كلية التربية

كلية معتمدة من الهيئة القومية لضمان جودة التعليم

إدارة: البحوث والنشر العلمي (المجلة العلمية)

=======

مقارنة الخصائص السيكومترية بين الاختبارات التحصيلية الإلكترونية والورقية

إعـداد

د / سوميه شكري محمد محمود

مدرس علم النفس التربوي

كلية التربية . جامعة المنيا

كلية التربية . جامعة الإمام عبد الرحمن بن فيصل

﴿ المجلد الرابع والثلاثون– العدد السادس – يونيه ٢٠١٨م ﴾

http://www.aun.edu.eg/faculty_education/arabic

ملخص البحث:

هدف البحث الحالي إلى مقارنة الخصائص السيكومترية بين الاختبار التحصيلي الإلكتروني والورقي، وذلك من خلال مقارنة الإحصاءات الخاصة بمعاملات صعوبة المفردات ومعاملات تمييزها وقيم الثبات والصدق التلازمي للاختبار التحصيلي عند تقديمه بثلاث طرق مختلفة: الطريقة الإلكترونية أثناء المحاضرة، والطريقة الإلكترونية عن بعد، والطريقة الورقية. وقد قارن البحث ثبات الاتساق الداخلي للاختبار بصوره الثلاثة محسوباً بطريقتين: طريقة التجزئة النصفية، وطريقة ألفا كرونباخ. كما قارن البحث الصدق التلازمي للاختبار بصوره الثلاثة محسوباً بمعامل ارتباط سبيرمان بين درجة الطلاب على الاختبار التحصيلي التكويني المستخدم واختبار نهاية الفصل الدراسي كمحك. وقد استخدم البحث اختباراً تحصيلياً تكوينياً من عشرين مفردة من نوع الاختيار من متعدد ذات أربع بدائل في مقرر "مهارات التفكير"، واستخدم المنهج التجريبي، حيث تم اختيار ثلاث شعب متكافئة في التحصيل من طالبات كلية التربية بجامعة الإمام عبد الرحمن بن فيصل.

أخذت المجموعة الأولى الاختبار إلكترونياً أثناء المحاضرة من خلال نظام البلاك بورد (Blackboard)، مع ضبط خيارات النظام، بحيث يتيح للممتحن محاولةً واحدة، وتم توقيت الاختبار بتحديد ٢٥ دقيقة كزمن للإجابة، وقُدِّمت الأسئلة للممتحنين بشكل عشوائي، كما وُزِّعت بدائل الإجابة بشكل عشوائي، وتم تقديم جميع الأسئلة في صفحة واحدة؛ مما يتيح للممتحن التنقل بين الأسئلة ومراجعة الإجابات.

وأخذت المجموعة الثانية الاختبار إلكترونياً عن بعد من خلال نظام البلاك بورد، وتم ضبط خيارات تقديم الاختبار بحيث يكمل الممتحن الاختبار في جلسة واحدة، مع حظر الرجوع؛ مما يمنع الممتحن من التنقل بين الأسئلة، وتم توقيت الاختبار بتحديد ٢٥ دقيقة كزمن للإجابة، وقُدِّمت الأسئلة للممتحنين بشكل عشوائي، كما وُزِّعت بدائل الإجابة بشكل عشوائي.

وأخذت المجموعة الثالثة الاختبار في صورة ورقية. وأظهرت النتائج وجود فروق ذات دلالة إحصائية بين معاملات صعوبة المفردات عند تقديمه إلكترونياً عن بعد، وعند تقديمه إلكترونياً أثناء المحاضرة أو ورقياً، وكانت الفروق لصالح الصورة الأولى "الإلكترونية عن بعد"، بينما لم توجد فروق ذات دلالة إحصائية بين معاملات تمييز مفردات الاختبار التحصيلي بصوره الثلاثة "إلكترونياً عن بعد وإلكترونياً أثناء المحاضرة وورقياً"، وتم الحصول على أفضل معامل ثبات للاختبار وأفضل معامل للصدق التلازمي عند تقديمه إلكترونياً أثناء المحاضرة، وكان أقل معامل للثبات وأقل معامل للصدق التلازمي عند تقديم الاختبار إلكترونياً عن بعد.

وفي ضوء هذه النتائج يوصي البحث بتقديم الاختبار إلكترونياً بتواجد المعلم، وعدم تقديمه إلكترونياً عن بعد، كما يوصي بتطوير إجراءات ضبط تقديم الاختبار عن بعد؛ ليتيح للمعلم مراقبة أداء الممتحنين أثناء تقديم الاختبار.

Abstract:

The present study aimed at comparing the psychometric characteristics of both the electronic achievement tests versus the pencil and paper ones through making a comparison of the statistical analysis of the difficulty index and its discrimination index as well as the reliability and the concurrent validity of the achievement test when presenting it in three different ways: computer based testing through the lecture, computer based testing at home and Paper-pencil testing. The co-efficient reliability was calculated by two different ways of the internal consistency: Cronbach's alpha and the split half-reliability coefficient, and calculating the concurrent validity of the test through calculating the coefficient correlation of students' scores on the formative achievement test and the summative one. The formative test consisted of twenty MCQ items with four alternatives in the Thinking Skills Course using the experimental research design. Three groups were selected from among the Faculty of Education for girls in Imam Abdulrahman Bin Faisal University. The first group was tested electronically using the Blackboard system during lectures making sure of giving students only one trial to answer the test. The allotted time was twenty five minutes. Shuffling questions as well as the alternatives where considered. All questions were presented in one page as to give students the chance to revise their answers before submission. The second group was tested asynchronously using the Blackboard system. Students were allowed to answer the test in one session and there were not allowed to go back to review questions. Shuffling questions as well as the alternatives where considered as to reduce the opportunities for cheating the answers to minimum, while the third group was tested conventionally (paper-pencil test). Results revealed that there was a statistical difference between difficulty indexes of the achievement test items in presenting it asynchronously and when presenting it while lecturing and conventionally (paper-pencil test),and there was no statistical difference between

discrimination indexes of the achievement test items in presenting it while lecturing and when presenting it asynchronously and conventionally (paper-pencil test) The value of the reliability coefficient and the concurrent validity was extremely high when conducting the test during lectures and it was extremely low when conducting the test asynchronously. In the light of these results, the present study recommends developing the procedures of managing tests asynchronously as to give teachers more opportunities to monitor students while conducting tests.

