارنة متوسط مستويات التحصيل للمجموعتين واستخلاص استنتاجات حول الاختلافات المحتملة في مدى كفاية النظامين التعليميين؛ فتكون دراسة (D) (Crocker & Aigina, 2008, 158-159).

وعلى الرغم من أن نظرية إمكانية التعميم تشير إلى دراسات (G) و(D) على أنها دراسات، إلا أنها تمثل أيضًا مرحلتين من التحليل يمكن تطبيقهما على نفس مجموعة البيانات، دراسة/تحليل (G) المرحلة الأولى التي يتم خلالها تقدير مكونات التباين للأوجه المختلفة، ودراسة/تحليل (D) المرحلة الثانية التي يتم خلالها تطبيق مكونات التباين لتقدير ثبات الدرجة لغرض معين (Clayson et al., 2021, 175).

٨- القرارات النسبية والقرارات المطلقة Relative decisions and absolute decisions:

يرغب متخذ القرار في اتخاذ نوعين من القرارات بناءً على القياس السلوكي: نسبية (مرجعية المعيار Norm-referenced) تركز على ترتيب الأفراد، ومطلقة (مرجعية المحك أو المجال Criterion- or domain-referenced) تركز على مستوى الأداء بغض النظر عن الرتبة (Webb & Shavelson, 2005, 717).

وتتعلق القرارات النسبية باتساق الدرجات في تصنيف الأفراد وفقًا للاختلافات في شخصياتهم وسلوكياتهم ومعارفهم ومهاراتهم، وتركز على الاختلافات النسبية في السمة المقاسة، فقد يرغب الفرد في معرفة درجته في اختبار ما مقارنة بدرجات زملائه الذين أكملوا الاختبار، بينما تتعلق القرارات المطلقة بفهرسة المستوى الدقيق أو المطلق لبناء الأفراد في مجال معين بغض النظر عن أداء الآخرين، فقد يكون الفرد مهتمًا أكثر بمعرفة ما إذا كانت درجاته في اختبار تحديد المستوى تعكس إتقانًا كافيًا للمهارات الأساسية التي تؤهله لأخذ الدورة التدريبية المطلوبة ، فتُتخذ قرارات القياس بناءً على المستوى المطلق للدرجات بغض

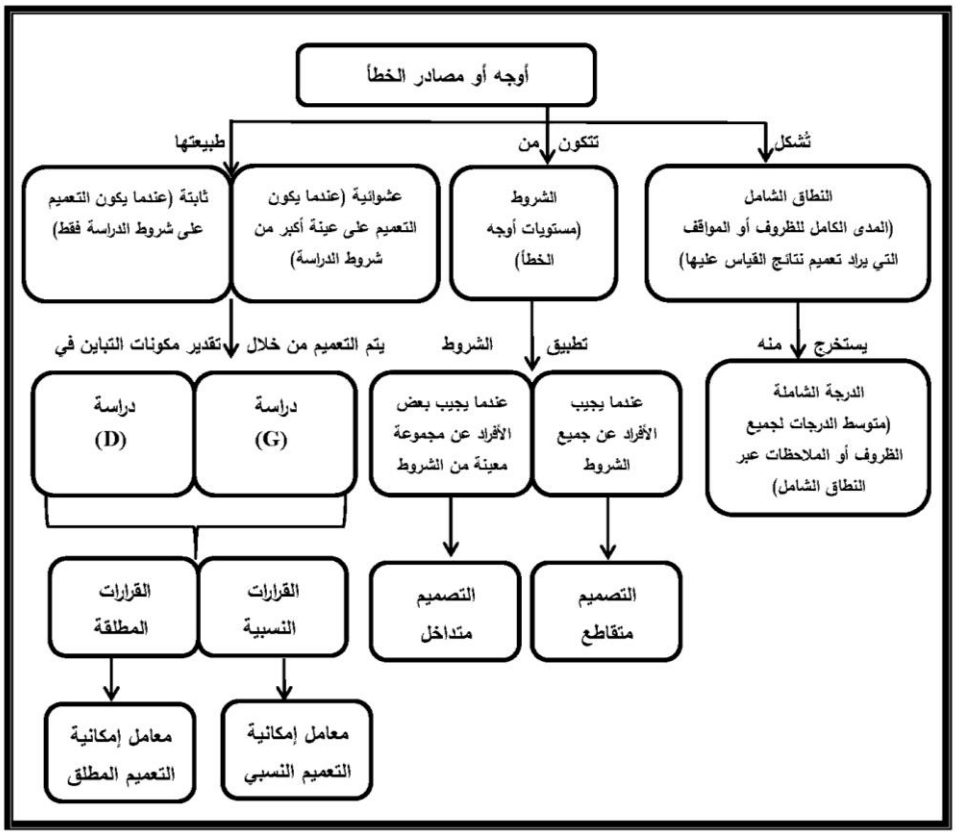
النظر عن الترتيب النسبي للمختبرين (Li et al., 2015, 1324 ؛ Vispoel et al., 2017, 3-4).

٩- معامل إمكانية التعميم النسبي ($E\rho^2$) Coefficients of generalizability ومعامل الاعتمادية أو معامل إمكانية التعميم المطلق أو معامل فاي (ϕ) (Phi-coefficient) Dependability coefficient:

تسمى المعاملات التي تتناول الثبات في سياق القرارات النسبية بمعاملات إمكانية التعميم النسبية، والقرارات المطلقة بمعاملات إمكانية التعميم المطلقة، ويعكس معامل ($E\rho^2$) الدرجة التي تحافظ بها موضوعات القياس على نظام الترتيب عبر الأوجه، بغض النظر عن التغييرات المحتملة في ارتفاع الدرجة الأولية، ويشبه معاملات الثبات في نظرية القياس الكلاسيكية، وتعكس معاملات (ϕ) كلًا من درجة الاتساق في نظام ترتيب موضوعات القياس والاتساق في ارتفاع الدرجة الأولية، وتكون مفيدة عندما تكون القيم الفعلية للدرجات التي تم الحصول عليها مهمة أو ذات معنى للفاحص، ومكونات التباين التي تعكس نظام ترتيب الأفراد هي التي تشارك فقط في تقديرات (G) للقرارات النسبية، بينما تشارك جميع مكونات التباين في تقديرات (G) للقرارات المطلقة (Yelboga & Tavsancil, 2010, 1850).

ويعرف معامل ($E\rho^2$) بمعامل الثبات الذي يقيم الاستقرار في الوضع النسبي أو ترتيب الأفراد، ويشبه معامل ألفا أو التجزئة النصفية المقدرة من خلال نظرية القياس الكلاسيكية، ويأخذ في اعتباره فقط خطأ القياس المرتبط بالموضوعات محل الاهتمام، أي أن الزيادات النسبية بين متوسط درجات المفردات أو الأفراد لا تقلل من الثبات، ويعرف معامل (ϕ) بمعامل الثبات الذي يقيم اتساق الدرجة أو المستوى المطلق للأداء، ويأخذ في اعتباره جميع المصادر المحتملة لخطأ القياس، أي أن الزيادات في التباين المرتبط بأي وجه تقلل من الثبات (Clayson et al., 2021, 175-176).

ويوضح الباحثان في شكل (١) العلاقة بين مفاهيم نظرية إمكانية التعميم:



شكل (١)
العلاقة بين مفاهيم نظرية إمكانية التعميم

رابعًا : تقدير الثبات باستخدام نظرية إمكانية التعميم :

يتضمن الثبات بشكل عام قياس الاتساق والتناقضات في الدرجات الملاحظة، وتعتبر نظرية إمكانية التعميم مناسبة بشكل خاص لمعالجة مثل هذه الأمور، حيث إنها تمكن الباحث من تحديد وتمييز مصادر التناقضات في الدرجات الملاحظة التي تنشأ عند إعادة تطبيق القياس في ظروف مختلفة (Brennan, 2011, 1).

ويُعرف معامل الثبات في نظرية إمكانية التعميم بالنسبة بين تباين الدرجات التي تعزى إلى الفرد والتباين الكلي الناتج عن كل من الفرد ومصادر أخطاء القياس المختلفة (Li et al., 2015, 1324).

وتُميز نظرية إمكانية التعميم بين نوعين من الثبات هما معامل إمكانية التعميم النسبي ($E\rho^2$) ومعامل إمكانية التعميم المطلق (ϕ)، حيث يركز ($E\rho^2$) على الاتساق الذي

يتم من خلاله تصنيف المختبرين أو المستجيبين بناءً على أدائهم في الاختبار أو المقياس من قبل اثنين أو أكثر من المقدرين المختلفين، بينما يهتم (ϕ) بالاتساق الذي حدث للدرجات حول نقطة قياس معينة (Dorathy et al., 2021, 174).

وبمجرد تقدير مكونات التباين، يقوم الباحثون بتقدير تباينات الخطأ ومعامل ($E\rho^2$) الذي يتضمن تباين الخطأ النسبي، ومعامل (ϕ) الذي يتضمن تباين الخطأ المطلق، ويكون تباين الخطأ النسبي (وبالتالي معامل ($E\rho^2$)) مناسبًا عندما يركز الاهتمام على ترتيب رتب الطلاب، ويتضمن جميع مكونات التباين التي تمثل التفاعلات بين الطلاب والأوجه، بينما يكون تباين الخطأ المطلق (وبالتالي معامل (ϕ)) مناسبًا عندما يركز الاهتمام على الدرجات الفعلية أو المطلقة للطلاب، ويشتمل على تباين الخطأ النسبي ومكونات التباين للأوجه نفسها (Tavakol& Brennan, 2013, 222).

وتتراوح قيم معاملي الثبات ($E\rho^2$) و(ϕ) بين (صفر إلى ١)، وتشير الدرجات الأعلى إلى ثبات أفضل، ويكون معامل ($E\rho^2$) منخفضًا عندما تكون ترتيبات الأفراد غير متسقة، ويكون معامل (ϕ) منخفضًا عندما تكون القياسات لنفس الأفراد غير متسقة، وبشكل عام، يجب أن تكون معاملات ($E\rho^2$) أعلى من معاملات (ϕ)، لأن تباين الخطأ المطلق يميل إلى أن يكون أكبر من تباين الخطأ النسبي، لأنه في القياس المطلق هناك مصادر محتملة أكثر لتباين الخطأ، ومن الناحية العملية، يمكن أن يكون الفرق بين تباين الخطأ المطلق والنسبي صغيرًا جدًا في غياب الفروق المنتظمة بين أوجه الخطأ أو التفاعلات بينها، وسوف يقترب تقدير معامل (ϕ) من تقدير معامل ($E\rho^2$) حيث إن تقديرات التباين للأوجه والتفاعلات بينها تقترب من الصفر (Cardinet et al., 2010, 6 ؛ Clayson et al., 2021, 176, 179).

ويمكن حساب معامل إمكانية التعميم النسبي (E_ρ^2) ومعامل إمكانية التعميم المطلق (ϕ) من المعادلتين التاليتين:

$$E\rho^2 = \frac{\sigma^2(\tau)}{\sigma^2(\tau) + \sigma^2(\delta)}$$

$$\phi = \frac{\sigma^2(\tau)}{\sigma^2(\tau) + \sigma^2(\Delta)}$$

حيث يتم تعريفهما بشكل عام من حيث تباين الدرجة الشاملة $\sigma^2(\tau)$ وتباين الخطأ، فيستخدم معامل (E_ρ^2) تباين الخطأ النسبي $\sigma^2(\delta)$ الذي يتجاهل مصادر التباين التي لا تؤثر

على ترتيب الأفراد، ويستخدم معامل (ϕ) تباين الخطأ المطلق ($\sigma^2(\Delta)$) الذي يتضمن جميع مصادر خطأ التباين التي تؤثر على القياسات المطلقة للدرجات، ويمثل كلا المعاملين نسبة تباين الدرجة الشاملة إلى تباين الدرجة الشاملة والخطأ (Clayson et al., 2021, 178؛ Brennan, 2001, 13).

ويمكن حساب معامل إمكانية التعميم النسبي والمطلق في حالة التصميم المتقاطع كليًا أحادي الوجه (مفحوص p× طريقة m) من المعادلتين الأولى والثانية، والتصميم المتقاطع كليًا ثنائي الوجه (مفحوصp× مهمةt× طريقةm) من المعادلتين الثالثة والرابعة:

$$E_p^2 = \frac{\sigma_p^2}{\sigma_p^2 + \frac{\sigma_{pm}^2}{n'_m}} \quad \dots\dots\dots\dots(1)$$

$$\phi = \frac{\sigma_p^2}{\sigma_p^2 + \frac{\sigma_m^2}{n'_m} + \frac{\sigma_{pm}^2}{n'_m}} \quad \dots\dots\dots(2)$$

$$E_p^2 = \frac{\sigma_p^2}{\sigma_p^2 + \frac{\sigma_{pt}^2}{n'_t} + \frac{\sigma_{pm}^2}{n'_m} + \frac{\sigma_{ptm}^2}{n'_t n'_m}} \quad \dots\dots\dots\dots(3)$$

$$\phi = \frac{\sigma_p^2}{\sigma_p^2 + \frac{\sigma_t^2 + \sigma_{pt}^2}{n'_t} + \frac{\sigma_m^2 + \sigma_{pm}^2}{n'_m} + \frac{\sigma_{tm}^2 + \sigma_{ptm}^2}{n'_t n'_m}} \quad \dots\dots(4)$$

حيث يشير (σ^2_p) إلى تباين الدرجة الشاملة، (σ^2_t) تباين المهام، (σ^2_m) تباين طرق التصحيح، (σ^2pt) تباين تفاعل (مفحوص– مهمة)، (σ^2pm) تباين تفاعل (مفحوص– طريقة)، (σ^2ptm) تباين تفاعل (مفحوص– مهمة– طريقة)، (n'_t) عدد معين من المهام، (n'_m) عدد معين من طرق التصحيح (العرايضة والشريفين، ٢٠٢٢أ، ٥٧٦).

ويتضح مما سبق أن حساب معامل إمكانية التعميم النسبي والمطلق يعتمد على تحديد أوجه الخطأ، وعدد مستويات كل وجه، ونوع التصميم متقاطع أم متداخل، وحساب تباينات الأوجه والتفاعلات بينها، وتباين الخطأ النسبي والمطلق، وتشير المعاملات المرتفعة إلى زيادة الثبات وإمكانية التعميم.

المحور الثاني: الحكمة الاختبارية:

أولًا: نشأة مفهوم الحكمة الاختبارية:

إن مفهوم الحكمة الاختبارية له تاريخ طويل نسبيًا في البحث التربوي, وتعود أصوله إلى عام (١٩٤٦) عندما أشار العالم **Cronbach** إلى أهميتها في حصول الطلاب على درجات أعلى من الدرجة المقدرة للمحتوى المعرفي الذي تم تحصيله، وهو ما يعني وجود مجموعة من القدرات المعرفية التي يوظفها الطلاب في الاختبار لتساعدهم في الحصول على درجات عالية بغض النظر عن معرفتهم بمحتوى المادة الدراسية (Otoum et al., 2015, 102).

وفي عام (١٩٥١) أكد **Thorndike** على أن الحكمة الاختبارية عامل مؤثر في ثبات الاختبار, ومصدر محتمل للتباين في درجاته, وسمة عامة ثابتة للمتقدم له, وأصبحت منذ ذلك الحين مصطلحًا يستخدم بشكل متكرر, وعلى الرغم من أن المهتمين بمجال الاختبارات اعترفوا بوجود الحكمة الاختبارية؛ إلا أنه لم يتم إجراء أي بحوث إمبريقية للتعرف عليها بدقة, ثم أوضح **Gibb (1964)** ندرة البحث حول هذا المفهوم باستثناء العمل المبكر لـ **Thorndike** وقدم تعريفًا عمليًا لها, وطور أداة لقياسها, ويُعد بحثه أول البحوث الكلاسيكية النظرية في هذا المجال, ثم قدم **Millman, Bishop, and Ebel (1965)** تصنيفًا شاملًا للحكمة الاختبارية بهدف أن يكون إطارًا للبحوث الإمبريقية المستقبلية, وفي غضون العقود التالية تم الاهتمام بالحكمة الاختبارية اهتمامًا كبيرًا في الأدب المهني, ولم يتم فحصها كمصدر من مصادر التباين الإضافية فقط؛ بل تم أيضا التحقق من مكوناتها, وارتباطها بمتغيرات أخرى, وقياسها, وقابليتها للتعليم والتطبيق (Haynes, 2011, 16؛ Mutua, 2012,1-2).

ونتيجة لارتباط الحكمة الاختبارية بالتحصيل الدراسي لدى الطلاب بما يعكس قدرتهم على إدارة وتنظيم موقف الاختبار واستخدام خصائصه وصيغه في الحصول على درجات مرتفعة؛ تعددت البحوث الارتباطية والتجريبية بعد ذلك حول هذا المتغير, واتجهت بعضها إلى دراسة العلاقات الارتباطية بين الحكمة الاختبارية كأحد السمات الشخصية للطلاب والتحصيل الدراسي, واتجه البعض الآخر إلى دراسة أثر تعلم مبادئها على التحصيل الدراسي, ولما كانت درجات الاختبار تتأثر بمستوى الحكمة الاختبارية لدى الطلاب؛ فإن على المعنيين بتطوير الاختبارات والمقاييس وتطبيقها ومتخذي القرار دراسة متغير الحكمة الاختبارية والتعرف على

مدى تأثيره على درجات الطلاب, كونه أحد مهددات صدق الاختبارات ومصدر من مصادر تباين درجاتهم, وقدرة معرفية يمتلكونها ويوظفونها في مواقف الاختبار, مما يجعل من الضروري توفير مقياس يتسم بالموضوعية للتعرف على درجة استخدامهم لها (حماد, ٢٠١٠, ٢٩٧– ٢٩٨؛ الطراونة, ٢٠٢٠, ١٤٩).

ثانيًا: مفهوم الحكمة الاختبارية:

تعددت تعريفات الباحثين عند تناولهم لمفهوم الحكمة الاختبارية تبعًا لوجهات نظر كل منهم في تحديد مكونات الحكمة الاختبارية من حيث كونها مجموعة من القدرات أو المهارات أو الاستراتيجيات, لذا يعرض الباحثان مفهوم الحكمة الاختبارية وفق التصنيفات التالية:

أ– تعريفات تناولت الحكمة الاختبارية كمجموعة من القدرات:

يُشير Houston (2005, 2) إلى الحكمة الاختبارية بأنها: قدرة الفرد على تحسين درجة اختباره من خلال التعرف على الإشارات الموجودة في مفردات وصيغ الاختبار أو الموقف الاختباري واستخدامها بشكل فعال, وهذه القدرة تكون مستقلة عن معرفته بمحتوى مادة الاختبار.

ويُعرفها Haynes (2011, 24) بأنها قدرة الفرد المختبر على استخدام الإشارات المتضمنة داخل محتوى أسئلة الاختبار لزيادة احتمال التوصل إلى الإجابات الصحيحة.

ويُوضحها Mutua (2012, 9) بأنها قدرة الفرد على الاستفادة من خصائص وصيغ الاختبار أو موقف إجراء الاختبار للحصول على درجة عالية مقارنة بالدرجة التي يمكن أن يحصل عليها نتيجة معلوماته العلمية فقط.

ويُعبر Olagoke (2012, 40) عنها بالقدرة على إظهار مهارات إجراء الاختبار في الاستفادة من خصائص وصيغ الاختبار وموقف إجراء الاختبار للحصول على درجة تتناسب مع القدرات المقاسة, فهي تشير إلى قدرة الفرد على الاستفادة من خصائص وشكل الاختبار للحصول على درجة عالية.

ويضع Roberson (2020, 4) تعريفًا للحكمة الاختبارية يتمثل في قدرة الفرد على التوصل إلى الإجابة الصحيحة لأسئلة الاختبار بشكل مستقل عن امتلاك المعرفة بالإجابة الصحيحة, وذلك من خلال التركيز على موقف الاختبار وخصائصه واستخدام مهارات مستقلة

عن محتوى الاختبار تمكنه من تمييز الإجابة الصحيحة لسؤال ما بسبب المعلومات الموجودة في الاختبار نفسه.

ويعرفها قنديل وموسى (٢٠٢٢, ١٣٨) بأنها قدرة معرفية مكتسبة لمجموعة من مهارات الاستفادة من خصائص موقف الاختبار, والتي يمارسها الفرد أثناء الاختبار لتحسين درجته عليه.

ب- تعريفات تناولت الحكمة الاختبارية كمجموعة من المهارات:

يشير طاحون (٢٠١٠, ٨٣) إلى الحكمة الاختبارية على أنها مجموعة من المهارات التي تسمح للطالب الذي يمتلكها بالاستفادة من خصائص الاختبار وبنيته وشكله والموقف الاختباري في الحصول على درجات مرتفعة في هذا الاختبار.

ويُحددها الزهراني (٢٠١٥, ٢٢٥) في مجموعة من المهارات والقدرات المعرفية التي توجد لدى الطالب, ويقوم بتوظيفها في مواقف الاختبار, للاستفادة من طريقة صياغة أسئلة الاختبار في الحصول على أعلى درجة ممكنة, وتتضمن مهارات يستخدمها الطالب قبل الإجابة وأثناء الإجابة وبعد الإجابة عن أسئلة الاختبار.

ويُعرفها محمد (٢٠١٨, ٧٦) بأنها مجموعة من المهارات يستخدمها بعض الطلاب بفاعلية أكثر من البعض الآخر, وتعكس مدى وعيهم أثناء مواقف تناول الاختبارات, وتظهر من خلال حصولهم على درجات أعلي مما يمكنهم الحصول عليه بناءً على مستوى معرفتهم بالمحتوى الذي تقيسه مفردات تلك الاختبارات.

وتضع أبو الحسن (٢٠٢٢, ٥٥) تعريفًا للحكمة الاختبارية يتمثل في مهارة الطالب في التعامل مع الموقف الاختباري والاستفادة من خصائص الاختبار وظروفه للحصول على درجات أفضل من خلال الاستعداد للاختبار, وإدارة وقته, والتعامل مع ورقة الأسئلة, والتعامل مع ورقة الإجابة, والمراجعة.

ج- تعريفات تناولت الحكمة الاختبارية كمجموعة من الاستراتيجيات:

يُشير Cohen (2013, 4) إلى أن الحكمة الاختبارية تتمثل في استراتيجيات استخدام المعرفة بصيغ الاختبار والمعلومات الخارجية الأخرى للحصول على استجابات تزداد احتمال كونها صحيحة دون إشراك المعرفة المطلوبة للقدرة على الأداء.

ويعرفها Waiprakhon and Jaturapitakkul (2018, 152) بأنها استراتيجيات استخدام المعرفة الخاصة بصيغ الاختبار وبعض المعلومات الخارجية الأخرى في الإجابة عن مفردات الاختبار دون القيام بالعمليات المعرفية المتوقعة المتعلقة بمحتوى الاختبار, كاستخدام عمليات الحذف بدلًا من الاختيار الأعمى, أو استخدام إشارات في مفردات الاختبار للإجابة عن مفردة معينة, أو اختيار بديل معين لاحتوائه على كلمات توحي بالإجابة الصحيحة

ويعبر DeVore et al., (2016, 1) عنها بمجموعة من الاستراتيجيات المعرفية التي يستخدمها الطالب, والتي تهدف إلى تحسين درجاته في الاختبار بغض النظر عن موضوعات مادة الاختبار.