مقدمة:

يشهد العصر الحالي تطوراً معرفياً متسارعاً كمياً ونوعياً، مما يدفع المؤسسات التعليمية للبحث عن أدوات تمكنها من متابعة ذلك النمو المتلاحق للمعرفة، ولم يعد هناك بُد من دعم بيئات التعلم ببرمجيات الحاسوب وتكنولوجيا المعلومات؛ لما تقدمه من ميزات متعددة، أهمها سهولة الوصول للمعلومة، ودعم بيئة التعلم بمعينات بصرية وسمعية تمكن المتعلمين من معالجة المعلومات المعقدة، وتعين المعلمين على شرح المفاهيم المجردة، كما أنها تضيف إلى بيئة التعلم عاملي التشويق والإثارة، مما يحسن دافعية المتعلمين. ومن المتوقع أن تحل المدرسة الإلكترونية محل المدرسة التقليدية؛ لقدرتها على علاج الكثير من مشكلات التعليم في الدول الفقيرة؛ لأنها تتيح الخدمة التعليمية لعدد كبير من المتعلمين دون الحاجة لإمكانات لوجستية ضخمة، وتقضي على الحدود الجغرافية للمدرسة، وتتيح الفرصة للاستعانة بمعلمين أكفاء من أي مكان بالعالم. حيث تتيح الفصول الافتراضية الاتصال والتفاعل بين الطلاب والمعلم عبر بيئات جغرافية من مختلف أنحاء العالم وتدعم بيئة التعلم بالعديد من الوسائط وأدوات التعلم الفعالة (غريب، ٢٠١٤: ١٤٤). وكلما تحولت البيئات التعليمية إلى استخدام تقنية الحاسوب والمعلومات في التدريس والتعلم فإنه من الأجدى أن تتناسب البيئات الاختبارية مع البيئات التعليمية والتدريسية (الخزي والذكري،٢٠١١: ١٩٣).

وعلى الرغم من توفر العديد من تقنيات إعداد الاختبارات الإلكترونية وإتاحتها على شبكة الإنترنت إلا أنها لم تحظ بعد بالاهتمام من قبل التربويين، وقد يرجع ذلك إلى الاعتقاد بصعوبة إعدادها أو تطبيقها، أو إلى عدم الثقة في نتائجها، ومن ناحية أخرى هناك اتجاه سالب لدى المتعلمين تجاه الاختبارات الإلكترونية، فهناك نزعة لدى المتعلمين لمقاومة التغيير، لذا أصبح من الضروري دراسة الخصائص السيكومترية للاختبارات الإلكترونية مقارنة بالاختبارات الورقية التقليدية، مما يتيح لكل من التربويين والمتعلمين مزيداً من المعلومات حول هذه الاختبارات، وتقييمها بشكل موضوعي قائم على الأدلة العلمية، كما يتيح لمصممي تقنيات الاختبارات الإلكترونية المعرفة اللازمة لتطوير تلك التقنيات، والشروط اللازمة لجعل نتائج هذه الاختبارات ذات موثوقية عالية.

مشكلة البحث:

اهتم كثير من الباحثين في السنوات الأخيرة بالاختبارات الإلكترونية، وتنوع مجال هذا الاهتمام، فبعض الدراسات اهتمت بدراسة فاعلية الاختبارات الإلكترونية كأداة تعليمية تستخدم لتدريب المتعلمين ورفع مستوى أدائهم في الاختبارات التحصيلية، كما اهتم المختصون في مجال تكنولوجيا التعليم بالاختبارات الإلكترونية من حيث مقارنة جودة البرمجيات المختلفة المستخدمة

في إعداد هذه الاختبارات من الناحية الفنية والتقنية، وأعدوا برامج تعليمية لتمكين المعلمين من استخدام الاختبارات الإلكترونية ووضع معايير لجودة هذه الاختبارات.

أما من الناحية السيكومترية فهناك ندرة ملحوظة في الدراسات التي بحثت في الخصائص السيكومترية للاختبارات الإلكترونية مقارنة بالاختبارات الورقية التقليدية، خاصة أن هناك حالة من تباين الاتجاه حول فاعلية الاختبارات الإلكترونية في قياس التحصيل. فبعض المختصين يرون أن الاختبارات الإلكترونية تقدم العديد من المزايا؛ حيث يتم من خلالها إعداد المفردات وإخراجها بطريقة صحيحة وسهلة، وتقديمها للممتحنين، وتسجيل استجاباتهم، وتصحيحها، وتقديم التغذية المرتدة الفورية، ورصد درجات الاختبار، وتخزين النتائج. في حين يشكك البعض الآخر في دقة نتائج هذه الاختبارات، خاصة عند تطبيقها عن بعد؛ لأنها حينئذ تهيئ فرص الغش للممتحنين، فقد يقدم الاختبار على الشاشة شخص آخر غير الممتحن، لذا أصبح من الضروري دراسة تأثير طريقة تقديم الاختبار (ورقياً / إلكترونياً) على خصائصه السيكومترية.

وقد أعرب كل من خبراء التقييم والباحثين والممارسين والمستخدمين عن قلقهم بشأن إمكانية مقارنة الدرجات بين أسلوبي تقديم الاختبار إلكترونياً وورقياً (Wang. et al., 2008: 5). كما أن هناك العديد من الطلاب يزيد لديهم قلق الاختبار في حالة تطبيق الاختبارات الإلكترونية، خاصة إذا ما تم توقيت الاختبار، مما يثير التساؤل عن مدى تأثير ذلك على ثبات الاختبار وصدقه.

ولم يقتصر تأثير تقنيات الحاسوب والمعلومات في التعليم على الدول المتقدمة فحسب، إنما امتد أيضاً إلى مؤسسات التعليم في الشرق الأوسط، ولكن انحسرت هذه الاستخدامات في التعلم والتدريس، ولم تستخدم في قياس التحصيل؛ نظراً لظن التربويين بعدم مصداقيتها، مما جعل دراسة كفاءة الاختبارات الإلكترونية أمراً غاية في الأهمية، تفرضه الفجوة الحاصلة بين التوجه السريع في استخدام الحاسوب في عملية التعليم من جهة والبطء الشديد في اعتماد الاختبارات الإلكترونية من جهة أخرى (الخزي والذكري،٢٠١١: ١٦٩-١٧٠). ولاشك أن الاختبارات الإلكترونية وسيلة مهمة في تطوير عملية التقويم والعملية التعليمة بشكل عام؛ لأنها توفر الكثير من الجهد للمعلم، مما يهيئ له فرصة جيدة للاهتمام بأنشطة التعلم الأخرى، ولذلك هناك حاجة إلى مزيد من الدراسات لتقييم طرق تقديم الاختبارات الإلكترونية؛ لتوجيه المستخدمين إلى الطريقة الصحيحة؛ لاستخدامها والاستفادة منها.

تساؤلات البحث:

١. هل تختلف قيم معاملات صعوبة المفردات باختلاف طريقة تقديمه (إلكترونياً أثناء المحاضرة / إلكترونياً عن بعد / ورقياً)؟
٢. هل تختلف قيم معاملات تمييز المفردات باختلاف طريقة تقديمه (إلكترونياً أثناء المحاضرة / إلكترونياً عن بعد / ورقياً)؟
٣. هل تؤثر طريقة تقديم الاختبار التحصيلي (إلكترونياً أثناء المحاضرة / إلكترونياً عن بعد / ورقياً) على معامل ثباته؟
٤. هل تؤثر طريقة تقديم الاختبار التحصيلي (إلكترونياً أثناء المحاضرة / إلكترونياً عن بعد / ورقياً) على معامل صدقه التلازمي؟

فروض البحث:

١. لا توجد فروق ذات دلالة إحصائية عند مستوى دلالة ($\alpha \leq$ ٠.٠٥) بين معاملات صعوبة مفردات الاختبار التحصيلي عند تقديمه إلكترونياً أثناء المحاضرة وتقديمه إلكترونياً عن بعد، وتقديمه ورقياً.
٢. لا توجد فروق ذات دلالة إحصائية عند مستوى دلالة ($\alpha \leq$ ٠.٠٥) بين معاملات تمييز صعوبة مفردات الاختبار التحصيلي عند تقديمه إلكترونياً أثناء المحاضرة وتقديمه إلكترونياً عن بعد، وتقديمه ورقياً.
٣. لا تختلف قيمة معامل ثبات الاتساق الداخلي للاختبار التحصيلي عند تقديمه إلكترونياً أثناء المحاضرة وتقديمه إلكترونياً عن بعد، وتقديمه ورقياً.
٤. لا تختلف قيمة معامل الصدق التلازمي للاختبار التحصيلي عند تقديمه إلكترونياً أثناء المحاضرة وتقديمه إلكترونياً عن بعد، وتقديمه ورقياً.

أهداف البحث:

١. التعرف على أثر طريقة تقديم الاختبار التحصيلي على كل من الخصائص السيكومترية للاختبار (الصدق والثبات) والخصائص السيكومترية للفقرات (الصعوبة والتمييز).
٢. الكشف عن الطريقة الأفضل لتقديم الاختبار التحصيلي.
٣. تقييم مصداقية الاختبارات التحصيلية الإلكترونية عند تطبيقها عن بعد.

أهمية البحث:

أولاً- الأهمية النظرية:

يقدم البحث معلومات عن مدى مصداقية الاختبارات التحصيلية الإلكترونية، خاصة عند تقديم الاختبار عن بعد، فهناك ندرة في هذه المعلومات في الأدبيات السيكولوجية، حيث يقدم البحث أدلةً لتأكيد أو دحض الظن الشائع لدى التربويين بانعدام موثوقية الاختبارات الإلكترونية عند تطبيقها عن بعد، رغم توفر العديد من القيود، كضبط الزمن المتاح للحل، وظهور المفردة للممتحن مرة واحدة، وعدم قدرته للرجوع لها بعد تقديم الإجابة، وترتيب كل من المفردات وبدائلها بطريقة عشوائية، بحيث يختلف الترتيب تماماً بين الممتحنين.

ثانياً- الأهمية التطبيقية:

١. إمداد التربويين بالمعلومات اللازمة لاختيار الطريقة المناسبة لتقديم الاختبارات التحصيلية.
٢. توجيه اهتمام مصممي برمجيات الاختبارات الإلكترونية إلى تطوير خيارات تقديم الاختبارات بما يتناسب مع تحقيق الشروط السيكومترية للاختبار التحصيلي الجيد.
٣. التحقق من دقة نتائج الاختبارات التحصيلية الإلكترونية قبل الاندفاع في تعميمها على المراحل التعليمية المختلفة.

حدود البحث:

يتحدد البحث الحالي بما يلي:

١. **الأداة المستخدمة:** وهي اختبار تحصيلي تكويني في مقرر مهارات التفكير من ٢٠ مفردة من نوع الاختيار من متعدد، يقيس أربعة مستويات معرفية طبقاً لتصنيف بلوم للأهداف، هي: (التذكر / الفهم / التطبيق / التحليل).
٢. **عينة البحث:** وشملت ٨٣ طالبة ممن يدرسون المقرر بكلية التربية، بجامعة الإمام عبد الرحمن بن فيصل.
٣. **النظام المستخدم في تقديم الاختبار إلكترونياً:** وهو برنامج إدارة التعلم "البلاك بورد" (Blackboard).

التعريف الإجرائي لمصطلحات البحث:

١. **الاختبار التحصيلي:** هو أداة قياس تحدد درجة تحصيل الممتحن لمحتوى تعليمي محدد جيداً، من خلال إجابته عن عينة من المفردات تمثل الأهداف السلوكية للمحتوى، ويتم إعداده وفق طريقة منظمة من عدة خطوات تخضع لشروط وقواعد محددة.
٢. **الاختبار الإلكتروني:** هو اختبار يتم تقديمه والاستجابة على مفرداته وتصحيحها فورياً باستخدام برمجيات حاسوبية.
٣. **ثبات الاختبار:** هو الاتساق والدقة في نتائج الاختبار، وتعني الدقة خلو الدرجة المقدرة على الاختبار من مصادر الخطأ، ويعني الاتساق الاستقرار في علامات الفرد إذا أخذ الاختبار نفسه عدة مرات في نفس الظروف.
٤. **معامل ثبات الاتساق الداخلي للاختبار:** هو معامل يتم تقديره ليعبر عن درجة التجانس الداخلي بين فقرات الاختبار. وتم تقديره في هذا البحث باستخدام كل من معامل ألفا كرونباخ، ومعامل ثبات التجزئة النصفية المصحح بطريقة سبيرمان براون.
٥. **الصدق التلازمي للاختبار:** هو اتساق نتائج الاختبار مع نتائج محك خارجي مستقل يقيس السمة نفسها، ويتم تطبيقه بالتزامن مع الاختبار. وتم تقديره في هذا البحث من خلال حساب معامل ارتباط سبيرمان بين درجات الطالبات في الاختبار ودرجاتهم في الاختبار النهائي لنفس المقرر.
٦. **صعوبة المفردة:** هي نسبة الممتحنين الذين أجابوا عن المفردة إجابة صحيحة من المجموع الكلي للممتحنين.
٧. **تمييز المفردة:** هو قدرة الفقرة على التمييز بين الممتحنين المتقنين والممتحنين غير المتقنين. وتم حساب معاملات التمييز للفقرات باستخدام معامل ارتباط بوينت بايسيريال (Point biserial) بين درجات الممتحنين على الفقرة ودرجاتهم الكلية على الاختبار.