وتحدد كردي (٢٠٢٤, ٤٤) مفهوم الحكمة الاختبارية في مجموعة من الاستراتيجيات التي يستخدمها الطلاب عند الإجابة عن أسئلة الاختبار, بحيث تساعدهم على الحصول على درجة أعلى من المتوقع منهم, وذلك من خلال الاعتماد على مؤشرات خارجية للوصول إلى الإجابة الصحيحة.

ويتضح من خلال عرض وجهات نظر الباحثين لمفهوم الحكمة الاختبارية ما يلي:

– تُمثل الحكمة الاختبارية مجموعة من القدرات أو المهارات أو الاستراتيجيات التي يستخدمها الفرد في المواقف الاختبارية التي قد يتعرض لها, بما يساعده في الحصول على درجات مرتفعة أعلى من درجته المتوقعة في حالة اعتماده فقط على المعرفة الجزئية بمحتوى الاختبار.

– تُشكل الحكمة الاختبارية مجموعة من المكونات التي تعكس وعي الفرد بالمتطلبات اللازمة للتعامل بفاعلية مع الموقف الاختباري, بما يضمن التهيئة الذاتية للاختبار والاستفادة من خصائص وصيغ الاختبار في اكتشاف الإجابة الصحيحة ومراجعة وتقييم أدائه الاختباري.

– تُعد مكونات الحكمة الاختبارية أحد العوامل الشخصية التي قد تساعد الفرد المختبر في الحصول على مستويات تحصيلية مرتفعة مقارنة بالأفراد الذين تنقصهم مثل هذه المكونات.

وفي ضوء ذلك؛ يُعرف الباحثان الحكمة الاختبارية بأنها "قدرة الطالب على التعامل بفاعلية مع الموقف الاختباري من خلال الاستعداد لأداء الاختبار بما يضمن تهيئته نفسيًا وعقليًا للحصول على درجات مرتفعة, والإدارة الجيدة لوقت الاختبار بما يساعده على تنظيم

وقت الإجابة مع الاحتفاظ بالدقة, والتعامل المنظم مع ورقتي الأسئلة والإجابة بما يمكنه من فهم متطلبات كل سؤال وتنظيم عرض الإجابة, واستنتاجه للمؤشرات المستترة داخل أسئلة الاختبار بما يساعده على التوصل إلى الإجابة الصحيحة, وتحديد أوجه القوة والقصور في إجابته سواء من خلال الإجابات التي قدمها لكل سؤال أو من خلال تقييمه لأدائه الاختباري بشكل عام بعد الانتهاء من الاختبار".

ثالثًا: أهمية الحكمة الاختبارية والعوامل المؤثرة فيها:

يُعد استخدام الطالب للحكمة الاختبارية أحد العوامل المساعدة له للحصول على درجات مرتفعة في الاختبار مقارنة بالطالب الذي لا يمتلك مثل هذه المهارات, وبالتالي فإن استخدام مهارات الحكمة الاختبارية لا يُعد بديلًا عن تحصيل الطالب لمادة الاختبار واستيعابه للبناء المعرفي الذي تتكون منه المادة من خلال عمليات الاستذكار والمراجعة.

حيث يميل الطلاب الذين يمتلكون مستوى مرتفع من الحكمة الاختبارية إلى أن يكونوا أكثر نجاحًا أثناء إجراء الاختبارات لأنهم يشعرون براحة أكبر وتنظيم أفضل, ويكونوا قادرين على إكمال الاختبارات في الوقت المخصص لإجراء الاختبار, ويمتلكون القدرة على إدراة وقت الاختبار بشكل مناسب, والتحقق من التعليمات بعناية, ونادرًا ما يتركون معلومات مهمة أثناء إجاباتهم عن أسئلة الاختبار(Peng et al., 2014, 368).

ويُمكن للطلاب الذين يتمتعون بحكمة اختبارية عالية أن يتفوقوا في الأداء الاختباري على زملائهم ذوي القدرات المتساوية الذين يفتقرون للحكمة الاختبارية, فهم يمتلكون القدرة على الإجابة الصحيحة عن الأسئلة التي تحتوى على إشارات أو دلائل خارجية, وبالتالي يحصلون على درجات عالية دون معرفة جيدة بالمحتوى الذي يتم اختباره, ويتمتعون بمهارات تنظيم الوقت بشكل واعي, والتأني في قراءة الأسئلة والتعرف على أجزائها, واستبعاد أكبر عدد ممكن من البدائل في أسئلة الاختيار من متعدد, والتخمين الذكي بين البدائل المتبقية (Tavakoli& Samiani, 2014, 1877).

كما أن إعداد الطلاب للاختبارات والتدريب عليها من خلال تنظيم الدراسة وإعطائهم صورة مسبقة عن طبيعة كل اختبار ونوع الأسئلة، وكيفية البدء بالإجابة، وكيفية التعامل مع الاختبارات والأسئلة، واتباع التعليمات وكيفية الإجابة عنها وإنهاء الاختبار، يساعدهم على بناء الثقة في أنفسهم وتقليل القلق والخوف من الاختبارات ويكون لديهم اتجاهات إيجابية تجاهها(Otoum et al., 2015, 102) .

ويُعد تعريف الطلاب بمهارات واستراتيجيات الحكمة الاختبارية وكيفية إجراء الاختبارات وسيلة فعالة لتحسين الأداء الاختباري, خاصة بين الطلاب الذين تنقصهم الخبرة الكافية في إجراء الاختبارات المقننة التي تجرى على نطاق واسع (Russo, 2019, 12).

فالطالب الذي يمتلك مستوى مرتفع من الحكمة الاختبارية يمكنه النجاح والحصول على درجات عالية في الاختبار, لأن هذا المستوى المرتفع يؤدي إلى انخفاض درجة القلق والتوتر لديه أثناء الاختبار, مما يمكنه من إدارة وقت الاختبار بكفاءة, والتعامل مع ورقة الإجابة بفاعلية, واستخدام التخمين الذكي, وتجنب الوقوع في الأخطاء, ومن ثم يمكنه النجاح والحصول على درجة عالية في الاختبار(حسب الله, ٢٠١٩, ٤٤٠).

ويتضح مما سبق؛ أن تمتع الفرد بالحكمة الاختبارية أحد العوامل المهمة لحصوله على درجات مرتفعة مقارنة بزملائه الذين يمتلكون نفس مستوى المعرفة الخاصة بمحتوى الاختبار, والذي قد يرجع إلى انخفاض مستوى القلق لديه أثناء المواقف الاختبارية, وقدرته على إدارة وقته بكفاءة وتقليل التوتر الناتج عن الشعور بعدم كفاية وقت الاختبار, كما أن امتلاكه لمستوى مرتفع من الحكمة الاختبارية يجعله يفكر في إجابات بعض الأسئلة التي قد يصعب عليه حلها بطريقة فعالة تمكنه من كشف الإشارات وتذكر بعض المعلومات التي قد توجد في محتوى أسئلة أخرى, وتجنبه الوقوع في بعض الأخطاء من خلال اتباع تعليمات الاختبار بعناية والتعامل السليم مع ورقتي الأسئلة والإجابة, كما أنها تساعده في اكتشاف أخطائه التي قد يقع فيها من خلال عمليات المراجعة وتقويم الأداء, فيستفيد من ممارساته الفعالة في الاختبارات التالية, ويحاول تلافي أوجه القصور التي وقع فيها.

وبالرجوع إلى كل من (Mutua, 2012, 14 ؛ وادي, ٢٠١٣, ٣٠٢؛ المطيري, ٢٠١٤, ٢٠– ٢١) أمكن تلخيص العوامل المؤثرة في الحكمة الاختبارية لدى الأفراد في النقاط التالية:

١– القدرة العقلية: حيث إن الحكمة الاختبارية تتطلب ممارسة السلوك الذكي, فكلما زادت قدرة الفرد العقلية ساعده ذلك على ممارسة الاستدلال والاستنباط وغيرها من العمليات المعرفية المرتبطة بالاختبار والموقف الاختباري.

٢– الخبرة السابقة للفرد: حيث إن خبرة الفرد السابقة في تعامله مع المواقف الاختبارية, وما يكتسبه من خبرات بالتدريب والممارسة وتعرضه لمواقف اختبارية متنوعة تجعله

يكتسب قدرات ومهارات متعددة تساعده على الأداء الجيد في الاختبارات, لكن يجب الإشارة إلى أن مجرد امتلاك الفرد لخبرات سابقة فقط لا تضمن له النجاح مستقبلًا في الاختبارات التي يتعرض لها.

٣- الفترة الزمنية لآخر موقف اختباري تعرض له الفرد: حيث أشارت نتائج البحوث إلى انخفاض مستوى الحكمة الاختيارية لدى البالغين الذين مضى على تعرضهم للمواقف الاختبارية فترات زمنية طويلة.

٤- السمات الشخصية للفرد: فما يتمتع به الفرد من سمات شخصية يؤثر بشكل مباشر على ممارسات الحكمة الاختبارية, فتمتع الفرد بالضبط النفسي والاتزان الانفعالي وقدرته على التخلص من مشاعر القلق والثقة بالنفس وارتفاع مستوى الدافعية يؤثر بشكل إيجابي على ممارسته للحكمة الاختبارية.

٥- نوع المفردات الاختبارية: حيث إن مفردات الأسئلة الموضوعية بصفة عامة والاختيار من متعدد بصفة خاصة تكون أكثر عرضة لممارسة العديد من مكونات الحكمة الاختبارية, نظرًا لأن مفردات الاختيار من متعدد تحتوى على العديد من المكونات (جذع, وبدائل, ومشتتات), كما أنها الأكثر صعوبة في إعدادها, لذا تكون أكثر عرضة لظهور العديد من أوجه القصور في الإعداد.

رابعًا: مكونات الحكمة الاختبارية:

يشير Houston (2005, 2) إلى أن مفهوم الحكمة الاختبارية يتكون من مكونات متعددة حسب وجهات نظر الباحثين تساعد الأفراد على التوصل للإجابة الصحيحة دون الاعتماد على المعرفة أو الخبرة في مجال المحتوى المراد قياسه, ومن هذه المكونات مهارات إدارة الوقت، ومحاولة الإجابة عن جميع الأسئلة, والفهم الكامل لتعليمات الاختبار, وخصائص مفردات الاختبار نفسها, وقدرة الفرد على تحديد الإجابة الصحيحة من خلال إشارات المفردات كاستخدام المعلومات من مفردات أخرى والارتباطات النحوية وطول البدائل الصحيحة والبدائل الصحيحة المحددة.

ويصنف أبو هاشم (٢٠٠٨, ٢٢٦) مكونات الحكمة الاختبارية إلى خمس مهارات أساسية حددها في ضوء مراجعته لبعض البحوث والمقاييس العربية والأجنبية, وتتضمن:

الاستعداد للاختبار, وإدارة وقت الاختبار, والتعامل مع ورقة الأسئلة, والتعامل مع ورقة الإجابة, والمراجعة.

ويرى طاحون (٢٠١٠, ٩١) أن التصنيف الأفضل لمكونات الحكمة الاختبارية هو الذي يتضمن كل الاستراتيجيات التي يستفيد منها الطالب في الحصول على درجات مرتفعة في الاختبار, وقد صنف استراتيجيات الحكمة الاختبارية إلى سبع استراتيجيات تتضمن: التهيؤ للاختبار, وتنظيم الوقت, وتجنب الخطأ, والتعامل مع ورقة الإجابة, والمراجعة, والتخمين, ومراعاة قصد مصمم الاختبار.

ويحدد Haynes (2011, 6) مكونات الحكمة الاختبارية في تخصيص وتنظيم الوقت, وترتيب أولويات الإجابة عن الأسئلة, والقراءة بعناية والتركيز على العناصر المهمة في تعليمات الاختبار, واستدعاء المعلومات عن طريق استخدام معينات الذاكرة, والتقدم بشكل منظم وسريع خلال الاختبار, وإجراء التخمين المدعم بالمعلومات (التخمين الذكي), ومراجعة الإجابة, والسيطرة على موقف الاختبار.

ويذكر Mutua (2012, 19-22) أن بحوث Millman et al., (1965) أسهمت في وضع إطار مفاهيمي لمكونات الحكمة الاختبارية, حيث تم تصنيفها إلى فئتين: الفئة الأولى, عناصر مستقلة عن واضع الاختبار أو الهدف من الاختبار وتكون قابلة للتطبيق في معظم حالات الاختبار, وتضم هذه الفئة استراتيجيات استخدام الوقت, واستراتيجية تجنب الخطأ, واستراتيجية التخمين, واستراتيجية التفكير الاستنتاجي, والفئة الثانية, عناصر معتمدة على واضع الاختبار أو الهدف من الاختبار, وتضم هذه الفئة استراتيجية مراعاة القصد واستراتيجية استخدام الإشارات أو الأدلة.

كما يشير Olagoke (2012, 99) إلى وجود تصنيف ثلاثي للحكمة الاختبارية حسب بحوث كل من (Wenden, 1991؛ Walter and Siebert, 1990) يتمثل في ثلاث فئات رئيسة: الفئة الأولى, الاستراتيجيات المستخدمة قبل الإجابة عن الاختبار وتضم (كتابة ملخصًا لكل سؤال أولًا, وقراءة جميع الأسئلة في البداية ثم البدء بالأسئلة السهلة, وقراءة التعليمات بعناية, وتنظيم وتوزيع الوقت على الأسئلة حسب طولها وصعوبتها, وتكوين صورة ذهنية لكيفية الإجابة, ووضع خط تحت الكلمات الرئيسية في الأسئلة), الفئة الثانية, الاستراتيجيات المستخدمة أثناء الإجابة عن الاختبار وتضم (الإجابة عن الأسئلة بالترتيب

الزمني, ومراجعة كل سؤال مباشرة بعد الانتهاء من إجابته, واستثمار كل الوقت المتاح للاختبار, والكتابة الفورية لكل ما يتبادر إلى الذهن, والإجابة عن جميع الأسئلة حتى التي لا يعرفها), الفئة الثالثة, الاستراتيجيات المستخدمة بعد الإجابة عن الاختبار وتضم (مراجعة الإجابات لتصحيح الأخطاء الإملائية والنحوية, وإعادة قراءة جميع الأسئلة للتأكد من فهمها بشكل صحيح, ومراجعة كل المحتوى واللغة, وتجنب الوقوع في تغييرات الدقائق الأخيرة).

ويرى Dodeen (2015, 111) أن مكونات الحكمة الاختبارية تتضمن (التنبؤ بأسئلة الاختبار, وقراءة تعليمات الاختبار بعناية, وتصفح جميع أسئلة الاختبار قبل البدء في الإجابة, وإدارة وقت الاختبار, والإجابة عن أسئلة الاختيار من متعدد, والتخمين, والإجابة عن جميع الأسئلة, ومراجعة الاختبار), كما حدد مجموعة من الإجراءات لتعليم الطلاب هذه الاستراتيجيات.

ويحدد Gbore and Osakuade (2016, 36) مكونات الحكمة الاختبارية في ثمان مهارات رئيسية حسب أهميتها تتمثل في: الاستخدام الفعال للوقت، وأساليب الاستجابة المناسبة، وتجنب الخطأ، واستبعاد البدائل الخاطئة, وإدارة الأسئلة الصعبة، والاستعداد لأداء الاختبار, واستخدام أفكار وقصد مصمم الاختبار، وإدارة القلق.

ويصنف الخولي (٢٠١٨, ٤١٦) مكونات الحكمة الاختبارية إلى مهارات في اختبارات الاختيار من متعدد ومهارات في الاختبارات المقالية, حيث تتضمن مهارات الاختيار من متعدد: الربط بين الجذع والبدائل, وإدراك أوجه التشابه والاختلاف بين البدائل, والتصنيف والمفاضلة بين البدائل, وتتضمن مهارات الاختبارات المقالية: الانتباه لعناصر الإجابة, ومحاولة فهم الأسئلة والتريث في فهمها, والتفكير قبل الإجابة, والانتباه إلى تعليمات الاختبار, وتنظيم الإجابة, ومراعاة اتجاهات المعلم, والالتزام بزمن الاختبار, وترتيب أولوية الإجابة وتسلسلها, ومراجعة الإجابة.

وبنظرة فاحصة لمكونات الحكمة الاختبارية التي تناولها الباحثان في الأطر النظرية والبحوث السابقة يتضح ما يلي:

– وجود تداخل في بعض مكونات الحكمة الاختبارية, والذي قد ينتج إما بسبب اختلاف مسمى المكون من باحث إلى آخر (كالاستعداد للاختبار والتهيؤ للاختبار, وتنظيم الوقت

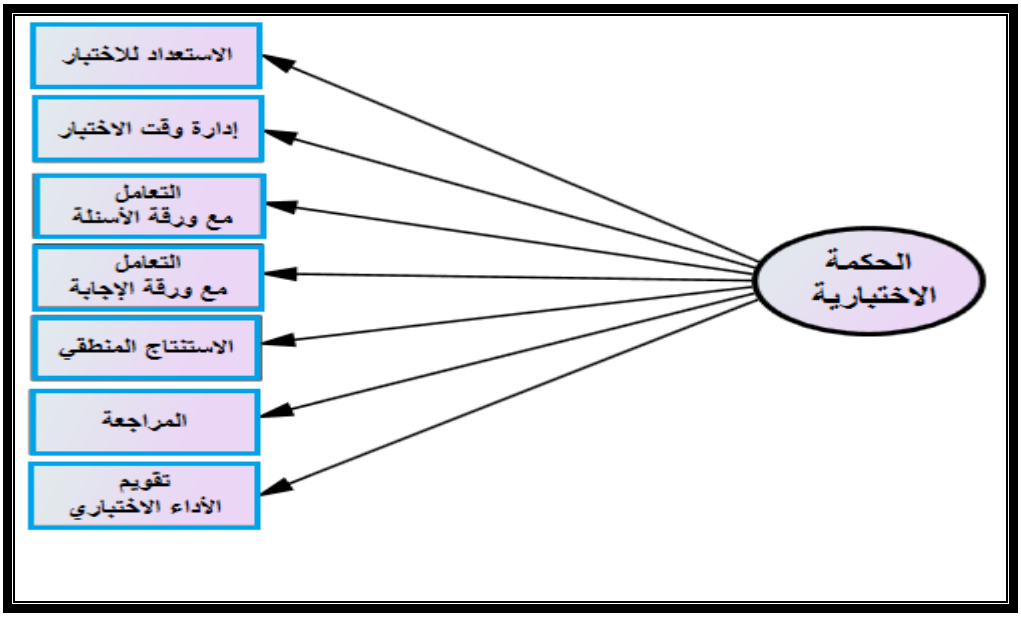
وإدارة الوقت, وتجنب الخطأ والتعامل مع ورقة الأسئلة), أو كون المكون يندرج تحت مكون آخر أكثر شمولًا (التخمين الذكي واستبعاد البدائل الخاطئة يندرجا تحت الاستنتاج المنطقي).

– اختلاف عدد مكونات الحكمة الاختبارية ومسمياتها باختلاف الهدف من البحث, فبعضها ركز على المكونات التي تستخدم بكثرة مع الاختبارات الموضوعية, وبعضها ركز على المكونات التي تستخدم بكثرة مع الاختبارات المقالية, وبعضها جمع بين المكونات التي تستخدم في كلا النوعين.

– انطلقت أغلب البحوث في وضع وتصنيف مكونات الحكمة الاختبارية من تصنيف (Millman et al., 1965), والذي يصنفها لمكونات مستقلة عن واضع الاختبار ومكونات معتمدة على واضع الاختبار.

– يمكن تصنيف مكونات الحكمة الاختبارية حسب الوقت الذي يستخدم فيه الفرد هذه المكونات إلى مكونات يستخدمها قبل الإجابة عن الاختبار, ومكونات يستخدمها أثناء الإجابة عن الاختبار, ومكونات يستخدمها بعد الإجابة عن الاختبار, ويندرج تحت كل منهم مجموعة من المكونات الفرعية.

وفي ضوء ما سبق؛ حدد الباحثان مكونات الحكمة الاختبارية في سبعة مكونات رئيسة, وقد حرصا على أن تكون مسميات هذه المكونات واضحة وغير متداخلة مع بعضها البعض, وارتباطها بأداء الاختبارات لطلاب المرحلة الثانوية, ومناسبتها لطبيعة الاختبارات التي يختبرون بها في الوقت الحالي, ويوضح شكل (٢) ذلك:



شكل (٢)
مكونات الحكمة الاختبارية

١- الاستعداد للاختبار Test readiness: قدرة الطالب على تدوين الملاحظات المهمة أثناء التعلم, ووضع جدول منظم لاستذكار ومراجعة جميع المقررات قبيل الاختبارات, واستخدامه للكلمات المفتاحية المساعدة على تذكر الحل أثناء الاختبار, وحل نماذج اختبارية سابقة للتدريب على الموقف الاختباري, والابتعاد عن حديث الزملاء الذين يظهرون قلقًا عاليًا من الاختبار, وتجنب مناقشة الزملاء حول الاختبار قبيل دخول قاعة الاختبار.

٢- إدارة وقت الاختبار Test time Management: قدرة الطالب على تنظيم وقت الاختبار والالتزام به, وتخصيص وقت محدد لإجابة كل سؤال ومراجعته بما يتناسب مع مكوناته, والإجابة عن أسئلة الاختبار في أسرع وقت مع المحافظة على الدقة والاتقان, وتجنب إضاعة وقت طويل للإجابة عن سؤال معين على حساب وقت الأسئلة الأخرى, وتجنب إضاعة الوقت في أمور غير مهمة أثناء الأداء على الاختبار, والتأكد من الوقت المحدد لكل سؤال من حين لآخر.

٣- التعامل مع ورقة الأسئلة Dealing with the test paper: قدرة الطالب على القراءة المتأنية لتعليمات الاختبار وفهمها جيدًا للالتزام بها أثناء الإجابة, وتكوين فكرة

عامة عن جميع الأسئلة بقراءتها جيدًا قبل البدء في الإجابة, وتحديد الكلمات والعناصر المهمة في السؤال لفهم المطلوب, ووضع خطوط على العناصر الأساسية المكونة للسؤال, وإعادة قراءة الأسئلة التي تبدو غامضة أكثر من مرة للتأكد من المطلوب.

٤- التعامل مع ورقة الإجابة **Dealing with the answer sheet**: قدرة الطالب على تركيز انتباهه أطول فترة ممكنة لإنتاج إجابات كاملة, وتعديل أفكاره حسب ما يطرأ على الذاكرة من معلومات أثناء الإجابة, وتنظيم وعرض إجاباته في تسلسل منطقي, وتقسيم إجابة كل سؤال في صورة عناصر أساسية, وإبراز العناصر الرئيسة عند الإجابة عن كل سؤال, والإجابة بخط واضح ومقروء, وتنظيم وتنسيق ورقة الإجابة, والالتزام بترقيم الإجابة عن الأسئلة بنفس ترقيمها في ورقة الأسئلة.

٥- الاستنتاج المنطقي **Logical reasoning**: قدرة الطالب على الاستفادة من طريقة صياغة مفردات الاختبار في إجابة بعضها البعض, واكتشافه للمؤشرات المستترة داخل صياغة بعض المفردات للوصول للإجابات الصحيحة, واستنتاجه للإجابة عن بعض الأسئلة من المعلومات التي قد تتوفر في أسئلة أخرى, فيستفيد من الكلمات المفتاحية الموجودة في نص السؤال, ويختار البدائل التي تتسق لغويًا مع نص السؤال, واستبعاد البدائل التي تبدو خاطئة أو غير المنطقية.