الإطار النظري:

لقد زادت العملية التعليمية تعقيداً نتيجة للزيادة في أعداد التلاميذ والتقدم العلمي والتكنولوجي الهائل والمتدفق بصورة متلاحقة (سمعان،٢٠١٢ :٤٤). وتعد الاختبارات الإلكترونية سلاحاً ذا حدين، فهي – من ناحية – وسيلة مهمة لدعم وتطوير العملية التعليمية؛ لأنها تقدم عدة مميزات تسهم في حل هذا التعقيد، ومن ناحية أخرى هناك شكوك حول مصداقيتها خاصة عند تقديمها خارج حجرة الصف. رغم أن تطبيق الاختبار الإلكتروني خارج حجرة الصف له ميزة كبيرة، بشرط توفر مجموعة من الضمانات لتحقق المصداقية؛ لأن ذلك يسمح بتوفير جزء من الوقت للمعلم والطلاب يمكن استثماره في أنشطة التعلم، كما أن تطبيق الاختبار عن بعد يعد نشاطاً تربوياً إذا ما استهدف تنمية الرقابة الذاتية لدى الطلاب، بالإضافة

إلى أن الاختبارات الإلكترونية جزء أساسي من التعلم عن بعد، والذي تتم فيه جميع أنشطة التعلم والتقويم عن بعد من خلال شبكة الإنترنت.

ويمثل الاختبار موقفاً مثيراً من مجموعة من المشكلات يتطلب لحلها مجموعة استجابات تمثل أدلة على امتلاك الممتحن لمستوى محدد من السمة المستهدفة بالقياس. وتتوقف فاعلية الاختبار على خصائصه السيكومترية المرتبطة بالصدق والثبات والقدرة على تمييز المستويات المختلفة للسمة المقاسة. وتتعدد أنواع الاختبار بتعدد محك التصنيف، وإذا اعتبرنا طريقة تقديم الاختبار للممتحنين محكاً للتصنيف ينتج عن ذلك نوعان: الاختبارات الورقية، والاختبارات الإلكترونية التي يمكن أن تقدم بعدة طرق مختلفة داخل الصف وخارجه.

وقد عرف مندور (٢٠١٣: ٣٩٧-٣٩٨) الاختبارات الإلكترونية على أنها الاختبارات التي تتم بواسطة تقنيات الكمبيوتر وشبكاته، فمن خلالها يتم القيام بكافة أنشطة الاختبارات، مثل: إعداد مهام التقييم، وعرضها على الطلاب، واستقبال الاستجابات، وتصحيحها، وتقديم تغذية مرتدة للمستجيبين، وتقدير درجاتهم، ورصد نتائج الاختبار، وتفسيرها، وتخزينها، وتوفير إجراءات الأمان حفاظاً على السرية.

وتضم الاختبارات الإلكترونية جميع الاختبارات التي يمكن تصميمها وإنتاجها من خلال الحاسوب، سواء كانت قائمة على الإنترنت WBT أو قائمة على الكمبيوتر CBT (غريب، ٢٠١٤: ١٤٩)، حيث يتم عرض بنود الاختبار على الشاشة، ويجيب الطالب عنها مباشرة باستخدام أدوات الإدخال (الخزي وذكري،٢٠١١: ١٧١).

وفي هذا السياق يجب أن نميز بين الاختبارات الإلكترونية والاختبارات المكيفة أو التفريدية، حيث إن الاختبارات المكيفة هي نوع خاص من الاختبارات الإلكترونية، يتم فيه اختيار المفردات التي تقدم للممتحن بناء على استجابته للمفردات السابقة، ولذلك يشترط أن تعتمد الاختبارات المكيفة على مفردات مدرجة طبقاً لأحد نماذج نظرية الاستجابة للمفردة، فهي تعتمد على التقدير الآني لقدرة الممتحن من خلال استجابته على مفردة محددة، ثم يقوم النظام بسحب مفردات تناسب هذا المستوى من القدرة؛ بهدف الوصول إلى أعلى درجة ممكنة من دقة تقدير القدرة في أقل زمن ممكن وباستخدام أقل عدد من المفردات، وبذلك يختلف الاختبار المستخدم من ممتحن لآخر، وهو ما يطلق عليه الاختبار المرن (المصري، ٢٠٠٢: ٢٠٤-٢٠٥).

وتتعدد برامج إدارة التعلم الإلكتروني التي تستخدم في بناء الدروس والبرامج التعليمية والاختبارات الإلكترونية على شبكة الإنترنت، وتعمل هذه البرامج على تصميم المحتوى التعليمي

باستخدام قوالب جاهزة، يمكن تعديلها وتركيبها بسهولة، بما يتناسب مع طبيعة المحتوى، ومن أشهر هذه البرامج: Moodle, A tutor, Docbo AT, Clar online, Dokeos, Interact (بدوي، ٢٠١٤: ١٤٧)

ويعد نظام إدارة التعلم "البلاك بورد" Blackboard واحداً من أقوى أنظمة إدارة التعلم الإلكترونية؛ إذ تستخدمه أكثر من ٣٦٠٠ مؤسسة تعليمية على مستوى العالم في تقديم خدمات تعليمية راقية للمعلم والطالب، كما ساعد كثيراً من المؤسسات التعليمية في نشر التعليم بقوة عن طريق الإنترنت لتميزه بالمرونة وقابليته للتطوير (شعيب، ٢٠٠٤: ١٨٣).

أنواع الاختبارات الإلكترونية:

يمكن تصنيف الاختبارات الإلكترونية في فئتين، فهناك الاختبار المعتمد على الحاسوب CBT حيث يتم تطبيق الاختبار دون الحاجة للاتصال بشبكة الإنترنت، والاختبار المعتمد على الشبكة NBT، حيث توفر الشبكة فرصة الاختبار عن بعد، ويعد نظام إدارة التعلم "البلاك بورد" المستخدم في البحث الحالي من النوع الأخير.

وتتعدد مميزات الاختبارات الإلكترونية؛ لأنها تتيح للممتحن الانتقال بسهولة من سؤال لآخر، وكذلك تعديل إجابته دون أن يترك أثراً للتعديل كما هو الحال في الاختبارات التقليدية (البلوي، ٢٠١٣: ٢٠٨)، كما توصلت دراسة التخاينة وأبو موسى (٢٠٠٩) إلى فاعلية الاختبارات الإلكترونية التكوينية في زيادة التحصيل الأكاديمي لطلاب كلية التربية وتحسين اتجاههم نحوها (التخاينة وأبو موسى، ٢٠٠٩: ١١٧).

وتقدم الاختبارات الإلكترونية للمعلم العديد من بدائل تصميم الاختبار، منها: اختيار نوع الأسئلة المناسب؛ حيث تصل أنواع المفردات التي يمكن أن يختار منها المعلم إلى عشرة أنواع، وتتيح إدراج الصور الملونة، أو مقاطع الصوت، أو الفيديو، بالإضافة لتنوع طرق عرض الأسئلة والإجابة عنها، سواء أكانت مرتبة وفق نظام معين أم بشكل عشوائي، بحيث يقدَّم الاختبار الواحد بأكثر من شكل، بما يحد من ظاهرة الغش (بدوي، ٢٠١٤: ١٤٨)، كما تقدم الاختبارات الإلكترونية مميزات هامة تتعلق بتكيف الأداء الاختباري لذوي الاحتياجات الخاصة، وأثناء تعلم اللغات الأجنبية، حيث تتيح التحكم في حجم الخط، وتحويل النص إلى صوت (Hansen & Mislevy, 2005: 214).

كما توفر الاختبارات الإلكترونية الوقت والجهد المستغرق في إخراج الاختبارات وطباعتها ونسخها وتوزيعها على الطلاب وتصحيحها ورصد درجاتها، بالإضافة إلى أن بعض البرمجيات

تقدم تحليلاً إحصائياً لمفردات الاختبار، وتفسيراً دقيقاً لمعاملات صعوبة المفردات ومعاملات التمييز، كما توفر وقت الطالب أثناء أداء الاختبار؛ فقد توصلت دراسة الخزي وذكري (٢٠١١: ١٧٢) إلى أن الوقت المستغرق في أداء الاختبار الإلكتروني أقل بشكل دال مقارنة بالوقت المستغرق في أداء الاختبار الورقي.

ورغم تلك المميزات تواجه الاختبارات الإلكترونية عدة انتقادات، فبالإضافة للشك السائد في مصداقيتها، يرى بعض المربون أن تقديم الاختبار إلكترونياً يمثل عبئاً معرفياً على الممتحن، مما يؤثر على أدائه في الاختبار، بينما بينت دراسة Prisacari & Danielson (2017) أن عدم وجود فرق معنوي في العبء المعرفي الذي ينتج عن الاختبارات يرجع إلى طريقة تقديمه (إلكترونياً أو ورقياً)، وأن الاختبار الإلكتروني يمكن تنفيذه في البيئات التعليمية دون أن يمثل ذلك عبئاً معرفياً على الممتحنين.

وتعد الاختبارات التحصيلية من أهم أدوات التقييم في النظام التعليمي؛ لما تتمتع به من موضوعية، ويتوقف نجاح الاختبار التحصيلي في تحقيق الهدف منه على خصائصه السيكومترية. وتشمل الخصائص السيكومترية للاختبار تلك المعلومات المتعلقة بصدقه وثباته، ويشير المفهوم الإحصائي لثبات الاختبار إلى نسبة التباين الحقيقي إلى التباين الكلي سواء كان مرتبطاً بالسمة المقيسة أو لا يرتبط بها، بينما يشير مفهوم الصدق إلى نسبة التباين الحقيقي المرتبط أو المنسوب للسمة المقاسة إلى التباين الكلي (عودة، ٢٠١٠: ٣٣٤– ٣٣٦). ويقصد بثبات الاختبار مدى الاتساق والدقة في نتائجه، أما الدقة فتعني خلو الدرجة المقدرة على الاختبار من مصادر الخطأ، وأما الاتساق فيعني الاستقرار في علامات الفرد إذا أخذ الاختبار نفسه عدة مرات في نفس الظروف.

ومع تعدد طرق حساب ثبات الاختبار، تم اختيار ثبات الاتساق الداخلي للاختبار في هذا البحث؛ نظراً لمناسبته للاختبارات التحصيلية، وتغلبه على مشكلات التذكر في طريقة إعادة التطبيق، وتأثير الملل والإرهاق على الأداء في طريقة الصور المتكافئة. ويعبر هذا النوع من الثبات عن تجانس الاختبار، وما إذا كانت تشترك مفرداته المختلفة في قياس السمة نفسها، ويقاس بعدة طرق: كمعامل الارتباط بين المفردة والدرجة الكلية للمقياس، والتجزئة النصفية، ومعامل ألفا كرونباخ، ومعامل كيودر– ريتشاردسون. وتستخدم طريقة معامل الارتباط بين المفردة والدرجة الكلية للاختبار عندما نهدف إلى تحديد المفردات التي تتنبأ بالدرجة الكلية للاختبار بشكل أفضل؛ بهدف تقليص طول الاختبار لأسباب متعلقة بالزمن المتاح للتطبيق، حيث يتم حساب معاملات الارتباط بين درجات المفحوصين على المفردة والدرجة الكلية على

الاختبار بعد استبعاد هذه المفردة. أما في طريقة التجزئة النصفية فيتم حساب معامل الارتباط بين درجات الممتحنين على نصفي الاختبار، ومن عيوب هذه الطريقة أن معامل الثبات الناتج يكون منخفضاً نتيجة التجزئة. أما في كل من طريقة ألفا كرونباخ وطريقة كيودر – ريتشاردسون فيتم حساب متوسط جميع معاملات ثبات التجزئة النصفية الممكنة للاختبار، ومن مميزات هذه الطرق إمكانية تحديد المفردات التي تؤدي إلى انخفاض الاتساق الداخلي للاختبار، ويعتمد تفسير قيمة معامل ألفا كرونباخ على عدد المفردات، فكلما زاد طول الاختبار احتجنا قيمة أعلى لألفا كرونباخ كدليل على ثباته، ومن ناحية أخرى فإن ارتفاع قيمة ألفا كرونباخ إلى ٠.٩ يعد مؤشراً لتكرار المفردات، أو بمعنى آخر عدم استقلالية مفردات الاختبارات. (Todd & Bradely, 2001: 29- 30)