٦- المراجعة **Review**: قدرة الطالب على إصدار حكم على قيمة الإجابات التي أعطاها لكل سؤال عن طريق الفحص الواعي لإجابته عن أسئلة الاختبار, والتأكد من صحتها ومدى مناسبتها للمطلوب في كل سؤال, وكشف ما بها من أوجه قصور, وتحديد نوع الخطأ إن وجد وتعديله, والتأكد من عدم ترك أية أسئلة بدون إجابة, والتأني في مراجعة الإجابات دون الوقوع في خطأ تغيير الإجابة قبيل انتهاء وقت الاختبار.

٧- تقويم الأداء الاختباري **Evaluation of test performance**: قدرة الطالب على تحديد مواطن القوة والضعف في أدائه على الاختبار الذي مر به, وتحديد أي أساليب الاستذكار التي استفاد منها في الإجابة وأيها غير مناسبة وتعديلها, وتحديد أي أنماط الأسئلة التي أخفق فيها وأسبابها, وذلك بمراجعة الأسلوب المتبع في الاجابة عن الأسئلة الموضوعية, ومراجعة طريقته في الإجابة عن الأسئلة المقالية, وتحديد أسباب الإخفاق في مراجعة إجابة الأسئلة قبل تسليم ورقة الإجابة.

خامسًا: قياس الحكمة الاختبارية:

يُعد Gibb (1964) أول من قام بمحاولة بناء أداة صادقة لقياس الحكمة الاختبارية والتي عرفها بأنها "القدرة على التفاعل بشكل مفيد مع الإشارات الثانوية في الاختبار", ثم في العام التالي قدم Millman, at el., (1965) تعريفًا أكثر شمولًا للحكمة الاختبارية تمثل في "قدرة الفرد على استخدام خصائص وصيغ الاختبار والاستفادة من موقف إجراء الاختبار في الحصول على درجات مرتفعة", والذي أصبح التعريف الأكثر اقتباسًا في البحوث والدراسات التالية مع تغييرات طفيفة فقط في التعريف (Mutua, 2012, 18- 19).

واقترح Gibb (1964) بناءً على تعريفه للحكمة اختبارًا مكونًا من (٧٠) مفردة من نوع الاختيار من متعدد لفحص افتراضه بأن الحكمة الاختبارية تمثل القدرة على اكتشاف سبعة أنواع من الإشارات الثانوية التي قد تحتوى عليها مفردات الاختبار والتي تتمثل في: الارتباطات اللفظية بين الجذع والبدائل, والبدائل غير الصحيحة التي لا ترتبط بالجذع, وإدراك كلمات مثل الكل– لا شيء– أبدًا– دائمًا, والبدائل الصحيحة الأكثر دقة في المعنى, والبدائل الصحيحة الأكثر طولًا من البدائل الخاطئة, والإشارات النحوية واللغوية, ووجود مفردات تكشف عن إجابات مفردات أخرى (Haynes, 2011, 16).

وقدم Millman, at el., (1965) تصنيفًا واسع النطاق للحكمة الاختبارية يوضح بالتفصيل العناصر التي يمكن أن يستخدمها الأفراد في المواقف الاختبارية, حيث يضم هذا التصنيف فئتين من العناصر: الأولى تضم العناصر المستقلة عن بنية الاختبار كإدارة وقت الاختبار, وتجنب الأخطاء, والتفكير الاستنتاجي, والثانية تضم العناصر المعتمدة على بنية الاختبار كالاستفادة من الإشارات والمعلومات الموجودة في بعض الأسئلة في التوصل للإجابة الصحيحة (Roberson, 2020, 6).

وعلى إثر ذلك ظهرت وتنوعت التصنيفات التي هدفت إلى التعرف على الحكمة الاختبارية وتحديد مكوناتها, والتي صممت في ضوئها أدوات متنوعة من اختبارات ومقاييس هدفت إلى تحديد الحكمة الاختبارية في الواقع التعليمي, ومعرفة درجة توافرها وممارستها في العملية التعليمية, وأثرها على نتائج التحصيل وعلاقتها بمتغيرات أخرى (المطيري, ٢٠٢٤, ٢٦).

ومن خلال اطلاع الباحثان على الدراسات والبحوث السابقة التي استهدفت قياس الحكمة الاختبارية يمكن تقسيم طرق قياسها إلى:

- اختبارات أداء: حيث تكون مفرداتها من نوع الاختيار من متعدد في محتوى تحصيلي حر (خارج المحتوى الدراسي), ويتم بناء هذه المفردات بحيث تشتمل على ثغرات في بنائها, وتكون الاستجابة عليها ثنائية من نوع (صفر, ١), ويطلب من المستجيب التوصل إلى الاستجابة الصحيحة بناء على قدرته على التخمين الذكي والاستنتاج الاستنباطي والتوصل إلى الإشارات التي قد تحتوى عليها بعض المفردات, كما في دراسة كل من (Houston, 2005؛ الخولي, ٢٠١٨؛ Roberson,2020).
- مقاييس تقرير ذاتي: حيث تكون مفرداتها من نوع عبارات التقرير الذاتي حول ما يعتقده الفرد حول استخدامه لمكونات الحكمة الاختبارية, ويطلب من المستجيب اختيار البديل الذي ينطبق عليه حسب تدريج ليكرت, وقد اعتمدت معظم الدراسات والبحوث السابقة على هذه الطريقة مثل دراسة كل من (Mutua, 2012؛ Tavakoli& Samiani, 2014؛ الزهراني, ٢٠١٥؛ Gbore & Osakuade, 2016؛ كردي, ٢٠٢٤).

وقد اعتمد الباحثان في قياس الحكمة الاختبارية على الطريقة الثانية (عبارات التقرير الذاتي), نظرًا لأنها أكثر الطرق استخدامًا في البحوث والدراسات السابقة, كما أن صالحة لقياس مكونات الحكمة الاختبارية في الاختبارات المقالية والموضوعية وعدم اقتصارها على نوع الاختيار من متعدد فقط, بالإضافة إلى قدرتها على قياس مدى واسع من مكونات الحكمة الاختبارية لدى الأفراد.

دراسات وبحوث سابقة:

المحور الأول: دراسات وبحوث سابقة استخدمت نظرية إمكانية التعميم في تقدير ثبات بعض الاختبارات والمقاييس النفسية والتربوية:

أجرى Ogunka and Orluwene (2019) بحثًا هدف إلى استخدام نظرية إمكانية التعميم في تقدير مكونات التباين والتحقق من ثبات درجات اختبار في اللغة الإنجليزية، وبلغ عدد المشاركين في البحث (١٥٣) طالبًا من طلاب المرحلة الثانوية، وتمثلت أداة البحث في الأسئلة المقالية لاختبار في اللغة الإنجليزية الذي أجراه المجلس الوطني للامتحانات لشهر نوفمبر/ ديسمبر ٢٠١٨، وتم تحليل البيانات إحصائيًا باستخدام برنامج

SPSS من خلال النموذج الخطي العام عن طريق مكونات التباين بطريقة MINQUE للتصميم ثنائي الوجه المتقاطع كليًا (الفرد× المفردة× المقدر)، وتوصلت نتائج البحث إلى أن مكونات التباين بلغت بالنسبة لتفاعل الفرد× المفردة× المقدر (٥,٢٥٩) بنسبة (٤٧,٢٢١٪)، وتفاعل الفرد× المفردة (٤,٢٥٣) بنسبة (٣٨,١٧٪)، والمقدر (٠,٨٠) بنسبة (٧,١٨٪)، وتفاعل الفرد× المقدر (٠,٤١٥) بنسبة (٣,٧٢٪)، والفرد (٠,٣٩٨) بنسبة (٣,٥٧٪)، والمفردة (٠,٠٢٤) بنسبة (٠,٢١٪)، وتفاعل المفردة × المقدر (–٠,٠٠٨) بنسبة (–٠,٠٧١٪)، وبلغت قيم تباين الخطأ النسبي (٥,١١٤٦)، وتباين الخطأ المطلق (٠,٤٢٢٢)، ومعامل إمكانية التعميم النسبي (٠,٥١٩٨)، ومعامل إمكانية التعميم المطلق (٠,٩٢٩).

وقام طباع (٢٠٢٠) ببحث هدف إلى تقدير ثبات اختبار تقييم كفاءة الطلاب في الرياضيات باستخدام نظرية إمكانية التعميم، وبلغ عدد المشاركين في (٣٣١) تلميذًا من تلاميذ الصف الرابع الابتدائي، وتمثلت أداة البحث في اختبار تحصيلي مكون من تسع مهمات مركبة في مجال الأعداد والحساب في الرياضيات موزعة على ثلاث صيغ: ثلاث مهمات محكمة البناء، وثلاث مهمات غير محكمة البناء، وثلاث مهمات ذات معلومات مشوشة، واستخدم البحث برنامج EduG لتحليل البيانات للتصميم ثنائي الوجه المتقاطع كليًا (التلميذ× المهمة× المقدر)، وتوصلت نتائج البحث إلى وجود مصادر خطأ كبيرة راجعة إلى أثر تفاعل التلميذ مع المهمة والتأثير الرئيسي للمهمة، حيث بلغت مكونات التباين بالنسبة لتفاعل التلميذ× المهمة (٠,٨٥٥) بنسبة (٤٢,٩٪)، والتلاميذ (٠,٤٤٩) بنسبة (٢٢,٥٪)، والبواقي (٠,٣٣٣) بنسبة (١٦,٧٪)، والمهمة (٠,٢٧٧) بنسبة (١٣,٩٪)، والمقدر (٠,٠٣٢) بنسبة (١,٦٪)، وتفاعل التلميذ× المقدر (٠,٠٢٧) بنسبة (١,٤٪)، وتفاعل المهمة× المقدر (٠,٠١٨) بنسبة (٠,٩٪)، وبلغت قيم تباين الخطأ النسبي (٠,١١٦)، وتباين الخطأ المطلق (٠,١٥٨)، ومعامل إمكانية التعميم النسبي (٠,٧٩)، ومعامل إمكانية التعميم المطلق (٠,٧٤)، ويجب زيادة عدد المهمات وليس عدد المقدرين لضمان مستويات ثبات مقبولة.

وهدف بحث Orluwene and Memory (2020) إلى استخدام نظرية إمكانية التعميم للتحقق من ثبات درجات اختبار في الكيمياء، وبلغ عدد المشاركين في البحث (٧٤)

طالبًا من طلاب المرحلة الثانوية، وتمثلت أداة البحث في الأسئلة الموضوعية لاختبار الثانوية العامة في الكيمياء الذي تم تطبيقه في مايو/ يونيو٢٠١٨، وتم تحليل البيانات إحصائيًا باستخدام التحليل العاملي للتباين بواسطة إجراء VARCOMP عبر النموذج أحادي المتغير للتصميم ثنائي الوجه المتقاطع كليًا (الطالب× المفردة× المقدر)، وتوصلت نتائج البحث إلى أن مكونات التباين بلغت بالنسبة لتفاعل الفرد× المفردة (١٥,٢٧) بنسبة (٧٥,٠٤٪)، وتفاعل الفرد× المفردة× المقدر (٣,٢٢١) بنسبة (١٥,٨٣٪)، والمفردة (١,٢٨٥) بنسبة (٦,٣١٪)، والمقدر (٠,٠٣٩) بنسبة (٠,١٩٢٪)، وتفاعل الفرد× المقدر (٠,٠٣٧٨) بنسبة (٠,١٨٦٪)، وتفاعل المفردة × المقدر (٠,٠٢٠) بنسبة (٠,٠٩٨٪)، ومعامل إمكانية التعميم النسبي (٠,٧٩٥)، ومعامل إمكانية التعميم المطلق (٠,٨٠٦).

وأجرى Akindahunsi and Afolabi (2021) بحثًا هدف إلى استخدام نظرية إمكانية التعميم للتحقق من ثبات درجات اختبار في اللغة الإنجليزية، وبلغ عدد المشاركين في البحث (٣١١١٣٨) طالبًا من طلاب الشهادة الثانوية، وتمثلت أداة البحث في الأسئلة الموضوعية لاختبار في اللغة الإنجليزية الذي أجراه المجلس الوطني للامتحانات لشهر يونيو /يوليو ٢٠١٧، وتم تحليل البيانات إحصائيا باستخدام برنامج R للتصميم أحادي الوجه المتقاطع كليًا (الفرد× المفردة)، وتوصلت نتائج البحث إلى أن مكونات التباين بلغت بالنسبة للمفردات (٠,٠٧٤٧) بنسبة (٣١,٦٠٪)، وللأفراد (٠,٠١٤٢) بنسبة (٦٪)، وبلغ معامل إمكانية التعميم النسبي (٠,٩٠) ومعامل إمكانية التعميم المطلق (٠,٨٧) مما يدل على ثبات الاختبار بدرجة مرتفعة، كما أشارت نتائج البحث أيضًا إلى أن الانخفاض في عدد المفردات أدى إلى انخفاض في معاملات إمكانية التعميم النسبي والمطلق في دراسة (D).

وقام Dorathy et al. (2021) ببحث هدف إلى التحقق من ثبات مقياس التفكير الناقد باستخدام نظرية إمكانية التعميم، وبلغ عدد المشاركين في البحث (١٠٤١) طالبًا جامعيًا، وتمثلت أداة البحث في مقياس التفكير الناقد، وتم تحليل البيانات باستخدام تحليل التباين وحساب الأخطاء النسبية والمطلقة ومعامل إمكانية التعميم المطلق للتصميم ثنائي الوجه المتقاطع كليًا (الفرد× المفردة× فترات التطبيق)، وتوصلت نتائج البحث إلى أن مقياس التفكير الناقد له مكونات تباين جيدة، حيث بلغت بالنسبة للخطأ (٠,٢٠٢) بنسبة (٩٥,٣٪)، وفترات التطبيق (٠,٠٠٧) بنسبة (٣,٣٠٪)، والمفردات (٠,٠٠٣) بنسبة (١,٤٢٪)، وبلغ

تباين الأخطاء النسبية والمطلقة (٠,٠٠١٤، ٠,٠١١١) على الترتيب، ومعامل إمكانية التعميم المطلق (٠,٦٣).

وهدف بحث العرايضة والشريفين (٢٠٢٢ب) إلى الكشف عن إسهام مصادر التباين المتعددة في ثبات اختبار في الرياضيات باستخدام نظرية إمكانية التعميم، وذلك بتقدير حجم تباين الخطأ المفسر من الأبعاد (المهام، والمصححين، والفترات الزمنية) في التباين الكلي، والشروط التي يمكن بها الوصول لمستويات أفضل لثبات الاختبار، وبلغ عدد المشاركين في البحث (٢٤٣) مفحوصًا من الصف الخامس الأساسي بواقع (١٢) مدرسة حكومية، وتمثلت أداة البحث في اختبار في الرياضيات تكون من (١٢) مهمة مركبة في مجال الأعداد والعمليات عليها موزعة بالتساوي على أربع صيغ من المهام (التطبيق، والاستدلال، والانتقاء، والرأي)، واستخدم البحث برنامج EduG لتحليل البيانات للتصميم ثلاثي الوجه المتقاطع كليًا (المفحوص× المهمة× المصحح× الفترة)، وتوصلت نتائج البحث إلى أن مكونات التباين بلغت بالنسبة للمفحوص (٠,٤٩٥) بنسبة (٢٤,٩٪)، وتفاعل (مفحوص– مهمة– فترة) (٠,٤٤) بنسبة (٢٢,١٪)، وتفاعل (مفحوص– مهمة) (٠,٣٥٦) بنسبة (١٧,٩٪)، وتفاعل (مفحوص– مهمة– مصحح– فترة) (٠,٢٨) بنسبة (١٤,١٪)، والمهمة (٠,٢٥١) بنسبة (١٢,٦٪)، كما بلغ معامل إمكانية التعميم النسبي والمطلق (٠,٨٦–٠,٨٢) على الترتيب، وأن إدراج بعد الفترة الزمنية أسهم في تحسين معاملات التعميم، كما أن زيادة عدد المهام والفترات الزمنية يسهم في ارتفاع معاملات التعميم، وزيادة عدد المصححين لا يرفع من معاملات التعميم بشكل جوهري.

وأجرت الحجاجي (٢٠٢٣) بحثًا هدف إلى الكشف عن الأهمية النسبية لجميع مصادر التباين لتقدير ثبات درجات المهام الأدائية في مقرر العلوم، وتحديد عدد مستويات الأوجه وأثره على ثبات درجات الطلاب، وبلغ عدد المشاركين في البحث (١١٥) تلميذة من تلميذات الصف الثاني المتوسط بمدينة مكة المكرمة، وتمثلت أداة البحث في سلم تقدير كلي وسلم تقدير وصفي تحليلي لتقييم ثلاث مهام أدائية، واستخدم البحث برنامج EduG لتحليل البيانات وحساب التباينات لتصميم البحث ثلاثي الوجه المتقاطع كليًا المتضمّن (التلاميذ× المهام× طريقة التقدير× المقيمين)، وتوصلت نتائج البحث إلى أن أكبر نسبة للتباين تأثيرًا على معاملات إمكانية التعميم في الدرجات الشاملة ترجع إلى التلاميذ حيث بلغت (٢١,٣٨)

بنسبة (٦٠,٨٪)، يليها تباين تفاعل (التلاميذ × المهام) (٦,٩٤) بنسبة (١٩'٨٪)، وأن أكبر مكون لتباين الخطأ المنتظم في القياس النسبي والمطلق يرجع إلى تفاعل التلاميذ مع المهام بنسبة (٧٦,٨٪، ٦١,٤٪)، وبلغت قيم معاملات إمكانية التعميم النسبي والمطلق (٠,٨٨، ٠,٨٥)على الترتيب، وأن زيادة عدد شروط أوجه المهام الأدائية يسهم بشكل أكبر في رفع قيم معاملات إمكانية التعميم النسبية والمطلقة مقارنة بزيادة طرق التقدير وعدد المقيمين.

وقام كامل (٢٠٢٣) ببحث هدف إلى التحقق من ثبات درجات اختبار المواقف لتقييم أداء الطالب المعلم في ضوء نظرية إمكانية التعميم، ومعرفة مكونات التباين الأكثر تأثيرًا على ثبات درجات الاختبار، ومعرفة أفضل طرق تحسين إجراءات القياس وأكثرها فعالية في تصميمات القياس للحصول على درجات قابلة للتعميم، وبلغ عدد المشاركين في البحث (٢٧١) طالبًا وطالبة من طلاب الفرقة الرابعة بكلية التربية جامعة السويس متوسط أعمارهم الزمنية (٢١,٦) عامًا بانحراف معياري (٠,٧٣٤)، وتمثلت أداة البحث في اختبار المواقف لتقييم أداء الطالب المعلم، واستخدم البحث برنامج EduG لتحليل البيانات لتصميم البحث ثلاثي الوجه المتقاطع كليًا المتضمّن (الأفراد× البعد× المصحح× الفترة الزمنية)، وتوصلت نتائج البحث إلى أن مكونات التباين بلغت بالنسبة لتفاعل الطالب× البعد (٢,٨٦) بنسبة (٤٠,٧٪)، والطالب (٢,٦٨) بنسبة (٣٨,١٪)، وتفاعل طالب× البعد× الفترة (٠,٨١) بنسبة (١١,٦٪)، وبلغت قيم معاملات إمكانية التعميم النسبي والمطلق (٠,٨٤، ٠,٨٤) على الترتيب مما يدل على ثبات درجاته، وتؤدي زيادة عدد الأبعاد وعدد فترات التطبيق إلى زيادة معاملات إمكانية التعميم النسبية والمطلقة أفضل من زيادة عدد المصححين.

وهدف بحث Medvedeva et al. (2023) إلى تطبيق نظرية إمكانية التعميم للتحقق من ثبات اختبار التعرف وتقييم مصادر محددة لخطأ القياس المرتبط باللغات والكلمات المختلفة المستخدمة لتقييم أداء الذاكرة، وبلغ عدد المشاركين في البحث (٨٠) مشاركًا، بواقع (٢١) ذكرًا و(٥٩) أنثى، وتكونت أداة البحث من اختبار التعرف على الكلمات باللغتين الإنجليزية والروسية، واستخدم البحث برنامج EduG لتحليل البيانات للتصميم ثنائي الوجه المتداخل (الفرد: اللغة× الكلمات)، وتوصلت نتائج البحث إلى أن أكبر مكون للتباين يرجع إلى الأفراد: اللغة× الكلمات حيث بلغ (٠,١٩١) بنسبة (٩١,٢٪) من إجمالي التباين، وأن التباين بين الأفراد داخل المجموعات اللغوية ضئيل حيث بلغ (٠,٠١٦) بنسبة (٧,٥٪)، وأن

نتائج دراسة التعميم بالنسبة إلى اللغتين الإنجليزية والروسية متشابهة جدًا، مما يشير إلى وجود اختلافات ضئيلة بسبب اللغة، وبلغ معامل إمكانية التعميم النسبي والمطلق (٠,٩٢- ٠,٩٢) على الترتيب، وتؤدي زيادة عدد الكلمات إلى زيادة معاملات إمكانية التعميم النسبية والمطلقة.

وأجرى Sanz-Fernández et al. (2024) بحثًا هدف إلى تقدير مصادر التباين وتصميم القياس الأفضل والتحقق من ثبات بعض المقاييس النفسية كالتدفق والدافعية والثقة بالنفس والمهارات النفسية لدى الرياضيين، وبلغ عدد المشاركين في البحث (٤٧٠) مشاركًا، بواقع (٢٦٦) ذكرًا و(٢٠٤) أنثى، تراوحت أعمارهم الزمنية بين (١٤–٧٠) عامًا، ومتوسط أعمارهم (٣٢,١ عامًا بانحراف معياري ١٣,٥)، وتكونت أدوات البحث من مقاييس التدفق، والدافعية، والمهمة وتوجيه الأنا، والثقة بالنفس، والمقياس النفسي للأداء الرياضي، واستخدم البحث برنامجي SAGT, SAS لتحليل البيانات لتصميمات متقاطعة مختلفة تشتمل على الأوجه (المشاركون، وشكل الاستجابة، ونوع الرياضة، والمقاييس، والعامل، والنوع) وتوصلت نتائج البحث إلى تمتع المقاييس بدرجة عالية من الثبات والتعميم في جميع التصميمات المختبرة، وأن أكبر مكون للتباين يرجع إلى نوع الرياضة حيث بلغ (١٧٧٨,٦٦) بنسبة (٣٩,٠٢٪) من إجمالي التباين في التصميم الأول، وفي التصميم الثاني أيضًا فبلغ مكون التباين (١٧٧٥,٤٠) بنسبة (٣٨,٩٣٪)، بينما أكبر مكون للتباين في التصميم الثالث يرجع إلى الأفراد حيث يساوي (٢٠,٤٧) بنسبة (٦٥,٨٦٪).

ويلاحظ الباحثان من خلال عرض دراسات وبحوث المحور الأول أنها هدفت إلى التحقق من ثبات بعض الاختبارات والمقاييس النفسية والتربوية باستخدام نظرية إمكانية التعميم، وتراوحت أعداد المشاركين فيها بين (٧٤–٣١١١٣٨) مشاركًا من تلاميذ مرحلة التعليم الأساسي وطلاب المرحلتين الثانوية والجامعية والرياضيين، واعتمد معظمها على التصميمات المتقاطعة كليًا أحادية أو ثنائية أو ثلاثية الوجه عدا بحث Medvedeva et al. (2023) الذي استخدم التصميم المتداخل ثنائي الوجه، وتم اختيار الطلاب والمفردات والمقدرين أو المصححين والمهام وعدد الأبعاد أو العوامل وفترات التطبيق وطريقة التقدير وعدد الكلمات ونوع اللغة وشكل الاستجابة ونوع الرياضة والمقياس والنوع كأوجه في هذه الدراسات، وركزت معظم هذه الدراسات على استخدام أحد الاختبارات التحصيلية للتحقق من

ثباتها باستخدام النظرية عدا بحث Dorathy et al. (2021) الذي استخدم أحد المقاييس النفسية وهو مقياس التفكير الناقد وبحث Sanz-Fernández et al. (2024) الذي طبق مقاييس التدفق والدافعية والثقة بالنفس، واعتمدت هذه الدراسات على بعض البرامج الإحصائية لتحليل البيانات باستخدام هذه النظرية ومن هذه البرامج EduG، SPSS، R، SAGT, SAS، ويوجد اختلاف في نتائج هذه الدراسات حيث توصل بعضها إلى أن أكبر مكون للتباين يرجع إلى التفاعلات الثنائية أو الثلاثية بين الأوجه، أو إلى أحد الأوجه كالطلاب أو المفردات، أو إلى الخطأ، وكانت معظم قيم معاملات إمكانية التعميم النسبية والمطلقة مرتفعة، كما أشارت نتائجها إلى أن زيادة عدد أحد الأوجه يؤدي إلى ارتفاع قيم معاملات إمكانية التعميم النسبية والمطقة بطريقة أفضل من زيادة وجه آخر، مثل زيادة عدد المهام في بحث طباع (٢٠٢٠)، وعدد المهام وفترات التطبيق في بحث العرايضة والشريفين (٢٠٢٢ب)، وعدد شروط أوجه المهام الأدائية في بحث الحجاجي (٢٠٢٣)، وعدد الأبعاد وفترات التطبيق في بحث كامل (٢٠٢٣)، وعدد الكلمات في بحث Medvedeva et al. (2023).

المحور الثاني: دراسات وبحوث سابقة تناولت حساب الخصائص السيكومترية لمقاييس الحكمة الاختبارية:

هدف بحث الزهراني (٢٠١٥) إلى إعداد مقياس للحكمة الاختبارية لدى طلاب الجامعة والتحقق من خصائصه السيكومترية, والتعرف على درجة استخدام طلاب الجامعة للحكمة الاختبارية, والكشف عن الفروق فيها تبعًا لمتغيرات النوع والتخصص والفرقة الدراسية, واستخدم الباحث المنهج الوصفي, وبلغ عدد المشاركين في البحث (٥٩٣) طالبًا وطالبة من طلاب جامعة أم القرى, وتمثلت أداة البحث في مقياس الحكمة الاختبارية (إعداد الباحث) حيث تكون من (٥٠) عبارة, وتم استخدام معاملات ارتباط بيرسون لحساب الاتساق الداخلي للمقياس, وصدق المحكمين والتحليل العاملي الاستكشافي والصدق التلازمي لحساب الصدق, وطريقة ألفا كرونباخ وإعادة التطبيق لحساب ثبات المقياس, وتوصلت نتائج البحث إلى اتصاف مقياس الحكمة الاختبارية بخصائص سيكومترية مناسبة من حيث الصدق والثبات وفق النظرية الكلاسيكية تجعله صالحًا للاستخدام لقياس الحكمة الاختبارية لدى طلاب الجامعة.

وأجرى Opesemowo at el. (2018) بحثًا هدف إلى بناء وتطوير مقياس الحكمة الاختبارية لدى الطلاب, وتحديد صدق التكوين الفرضي للمقياس وثباته, واعتمد البحث على المنهج الوصفي, وبلغ عدد المشاركين في البحث (٧٥٠) طالبًا وطالبة من طلاب المرحلة الثانوية بنيجريا, وتمثلت أداة البحث في مقياس الحكمة الاختبارية والذي تكون من (٣٢) عبارة, وتم تحليل البيانات باستخدام تحليل الثبات باستخدام معامل ألفا كرونباخ ومعاملات سبيرمان- براون وجتمان للتجزئة النصفية, وتحليل الصدق باستخدام التحليل العاملي الاستكشافي, وتوصلت نتائج البحث إلى تحسن الخصائص السيكومترية للمقياس بعد حذف (٨) عبارات, حيث بلغت قيم معاملات ألفا كرونباخ وسبيرمان- براون وجتمان بعد الحذف (٠,٧٨- ٠,٧٧- ٠,٧٦) على الترتيب, كما توصلت نتائج البحث إلى تمتع المقياس بمعاملات صدق وثبات جيدة تجعله صالحًا لقياس الحكمة الاختبارية لدى طلاب المدارس الثانوية.

وقام Thoma and Koller (2018) ببحث هدف إلى إعداد اختبار للحكمة الاختبارية باللغة الألمانية, ودراسة كل من تأثير المعرفة الموضوعية والتدريب على الحكمة الاختبارية على نتائج الاختبار المطور, واستخدم الباحثان المنهج الوصفي والتجريبي, وبلغ عدد المشاركين في البحث (٢٥٢) طالبًا من طلاب الجامعة, وتمثلت أداة البحث في اختبار الحكمة الاختبارية مكون من (٢٤) مفردة من نوع الاختبار من متعدد, وتوصلت نتائج البحث إلى تمتع اختبار الحكمة الاختبارية الذي طوره الباحثان بخصائص سيكومترية جيدة (بالاعتماد على التحليلات العاملية للتحقق من الصدق, ومعامل ألفا كرونباخ للتحقق من الثبات) تتوافق مع الجودة السيكومترية للاختبارات الدولية الحالية, كما أظهرت نتائج البحث أن الأشخاص الذين لديهم معرفة موضوعية أو الذين تلقوا تدريبًا أجابوا عن أسئلة الاختبار بشكل صحيح أكثر من أقرانهم الذين لا يمتلكون هذه المعرفة أو لم يتعرضوا للتدريب.

وهدف بحث الطراونة (٢٠٢٠) إلى تطوير مقياس الحكمة الاختبارية ومطابقة فقراته مع نموذج أندريش وفق نظرية الاستجابة للمفردة, واستخدم الباحث المنهج الوصفي، وبلغ عدد المشاركين في البحث (٢٣٨٧) طالبًا وطالبة من طلاب الجامعة الأردنية, وتمثلت أداة البحث في مقياس حكمة الاختبار (إعداد الباحث) والذي تكون من (٥٩) عبارة, واتبع الباحث مجموعة من الإجراءات لإخراج المقياس في صورته النهائية تمثلت في التحقق من مطابقة

استجابات المشاركين في البحث لافتراضات نموذج أندريش باستخدام محك المطابقة الداخلية والخارجية, ومطابقة فقرات المقياس لافتراضات نظرية الاستجابة للمفردة (أحادية البعد, والاستقلال الموضعي, ومنحنى خصائص الفقرة, والتحرر من السرعة), وتوصلت نتائج البحث إلى أن قياس الحكمة الاختبارية يتصف بدلالات صدق (صدق محتوى, تحليل عاملي استكشافي, صدق تلازمي) وثبات (ألفا كرونباخ, معامل الفصل وفق نموذج أندريش) مقبولة.

وأجرى نصار (٢٠٢١) بحثًا هدف إلى تحديد البنية العاملية للحكمة الاختبارية لدى طلبة الصف الأول الثانوي في الأردن, والكشف عن مستوى الحكمة الاختبارية لديهم, واستخدم الباحث المنهج الوصفي التحليلي, وبلغ عدد المشاركين في البحث (١٠٠٠) طالبًا وطالبة من طلاب الصف الأول الثانوي بالأردن, وتمثلت أدوات البحث في مقياس الحكمة الاختبارية (إعداد الباحث) والذي تكون من (٨٨) عبارة, واختبار كاتل للذكاء, وتوصلت نتائج البحث إلى استخراج ثمانية عوامل للحكمة الاختبارية وفقًا لنتائج التحليل العاملي الاستكشافي, وقد أكدت نتائج التحليل العاملي التوكيدي هذه العوامل, حيث وجدت مؤشرات مطابقة جيدة للنموذج, ووجدت مؤشرات صدق وثبات مقبولة للمقياس, كما توصلت نتائج البحث إلى أن مستوى الحكمة الاختبارية كان أقل من المتوسط في الأبعاد والدرجة الكلية للمقياس لدى المشاركين في البحث.

وقامت أبو الحسن (٢٠٢٢) ببحث هدف إلى التعرف على البنية العاملية الاستكشافية والتوكيدية لمقياس الحكمة الاختبارية لدى طلبة جامعة حماة, واستخدمت الباحثة المنهج الوصفي, وبلغ عدد المشاركين في البحث (٤٧٢) طالبًا وطالبة من طلاب جامعة حماة, وتمثلت أداة البحث في مقياس الحكمة الاختبارية (إعداد أبو هاشم, ٢٠٠٨) والمكون من (٥٠ عبارة), وقد تم حذف (٦) مفردات أثناء التحقق من صدق وثبات مفردات المقياس, وتوصلت نتائج التحليل العاملي الاستكشافي من الدرجة الأولى عن وجود خمسة عوامل مقبولة تفسر نسبة (٦٣,٣٩١) من التباين الكلي, وبإجراء التحليل العاملي من الدرجة الثانية تبين وجود عامل واحد رئيس تتشبع عليه أبعاد المقياس وفسر نسبة (٥٦,٤٢٩) من التباين الكلي, كما أظهرت نتائج التحليل العاملي التوكيدي وجود تطابق بين النموذج المفترض والنموذج الواقعي, كما توصلت نتائج البحث إلى أن قياس الحكمة الاختبارية

بصورته السورية المعدلة والمكونة من (٤٤) عبارة يتصف بمؤشرات صدق (صدق المحتوى, والصدق البنيوي, والصدق التمييزي) وثبات (التجزئة النصفية, ومعامل ألفا كرونباخ) مرتفعة.

وأجرى قنديل وموسى (٢٠٢٢) بحثًا هدف إلى تصميم مقياس الحكمة الاختبارية لدى طلاب كلية التربية الرياضية مدينة جامعة السادات, واستخدم الباحثان المنهج الوصفي, وبلغ عدد المشاركين في البحث (٦٠٠) طالبًا وطالبة من طلاب كلية التربية الرياضية, وتمثلت أداة البحث في مقياس الحكمة الاختبارية (إعداد الباحثان) حيث تكون من (٧٧) عبارة, وتوصلت نتائج البحث إلى تمتع مقياس الحكمة الاختبارية الذي صممه الباحثان في ضوء ستة أبعاد بخصائص سيكومترية جيدة باستخدام صدق المحكمين والاتساق الداخلي وثبات ألفا كرونباخ, كما أظهرت نتائج البحث أن غالبية طلاب كلية التربية الرياضية جامعة مدينة السادات يمتلكون مستويات مرتفعة من الحكمة الاختبارية عند أداء الاختبارات.

ويستخلص الباحثان من خلال عرض دراسات وبحوث المحور الثاني أنها هدفت إلى التحقق من الخصائص السيكومترية لمقياس الحكمة الاختبارية في ضوء نظرية القياس الكلاسيكية أو نظرية الاستجابة للمفردة، ولم تحاول إحدى هذه الدراسات التحقق من هذه الخصائص في ضوء نظرية إمكانية التعميم، واعتمدت هذه الدراسات على المنهج الوصفي، وتراوحت أعداد المشاركين فيها بين (٢٥٢– ٢٣٨٧) مشاركًا من طلاب المرحلتين الثانوية والجامعية، وتراوحت أعداد عبارات مقياس الحكمة الاختبارية فيها بين (٣٢–٨٨) عبارة، وتوصلت نتائج هذه الدراسات إلى تمتع مقياس الحكمة الاختبارية بخصائص سيكومترية جيدة حيث تم التحقق فيها من الصدق باستخدام طريقة صدق المحكمين وصدق المحتوى والتحليل العاملي الاستكشافي والتوكيدي والصدق التلازمي والتمييزي والبنيوي، كما تم التحقق من الاتساق الداخلي، والثبات بطرق ألفا كرونباخ، وكيودر وريتشاردسون (٢٠) والتجزئة النصفية, وبالتالي اعتمدت هذه الدراسات على قياس مصدر واحد فقط من مصادر أخطاء القياس دون الاهتمام بالمصادر الأخرى أو التفاعل بين هذه المصادر.

منهج البحث:

استخدم الباحثان المنهج الوصفي التحليلي للكشف عن مصادر التباين الأكثر تأثيرًا على ثبات درجات مقياس الحكمة الاختبارية, وبحث أثر زيادة كل من مصدر من مصادر التباين (الأوجه) على ثبات مقياس الحكمة الاختبارية لدى طلاب المرحلة الثانوية.

المشاركون في البحث:

بلغ عدد المشاركين في البحث الحالي (٣٠٠) مشاركًا من طلاب وطالبات المرحلة الثانوية, وقد تراوحت أعمارهم الزمنية ما بين (١٤- ٢٠) عام, بمتوسط حسابي (١٦,٨٢٢), وانحراف معياري (١,٢٣٤), ويوضح جدول (٢) وصف المشاركين في البحث في ضوء متغيرات النوع والصف ونوع التعليم:

جدول (٢)
وصف المشاركين في البحث حسب متغيرات النوع والصف ونوع التعليم

المتغيرات	التصنيف	التكرار	النسبة المئوية	المجموع
النوع	ذكور	١١٥	٪٣٨,٣	٣٠٠
	إناث	١٨٥	٪٦١,٧	
الصف الدراسي	الأول الثانوي	٨٣	٪٢٧,٧	٣٠٠
	الثاني الثانوي	٧٦	٪٢٥,٣	
	الثالث الثانوي	١٤١	٪٤٧	
نوع التعليم	أزهري	١٦٥	٪٥٥	٣٠٠
	عام	١٣٥	٪٤٥	

أداة البحث: مقياس الحكمة الاختبارية (إعداد الباحثين):

قام الباحثان بإعداد مقياس الحكمة الاختبارية وفق الخطوات التالية:

١- تحديد الهدف من المقياس: هدف المقياس إلى قياس الحكمة الاختبارية لدى طلاب المرحلة الثانوية.

٢- الاطلاع على بعض الأدبيات والدراسات السابقة التي تناولت الحكمة الاختبارية بالبحث والدراسة, وحاولت التعرف على مفهومها وتحديد مكوناتها, مثل دراسة كل من (Houston, 2005؛ Hayati & Ghojogh, 2008؛ Cohen, 2013؛ Dodeen,2015؛ Gbore & Osakuade, 2016؛ Waiprakhon & Jaturapitakkul, 2018؛ Roberson, 2020؛ قنديل, موسى, ٢٠٢٢؛ كردي, ٢٠٢٤).

٣- الاطلاع على بعض المقاييس التي استهدفت قياس الحكمة الاختبارية: وذلك لتحديد مكونات الحكمة الاختبارية التي اشتملت عليها هذه المقاييس, مثل المقاييس

المستخدمة في دراسة كل من (Houston, 2005 ؛ أبو هاشم, ٢٠٠٨؛ حماد, ٢٠١٠؛ Mutua, 2012؛Tavakoli& Samiani, 2014؛ Peng et al., 2014؛ الزهراني, ٢٠١٥, Otoum et al., 2015؛ Gbore & Osakuade, 2016 ؛ القرشي, ٢٠١٧؛ الخولي, ٢٠١٨؛ الطراونة, ٢٠٢٠؛ Roberson, 2020 ؛ قنديل وموسى, ٢٠٢٢؛ سليمان, ٢٠٢٣, كردي, ٢٠٢٤), وقد تباينت مسميات مكونات الحكمة الاختبارية حسب وجهات نظر الباحثين في تناولهم للحكمة الاختبارية, إلا أن أغلب هذه المكونات تدور حول بعض المكونات التي يستخدمها الطالب قبل الإجابة عن الاختبار, وأثناء الاختبار, وبعد الاختبار, وبكل منهم مكونات فرعية, وقد قام الباحثان أثناء بناء المقياس باختيار مسميات المكونات الأكثر وضوحًا وشمولًا والتي لا تتداخل مع بعضها البعض, وقاما باختيار هذه المكونات كونها هي الأكثر تكرارًا في الدراسات والمقاييس السابقة التي اطلع عليها الباحثان حتى وان اختلفت مسمياتها بين هذه الدراسات, وارتباطها بالأداء الاختباري لطلاب المرحلة الثانوية, ومناسبتها للعمر الزمني للمشاركين وطبيعة الاختبارات التي تجرى عليهم في المرحلة الثانوية في الوقت الحالي, والتي تتنوع ما بين أسئلة موضوعية ومقالية.

٤- صياغة التعريفات الإجرائية لمفهوم الحكمة الاختبارية ومكوناتها المتمثلة في (الاستعداد للاختبار, وإدارة وقت الاختبار, والتعامل مع ورقة الأسئلة, والتعامل مع ورقة الإجابة, والاستنتاج المنطقي, والمراجعة, وتقويم الأداء الاختباري).

٥- صياغة مجموعة من العبارات تعبر عن مكونات الحكمة الاختبارية لدى طلاب المرحلة الثانوية: حيث قام الباحثان بصياغة (٧٠) عبارة تعبر عن مكونات الحكمة الاختبارية, منهم (٥٧) عبارة إيجابية و(١٣) عبارة سلبية, وقد خُصصت (١٠) عبارات لكل مكون تعبر عن التعريف الإجرائي له, وأمام كل عبارة خمسة بدائل للاستجابة (دائمًا, غالبًا, أحيانًا, نادرًا, أبدًا), حيث يقوم المستجيب باختيار استجابة واحدة فقط وهي التي يرى أنها تنطبق علية بدرجة كبيرة وذلك بوضع علامة (√) تحت الاختيار المناسب له, وقد راعى الباحثان عند صياغة عبارات المقياس أن تكون العبارات مناسبة لطلاب المرحلة الثانوية, وأن تكون مرتبطة ارتباطًا مباشرًا بمؤشرات الحكمة الاختبارية والمكون الذي وردت ضمنه, وأن تتضمن في محتواها مؤشرات للمواقف السلوكية التي يمكن أن يتعرض لها طلاب المرحلة

الثانوية قبل وأثناء وبعد أدائهم على الاختبارات الأكاديمية، ويوضح جدول (٣) الصورة الأولية للمقياس.

جدول (٣)
توزيع عبارات المقياس على مكونات الحكمة الاختبارية (الصورة الأولية)

م	المكونات	أرقام العبارات
١	الاستعداد للاختبار	١، ٨، ١٥، ٢٢، ٢٩، ٣٦، ٤٣، ٥٠، ٥٧، ٦٤
٢	إدارة وقت الاختبار	٢، ٩، ١٦، ٢٣، ٣٠، ٣٧، ٤٤، ٥١، ٥٨، ٦٥
٣	التعامل مع ورقة الأسئلة	٣، ١٠، ١٧، ٢٤، ٣١، ٣٨، ٤٥، ٥٢، ٥٩، ٦٦
٤	التعامل مع ورقة الإجابة	٤، ١١، ١٨، ٢٥، ٣٢، ٣٩، ٤٦، ٥٣، ٦٠، ٦٧
٥	الاستنتاج المنطقي	٥، ١٢، ١٩، ٢٦، ٣٣، ٤٠، ٤٧، ٥٤، ٦١، ٦٨
٦	المراجعة	٦، ١٣، ٢٠، ٢٧، ٣٤، ٤١، ٤٨، ٥٥، ٦٢، ٦٩
٧	تقويم الأداء الاختباري	٧، ١٤، ٢١، ٢٨، ٣٥، ٤٢، ٤٩، ٥٦، ٦٣، ٧٠

وتم تمييز العبارات السلبية في جدول (٣) بوضع خط تحت أرقام عباراتها.

٦- صياغة تعليمات المقياس: صاغ الباحثان تعليمات المقياس في نقاط محددة حسب طبيعة المقياس, حيث اشتملت تعليمات المقياس على عدد عباراته, وعدد بدائل الاستجابة, وكيفية الاستجابة عن العبارات, ومكان تدوين الاستجابة عن كل عبارة, كما قُدم للمستجيب مثالًا توضيحيًا في البداية لكيفية الاستجابة عن عبارات المقياس.

٧- إعداد طريقة تقدير الدرجات: يتم تقدير درجات المقياس وفق تدريج خماسي بحيث تُعطي الدرجة (٥) إذا اختار المستجيب (دائمًا), والدرجة (٤) إذا اختار (غالبًا), والدرجة (٣) إذا اختار (أحيانًا), والدرجة (٢) إذا اختار (نادرًا), والدرجة (١) إذا اختار (أبدًا), وذلك في كل عبارات المقياس باستثناء العبارات السلبية, وهي العبارات أرقام (١٣- ١٥- ١٧-١٨-٢٨-٢٩-٣٢- ٣٣- ٤١-٤٤- ٤٥- ٦٣-٦٥) حيث تصحح بطريقة عكسية, وبذلك تكون أعلى درجة يحصل عليها المستجيب في المقياس= ٧٠× ٥ = ٣٥٠ درجة , وأدني درجة= ٧٠× ١ = ٧٠ درجة.