ويرتبط صدق الاختبار بثباته، فلكي يكون الاختبار صادقاً يجب أن يكون ثابتاً، والعكس ليس صحيحاً دوماً؛ فالاختبار الثابت ليس بالضرورة أن يكون صادقاً، فقد يكون الاختبار ثابتاً بينما لا يقيس ما يدعي قياسه. وهناك علاقة بين صدق المحتوى والاتساق الداخلي للاختبار، فالاختبار ذو الاتساق الداخلي المرتفع قد يعاني من الإطناب في مفرداته، وقد يشير ذلك إلى عدم تمثيل المفردات جميع الجوانب المختلفة للسمة، لذا يجب أن تتراوح القيم المقبولة لمعامل ألفا كرونباخ بين ٠.٧ إلى ٠.٨ (Todd & Bradely, 2001: 34). ومع تنوع طرق حساب صدق الاختبارات، فإن طريقة صدق المحتوى تعد الأنسب للتحقق من صدق الاختبارات التحصيلية؛ حيث يتم فيها مراجعة مفردات الاختبار، والتأكد من تمثيلها للمحتوى المستهدف بالقياس، ومطابقتها للأهداف السلوكية المقاسة. كما يمكن استخدام طريقة الصدق المرتبط بمحك، التي تعتمد على التحقق من صدق الاختبار من خلال حساب الارتباط بين نتائج الاختبار ونتائج اختبار آخر يقيس نفس السمة، وينقسم الصدق المرتبط بمحك إلى نوعين: الأول هو الصدق التنبؤي، وذلك عندما يكون المحك هو أداء الممتحنين في مهام مستقبلية مرتبطة بالسمة التي يقيسها الاختبار، والثاني هو الصدق التلازمي، وذلك عندما يكون المحك هو أداء الممتحنين على مهام آنية مرتبطة بالسمة التي يقيسها الاختبار. وقد استخدام البحث طريقة صدق المحتوى للتحقق من صدق الاختبار المستخدم وتمثيله للمحتوى المستهدف بالقياس، واستخدام طريقة الصدق التلازمي لدراسة تأثير طريقة تطبيق الاختبار على صدقه.

أهم الدراسات السابقة:

شغلت الاختبارات الإلكترونية اهتمام العديد من الباحثين، وتنوع مجال ذلك الاهتمام كماً وكيفاً، فاهتمت معظم الدراسات بمقارنة أداء الممتحنين بين الاختبارات الإلكترونية والاختبارات الورقية، ومن هذه الدراسات دراسة(Ozalp & Cagiltay (2010 التي هدفت إلى بحث تأثير طريقة تقديم الاختبار (إلكترونياً / ورقياً) على أداء الطلبة في اختبارات مقرر الكيمياء، وبينت النتائج عدم وجود فرق معنوي بين أداء الطلبة على الاختبارات الورقية والاختبارات الإلكترونية. وتتفق نتائج هذه الدراسة مع دراسة (Anakwe (2008 التي هدفت إلى بحث تأثير طريقة تقديم الاختبار (إلكترونياً / ورقياً) على أداء الطلبة في اختبارات مقرر المحاسبة. كما قامت دراسة (Wang, et al. 2008) بعمل تحليل بعدي لنتائج الدراسات التي اهتمت بمقارنة أداء الطلاب في اختبارات القراءة من مرحلة رياض الأطفال حتى الصف الثاني عشر في الولايات المتحدة الأمريكية على كل من الاختبارات الإلكترونية والاختبارات الورقية، وكشفت الدراسة أنه لم يكن لطريقة تقديم الاختبار (إلكترونياً / ورقياً) أي تأثير ذي دلالة إحصائية في أداء الممتحنين في اختبار تحصيل القراءة في هذه المرحلة.

كما هدفت دراسة (Maguire, et al.,2010) إلى مقارنة أداء مجموعتين من الطلاب على الاختبارات الإلكترونية والاختبارات الورقية، وأشارت النتائج إلى أن الطلاب الذين أكملوا جميع التقييمات إلكترونياً كان أداؤهم أعلى بكثير من أداء الطلاب الذين استكملوا جميع التقييمات ورقياً. أما بالنسبة للطلاب ذوي صعوبات التعلم، فقد بينت نتائج دراسة (Flowers, et al., 2011) أن أداء هؤلاء الطلاب كان مرتفعاً بشكل دال في حال أخذ الاختبار ورقياً.

وهدفت دراسة الخزى والزكري (٢٠١١) إلى اختبار تكافؤ الاختبارات الإلكترونية والاختبارات الورقية في قياس التحصيل الجامعي، وأثر تناول الاختبار إلكترونياً على الاتجاه نحو الاختبارات الإلكترونية، وقد بينت النتائج تكافؤ الاختبارات الإلكترونية والورقية في قياس التحصيل الدراسي، مع وجود فروق ذات دلالة إحصائية في الوقت اللازم لأداء الاختبارات لصالح الاختبارات الإلكترونية، وتحسن اتجاه الطلاب نحو الاختبارات الإلكترونية بعد خبرتهم بها. كما توصلت دراسة (Piaw (2012 إلى عدم وجود فروق دالة بين درجات المفحوصين في اختبار للدافعية وآخر للتحصيل ترجع إلى طريقة تقديم الاختبار ورقياً أو إلكترونياً.

واهتمت فئة أخرى من الدراسات بالجانب التقني والفني للاختبارات الإلكترونية، حيث هدفت دراسة هنداوي (٢٠١٠) إلى بحث أثر ثلاث طرق لتصميم الاستجابة في الاختبارات الإلكترونية (استجابة النقطة النشطة، استجابة إدخال النص، استجابة المنطقة المستهدفة) على معدل الأداء الفوري والمؤجل، وتوصلت الدراسة إلى عدم وجود فروق بين التصميمات الثلاث في الأداء الفوري، بينما كان تصميم استجابة المنطقة المستهدفة أكثر فاعلية من التصميمين الآخرين في الأداء المؤجل، ومن النتائج المثيرة للاهتمام في هذه الدراسة إمكانية التحكم في خصائص المفردات وتصميم الاستجابة في الاختبار الإلكتروني؛ لزيادة تفاعل الممتحن مع المفردات أثناء الاستجابة وتقديم التغذية الفورية؛ مما يزيد من بقاء التعلم في الذاكرة، وينقل مكانة الاختبار في العملية التعليمية من مجرد أداة للقياس إلى أداة للتعلم.

كما هدفت دراسة بدوي (٢٠١٤) إلى دراسة فاعلية برنامج مقترح في التعليم الإلكتروني لتنمية مهارات تصميم الاختبارات الإلكترونية والاتجاه نحو التقويم الإلكتروني لدي طلاب الدراسات العليا، وتوصلت الدراسة إلى فاعلية البرنامج المقترح كما ظهر ذلك من خلال تطور كل من الأداء العملي والتحصيلي في كفاية مهارات تصميم وإنتاج الاختبارات الإلكترونية، وكذلك تحسن اتجاه العينة نحو التقويم الإلكتروني.

وهدفت دراسة آل ملوذ والشربيني (٢٠١٥) إلى محاولة التعرف على مدى تحقق معايير الجودة في الاختبارات الإلكترونية من منظور أعضاء هيئة التدريس والطالبات في جامعة الملك خالد، وأوصت الدراسة بوضع لائحة تتضمن ضوابط إجراء الاختبارات الإلكترونية ومعايير جودتها، ورفع الكفاءة المهنية لأعضاء هيئة التدريس في إعدادها، وإنشاء لجنة للاختبارات والتقويم الإلكتروني بجميع كليات الجامعة، تكون مهمتها: فحص الاختبارات الإلكترونية، وتقييم مدى تضمنها معايير الجودة، وتصميم اختبارات إلكترونية توجه الطالب نحو التعلم وتكشف عن إبداعاته، وتفعيل الأنشطة الإلكترونية لتقويم الطلاب بشكل مستمر.

كما هدفت دراسة آل جديع (٢٠١٧) إلى الكشف عن اتجاهات أعضاء هيئة التدريس في جامعة تبوك نحو تطبيق نظام الاختبارات الإلكترونية، ومعرفة المعوقات التي تحول بين عضو هيئة التدريس وتطبيق هذا النوع من الاختبارات، وأشارت نتائج هذه الدراسة إلى وجود اتجاهات إيجابية لدى أعضاء هيئة التدريس نحو نظام الاختبارات الإلكترونية، حيث أكد أعضاء هيئة التدريس أن هذا النوع من الاختبارات يوفر عليهم الوقت والجهد والمال، إلا أن هناك بعض المعوقات التي تحول دون التوسع في استخدام هذا النوع من الاختبارات.

كما اهتمت فئة أخرى من الدراسات بدراسة تأثير عدة عوامل مختلفة على الأداء في الاختبارات الإلكترونية، فتناولت دراسة الخزي (٢٠١٠) أثر قلق الاختبار والجنس وتخصص

الطلاب وخبرتهم في استخدام الحاسوب على أدائهم في الاختبارات الإلكترونية، وتوصلت الدراسة إلى أن التخصص يؤثر على أداء الطلاب في الاختبارات الإلكترونية، وأن هناك علاقة ارتباط عكسية بين خبرة الطلاب بالحاسوب وقلق الاختبار الإلكتروني، وأوصت الدراسة بتعميم الثقافة التكنولوجية في مجال التقويم.

وقامت دراسة Prisacari & Danielson (2017) بدراسة تأثير طريقة تقديم الاختبار على العبء المعرفي الناتج عن أداء الاختبار، وأوضحت النتائج عدم وجود فروق في العبء المعرفي لأداء الاختبار ترجع لطريقة تقديمه.

ومن الدراسات التي اهتمت بمقارنة الخصائص السيكومترية للاختبارات الإلكترونية والاختبارات الورقية دراسة Retnawati (2010) حيث هدفت إلى مقارنة دقة درجات اختبار الكفاءة في اللغة الإنجليزية عند تقديمه ورقياً PPT وعند تقديمه إلكترونياً CPT، فتم حساب ثبات درجات الاختبار بناء على النظرية الكلاسيكية في القياس، وحساب دالة معلومات الاختبار بناء على النظرية الحديثة، وتوصلت الدراسة إلى تشابه تقديرات الثبات للاختبارات الورقية والإلكترونية، وكذلك تشابه دالة المعلومات للعديد من الاختبارات الفرعية للاختبار، إلا أن تقدير دالة المعلومات كانت أفضل في الاختبارات الإلكترونية بالنسبة للمستوى المتوسط من القدرة، وكانت دالة المعلومات أفضل في الاختبارات الورقية بالنسبة للمستويات المتطرفة من القدرة (Retnawati,2010: 135). واتفقت مع نتائج هذه الدراسة دراسة المومني (٢٠١٢) التي قارنت بين معاملات صدق الاختبار وثباته عند تقديمه ورقياً وإلكترونياً، حيث تم تقديم اختبار للغة الإنجليزية للصف الخامس الابتدائي بثلاث طرق مختلفة هي: الطريقة الورقية والطريقة الإلكترونية مع ثبات ترتيب الفقرات بالنسبة للممتحنين، والطريقة الإلكترونية مع اختلاف ترتيب الفقرات بالنسبة للممتحنين، وتوصلت الدراسة إلى عدم وجود فروق ذات دلالة إحصائية في دالة المعلومات التي يقدمها الاختبار أو معاملات الثبات ترجع لطريقة تقديمه، بينما كانت هناك فروق بين مستويات القدرة لصالح الطريقة الثالثة، وتتعارض نتائج هذه الدراسة مع دراسة حجازي والشريف(٢٠١٦) التي هدفت إلى تقصي الخصائص السيكومترية لاختبار تحصيلي في الرياضيات بصورتيه الورقية والإلكترونية، حيث أشارت النتائج إلى وجود فرق ذي دلالة إحصائية بين معاملي الصدق المحكي للاختبار لصالح الصورة الورقية، بينما لا توجد فروق ذات دلالة إحصائية بين معاملات ثبات الاختبار عند تقديمه ورقياً أو إلكترونياً.

يتضح من خلال الدراسات السابقة اهتمام معظمها بمقارنة الأداء في الاختبارات الإلكترونية والورقية، وأن هناك ندرة في الدراسات التي قارنت الخصائص السيكومترية للاختبارات الورقية والاختبارات الإلكترونية، مع وجود تعارض في نتائج هذه الدراسات.

إجراءات البحث:

عينة البحث:

تكونت عينة الدراسة من (٨٣) طالبة، هن مجموع الطالبات المسجلات في ثلاث شعب لمقرر "مهارات التفكير" بالفصل الدراسي الأول ٢٠١٧ م، ويبين الجدول التالي توزيع أفراد العينة طبقاً لطريقة تقديم الاختبار، ونظراً لأن الخصائص السيكومترية للمفردات تتأثر بخصائص عينة الممتحنين، فقد تم اختيار ثلاث مجموعات متكافئة من الطالبات، وتم التحقق من التكافؤ بإجراء البديل اللابارامتري لاختبار تحليل التباين الأحادي، وهو اختبار كروسكال ويلز؛ وذلك لقياس دلالة الفروق بين المجموعات في التحصيل العام، كما يقاس بالمعدل العام للتحصيل GPA. ويتضح من الجدول التالي عدم وجود فروق ذات دلالة إحصائية بين المجموعات الثلاث في معدل التحصيل العام، مما يدل على تكافؤ المجموعات.

جدول (١)

نتائج اختبار كروسكال ويلز لدلالة الفروق بين المجموعات في المعدل العام للتحصيل

المجموعة	طريقة تقديم الاختبار	حجم العينة	متوسط الرتب	كا^{2}	مستوى الدلالة
الأولى	إلكترونياً أثناء المحاضرة	٢٣	٤٣.٨٣	٠.٣٣٣	٠.٨٤٧
الثانية	إلكترونياً عن بعد	٢٩	٤٢.٠٧		
الثالثة	ورقياً	٣١	٤٠.٢٤		

منهج البحث:

للإجابة عن أسئلة البحث تم استخدام المنهج التجريبي، حيث تم اختيار ثلاث مجموعات من الطالبات اللاتي يدرسن المقرر، وتم إعداد اختبار تكويني موحد يقيس التحصيل في مقرر مهارات التفكير، وتم تقديم الاختبار للمجموعة الأولى إلكترونياً أثناء المحاضرة، وأخذت المجموعة الثانية الاختبار إلكترونياً عن بعد، وأخذت الثالثة الاختبار ورقياً.