٨- تحكيم المقياس من خلال عرضه على مجموعة من المحكمين[*]: عرض الباحثان المقياس في صورته الأولية على (٩) محكمين من أساتذة علم النفس التعليمي والإحصاء التربوي, وذلك لإعطاء تغذية راجعة حول وضوح ودقة تعليمات المقياس, وملاءمة عبارات المقياس للتعريف الإجرائي للمتغير, وملاءمة عبارات المقياس لطلاب المرحلة الثانوية, وتمثيل عبارات كل مكون من مكونات المتغير للتعريف الإجرائي له, ودقة صياغة عبارات المقياس وكفاية عددها, وإضافة أو حذف أو تعديل ما يرونه مناسبًا لعبارات المقياس, وقد بلغت نسب اتفاق المحكمين حول محاور التحكيم باستخدام معادلة هولستي Holsti (٠,٩٣٨), وهو معامل اتفاق مرتفع, كما تراوحت نسب اتفاق المحكمين على عبارات المقياس ما بين (٨٨,٨٨٪ : ١٠٠٪) وهي نسب اتفاق مرتفعة, وقد أخذ الباحثان بملاحظات وآراء السادة المحكمين على عبارات المقياس, وقاما بإجراء بعض التعديلات التي اقترحوها, ويوضح جدول (٤) بعض عبارات المقياس قبل التعديل وبعده في ضوء آراء السادة المحكمين:

جدول (٤)
بعض عبارات مقياس الحكمة الاختبارية قبل التعديل وبعده في ضوء آراء السادة المحكمين

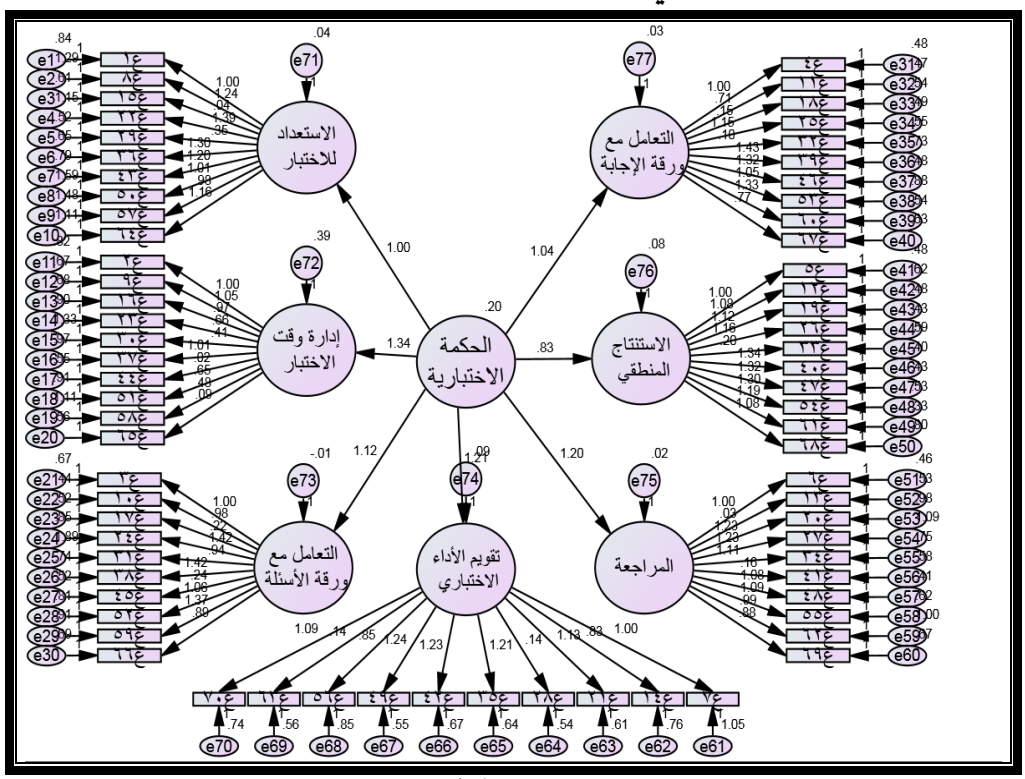
العبارة قبل التحكيم	العبارة بعد التحكيم
أقوم بحل نماذج اختبارات الأعوام السابقة للتدريب على الموقف الاختباري.	أحل نماذج اختبارات الأعوام السابقة للتدريب على الموقف الاختباري.
أمتنع عن النقاش حول موضوعات المادة قبل دخول الاختبار.	أمتنع عن النقاش حول موضوعات المادة مع زملائي قبل دخول الاختبار.
أحرص على تخصيص جزء من وقت الاختبار للمراجعة.	أخصص جزءًا من وقت الاختبار للمراجعة.
أحاول الاستفادة من الكلمات المفتاحية الموجودة في أسئلة الاختبار.	أستفيد من الكلمات المفتاحية الموجودة في أسئلة الاختبار.
أستبعد البدائل التي لا تتوافق لغويًا مع جذع السؤال.	أستبعد البدائل التي لا تتوافق لغويًا مع نص السؤال.
أحاول استخدام المنطق استنتاج إجابات الأسئلة التي لا أعرفها.	أستخدم المنطق في استنتاج إجابات الأسئلة التي لا أعرفها.

[*] يتوجه الباحثان بالشكر والتقدير إلى السادة الأساتذة أعضاء هيئة التدريس بقسم علم النفس التعليمي والإحصاء التربوي (أ.د/ حسين عبد العزيز الدريني، أ.د/ سيف الدين يوسف عبدون، أ.د/ أحمد مهدي مصطفي، أ.د/ محمد مصطفى مصطفى الديب، أ.د/ إبراهيم سيد أحمد عبد الواحد، أ.د/ عادل عبد المعطي الأبيض) بكلية التربية بنين جامعة الأزهر بالقاهرة، و(أ.د/ سهير أنور محفوظ، أ.د/ محمد إسماعيل حميدة) بقسم علم النفس التربوي بكلية التربية جامعة عين شمس، وأ.د حسني زكريا النجار بقسم علم النفس التربوي بكلية التربية جامعة كفر الشيخ لتحكيمهم أداة البحث.

٩- تطبيق المقياس على عينة البحث والمكونة من (٣٠٠) طالبًا وطالبة من طلاب المرحلة الثانوية, وذلك لحساب الخصائص السيكومترية للمقياس وذلك على النحو التالي:

أ- صدق المقياس:

نظرًا لأن الباحثان قاما ببناء المقياس بعد الاطلاع على الأطر النظرية والدراسات والبحوث السابقة التي استهدفت قياس الحكمة الاختبارية وتحديد مكوناتها, واتفاق معظم هذه الدراسات والبحوث حول بنية هذا المفهوم, وبالتالي فإن البنية الأساسية للمقياس محددة مسبقًا في ضوء ما اتفقت عليه هذه الدراسات والبحوث, لذا تم إجراء التحليل العاملي التوكيدي باستخدام برنامج Amos V.23 للتأكد من صحة نموذج البنية العاملية لمقياس الحكمة الاختبارية لطلاب المرحلة الثانوية، حيث تم حساب معاملات الانحدار المعيارية، ومعاملات الانحدار اللامعيارية، والخطأ المعياري، والقيمة الحرجة التي تعادل قيمة "ت" ودلالتها, كما في شكل (٣) وجدولي (٥)، (٦) وذلك بعد تطبيق المقياس على (٣٠٠) طالبًا وطالبة من طلاب المرحلة الثانوية:



شكل (٣)

مسار التحليل العاملي التوكيدي لمقياس الحكمة الاختبارية لطلاب المرحلة الثانوية

جدول (٥)
معاملات الانحدار المعيارية واللامعيارية والقيمة الحرجة ودلالتها للتحليل العاملي التوكيدي لمكونات مقياس الحكمة الاختبارية (ن=٣٠٠)

المتغير	المكون	معاملات الانحدار المعيارية	معاملات الانحدار اللامعيارية	الخطأ المعياري	القيمة الحرجة	مستوى الدلالة
الحكمة الاختبارية	الاستعداد للاختبار	٠,٩١٩	١			
	إدارة وقت الاختبار	٠,٦٩١	١,٣٣٧	٠,١٤٦	٩,١٦٧	***
	التعامل مع ورقة الأسئلة	٠,٩٦٦	١,١١٩	٠,١٢٣	٩,٠٩٠	***
	التعامل مع ورقة الإجابة	٠,٩٣٩	١,٠٣٩	٠,١١٠	٩,٤٢٧	***
	الاستنتاج المنطقي	٠,٨٠٤	٠,٨٣٤	٠,٠٩٤	٨,٨٥٣	***
	المراجعة	٠,٩٦١	١,١٩٥	٠,١٢١	٩,٨٨٣	***
	تقويم الأداء الاختباري	٠,٨٨١	١,٢١٣	٠,١٣٩	٨,٧١٤	***

يتضح من جدول (٥) أن معاملات الانحدار اللامعيارية لمكونات مقياس الحكمة الاختبارية جاءت قيمها الحرجة دالة عند مستوى دلالة (٠,٠٠١)، مما يدل على صحة نموذج البنية العاملية لمكونات مقياس الحكمة الاختبارية.

جدول (٦)
معاملات الانحدار المعيارية واللامعيارية والقيمة الحرجة ودلالتها للتحليل العاملي التوكيدي لعبارات مقياس الحكمة الاختبارية (ن=٣٠٠)

المكون	العبارة	معاملات الانحدار المعيارية	معاملات الانحدار اللامعيارية	الخطأ المعياري	القيمة الحرجة	مستوى الدلالة
الاستعداد للاختبار	١	٠,٤٦٧	١			
	٨	٠,٤٧٠	١,٢٣٥	٠,١٤٥	٨,٥٠٣	***
	١٥	٠,٠٢٥	٠,٠٣٩	٠,٠٧٠	٠,٥٦١	٠,٥٧٥
	٢٢	٠,٥٣٥	١,٣٩٠	٠,١٥١	٩,١٩٤	***
	٢٩	٠,٢٣١	٠,٣٤٩	٠,٠٧١	٤,٩١٩	***
	٣٦	٠,٦٢١	١,٣٠٢	٠,١٣١	٩,٩٤٩	***
	٤٣	٠,٥٥٠	١,١٩٩	٠,١٢٨	٩,٣٤٤	***
	٥٠	٠,٣٦٥	١,٠١٠	٠,١٤١	٧,١٤٣	***
	٥٧	٠,٣٦٦	٠,٩٧٩	٠,١٣٧	٧,١٦٣	***
	٦٤	٠,٤٧٤	١,١٦٣	٠,١٣٦	٨,٥٥٥	***
إدارة وقت	٢	٠,٦٩٢	١			

المكون	العبارة	معاملات الانحدار المعيارية	معاملات الانحدار اللامعيارية	الخطأ المعياري	القيمة الحرجة	مستوى الدلالة
الاختبار	٩	٠,٧٤٥	١,٠٥٣	٠,٠٦٧	١٥,٧٥٩	***
	١٦	٠,٧١٢	٠,٩٦٦	٠,٠٦٤	١٥,١٨١	***
	٢٣	٠,٥٢٠	٠,٦٦٣	٠,٠٥٨	١١,٤٠٧	***
	٣٠	٠,٢٩٥	٠,٤١٠	٠,٠٦٢	٦,٦٠٨	***
	٣٧	٠,٦٦٤	١,٠٠٦	٠,٠٧٠	١٤,٢٨٠	***
	٤٤	٠,٠٢٩	٠,٠٢٤	٠,٠٣٨	٠,٦٤٨	٠,٥١٧
	٥١	٠,٥١٢	٠,٦٥٤	٠,٠٥٨	١١,٢٤٧	***
	٥٨	٠,٣٦٧	٠,٤٧٧	٠,٠٥٨	٨,١٦٤	***
	٦٥	٠,٢٠٢	٠,٠٨٨	٠,٠٣٣	٢,٦٦٤	**
التعامل مع ورقة الأسئلة	٣	٠,٥١٨	١			
	١٠	٠,٥٩٢	٠,٩٨٥	٠,٠٨٨	١١,١٣٤	***
	١٧	٠,٢٤٨	٠,٢١٨	٠,٠٦٣	٣,٤٦٢	***
	٢٤	٠,٦٠٨	١,٤٢٤	٠,١٢٦	١١,٣١٥	***
	٣١	٠,٣٢٣	٠,٩٤٤	٠,١٣٣	٧,٠٨٣	***
	٣٨	٠,٦٣٣	١,٤٢٠	٠,١٢٢	١١,٥٩١	***
	٤٥	٠,٢٦٢	٠,٢٣٩	٠,٠٦٣	٣,٧٨٤	***
	٥٢	٠,٥٠٦	١,٠٦٤	٠,١٠٦	١٠,٠٣٥	***
	٥٩	٠,٥٧٩	١,٣٦٥	٠,١٢٤	١٠,٩٨١	***
	٦٦	٠,٤٧١	٠,٨٩٤	٠,٠٩٤	٩,٥٤٩	***
التعامل مع ورقة الإجابة	٤	٠,٥٨٤	١			
	١١	٠,٤٥٨	٠,٧١٣	٠,٠٧٤	٩,٦٧٩	***
	١٨	٠,٢٠٣	٠,١٥٢	٠,٠٥٨	٢,٦٢٢	**
	٢٥	٠,٦٣٣	١,١٥٣	٠,٠٩٣	١٢,٤٣٤	***
	٣٢	٠,٠٦٥	٠,٠٩٨	٠,٠٦٥	١,٥١٣	٠,١٣٠
	٣٩	٠,٦٣٩	١,٤٣٠	٠,١١٤	١٢,٥٢٤	***
	٤٦	٠,٦٨٧	١,٣١٨	٠,١٠٠	١٣,١٦٨	***
	٥٣	٠,٤٨٧	١,٠٥٠	٠,١٠٣	١٠,١٧٨	***
	٦٠	٠,٦٧١	١,٣٣٤	٠,١٠٣	١٢,٩٦١	***
	٦٧	٠,٤٣٤	٠,٧٧١	٠,٠٨٣	٩,٢٤٦	***
الاستنتاج المنطقي	٥	٠,٥٥٦	١			
	١٢	٠,٥٣٨	١,٠٧٨	٠,١٠٢	١٠,٥٣٧	***
	١٩	٠,٦٠١	١,١٢٢	٠,٠٩٨	١١,٣٨٦	***
	٢٦	٠,٦٣٣	١,١٥٥	٠,٠٩٨	١١,٧٧٤	***
	٣٣	٠,٢٢٣	٠,٢٠٣	٠,٠٧٣	٢,٧٩٦	**
	٤٠	٠,٧٠٠	١,٣٣٦	٠,١٠٦	١٢,٥٥١	***
	٤٧	٠,٦٨٦	١,٣٢٢	٠,١٠٧	١٢,٣٩٣	***
	٥٤	٠,٦٤٢	١,٣٠١	٠,١١٠	١١,٨٨٣	***
	٦١	٠,٦٩٥	١,١٨٦	٠,٠٩٥	١٢,٤٩١	***
	٦٨	٠,٤٨٨	١,٠٧٦	٠,١١٠	٩,٧٩٦	***
المراجعة	٦	٠,٦٣٧	١			

المكون	العبارة	معاملات الانحدار المعيارية	معاملات الانحدار اللامعيارية	الخطأ المعياري	القيمة الحرجة	مستوى الدلالة
	١٣	٠,٠٢٢	٠,٢٨	٠,٠٥٦	٠,٥٠٦	٠,٦١٣
	٢٠	٠,٥٦٩	١,٢٢٨	٠,١٠٠	١٢,٣٠٢	***
	٢٧	٠,٥٥٢	١,٢٣٤	٠,١٠٣	١١,٩٧٩	***
	٣٤	٠,٥٨١	١,١٠٦	٠,٠٨٨	١٢,٥٣٦	***
	٤١	٠,٢١٨	٠,١٦٣	٠,٠٥٩	٢,٧٥٦	**
	٤٨	٠,٦٨٤	١,٠٧٩	٠,٠٧٥	١٤,٣٤٨	***
	٥٥	٠,٦١٣	١,٠٨٩	٠,٠٨٣	١٣,١٠٤	***
	٦٢	٠,٤٨٢	٠,٩٨٧	٠,٠٩٣	١٠,٦٤٦	***
	٦٩	٠,٥١٤	٠,٨٨١	٠,٠٧٨	١١,٢٦٦	***
تقويم الأداء الاختباري	٧	٠,٥١٦	١			
	١٤	٠,٥٠٧	٠,٨٣٠	٠,٠٨٦	٩,٦٥١	***
	٢١	٠,٦٦٤	١,١٢٦	٠,٠٩٩	١١,٤٠٣	***
	٢٨	٠,١١٧	٠,١٤٠	٠,٠٥٣	٢,٦٥٦	**
	٣٥	٠,٦٨٥	١,٢١٣	٠,١٠٥	١١,٥٩٤	***
	٤٢	٠,٦٨٠	١,٢٢٩	٠,١٠٦	١١,٥٥٠	***
	٤٩	٠,٧١٩	١,٢٤٤	٠,١٠٥	١١,٨٩٤	***
	٥٦	٠,٤٩٦	٠,٨٥١	٠,٠٨٩	٩,٥٠٥	***
	٦٣	٠,٢٢٨	٠,١٤٣	٠,٠٥٣	٢,٦٨٧	**
	٧٠	٠,٦١٤	١,٠٨٦	٠,١٠٠	١٠,٩٠٤	***

يتضح من جدول (٦) أن جميع معاملات الانحدار اللامعيارية جاءت قيمها الحرجة دالة عند مستوى دلالة (٠,٠٠١– ٠,٠١) مما يدل على صدق نموذج البنية العاملية لعبارات مقياس الحكمة الاختبارية لطلاب المرحلة الثانوية, وذلك باستثناء العبارات أرقام (١٣، ١٥، ٣٢، ٤٤) حيث جاءت القيمة الحرجة لكل منهم غير دالة إحصائيًا, وبالتالي تم حذفهم وأصبح المقياس مكون من (٦٦) عبارة.

كما قام الباحثان باستخراج مؤشرات المطابقة للتأكد من حسن مطابقة النموذج كما في جدول (٧).

جدول (٧)
مؤشرات مطابقة نموذج التحليل العاملي التوكيدي لمقياس الحكمة الاختبارية (ن=٣٠٠)

م	مؤشرات المطابقة	قيمة المؤشر	المدى المقبول للمؤشر	القرار
١	النسبة بين X2 ودرجات الحرية DF/ CMIN	١,٧٠٥	أقل من (٥)	مقبول
٢	جذر متوسطات مربع البواقي (RMR)	٠,٠٤٢	الاقتراب من الصفر	مقبول
٣	مؤشر حسن المطابقة (GFI)	٠,٨٢٦	صفر إلى ١	مقبول
٤	مؤشر حسن المطابقة المصحح بدرجات الحرية (AGFI)	٠,٨٠٩	صفر إلى ١	مقبول
٥	مؤشر المطابقة المعياري (NFI)	٠,٧١٧	صفر إلى ١	مقبول
٦	مؤشر المطابقة النسبي (RFI)	٠,٧٠٥	صفر إلى ١	مقبول
٧	مؤشر المطابقة المتزايد (IFI)	٠,٨١٩	صفر إلى ١	مقبول
٨	مؤشر توكر لويس (TLI)	٠,٨٠٨	صفر إلى ١	مقبول
٩	مؤشر المطابقة المقارن (CFI)	٠,٨١٨	صفر إلى ١	مقبول
١٠	جذر متوسط مربع خطأ الاقتراب (RMSEA)	٠,٠٤٣	٠,٠٨ فأقل	مقبول

يتضح من جدول (٧) أن جميع قيم مؤشرات المطابقة جاءت في المدى المقبول, مما يدل على مطابقة نموذج التحليل العاملي التوكيدي لمقياس الحكمة الاختبارية لطلاب المرحلة الثانوية.

ب- الاتساق الداخلي:

قام الباحثان بتطبيق المقياس على (٣٠٠) مشاركاً من طلاب المرحلة الثانوية, وذلك لحساب الاتساق الداخلي لعبارات المقياس عن طريق حساب معاملات الارتباط بين درجة كل عبارة والدرجة الكلية للمكون الذي تنتمي إليه, ومعاملات الارتباط بين درجة كل عبارة والدرجة الكلية للمقياس، ومعاملات الارتباط بين درجات المكونات وبعضها البعض وبينها وبين والدرجة الكلية للمقياس, والجداول (٨)، (٩)، (١٠) توضح ذلك:

جدول (٨)
معاملات الارتباط بين درجة كل عبارة والدرجة الكلية للمكون الذي تنتمي إليه لمقياس الحكمة الاختبارية (ن= ٣٠٠)

م	معامل الارتباط	م	معامل الارتباط	م	معامل الارتباط	م	معامل الارتباط	م	معامل الارتباط
١	**٠,٤٩٧	١٥	**٠,٧٢٠	٢٩	حذفت سابقًا	٤٣	**٠,٥٥٢	٥٧	**٠,٥٣٦
٢	**٠,٧١٩	١٦	**٠,٢٦٢	٣٠	**٠,٤٣٥	٤٤	حذفت سابقًا	٥٨	**٠,٤٥٨
٣	**٠,٥٦٩	١٧	**٠,٢٣٢	٣١	**٠,٤٩٨	٤٥	**٠,٢٢٢	٥٩	**٠,٦٥٢
٤	**٠,٦٢٣	١٨	**٠,٦٦٥	٣٢	حذفت سابقًا	٤٦	**٠,٧٥٥	٦٠	**٠,٧٢٦
٥	**٠,٥٨٣	١٩	**٠,٦٦٠	٣٣	**٠,٢٤١	٤٧	**٠,٧٠٦	٦١	**٠,٦٨٧
٦	**٠,٦٦٣	٢٠	**٠,٦٨٧	٣٤	**٠,٦٣٥	٤٨	**٠,٦٦٦	٦٢	**٠,٥٨٤
٧	**٠,٦٠٥	٢١	**٠,٦٤١	٣٥	**٠,٧١٩	٤٩	**٠,٧٤٠	٦٣	**٠,٢٢٩
٨	**٠,٦٢٢	٢٢	**٠,٥٧٦	٣٦	**٠,٥٨٢	٥٠	**٠,٥٤٢	٦٤	**٠,٥٩٢
٩	**٠,٧٤٤	٢٣	**٠,٦٨٧	٣٧	**٠,٧١٥	٥١	**٠,٦٠٢	٦٥	**٠,١٩٦
١٠	**٠,٦٠٣	٢٤	**٠,٦٢٩	٣٨	**٠,٦٤٧	٥٢	**٠,٥٧٢	٦٦	**٠,٥٣٧
١١	**٠,٥٢٦	٢٥	**٠,٦٨٤	٣٩	**٠,٦٨١	٥٣	**٠,٥٧٨	٦٧	**٠,٥٢١
١٢	**٠,٦٢٠	٢٦	**٠,٦٣٨	٤٠	**٠,٧٤٠	٥٤	**٠,٦٨٨	٦٨	**٠,٥٨٦
١٣	حذفت سابقًا	٢٧		٤١	**٠,٢٣٨	٥٥	**٠,٦٩٠	٦٩	**٠,٥٧٥
١٤	**٠,٥٧٥	٢٨	**٠,٢٤٧	٤٢	**٠,٧٠٩	٥٦	**٠,٥٧٠	٧٠	**٠,٦٤٧

يتضح من جدول (٨) أن جميع قيم معاملات الارتباط بين درجة كل عبارة والدرجة الكلية للمكون الذي تنتمي إليه دالة إحصائيًا عند مستوى دلالة (٠,٠١)، وهو ما يدل على الاتساق الداخلي لعبارات المقياس.

جدول (٩)
معاملات الارتباط بين درجة كل عبارة والدرجة الكلية لمقياس الحكمة الاختبارية (ن= ٣٠٠)

م	معامل الارتباط	م	معامل الارتباط	م	معامل الارتباط	م	معامل الارتباط	م	معامل الارتباط
١	**٠,٤٣١	١٥	حذفت سابقًا	٢٩	**٠,٢٤٢	٤٣	**٠,٥٣٥	٥٧	**٠,٣٨٩
٢	**٠,٤٢٢	١٦	**٠,٤٨٩	٣٠	**٠,٣٠٠	٤٤	حذفت سابقًا	٥٨	**٠,٥٤٤
٣	**٠,٥٠٨	١٧	**٠,١٧١	٣١	**٠,٣٧١	٤٥	**٠,١٩٧	٥٩	**٠,٥٩٢
٤	**٠,٥٣٠	١٨	**٠,١٦٢	٣٢	حذفت سابقًا	٤٦	**٠,٦٠٣	٦٠	**٠,٦٠٠
٥	**٠,٥٢٩	١٩	**٠,٤٨٧	٣٣	**٠,١٨٢	٤٧	**٠,٥٧٥	٦١	**٠,٦١٠
٦	**٠,٥٩٣	٢٠	**٠,٥٦١	٣٤	**٠,٥٧٣	٤٨	**٠,٦٥٧	٦٢	**٠,٤٨٣
٧	**٠,٤٩٩	٢١	**٠,٦١٤	٣٥	**٠,٦٠٦	٤٩	**٠,٦٤٠	٦٣	**٠,١٦٥
٨	**٠,٤٤٠	٢٢	**٠,٥١٩	٣٦	**٠,٥٨٤	٥٠	**٠,٣٩٦	٦٤	**٠,٤٨١
٩	**٠,٥٠٤	٢٣	**٠,٥٧٠	٣٧	**٠,٥٣١	٥١	**٠,٥٦٣	٦٥	**٠,١٧٧
١٠	**٠,٥٦٧	٢٤	**٠,٥٩٩	٣٨	**٠,٦٢٦	٥٢	**٠,٥٠٦	٦٦	**٠,٥٠٠
١١	**٠,٤٤٣	٢٥	**٠,٦٠٦	٣٩	**٠,٥٨٨	٥٣	**٠,٥٠٥	٦٧	**٠,٤١٩
١٢	**٠,٤٢٧	٢٦	**٠,٤٧٨	٤٠	**٠,٥٤٦	٥٤	**٠,٥١٤	٦٨	**٠,٤٢٣
١٣	حذفت سابقًا	٢٧	**٠,٥٥٦	٤١	**٠,١٧٤	٥٥	**٠,٥٦٣	٦٩	**٠,٥١٠
١٤	**٠,٤٦٨	٢٨	**٠,١٦٦	٤٢	**٠,٦١٩	٥٦	**٠,٥٠٥	٧٠	**٠,٥٩٨

يتضح من جدول (٩) أن جميع قيم معاملات الارتباط بين درجة كل عبارة والدرجة الكلية لمقياس الحكمة الاختبارية دالة إحصائيًا عند مستوى دلالة (٠,٠١)، وهو ما يدل على الاتساق الداخلي لعبارات المقياس.

جدول (١٠)
معاملات الارتباط بين درجات مكونات مقياس الحكمة الاختبارية وبعضها البعض والدرجة الكلية للمقياس (ن= ٣٠٠)

م	المكون	مكون١	مكون٢	مكون٣	مكون٤	مكون٥	مكون٦	مكون٧	الدرجة الكلية
١	الاستعداد للاختبار	-							
٢	إدارة وقت الاختبار	٠,٦٢١**	-						
٣	التعامل مع ورقة الأسئلة	٠,٦٥٨**	٠,٦٤٤**	-					
٤	التعامل مع ورقة الإجابة	٠,٦٢١**	٠,٥٥٠**	٠,٧٦٤**	-				
٥	الاستنتاج المنطقي	٠,٥١٤**	٠,٤٦٠**	٠,٦٣٥**	٠,٦٧٣**	-			
٦	المراجعة	٠,٦٣٠**	٠,٦٠٥**	٠,٧٥٤**	٠,٧١٨**	٠,٦٥٤**	-		
٧	تقويم الأداء الاختباري	٠,٦٨٩**	٠,٦٣٩**	٠,٦٨٩**	٠,٦٥٤**	٠,٥٥٩**	٠,٧٢٣**	-	
-	الدرجة الكلية	٠,٨١٥**	٠,٧٨٥**	٠,٨٨١**	٠,٨٤٧**	٠,٧٦٧**	٠,٨٧١**	٠,٨٥٦**	-

يتضح من جدول (١٠) أن جميع قيم معاملات الارتباط بين درجة المكونات وبعضها البعض وبينها وبين الدرجة الكلية للمقياس دالة عند مستوى دلالة (٠,٠١)، مما يشير إلى الاتساق الداخلي لمكونات المقياس.

ج– ثبات المقياس:

استخدم الباحثان معاملات ثبات أوميجا لماكدونالد للتحقق من ثبات البنية العاملية لمكونات مقياس الحكمة الاختبارية والدرجة الكلية, وجدول (١١) يوضح ذلك:

جدول (١١)
معاملات ثبات أوميجا للمكونات والدرجة الكلية لمقياس الحكمة الاختبارية (ن= ٣٠٠)

م	المكون	معامل ثبات أوميجا لماكدونالد
١	الاستعداد للاختبار	٠,٧٧١
٢	إدارة وقت الاختبار	٠,٧٨٠
٣	التعامل مع ورقة الأسئلة	٠,٧٣٦
٤	التعامل مع ورقة الإجابة	٠,٧٨٣
٥	الاستنتاج المنطقي	٠,٨٣٣
٦	المراجعة	٠,٧٨٣
٧	تقويم الأداء الاختباري	٠,٨٠٩
-	الدرجة الكلية	٠,٩٥٠

يتضح من جدول (١١) أن معاملات ثبات أوميجا لمكونات الحكمة الاختبارية تراوحت ما بين (٠,٧٣٦–٠,٨٣٣), وللدرجة الكلية بلغت (٠,٩٥٠) وهي معاملات ثبات مرتفعة، مما يشير إلى الثقة في النتائج التي يمكن التوصل إليها من خلال المقياس.

ويوضح جدول (١٢) الصورة النهائية لمقياس الحكمة الاختبارية بعد حذف (٤) عبارات في التحليل العاملي التوكيدي، وبالتالي يكون عدد عبارات المقياس في الصورة النهائية (٦٦) عبارة, وتكون أعلى درجة يحصل عليها المستجيب في المقياس= ٦٦× ٥ = ٣٣٠ درجة , وأدني درجة= ٦٦× ١ = ٦٦ درجة.

جدول (١٢)
توزيع عبارات المقياس على مكونات الحكمة الاختبارية (الصورة النهائية)

م	المكونات	أرقام العبارات
١	الاستعداد للاختبار	١، ٨، ٢٠، ٢٧، ٣٣، ٤٠، ٤٦، ٥٣، ٦٠
٢	إدارة وقت الاختبار	٢، ٩، ١٤، ٢١، ٢٨، ٣٤، ٤٧، ٥٤، ٦١
٣	التعامل مع ورقة الأسئلة	٣، ١٠، ١٥، ٢٢، ٢٩، ٣٥، ٤١، ٤٨، ٥٥، ٦٢
٤	التعامل مع ورقة الإجابة	٤، ١١، ١٦، ٢٣، ٣٦، ٤٢، ٤٩، ٥٦، ٦٣
٥	الاستنتاج المنطقي	٥، ١٢، ١٧، ٢٤، ٣٠، ٣٧، ٤٣، ٥٠، ٥٧، ٦٤
٦	المراجعة	٦، ١٨، ٢٥، ٣١، ٣٨، ٤٤، ٥١، ٥٨، ٦٥
٧	تقويم الأداء الاختباري	٧، ١٣، ١٩، ٢٦، ٣٢، ٣٩، ٤٥، ٥٢، ٥٩، ٦٦

وتم تمييز العبارات السلبية في جدول (١٢) بوضع خط تحت أرقام عباراتها.

خطوات تنفيذ البحث:

لتنفيذ البحث اتبع الباحثان الخطوات الآتية:

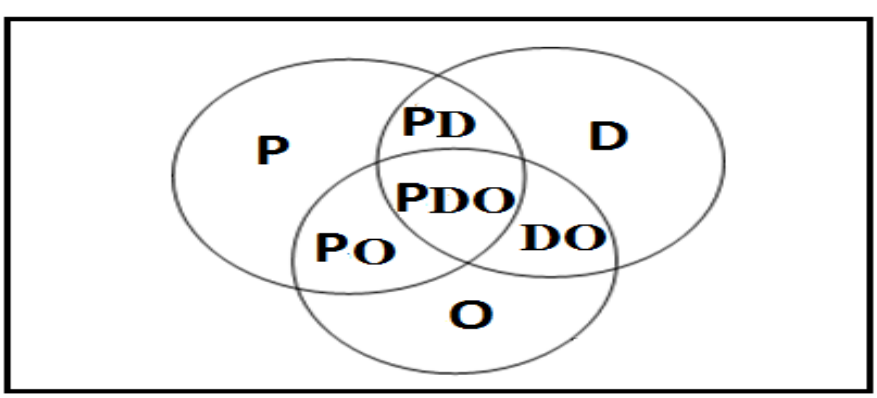
١- الاطلاع على الأطر النظرية والدراسات والبحوث السابقة التي تناولت المفاهيم الأساسية للبحث (نظرية إمكانية التعميم, الحكمة الاختبارية), وكتابة الإطار النظري للبحث.

٢- إعداد مقياس الحكمة الاختبارية في ضوء الإطار النظري والدراسات والبحوث السابقة, ثم عرضه على السادة المحكمين.

٣- تحديد أوجه أو مصادر تباين الخطأ والتي شملت عدد الأبعاد المكونة لمقياس الحكمة الاختبارية وفترات التطبيق.

٤- تحديد مستويات الأوجه حيث يشتمل وجه عدد الأبعاد على سبعة مستويات (من البعد الأول إلى البعد السابع)، ويشتمل وجه فترات التطبيق مستويين (الفترة الأولى, الفترة الثانية).

٥- اختيار تصميم البحث المناسب وهو التصميم ثنائي الوجه المتقاطع كليًا (الطلاب Persons (P)× عدد الأبعاد Dimensions (D) × فترات التطبيق Occasions (O)) حيث إنه تم تطبيق جميع مفردات المقياس على جميع المشاركين في البحث، ويوضح شكل (٤) تصميم البحث:



شكل (٤)
رسم توضيحي للتصميم ثنائي الوجه المتقاطع كليًا (الطلاب× عدد الأبعاد× فترات التطبيق)

وفي ضوء هذا التصميم تكون معادلتي حساب معامل إمكانية التعميم النسبي ($E\rho^2$) ومعامل إمكانية التعميم المطلق (ϕ) على النحو التالي:

$$E_P^2 = \frac{\sigma_P^2}{\sigma_P^2 + \frac{\sigma_{PD}^2}{n'_D} + \frac{\sigma_{PO}^2}{n'_O} + \frac{\sigma_{PDO}^2}{n'_D n'_O}}$$

$$\phi = \frac{\sigma_P^2}{\sigma_P^2 + \frac{\sigma_D^2 + \sigma_{PD}^2}{n'_D} + \frac{\sigma_O^2 + \sigma_{PO}^2}{n'_O} + \frac{\sigma_{DO}^2 + \sigma_{PDO}^2}{n'_D n'_O}}$$

حيث يشير (σ^2_p) إلى تباين الدرجة الشاملة، (σ^2_D) تباين الأبعاد، (σ^2_O) تباين فترات التطبيق، (σ^2PD) تباين تفاعل (الطلاب×الأبعاد)، (σ^2PO) تباين تفاعل (الطلاب×فترات التطبيق)، (σ^2PDO) تباين تفاعل (الطلاب×الأبعاد× فترات التطبيق)، (n'_O) عدد الأبعاد، (n'_O) عدد فترات التطبيق.

٦- تطبيق مقياس الحكمة الاختبارية على (٣٠٠) مشاركًا من طلاب وطالبات المرحلة الثانوية العامة والأزهرية, ثم إعادة تطبيق المقياس بفاصل زمني (٢٠ يوم) على نفس عدد المشاركين الذين استجابوا عن عبارات المقياس في التطبيق الأول.

٧- حساب الخصائص السيكومترية للمقياس في ضوء نظرية القياس التقليدية والمتمثلة في التحليل العاملي التوكيدي، والاتساق الداخلي، ومعاملات ثبات أوميجا لماكدونالد.

٨- إجراء المعالجات الإحصائية في ضوء نظرية إمكانية التعميم للإجابة عن تساؤلات البحث.

٩- عرض ومناقشة وتفسير نتائج البحث في ضوء الإطار النظري ونتائج الدراسات والبحوث السابقة وخصائص المشاركين.

الأساليب الإحصائية المستخدمة:

تم استخدام الأساليب الإحصائية المناسبة لنظرية إمكانية التعميم والمتمثلة في: تحليل التباين، وتباين التمييز (تباين الدرجة الشاملة)، وتباين الخطأ النسبي، وتباين الخطأ المطلق، ومعامل إمكانية التعميم النسبي، ومعامل إمكانية التعميم المطلق، لتحليل دراسة التعميم والقرار وذلك باستخدام برنامج (EduG) V.6.1-e في ضوء تصميم البحث المحدد,

بالإضافة إلى اختبار كوكرون للمقارنة بين معاملات إمكانية التعميم النسبية والمطلقة في ضوء اختلاف حجم العينة.

نتائج البحث ومناقشتها وتفسيرها:

نتائج التساؤل الأول وتفسيره:

ينص التساؤل الأول على " ما مصادر التباين (الطلاب, وعدد الأبعاد, وفترات التطبيق) الأكثر تأثيرًا على ثبات درجات مقياس الحكمة الاختبارية لدى طلاب المرحلة الثانوية"؟

وللإجابة عن هذا التساؤل ؛ تم تقدير مصادر التباين، والتفاعل بين هذه المصادر ونسب تأثيرها على الدرجة الشاملة في أداء الطلاب، ومعاملات إمكانية التعميم النسبية والمطلقة، ويوضح جدول (١٣) نتائج تقدير مصادر التباين لتصميم البحث الثنائي المتقاطع كليًا (الطلاب× عدد الأبعاد × فترات التطبيق).

جدول (١٣)
نتائج تحليل التباين للتصميم (الطلاب× عدد الأبعاد× فترات التطبيق)
لمقياس الحكمة الاختبارية (ن= ٣٠٠)

مصادر التباين	مجموع المربعات	درجات الحرية	متوسط المربعات	المكونات				
				العشوائية	المختلطة	المصححة	نسبة التباين	الخطأ المعياري
الطلاب (P)	٧٥٩٠٤,٤٢٧	٢٩٩	٢٥٣,٨٦٠	١٦,٢٤٩	١٦,٢٤٩	١٦,٢٤٩	٪٣٧,٣	١,٤٨٥
عدد الأبعاد (D)	٤٨٤١٨,٣٢٩	٦	٨٠٦٩,٧٢١	١٣,٣٩٥	١٣,٣٩٥	١٣,٣٩٥	٪٣٠,٨	٦,٧٢٤
فترات التطبيق (O)	٠,٠٩٥	١	٠,٠٩٥	٠,٠٢١	٠,٠٢١	٠,٠٢١	٪٠,٠	٠,٠٠٧
الطلاب× عدد الأبعاد (P× D)	٢٢٦٥٨,٠٩٩	١٧٩٤	١٢,٦٢٩	٠,٧٧١	٠,٧٧١	٠,٧٧١	٪١,٨	٠,٢٨٠
الطلاب× فترات التطبيق (P× O)	٧٤٢٢,٦١٩	٢٩٩	٢٤,٨٢٤	١,٩٦٢	١,٩٦٢	١,٩٦٢	٪٤,٥	٠,٢٩٣
عدد الأبعاد× فترات التطبيق (D×O)	١٨٤,٨٣٨	٦	٣٠,٨٠٦	٠,٠٦٥	٠,٠٦٥	٠,٠٦٥	٪٠,٢	٠,٠٥١

مصادر التباين	مجموع المربعات	درجات الحرية	متوسط المربعات	المكونات				
				العشوائية	المختلطة	المصححة	نسبة التباين	الخطأ المعياري
الطلاب× عدد الأبعاد× فترات التطبيق (P×D ×O)	١٩٨٩٠,٤٤٧	١٧٩٤	١١,٠٨٧	١١,٠٨٧	١١,٠٨٧	١١,٠٨٧	٪٢٥,٥	٠,٣٦٩
المجموع	١٧٤٤٧٨,٨٥٦	٤١٩٩					٪١٠٠	

يتضح من جدول (١٣) أن أكبر نسبة للتباين تأثيرًا على معاملات إمكانية التعميم في الدرجات الشاملة ترجع إلى (الطلاب) حيث بلغت نسبة التباين (٣٧,٣٪) من إجمالي التباين الكلي، يليها تباين (عدد الأبعاد) بنسبة (٣٠,٨٪), ثم تباين تفاعل (الطلاب× عدد الأبعاد× فترات التطبيق) بنسبة (٢٥,٥٪), ثم تباين تفاعل (الطلاب× فترات التطبيق) بنسبة (٤,٥٪), ثم تباين (الطلاب× عدد الأبعاد) بنسبة (١,٨٪), ثم تباين (عدد الأبعاد× فترات التطبيق) بنسبة (٠,٢٪), ثم تباين (فترات التطبيق) بأقل نسبة للتباين (٠,٠٪) تأثيرًا على معاملات إمكانية التعميم في الدرجات الشاملة.

كما تم تحليل دراسة التعميم G-Study لتصميم البحث؛ بهدف حساب نسبة تباين الخطأ النسبي والمطلق لكل مصدر من مصادر التباين؛ واستخراج قيم معاملات إمكانية التعميم النسبي والمطلق, ويوضح جدول (١٤) نتائج تحليل دراسة التعميم لتصميم البحث الثنائي المتقاطع كليًا (الطلاب× عدد الأبعاد× إعادة التطبيق).

جدول (١٤)
نتائج تحليل دراسة التعميم للتصميم (الطلاب× عدد الأبعاد× فترات التطبيق)
لمقياس الحكمة الاختبارية (ن= ٣٠٠)

مصادر التباين	تباين التمييز (تباين الدرجة الشاملة)	تباين الخطأ النسبي	نسبة تباين الخطأ النسبي	تباين الخطأ المطلق	نسبة تباين الخطأ المطلق
الطلاب (P)	١٦,٢٤٩	.....	.....	.....	.....
عدد الأبعاد (D)	.....	.....	.....	١,٩١٣	٥٠,٣٪
فترات التطبيق (O)	.....	.....	.....	٠,٠٠٠	٠,٠٪
الطلاب× عدد الأبعاد (P× D)	.....	٠,١١٠	٥,٩٪	٠,١١٠	٢,٩٪
الطلاب× فترات التطبيق (O ×P)	.....	٠,٩٨١	٥٢,١٪	٠,٩٨١	٢٥,٨٪
عدد الأبعاد× فترات التطبيق (D×O)	.....	.....	.....	٠,٠٠٤	٠,١٪
الطلاب× عدد الأبعاد× فترات التطبيق (P×D ×O)	.....	٠,٧٩١	٤٢,٠٪	٠,٧٩١	٢٠,٨٪
مجموع التباينات	١٦,٢٤٩	١,٨٨٣	١٠٠٪	٣,٨٠١	١٠٠٪
الانحراف المعياري	٤,٠٣١				
الخطأ المعياري النسبي	١,٣٧٢				
الخطأ المعياري المطلق	١,٩٤٩				
معامل إمكانية التعميم النسبي	٠,٩٠				
معامل إمكانية التعميم المطلق	٠,٨١				

يتضح من جدول (١٤) أن أكبر مكون لتباين الخطأ في القياس النسبي يرجع إلى تفاعل (الطلاب × فترات التطبيق) بنسبة بلغت (٥٢,١٪) من إجمالي نسب التباين الكلي؛ مما يدل على اختلاف ترتيب درجات الطلاب باختلاف فترات التطبيق، ثم تأتي نسبة تباين الخطأ النسبي الراجع إلى تفاعل (الطلاب× عدد الأبعاد× فترات التطبيق) بنسبة (٤٢٪)؛ مما يدل على اختلاف ترتيب درجات الطلاب باختلاف عدد الأبعاد وفترات التطبيق، في حين جاء مكون تباين الخطأ النسبي الراجع إلى تفاعل (الطلاب× عدد الأبعاد) بنسبة (٥,٩٪) وهي

نسبة منخفضة مقارنة بالنسب الأخرى, كما يلاحظ أن قيم نسب باقي مكونات التباين جاءت منخفضة جدًا لا تؤثر على إمكانية التعميم.

ويتضح أيضًا من جدول (١٤) أن أكبر مكون لتباين الخطأ في القياس المطلق يرجع إلى (عدد الأبعاد) بنسبة بلغت (٥٠,٣٪) من إجمالي نسب التباين الكلي؛ مما يدل على اختلاف متوسط درجات الطلاب باختلاف عدد الأبعاد، ثم تأتي نسبة تباين الخطأ المطلق الراجع إلى تفاعل (الطلاب× فترات التطبيق) بنسبة (٢٥,٨٪)؛ مما يدل على اختلاف متوسط درجات الطلاب باختلاف فترات التطبيق، ثم تأتي نسبة تباين الخطأ المطلق الراجع إلى تفاعل (الطلاب× عدد الأبعاد× فترات التطبيق) بنسبة (٢٠,٨٪)؛ مما يدل على اختلاف متوسط درجات الطلاب باختلاف كل من عدد الأبعاد وفترات التطبيق, في حين جاءت مكونات تباين الخطأ المطلقة الراجعة إلى كل من (الطلاب× عدد الأبعاد) و(عدد الأبعاد× فترات التطبيق) على الترتيب (٢,٩٪– ٠,١٪) وهي نسبة منخفضة جدًا مقارنة بالنسب الأخرى, كما يلاحظ أن قيم نسب باقي مكونات التباين جاءت منخفضة جدًا لا تؤثر على إمكانية التعميم.

كما يشير جدول (١٤) إلى أن قيمة معامل إمكانية التعميم النسبي بلغت (٠,٩٠), وقيمة معامل إمكانية التعميم المطلق بلغت (٠,٨١), وهي قيم أعلى من الحد الأدنى للقيمة التي يمكن الوثوق بها (٠,٨٠), مما يدل على أن درجات الطلاب على مقياس الحكمة الاختبارية في تتسم بإمكانية التعميم.

وتتفق هذه النتائج مع ما توصلت إليه دراسة كل من (Ogunka& Orluwene,2019؛ وطباع, ٢٠٢٠؛ Orluwene and Memory, 2020؛ العرايضة والشريفين, ٢٠٢٢ب ؛Medvedeva et al., 2023) من أن أكبر مكونات لتباين الخطأ النسبي والمطلق يرجع إلى التفاعلات الثنائية أو الثلاثية بين الأوجه التي اهتمت بها كل دراسة.

وتتفق هذه النتائج أيضًا مع ما توصلت إليه دراسة Dorathy et al., (2021) من وجود نسبة تباين ضئيلة ترجع إلى فترات التطبيق, وما توصلت إليه دراسة Sanz-Fernández et al. (2024) من وجود نسبة تباين كبيرة ترجع إلى الأفراد المشاركين في البحث.

كما جاءت قيم معاملي إمكانية التعميم النسبي والمطلق جيدة ومتوافقة مع ما توصلت إليه دراسة كل من (Akindahunsi & Afolabi 2021؛ الحجاجي, ٢٠٢٣؛ كامل, ٢٠٢٣؛ Medvedeva et al.,2023) في الحصول على قيم معاملات إمكانية تعميم جيدة, وتختلف مع ما توصلت إليه دراسة كل من (طباع, ٢٠٢٠؛ Dorathy et al., 2021) في الحصول على قيم معاملات تعميم منخفضة عن الحد المقبول لمعاملات إمكانية التعميم النسبية والمطلقة (٠,٨٠), كما أن الحصول على قيم معاملات ثبات جيدة لمقياس الحكمة الاختبارية يتفق مع ما توصلت إليه دراسة كل من (الزهراني, ٢٠١٥؛ Opesemowo at el., 2018؛ Thoma& Koller, 2018؛ الطراونة, ٢٠٢٠؛ أبو الحسن, ٢٠٢٢) من تمتع مقياس الحكمة الاختبارية بمعاملات صدق وثبات جيدة وفق النظرية الكلاسيكية في القياس ونظرية الاستجابة للمفردة .

ويفسر الباحثان زيادة نسبة خطأ القياس النسبي والمطلق الراجع إلى تفاعل (الطلاب× فترات التطبيق) إلى اختلاف ترتيب درجات الطلاب في التطبيق الأول والثاني, أي أن الطلاب استجابوا على عبارات المقياس بشكل أفضل في التطبيق الثاني عن التطبيق الأول, أو استجابوا على بعض العبارات بشكل أفضل في التطبيق الأول عن التطبيق الثاني, والذي قد يرجع إلى انتقال أثر التطبيق على المقياس أو زيادة ألفتهم بعبارات المقياس في حالة التطبيق الثاني, أو الشعور بالملل ونقص دافعيتهم للاستجابة على عبارات المقياس في التطبيق الثاني نتيجة مرورهم بهذه الخبرة في التطبيق الأول, كما يفسر الباحثان وجود خطأ قياس نسبي ومطلق راجع إلى تفاعل (الطلاب× عدد الأبعاد× فترات التطبيق) إلى اختلاف ترتيب درجات الطلاب داخل أبعاد المقياس بصفة عامة أو خلال فترتي التطبيق, والذي قد يرجع إلى تبني الطالب أسلوب للاستجابة على عبارات أبعاد المقياس لتحسين أدائهم في التطبيق الثاني, خاصة مع الطلاب الذين يمتلكون مستوى مرتفع من الحكمة الاختبارية, وقد يرجع أيضًا إلى تفاوت الخبرات السابقة للطلاب أنفسهم حول مكونات الحكمة الاختبارية, وبالتالي تختلف استجاباتهم من بعد إلى بعد آخر, ومن فترة إلى أخرى, والذي يؤثر بدوره على تباين درجات الطلاب.

ويدعم ذلك ما أشار إليه فرج (٢٠٠٧, ٣١٠– ٣١١) من وجود مصادر تباين خطأ ترجع إلى التذبذب العشوائي في أداء الفرد في فترتي التطبيق, حيث إن الأداء النفسي لا

يحدث من خلال قوالب صارمة, فتعرض الفرد لتغيرات صحية أو معنوية قد تجعله تلقائيًا في إحدى فترتي التطبيق وقلقًا في الفترة الأخرى, كما أن طبيعة السمة المقاسة وطريقة قياسها قد تتأثر بالخبرات السابقة للفرد, وعلى الرغم من أن عددًا من هذه العوامل يمكن التحكم فيه وضبطه؛ إلا أنه لا يمكن إغفال تأثير مثل هذه العوامل بصورة تامة.

ويُرجع الباحثان ارتفاع قيم معاملي التعميم النسبي والمطلق لمقياس الحكمة الاختبارية إلى ارتفاع تباين درجات عينة البحث (١٦,٢٤٩) مقارنة بتباين الخطأ النسبي (١,٨٨٣) وتباين الخطأ المطلق (٣,٨٠١), والذي قد يكون بسبب تنوع الخبرات السابقة المرتبطة بالأداء الاختباري للطلاب عينة البحث من حيث الصف الدراسي (الأول, والثاني, والثالث), ونوع التعليم (أزهري, وعام), حيث تختلف الاختبارات التي يتعرضون لها من عددها وأنواعها, وطريقة وضعها والإجابة عنها, والذي يؤدي إلى وجود تباين في مستويات الحكمة الاختبارية لدى الطلاب, ويدعم ذلك ما أشار إليه كل من (Mutua, 2012, 14 ؛ المطيري, ٢٠١٤, ٢٠– ٢١) من أن الخبرات السابقة للفرد في تعامله مع المواقف الاختبارية, والسمات الشخصية له, ونوع الاختبارات التحصيلية التي يختبرون بها من العوامل التي تؤثر في الحكمة الاختبارية لدى الطلاب.

وفي ضوء ذلك؛ فإن وجود تباين مرتفع في درجات الطلاب علي مقياس الحكمة الاختبارية مقارنة بتباين الخطأ النسبي والمطلق, أدى إلى الحصول على معاملات ثبات مناسبة وفق نظرية إمكانية التعميم نتيجة اعتماد حساب هذه المعاملات على تباين درجات الطلاب المستجيبين علي المقياس والذي يسمي بتباين التمييز (تباين الدرجة الشاملة), ويتفق ذلك مع ما أشار إليه السيد (٢٠٠٥, ٣٩٤) من ارتباط الثبات بصفة عامة ارتباطًا مباشرًا بتباين درجات المقياس, حيث تنخفض قيم معاملات الثبات عندما ينقص تباين الدرجات, وتزداد معاملات الثبات تبعً لزيادة التباين, وبما أن التباين يدل على فروق بين الأفراد في درجات المقياس فإن معامل ثبات درجات مقياس طُبق على مجموعة متجانسة من الأفراد (أقل تباينًا) ينقص قيمته العددية عن معامل ثبات درجات نفس المقياس عندما يُطبق على مجموعة أخرى أقل تجانسًا من المجموعة الأولى.

نتائج التساؤل الثاني وتفسيره:

ينص التساؤل الثاني على " ما أثر زيادة كل مصدر من مصادر التباين (عدد الأبعاد, وفترات التطبيق) على ثبات مقياس الحكمة الاختبارية لدى طلاب المرحلة الثانوية؟

وللإجابة عن هذا التساؤل؛ تم زيادة عدد الأبعاد، وفترات التطبيق؛ بهدف خفض مصادر تباين الخطأ المنتظم، ورفع قيم معاملات إمكانية التعميم، ومعرفة الأوجه الأكثر تأثيرًا في قيم هذه المعاملات؛ ويوضح جدول (١٥) دراسات القرار للتصميم ثنائي الوجه المتقاطع كليًا (الطلاب× عدد الأبعاد× فترات التطبيق) مدرجة بها قيم تباينات الخطأ النسبية والمطلقة، ومعاملات إمكانية التعميم النسبية والمطلقة لكل دراسة قرار بناءً على دراسة إمكانية التعميم الأولية.

جدول (١٥)
نتائج تحليل دراسات القرار للتصميم (الطلاب× عدد الأبعاد× فترات التطبيق) لمقياس الحكمة الاختبارية بتثبيت فترات التطبيق

م	مصادر التباين		تباين الخطأ النسبي المتوقع	الخطأ المعياري النسبي	تباين الخطأ المطلق المتوقع	الخطأ المعياري المطلق	معامل إمكانية التعميم النسبي	معامل إمكانية التعميم المطلق
	عدد الأبعاد	فترات التطبيق						
١	٧	٢	١,٨٨٣	١,٣٧٢	٣,٨٠١	١,٩٤٩	٠,٩٠	٠,٨١
٢	٨	٢	١,٧٧٠	١,٣٣٠	٣,٤٤٩	١,٨٥٧	٠,٩٠	٠,٨٢
٣	٩	٢	١,٦٨٢	١,٢٩٧	٣,١٧٤	١,٧٨١	٠,٩١	٠,٨٤
٤	١٠	٢	١,٦١٢	١,٢٦٩	٢,٩٥٥	١,٧١٩	٠,٩١	٠,٨٥

يتضح من جدول (١٥) أن زيادة عدد أبعاد مقياس الحكمة الاختبارية من (٧) أبعاد إلى (٨– ٩– ١٠) أبعاد لا يؤثر على معامل إمكانية التعميم النسبي, حيث ظلت قيمة معامل إمكانية التعميم النسبي كما هي (٠,٩٠) في حالة زيادة عدد الأبعاد إلى (٨) أبعاد, وارتفعت بقدر ضئيل جدًا إلى (٠,٩١) في حالة زيادة عدد الأبعاد إلى (٩–١٠) أبعاد, مما يشير إلى أن زيادة عدد أبعاد مقياس الحكمة الاختبارية لا يؤدي إلى تغيير كبير في قيم معاملات إمكانية التعميم النسبية.

كما يتضح أيضًا من جدول (١٥) أن زيادة عدد أبعاد مقياس الحكمة الاختبارية من (٧) أبعاد إلى (٨– ٩– ١٠) أبعاد يؤثر بقدر ضئيل على معامل إمكانية التعميم المطلق, حيث ارتفعت قيمة معامل إمكانية التعميم المطلق بقدر ضئل جدًا إلى (٠,٨٢) في حالة زيادة عدد الأبعاد إلى (٨) أبعاد, وارتفعت بقدر ضئيل إلى (٠,٨٤) و (٨٥) في حالة زيادة عدد الأبعاد

إلى (٩-١٠) أبعاد, مما يشير إلى أن زيادة عدد أبعاد مقياس الحكمة الاختبارية يؤدي إلى تغيير ضئيل في قيم معاملات إمكانية التعميم المطلقة.

وبشكل عام تشير النتائج السابقة إلى وجود قيم مرتفعة لمعاملات إمكانية التعميم النسبية والمطلقة لمقياس الحكمة الاختبارية في حالة وجود (٧) أبعاد أو أكثر, وأن زيادة عدد أبعاد مقياس الحكمة الاختبارية لا يؤثر بشكل كبير جدًا في ارتفاع قيم هذه المعاملات.

جدول (١٦)

نتائج تحليل دراسات القرار للتصميم (الطلاب× عدد الأبعاد× فترات التطبيق) لمقياس الحكمة الاختبارية بتثبيت عدد الأبعاد

م	مصادر التباين		تباين الخطأ النسبي المتوقع	الخطأ المعياري النسبي	تباين الخطأ المطلق المتوقع	الخطأ المعياري المطلق	معامل إمكانية التعميم النسبي	معامل إمكانية التعميم المطلق
	عدد الأبعاد	فترات التطبيق						
١	٧	٢	١,٨٨٣	١,٣٧٢	٣,٨٠١	١,٩٤٩	٠,٩٠	٠,٨١
٢	٧	٣	١,٢٩٢	١,١٣٦	٣,٢٠٩	١,٧٩١	٠,٩٣	٠,٨٤
٣	٧	٤	٠,٩٩٦	٠,٩٩٨	٢,٩١٢	١,٧٠٦	٠,٩٤	٠,٨٥
٤	٧	٥	٠,٨١٩	٠,٩٠٥	٢,٧٣٥	١,٦٥٣	٠,٩٥	٠,٨٦

يتضح من جدول (١٦) أن زيادة عدد فترات تطبيق مقياس الحكمة الاختبارية من (٢) إلى (٣-٤-٥) فترات يؤثر على معامل إمكانية التعميم النسبي, حيث ارتفعت قيمة معامل إمكانية التعميم النسبي إلى (٠,٩٣) في حالة زيادة فترات التطبيق إلى (٣) فترات, ثم ارتفعت بعد ذلك بقدر ضئيل جدًا إلى (٠,٩٤- ٠,٩٥) في حالة زيادة فترات التطبيق إلى (٤-٥) فترات, مما يشير إلى أن زيادة عدد فترات تطبيق مقياس الحكمة الاختبارية إلى (٣) فترات يؤدي إلى تغيير ملحوظ في قيمة معامل إمكانية التعميم النسبي, وتغيرات ضئيلة في حالة زيادة عدد فترات التطبيق إلى (٤-٥) فترات.

كما يتضح أيضًا من جدول (١٦) أن زيادة عدد فترات تطبيق مقياس الحكمة الاختبارية من (٢) إلى (٣-٤-٥) فترات يؤثر على معامل إمكانية التعميم المطلق, حيث ارتفعت قيمة معامل إمكانية التعميم المطلق إلى (٠,٨٤) في حالة زيادة فترات التطبيق إلى (٣) فترات, ثم ارتفعت بعد ذلك بقدر ضئيل جدًا إلى (٠,٨٥- ٠,٨٦) في حالة زيادة فترات التطبيق إلى (٤-٥) فترات, مما يشير إلى أن زيادة عدد فترات تطبيق مقياس الحكمة الاختبارية إلى (٣) فترات يؤدي إلى تغيير ملحوظ في قيمة معامل إمكانية التعميم المطلق, وتغيرات ضئيلة في حالة زيادة عدد فترات التطبيق إلى (٤-٥) فترات.

وبشكل عام تشير النتائج السابقة إلى وجود قيم مرتفعة لمعاملات إمكانية التعميم النسبية والمطلقة لمقياس الحكمة الاختبارية في حالة وجود فترتي تطبيق أو أكثر, وأن زيادة فترات تطبيق مقياس الحكمة الاختبارية إلى (٣) فترات يؤثر بشكل ملحوظ في ارتفاع قيم هذه المعاملات.

وتتفق هذه النتائج مع ما توصلت إليه دراسة كل من (طباع, ٢٠٢٠؛ Akindahunsi & Afolabi ,2021؛ العرايضة والشريفين, ٢٠٢٢ب؛ الحجاجي, ٢٠٢٣؛ كامل, ٢٠٢٣؛ Medvedeva et al., 2023) من وجود تأثيرات ضئيلة أو كبيرة لزيادة مستويات الأوجه على معاملات إمكانية التعميم النسبية والمطلقة, كما تتفق هذه النتائج مع ما توصلت إليه دراسة كل من (نصار, ٢٠٢١؛ أبو الحسن, ٢٠٢٢) من صدق وثبات البنية العاملية المقترحة لمقياس الحكمة الاختبارية وتمتع المقياس بمؤشرات مطابقة جيدة.

ويفسر الباحثان وجود تأثير ضئيل لزيادة عدد أبعاد مقياس الحكمة الاختبارية على معاملات إمكانية التعميم النسبية والمطلقة إلى وجود تباين مشترك بين الأبعاد المكونة للمقياس وارتباطها بدرجة كبيرة مع الدرجة الكلية وبالتالي يمكن ألا يؤدي زيادة عدد الأبعاد إلى ظهور تباينات إضافية تؤثر بشكل جوهري على معاملات إمكانية التعميم.

ويرجع الباحثان وجود تأثير ملحوظ لزيادة فترات التطبيق على معاملات إمكانية التعميم النسبية والمطلقة إلى أن تطبيق المقياس أكثر من مرة على الأفراد يساهم في كشف العديد من مصادر الخطأ المؤثرة في درجات الأفراد خلال فترات التطبيق, فالعوامل التي تؤثر في الفترة الزمنية بين التطبيق الأول والثاني قد لا تكون بنفس الدرجة والنوع في الفترة الزمنية بين التطبيق الثاني والثالث, وبالتالي تتباين درجات الأفراد مما يؤثر في قيم معاملات إمكانية التعميم, فبزيادة فترات التطبيق يمكن استكشاف مجموعة أوسع من الظروف التي من الممكن أن تؤثر على تباين درجات الطلاب المستجيبين على عبارات المقياس.

ويدعم ذلك ما أوصى به فرج (٢٠٠٧, ٣١٣) من عدم الاكتفاء بحساب ثبات الاختبارات والمقاييس على مدى فترة زمنية واحدة, بل على أكثر من فترة زمنية, حيث يوفر هذا الإجراء تقديرًا مناسبًا للثبات, واكتشاف مدى تأثره بالفترة الزمنية للتطبيق.

نتائج التساؤل الثالث:

ينص التساؤل الثالث على " ما أثر اختلاف حجم العينة (١٠٠– ٢٠٠–٣٠٠) على قيم معاملات إمكانية التعميم النسبية والمطلقة لمقياس الحكمة الاختبارية لدى طلاب المرحلة الثانوية"؟

وللإجابة عن هذا التساؤل ؛ تم تقدير نسب تباين الخطأ النسبي والمطلق لمصادر التباين، والتفاعل بين هذه المصادر ونسب تأثيرها، ومعاملات إمكانية التعميم النسبية والمطلقة عند أحجام العينة (١٠٠– ٢٠٠–٣٠٠)، ويوضح جدول (١٧) ذلك:

جدول (١٧)

نتائج تحليل دراسات التعميم للتصميم (الطلاب× عدد الأبعاد× فترات التطبيق) لمقياس الحكمة الاختبارية في ضوء حجم العينة (١٠٠-٢٠٠-٣٠٠)

مصادر التباين	تباين التمييز			تباين الخطأ النسبي (النسبة)			تباين الخطأ المطلق(النسبة)		
حجم العينة	١٠٠	٢٠٠	٣٠٠	١٠٠	٢٠٠	٣٠٠	١٠٠	٢٠٠	٣٠٠
الطلاب (P)	٧,٩٥٧	١٠,٥١٣	١٦,٢٤٩	.....	.....	.....	.....	.....	.....
عدد الأبعاد (D)	.....	.....	.....	.....	.....	.....	٢,٠٥١ (٪٣٧,٨)	٢,٠٩٦ (٪٤٩,٠)	١,٩١٣ (٪٥٠,٣)
فترات التطبيق (O)	.....	.....	.....	.....	.....	.....	٠,٠٠١ (٪٠,٠)	٠,٠٠٠ (٪٠,٠)	٠,٠٠٠ (٪٠,٠)
الطلاب× عدد الأبعاد (P×D)	.....	.....	.....	٠,١٠٩ (٪٣,٣)	٠,١٠٥ (٪٤,٨)	٠,١١٠ (٪٥,٩)	٠,١٠٩ (٪٢,٠)	٠,١٠٥ (٪٢,٥)	٠,١١٠ (٪٢,٩)
الطلاب× فترات التطبيق (P× O)	.....	.....	.....	٢,٢٨٧ (٪٦٨,١)	١,١٣٠ (٪٥٢)	٠,٩٨١ (٪٥٢,١)	٢,٢٨٧ (٪٤٢,٢)	١,١٣٠ (٪٢٦,٤)	٠,٩٨١ (٪٢٥,٨)
عدد الأبعاد× فترات التطبيق (D×O)	.....	.....	.....	.....	.....	.....	٠,٠٠٩ (٪٠,٢)	٠,٠٠٩ (٪٠,٢)	٠,٠٠٤ (٪٠,١)
الطلاب× عدد الأبعاد× فترات التطبيق (P× D ×O)	.....	.....	.....	٠,٩٦١ (٪٢٨,٦)	٠,٩٣٩ (٪٤٣,٢)	٠,٧٩١ (٪٤٢,٠)	٠,٩٦١ (٪١٧,٧)	٠,٩٣٩ (٪٢١,٩)	٠,٧٩١ (٪٢٠,٨)
مجموع التباينات	٧,٩٥٧	١٠,٥١٣	١٦,٢٤٩	٣,٣٥٨	٢,١٧٥	١,٨٨٣	٥,٤٢١	٤,٢٨١	٣,٨٠١
الانحراف المعياري				٢,٨٢٠	٣,٢٤٢	٤,٠٣١			

مصادر التباين	تباين التمييز	تباين الخطأ النسبي (النسبة)			تباين الخطأ المطلق(النسبة)
	الخطأ المعياري النسبي	١,٨٣٢	١,٤٧٥	١,٣٧٢	
	الخطأ المعياري المطلق	٢,٣٢٨	٢,٠٦٩	١,٩٤٩	
	معامل إمكانية التعميم النسبي	٠,٧٠	٠,٨٣	٠,٩٠	
	معامل إمكانية التعميم المطلق	٠,٥٩	٠,٧١	٠,٨١	

يتضح من جدول (١٧) ما يلي:

- **ارتفاع تباين التمييز بزيادة حجم العينة (١٠٠– ٢٠٠– ٣٠٠), حيث بلغت تباينات التمييز (٧,٩٥٧– ١٠,٥١٣– ١٦,٢٤٩) على الترتيب.**
- **انخفاض مجموع تباينات الخطأ النسبي بزيادة حجم العينة (١٠٠– ٢٠٠– ٣٠٠), حيث بلغ مجموع تباينات الخطأ النسبي (٣,٣٥٨– ٢,١٧٥– ١,٨٨٣) على الترتيب.**
- **انخفاض مجموع تباينات الخطأ المطلق بزيادة حجم العينة (١٠٠– ٢٠٠– ٣٠٠), حيث بلغ مجموع تباينات الخطأ المطلق (٥,٤٢١– ٤,٢٨١– ٣,٨٠١) على الترتيب.**
- **ارتفاع قيم معاملات إمكانية التعميم النسبي بزيادة حجم العينة (١٠٠– ٢٠٠– ٣٠٠), حيث بلغت قيم معاملات إمكانية التعميم النسبي (٠,٧٠– ٠,٨٣– ٠,٩٠) على الترتيب.**
- **ارتفاع قيم معاملات إمكانية التعميم المطلق بزيادة حجم العينة (١٠٠– ٢٠٠– ٣٠٠), حيث بلغت قيم معاملات إمكانية التعميم المطلق (٠,٥٩– ٠,٧١– ٠,٨١) على الترتيب.**

وتشير هذه النتائج بشكل عام إلى أن قيم معامل إمكانية التعميم النسبي والمطلق ترتفع بزيادة حجم العينة, مما يدل على أنه كلما زادت أحجام العينة ترتفع قيم معاملات إمكانية التعميم النسبي والمطلق لمقياس الحكمة الاختبارية لدى طلاب المرحلة الثانوية.

كما استخدم الباحثان اختبار كوكرون (Cocron) والذي يتبع توزيع اختبار (كا^{2}) للمقارنة بين معاملات إمكانية التعميم النسبية والمطلقة الناتجة من تحليل دراسات التعميم عند حجوم العينة (١٠٠– ٢٠٠–٣٠٠), ويوضح جدول (١٨) ذلك :

جدول (١٨)
مقارنة بين معاملات إمكانية التعميم النسبي والمطلق باستخدام اختبار Cocron

معامل الثبات	حجم العينة	قيم معاملات إمكانية التعميم	عدد المفردات	درجات الحرية	قيما كا٢	مستوى الدلالة	فترة الثقة من	فترة الثقة إلى
معامل إمكانية التعميم النسبي	١٠٠	٠,٧٠	٦٦	٢	**٥٢,٥٤	٠,٠١	٠,٦٠٩	٠,٧٧٨
	٢٠٠	٠,٨٣					٠,٧٩٤	٠,٨٦١
	٣٠٠	٠,٩٠					٠,٨٨٣	٠,٩١٥
معامل إمكانية التعميم المطلق	١٠٠	٠,٥٩	٦٦	٢	**٢٥,٦٢	٠,٠١	٠,٤٤٦	٠,٦٩٦
	٢٠٠	٠,٧١					٠,٦٤٩	٠,٧٦٤
	٣٠٠	٠,٨١					٠,٧٧٨	٠,٨٣٩

يتضح من جدول (١٨) وجود فروق ذات دلالة إحصائية بين معاملات إمكانية التعميم النسبية ترجع إلى اختلاف حجم العينة (١٠٠-٢٠٠-٣٠٠), حيث بلغت قيمة كا٢ (٥٢,٥٤), وهي قيمة دالة إحصائيًا عند مستوى دلالة (٠,٠١), كما يتضح وجود فروق ذات دلالة إحصائية بين معاملات إمكانية التعميم المطلقة ترجع إلى اختلاف حجم العينة (١٠٠-٢٠٠-٣٠٠), حيث بلغت قيمة كا٢ (٢٥,٦٢), وهي قيمة دالة إحصائيًا عند مستوى دلالة (٠,٠١), مما يشير إلى وجود أثر ذي دلالة إحصائية لحجم العينة على قيم معاملات إمكانية التعميم النسبية والمطلقة لمقياس الحكمة الاختبارية لدى طلاب المرحلة الثانوية.

ويفسر الباحثان ارتفاع معاملات إمكانية التعميم النسبية والمطلقة لمقياس الحكمة الاختبارية بزيادة حجم العينة إلى أنه كلما زاد حجم العينة كلما كانت أكثر تمثيلًا للمجتمع الأصل الذي سحبت منه, وبالتالي ستكون أكثر تنوعًا وتعكس بصورة أفضل التباينات بين أفراد المجتمع الأصلي, وبالتالي تقل أخطاء القياس المرتبطة بأسلوب المعاينة, كما أنه قد يؤدي إلى زيادة تباين درجات الأفراد, والذي يؤثر بدوره في استخراج قيم معاملات إمكانية التعميم.

كما يعزو الباحثان هذه النتيجة في ضوء أن حجم العينات الأكبر قد يعمل على تقليل نسب تأثير الأخطاء العشوائية المؤثرة على استجابات الطلاب على مقياس الحكمة الاختبارية, كما أن اختيار حجم عينة كبير يؤدي إلى تقليل الخطأ المتوقع للمعاينة ما قد يؤدي إلى انخفاض قيمة الخطأ المعياري للقياس.

ويدعم ذلك ما أشار إليه علام (٢٠٠٥, ١٨٠) من اعتماد القيم التقديرية لمعاملات الثبات بصفة عامة اعتمادًا كبيرًا على مدى الفروق الفردية بين الأفراد, فكلما ازداد تباين الدرجات الحقيقة للأفراد إزدادت قيم معاملات الثبات, كما يدعم هذه النتيجة ما أشار إليه Atilgan (2013, 215) من ارتفاع قيم معاملات إمكانية التعميم النسبية والمطلقة بزيادة حجم العينة, وأن أحجام العينة الصغيرة لا توفر تقديرًا غير متحيز لمعاملات إمكانية التعميم, ويمكن الحصول على تقديرات دقيقة وقوية لهذه المعاملات بزيادة حجم العينة.

توصيات البحث:

١- عقد دورات تدريبية للباحثين وأعضاء هيئة التدريس عن نظريات القياس (النظرية الكلاسيكية, ونظرية الاستجابة للمفردة, ونظرية إمكانية التعميم), وتوضيح أهم الفروق بينهم من حيث افتراضات كل نظرية, ومفاهيمها, ومنطلقاتها السيكومترية.

٢- عقد ورش عمل تدريبية للباحثين وأعضاء هيئة التدريس عن نظرية إمكانية التعميم، وكيفية استخدامها في تقدير ثبات الاختبارات والمقاييس النفسية والتربوية, وبرامجها الإحصائية, خاصة في ظل ندرة الأطر النظرية والبحوث السابقة التي اهتمت بتطبيقاتها الإحصائية في تحليل صدق وثبات الاختبارات والمقاييس النفسية.

٣- توجيه نظر الباحثين إلى ضرورة تقديم مؤشرات سيكومترية عن مصادر الخطأ التي قد تؤثر على ثبات المقاييس المستخدمة في بحوثهم بالاعتماد على معاملات نظرية إمكانية التعميم.

٤- ضرورة اهتمام المعنيين بالعملية التعليمية في المرحلة الثانوية باستخدام مقياس للحكمة الاختبارية للكشف على مستواها لدى الطلاب، ومن ثم يمكن تدريبهم على مكوناتها وأهم الاستراتيجيات التي يمكن أن يتبعها الطلاب قبل وأثناء وبعد الاختبار.

٥- توجيه نظر أعضاء هيئة التدريس بالمراحل الدراسية المختلفة بضرورة مراعاة فنيات صياغة المفردات الاختبارية عند إعداد الاختبارات التحصيلية الموضوعية والمقالية, بما يضمن الحصول على تقييم عادل بين الطلاب وفق ما بذلوه من استذكار ومراجعة لدروسهم وليس وفق امتلاكهم لمستوى مرتفع من الحكمة الاختبارية.

٦- الأخذ في الاعتبار مصادر تباين الخطأ المؤثرة على ثبات مقياس الحكمة الاختبارية لدى طلاب المرحلة الثانوية للحصول على معاملات إمكانية تعميم نسبية ومطلقة مرتفعة عند استخدام المقياس.

دراسات وبحوث مقترحة:

١- استخدام نظرية إمكانية التعميم في تقدير ثبات بعض المقاييس النفسية باختلاف مستوى الاستجابة (ثلاثي, خماسي, سباعي).

٢- تقدير معاملات ثبات مقياس الحكمة الاختبارية في ضوء نظريتي إمكانية التعميم والاستجابة للمفردة "دراسة مقارنة".

٣- تقدير معاملات ثبات مقياس الحكمة الاختبارية في ضوء نظرية إمكانية التعميم باستخدام برنامجي (EduG) و(GENOVA).

٤- تكافؤ البنية العاملية لمقياس الحكمة الاختبارية لدى طلاب المرحلة الثانوية في ضوء متغيري نوع التعليم (عام, أزهري), التخصص (علمي, أدبي).

٥- تحليل بروفيلات مكونات الحكمة الاختبارية لدى طلاب المرحلة الثانوية في ضوء بعض المتغيرات.

قائمة المراجع

أولًا: المراجع العربية

أبو الحسن, أسماء (٢٠٢٢). البنية العاملية الاستكشافية والتوكيدية لمقياس الحكمة الاختبارية: دراسة ميدانية على عينة من جامعة حماة. *مجلة كلية التربية للبنات, جامعة بغداد, ٣٣* (٤), ٥٣-٦٩.

أبو هاشم, السيد (٢٠٠٨). النموذج البنائي التنبؤي لمهارات الدراسة والحكمة الاختبارية والتحصيل الدراسي لدى طلاب المرحلة الثانوية. *مجلة كلية التربية, جامعة المنصورة, ٦٨* (١), ٢١٠-٢٧٠.

الثبيتي، علي (١٩٩٨). ثبات أدوات القياس في العلوم الإنسانية بين النظرية التقليدية للاختبار ونظرية التعميم. *مجلة مستقبل التربية العربية، المركز العربي للتعليم والتنمية، ٤* (١٥)، ١١٢-١٣٤.

الحجاجي، عائشة (٢٠٢٣). أثر عدد المهام الأدائية وطريقة التقدير والمقيمين في ثبات سلالم التقدير لأنشطة تقييم الأداء لمقرر العلوم باستخدام نظرية إمكانية التعميم. *المجلة التربوية، كلية التربية، جامعة سوهاج، ١١٦* (١١٦)، ٢٥٧-٢٨٥.

الحربي، خليل والحربي، عيد (٢٠١٧). مؤشرات الثبات باستخدام نظرية التعميم ومؤشرات صدق البناء لمقياس موهبة للإبداع. *مجلة جامعة طيبة للعلوم التربوية، كلية التربية، جامعة طيبة، ١٢* (٣)، ٤٢٥-٤٤١.

الخولي, هشام (٢٠١٨). الإسهام النسبي لأثر برنامج تدريبي لمهارات الحكمة الاختبارية على التحصيل الدراسي في ضوء تحمل المخاطرة وقلق الاختبار والدافعية الدراسية لدى طلاب الجامعة. *مجلة كلية التربية, جامعة بنها, ٢٩* (١١٣), ٤٠١- ٤٨٤.

الزهراني, محمد (٢٠١٥). الخصائص السيكومترية لمقياس الحكمة الاختبارية لدى طلاب الجامعة في البيئة السعودية. *المجلة العلمية لكلية التربية النوعية, ٣* (٤), ٢١٧- ٢٦٦.

السيد, فؤاد (٢٠٠٥). *علم النفس الإحصائي وقياس العقل البشري* (ط٣). دار الفكر العربي.

الطراونة, صبري (٢٠٢٠). تطوير مقياس لحكمة الاختبار ومطابقة فقراته مع نموذج أندريش. *مجلة مؤتة للبحوث والدراسات, سلسلة العلوم الإنسانية والاجتماعية, ٣٥* (٦), ١٤١-١٧٠.

الطريري، عبد الرحمن (١٩٩٧). *القياس النفسي والتربوي نظريته، أسسه، تطبيقاته*. مكتبة الرشد.

العبيدي, شيماء (٢٠٢٣). حكمة الاختبار لدى الطلبة الموهوبين. *مجلة جامعة الأنبار للعلوم الإنسانية*, ٢٠ (١), ٦٢٢- ٦٥٢.

العرايضة، محمد والشريفين، نضال (٢٠٢٢أ). استخدام نظرية التعميم للكشف عن مدى مساهمة مصادر التباين المتعددة في صدق اختبار في الرياضيات. *المجلة الأردنية في العلوم التربوية*، *١٨* (٣)، ٥٦٩-٥٨٣.

العرايضة، محمد والشريفين، نضال (٢٠٢٢ب). الكشف عن إسهام مصادر التباين المتعددة في ثبات اختبار في الرياضيات باستخدام نظرية التعميم. *مجلة العلوم التربوية، كلية التربية، جامعة الملك سعود*، *٣٤* (٢)، ١٩٣-٢١٨.

الغرايبة، عبد الهادي وبني أحمد، أحمد (٢٠٢٣). تقدير الثبات في الاختبارات بفقرات ثنائية التدريج ومتعددة التدريج وفق النظرية الكلاسيكية في القياس ونظرية التعميم: دراسة مقارنة. *المجلة الأردنية للعلوم التطبيقية، سلسلة العلوم الإنسانية، جامعة العلوم التطبيقية الخاصة*، ٣٦ (١)، ٩٦-١١٠.

القرشي, خديجة (٢٠١٧). الحكمة الاختبارية وعلاقتها بأنماط التفكير وقلق الاختبار لدى طالبات كلية التربية بجامعة الطائف. *المجلة التربوية, كلية التربية جامعة سوهاج, ٤٧*, ٣٢٥-٣٨٧.

المطيري, عبدالرحمن (٢٠٢٤). الحكمة الاختبارية: المفهوم, النشأة, الاستراتيجيات والأدوات (قراءة تأصيلية). *مجلة البحوث التربوية والنفسية*, ٢١ (٨١), ١- ٤٣.

أمين، صبري (٢٠١٦). ثبات تقييمات الكفاءة الاجتماعية لدى الأطفال باستخدام النسخة العربية لمقياس الكفاءة الاجتماعية المعدل في ضوء نظرية إمكانية التعميم. *مجلة العلوم التربوية، كلية الدراسات العليا للتربية، جامعة القاهرة*، ٢٤ (٢)، ١٢٩-١٨١.

حسب الله, عبدالعزيز (٢٠١٩). إدارة الوقت وحكمة الاختبار وبعض المتغيرات الديموجرافية كمنبئات بالتسويف الأكاديمي لدى طلاب كلية التربية. *مجلة كلية التربية, جامعة أسيوط*, ٣٥ (٩), ٤٣٥-٥٣٢.

حماد, ديانا (٢٠١٠). علاقة الحكمة الاختبارية بالأداء الناتج من اختبار تحصيلي ذو اختيار من متعدد مبني وفق نموذج راش لدى طالبات كلية التربية للبنات الأقسام الأدبية بجامعة أم القرى. *دراسات عربية في التربية وعلم النفس، رابطة التربويين العرب*, ٤ (٤), ٢٩٧- ٣٣٨.

سعادة, مروة وعيسى, أسماء (٢٠٢٢). الإسهام النسبي للتحكم الانتباهي واليقظة العقلية في التنبؤ بمهارات الحكمة الاختبارية لدى الطالبات المعلمات بكلية الاقتصاد المنزلي جامعة المنوفية. *مجلة بحوث التربية النوعية, جامعة المنصورة,* (٦٧), ٢١٩– ٢٦٣.

سليمان، أمين (٢٠٠٩). *القياس والتقويم في العلوم الإنسانية أسسه وأدواته وتطبيقاته*. دار الكتاب الحديث.

سليمان, معن (٢٠٢٣). ممارسة استراتيجيات حكمة الاختبار لدى طلبة المرحلة الأساسية في مدارس مديرية التربية والتعليم في لواء بني كنانة بمحافظة إربد في ضوء بعض المتغيرات الديموجرافية. *المجلة العلمية لكلية التربية, جامعة أسيوط, ٣٩*(٨), ١٨٣– ٢٢٠.

طاحون, حسين (٢٠١٠). فاعلية برنامج قائم على استراتيجيات الحكمة الاختبارية على كل من التحصيل الدراسي وقلق الاختبار لدى طلاب الدراسات العليا بكلية التربية– جامعة عين شمس. *المجلة المصرية للدراسات النفسية، الجمعية المصرية للدراسات النفسية, ٢٠* (٦٩), ٧٩–١٢٦.

طباع، فاروق (٢٠٢٠). استخدام نظرية إمكانية التعميم في تقدير ثبات اختبار تقييم كفاءة الرياضيات لدى طلاب السنة الرابعة ابتدائي. *المجلة الأردنية في العلوم التربوية، ١٦* (١)، ١–١٨.

طباع، فاروق ودعيدش، عبد السلام (٢٠٢٣). مصادر تباين الخطأ المؤثرة على ثبات وصدق تقييمات الكفاءة وفق نظرية إمكانية التعميم: مراجعة للبحوث. *مجلة الروائز، كلية العلوم الإنسانية والاجتماعية، جامعة باتنة ١ الحاج لخضر بالجزائر، ٧* (١)، ٨–٢٠.

علام, صلاح الدين (٢٠٠٥). *نماذج الاستجابة للمفردة الاختبارية أحادية البعد ومتعددة الأبعاد وتطبيقاتها في القياس النفسي والتربوي*. دار الفكر العربي.

علام، صلاح الدين (٢٠٠٦). *القياس والتقويم التربوي والنفسي أساسياته وتطبيقاته وتوجهاته المعاصرة*. دار الفكر العربي.

فرج, صفوت (٢٠٠٧). *القياس النفسي* (ط٦). مكتبة الأنجلو المصرية.

قنديل, وائل وموسى, حنان (٢٠٢٢). تصميم مقياس الحكمة الاختبارية لدى طلاب كلية التربية الرياضية جامعة مدينة السادات. *مجلة نظريات وتطبيقات التربية البدنية وعلوم الرياضة, كلية التربية جامعة مدينة السادات, ٣٧* (١), ١٣٢–١٧٠.

كامل، جوزيف (٢٠٢٣). ثبات اختبار المواقف لتقييم أداء الطالب المعلم في ضوء نظرية إمكانية التعميم. *المجلة المصرية للدراسات النفسية، الجمعية المصرية للدراسات النفسية، ٣٣* (١١٨)، ٤٦٣-٥٠٠.

كردي, ساندي (٢٠٢٤). الاسهامات النسبية للمعتقدات المعرفية والحكمة الاختبارية والذكاء الانفعالي والتسويف الأكاديمي للتنبؤ بقلق الاختبار لدى طلبة المرحلة الثانوية في المدينة المنورة. *مجلة الدراسات والبحوث التربوية, ٤*(١٠), ٣١- ٧٦.

محمد, محمد (٢٠١٨). الإسهام النسبي لمهارات الحكمة الاختبارية في التنبؤ بدرجات عينة من طلاب الثانوية العامة على اختبار البوكليت. *دراسات عربية في التربية وعلم النفس, رابطة التربويين العرب*، (٩٤), ٦٩- ١٠٨.

نصار, علي (٢٠٢١). *البنية العاملية للحكمة الاختبارية لدى طلاب الصف الأول الثانوي في الأردن* [رسالة دكتوراه غير منشورة]. كلية التربية, جامعة المنصورة.

وادي, عفاف (٢٠١٣). مهارات الحكمة الاختبارية وعلاقتها بالتحصيل الدراسي لدى طلبة كلية التربية/ ابن الهيثم للعلوم الصرفة. *مجلة البحوث التربوية والنفسية*,(٣٧), ٢٩٥- ٣٢٢.

ثانيًا: المراجع الأجنبية

Akindahunsi, O., & Afolabi, E. (2021). Using generalizability theory to investigate the reliability of scores assigned to students in English language examination in Nigeria. *Journal of Measurement and Evaluation in Education and Psychology, 12* (2), 147-162.

Albano, T. (2018). *Introduction to educational and psychological measurement using R*. University of Nebraska– Lincoln: Educational Psychology Papers and Publications.

Atilgan, H. (2013).Sample size for estimation of G and Phi coefficients in generalizability theory. *Eurasian Journal of Educational Research*, 51, 215-227

Brennan, R. (2001). *Generalizability theory: Statistics for social science and public policy*. Springer-Verlag.

Brennan, R. (2003). *Coefficients and Indices in Generalizability Theory.* Center for Advanced Studies in Measurement and Assessment, CASMA Research Report (Number 1), College of Education, University of Iowa.

Brennan, R. (2011). Generalizability theory and classical test theory. *Applied Measurement In Education, 24*, 1-21.

Bumbalkova, E. (2021). Test-taking strategies in second language receptive skills tests: A Literature Review. *International Journal of Instruction, 14*(2), 647-664.

Cardinet, J., Johnson, S., & Pini, G. (2010). *Applying generalizability theory using EduG*. Routledge - Taylor & Francis Group.

Clayson, P., Carbine, K., Baldwin, S., Olsen, J., & Larson, M. (2021). Using generalizability theory and the ERP Reliability Analysis (ERA) Toolbox for assessing test-retest reliability of ERP scores part 1: Algorithms, framework, and implementation. *International Journal of Psychophysiology, 166*, 174-187.

Cohen, A. D. (2013). Using test-wiseness strategy research in task development. In A. J. Kunnan (Ed.), *The companion to language assessment: Approaches and development, Part 7: Assessment development*. Wiley and Sons.

Crocker, L. & Algina, J. (2008). *Introduction to classical and modern test theory*. Cengage Learning.

DeVore,S., Stewart, J. & Stewart,G. (2016). Examining the effects of test wiseness in conceptual physics evaluations. *Physical Review Physics Educational Research, 12*(020138), 1-11.

Dodeen, H. (2015). Teaching test-taking strategies: Importance and techniques. *Psychology Research, 5*(2), 108-113.

Domino, G. & Domino, M. (2006). *Psychological testing: An introduction* (2nd ed.). Library of Congress.

Dorathy, S. Amadioha, A. & Orluwene, G. (2021). Application of generalizability theory in the estimation of dependability of critical thinking scale for University students. *Scholars Journal of Physics, Mathematics and Statistics, 8*(9), 171- 178.

Gbore, L. & Osakuade, J. (2016). Effects of testwiseness training in Mathematics on adolescent secondary school students' test anxiety in Ondo State, Nigeria. *Journal of Education and Practice, 7* (10), 34-39.

Hayati, A. & Ghojogh, A. (2008). Investigating the influence of proficiency and gender on the use of selected Test-wiseness strategies in Higher Education. *English Language Teaching, 1* (2), 169-181.

Haynes,P .(2011).*The Effect of testwiseness on self-efficacy and mathematics performance of Middle School Students with learning disabilities* [Doctoral dissertation, Virginia Commonwealth University]. ProQuest Dissertations and Theses Global.

Heitman, R., Kovaleski, J. & Pugh, S. (2009). Application of generalizability theory in estimating the reliability of Ankle-complex Laxity measurement. *Journal of Athletic Training, 44*(1), 48–52.

Houston, S. E. (2005). *Test-wiseness training: An investigation of the impact of test-wiseness in an employment setting* [Doctoral dissertation, The Graduate Faculty of the University of Akron]. ProQuest Dissertations and Theses Global.

Huang, C. (2009). Magnitude of task-sampling variability in performance assessment: a meta-analysis. *Educational and Psychological Measurement, 69* (6), 887- 912.

Jackson, D., Michaelides, G., Dewberry, C. Englert, p. (2022). Clarifying the scope of generalizability theory for multifaceted assessment. *New Zealand Journal of Psychology, 51*(2),53-64.

Li, M., Shavelson, R., Yin,Y., & Wiley, E. (2015). Generalizability theory, In R. Cautin & S. Lilienfeld (Eds.), *The Encyclopedia of clinical psychology* (1322-1341). Wiley.

Medvedeva, A., Galli, G., Petrovskaya, A.& Medvedev, O.(2023). A novel application of generalizability theory to evaluate the reliability of the recognition memory test. *Current Psychology, 42*,17020–17032.

Mustapha, A. (2014). Relationship between secondary school student's level of testwiseness and their performance in Mathematics in Bauchi Metropolis of Bauchi State- Nigeria, IOSR .*Journal Of Humanities And Social Science, 19*(3), 41-44.

Mutua, B. M. (2012). *Nature and practice of test wiseness the case of Kenyan schools* [Master's thesis, School of Education, University of Nairobi]. ProQuest Dissertations and Theses Global.

Ogunka, R.,& Orluwene, G.(2019). Application of generalizability theory in estimation of variance components in National Examination Council essay questions in English Language. *The African Journal of Behavioural and Scale Development Research, 1*(2), 1-6.

Olagoke,F. (2012). *Effects of testwiseness training on test-anxiety and achievement in mathematics among selected secondary school students*

in Ekiti state [Doctoral dissertation, University of Lagos]. ProQuest Dissertations and Theses Global.

Opesemowo, O., Afolabi, E., Awopeju, O.& Oluwatimilehin, T. (2018). Development of a scale for measuring students' testwiseness in senior secondary school examination in Nigerian. *International Journal of Research, 5*(19), 465-473.

Orluwene, G., & Memory, Q.(2020). Application of generalizability theory in reliability assessment of chemistry essay achievement test. *EPRA International Journal of Research and Development, 5* (4), 328-339.

Otoum, A., Khalaf, H. Bajbeer, A. & Hamad, H. (2015). The level of test–wiseness for the students of arts and science faculty at sharourah and its relationship with some variables. *Journal of Education and Practice, 6*(29), 102- 113.

Peng, Y., Hong, E., & Mason, E. (2014). Motivational and cognitive test-taking strategies and their influence on test performance in mathematics. *Educational Research and Evaluation, 20*(5), 366-385.

Roberson, D. (2020). *Test-wiseness and background knowledge: Their relative contributions to high test performance* [Master's thesis, Mississippi State University]. ProQuest Dissertations and Theses Global.

Russo, J. (2019). The impact of a short test-wiseness intervention on standardized numeracy assessment scores: A cautionary tale about using NAPLAN growth data to evaluate primary schools. *Networks: An Online Journal for Teacher Research, 21*(2), 1-16.

Sanders, R. (2005). Generalizability theory: Estimation, In B. Everitt & D. Howell (Eds.), *Encyclopedia of statistics in behavioral science* (Vol 2, 711-717). John Wiley & Sons, Ltd.

Sanz-Fernández, C., Morales-Sánchez, V., Castellano, J., & Mendo, A. (2024). Generalizability theory in the evaluation of psychological profile in track and field. *Sports, 12*(27), 1-11.

Tavakoli, E.& Samiani, S. (2014). Test-wiseness strategies in PBTs and IBTs: The case of EFL test takers, who benefits more?. *Procedia - Social and Behavioral Sciences*, 98, 1876 – 1884.

Tavakol, M., & Brennan, R. (2013). Medical education assessment: a brief overview of concepts in generalizability theory. *International Journal of Medical Education, 3*, 221-222.

Teker, G., Guler, N, & Uyanik, G. (2015). Comparing the effectiveness of SPSS and EduG using different designs for generalizability theory. *Educational Sciences: Theory & Practice, 15*(3), 635-645.

Thoma, G. & Koller, O. (2018). Test-wiseness: A neglected construct? Empirical findings on testing and trainability of test-wiseness. *Zeitschrift for Bildungsforsch*, 8, 63–80 .

Vispoel, W., Lee, H., Chen, T.,& Hong, H. (2023). Extending applications of generalizability theory-based bifactor model designs. *Psych, 5*, 545-575.

Vispoel, W., Morris, C., & Kilinc, M. (2017). Practical applications of generalizability theory for designing, evaluating, and improving psychological assessments. *Journal of Personality Assessment*, 1-15.

Vispoel, W., Morris, C., & Kilinc, M. (2018). Applications of generalizability theory and their relations to classical test theory and structural equation modeling. *Psychological Methods, 23* (1), 1-26.

Waiprakhon, T., & Jaturapitakkul, N. (2018). Test-taking strategies used in the reading section of the test of English for Thai engineers and technologists: A computer-based ESP test. *A Journal of PASAA,* 55, 147- 177.

Webb, N., & Shavelson, R. (2005). Generalizability theory: Overview, In B. Everitt & D. Howell (Eds.), *Encyclopedia of statistics in behavioral science* (Vol 2, 717-719). John Wiley & Sons, Ltd.

Yelboga, A., & Tavsancil, E. (2010). The examination of reliability according to classical test and generalizability on a job performance scale. *Educational Sciences: Theory & Practice, 10* (3), 1847-1854.

Yin, Y. & Shavelson, R. (2008). Application of generalizability theory to concept map assessment research. *Applied Measurement in Education, 21* (3), 273-291.

Zhao, Y., Zhu, K., Zhang, J., Liu, Z., & Wang, L.(2024). Exploring the measurement of psychological resilience in Chinese civil aviation pilots based on generalizability theory and item response theory. *Scientific Reports, 14* (1856), 1-12.