أداة البحث:

تتمثل أداة البحث في اختبار تحصيلي تكويني في مقرر مهارات التفكير. تم إعداد الاختبار من خلال الخطوات التالية:

١. **تحديد الغرض من الاختبار:** التقويم التكويني لتحصيل الطالبات في مقرر مهارات التفكير.
٢. **تحديد مجال الاختبار:** يغطى الاختبار موضوعات (التفكير الناقد / التفكير الإبداعي / استراتيجية حل المشكلات / استراتيجية اتخاذ القرار / عوامل نجاح تعليم التفكير).
٣. **إعداد جدول مواصفات الاختبار:** حيث تم تحديد الوزن النسبي لكل موضوع بناءً على الزمن المستغرق في التدريس، وتم تحديد الأهمية النسبية لمستويات الأهداف بناءً على نواتج التعلم، وقد اقتصرت المفردات على قياس الأهداف من مستويات التذكر والفهم والتطبيق والتحليل، التي اختلفت أوزانها النسبية تبعاً لكل موضوع من موضوعات المحتوى. ويوضح الجدول التالي جدول مواصفات الاختبار التكويني المستخدم.

جدول (٢)

جدول مواصفات الاختبار التحصيلي

المحتوى	التذكر	الفهم	التطبيق	التحليل	المجموع
مهارات التفكير الناقد (٣٠%)	_	(١٠%) مفردتان	(١٠%) مفردتان	(١٠%) مفردتان	(٣٠%) ٦ مفردات
مهارات التفكير الإبداعي (٣٠%)	_	(١٠%) مفردتان	(١٠%) مفردتان	(١٠%) مفردتان	(٣٠%) ٦ مفردات
استراتيجية حل المشكلات (١٥%)	(٥%) مفردة	(٥%) مفردة	(٥%) مفردة	_	(١٥%) ٣ مفردات
استراتيجية اتخاذ القرار (١٥%)	(٥%) مفردة	(٥%) مفردة	(٥%) مفردة	_	(١٥%) ٣ مفردات
عوامل نجاح تعليم التفكير (١٠%)	_	_	(١٠%) مفردتان	_	(١٠%) مفردتان
المجموع	(١٠ %) مفردتان	(٣٠ %) ٦ مفردات	(٤٠ %) ٨ مفردات	(٢٠ %) ٤ مفردات	(١٠٠ %) ٢٠ مفردة

٤. **كتابة مفردات الاختبار:** تم كتابة (٢٠) مفردة من نوع الاختيار من متعدد، مع مراعاة شروط الصياغة الجيدة.
٥. **التحقق من صدق محتوى الاختبار:** تم التحقق من صدق محتوى الاختبار بمراجعة ثلاثة من أعضاء هيئة التدريس لجدول المواصفات، والحكم على مدى مطابقة المفردات للأهداف المستهدفة بالقياس، ومراجعة المفردات، والتأكد من توفر شروط الصياغة الجيدة لمفردات الاختيار من متعدد.
٦. **إخراج الاختبار:** تم إخراج الاختبار في ثلاث صور:

أولاً: الصورة الإلكترونية باستخدام نظام البلاك بورد أثناء المحاضرة. وتم تحديد خيارات التقديم التي تناسب أداء الاختبار أثناء المحاضرة، وهي:

أ . إظهار البدائل بترتيب عشوائي.

ب . فرض الإكمال: أي ينبغي على المتعلم إنهاء الاختبار في جلسة واحدة وإرساله، ولا يسمح له بإعادة المحاولة.

ج . تعيين مؤقت، وهو ٢٥ دقيقة.

د . تقديم كل المفردات في وقت واحد، أي تظهر جميع الأسئلة للممتحن على شاشة الحاسوب مرة واحدة مما يسمح له بالتنقل بين الأسئلة واختيار نقطة البداية ومراجعة الإجابات.

هـ . ترتيب عشوائي للمفردات؛ فلا تظهر المفردات بالترتيب نفسه لدى جميع الممتحنين.

ثانياً: الصورة الإلكترونية باستخدام نظام البلاك بورد عن بعد. وتم تحديد خيارات التقديم التي تناسب أداء الاختبار عن بعد خارج وقت المحاضرة، حيث تم تحديد خيارات أكثر تقييداً ، وتمثلت هذه الخيارات فيما يلي:

أ . إظهار البدائل بترتيب عشوائي.

ب . فرض الإكمال: أي ينبغي على المتعلم إنهاء الاختبار في جلسة واحدة وإرساله، ولا يسمح له بإعادة المحاولة.

ج . تعيين مؤقت، وهو ٢٥ دقيقة.

د . تقديم كل مفردة على حدة، وحظر الرجوع، أي يظهر للممتحن على شاشة الحاسوب مفردة واحدة في كل مرة، وعليه إرسال الإجابة عن المفردة لتظهر المفردة التالية، ولا يُسمح للممتحن بالرجوع إلى المفردة السابقة مرة أخرى.

هـ . ترتيب عشوائي للمفردات؛ فلا تظهر المفردات بالترتيب نفسه لدى جميع الممتحنين؛ مما يقلل فرص العش.

ثالثاً: الصورة الورقية. وتم توزيع موقع البديل الصحيح في كل مفردة بشكل متساوٍ على مفردات الاختبار، مع كتابة تعليمات الاختبار بشكل واضح.

تطبيق الأدوات: تم تطبيق الاختبار بصوره الثلاثة: الإلكترونية أثناء المحاضرة، والإلكترونية عن بعد، والورقية، في الموعد المحدد للاختبار الفصلي في الخطة الدراسية للمقرر.

أساليب المعالجة الإحصائية المستخدمة: (وقد تم إجراء جميع الأساليب بالاستعانة ببرنامج SPSS)

١. اختبار كروسكال ويلز للمقارنة بين متوسطات رتب المجموعات في معدل التحصيل العام.
٢. معامل الارتباط الثنائي المتسلسل لحساب معاملات تمييز مفردات الاختبار في الطرق الثلاث لتقديم الاختبار.
٣. معادلة كرونباخ ألفا لحساب معامل ثبات الاختبار.
٤. معادلة سبيرمان – براون لحساب معامل ثبات التجزئة النصفية للاختبار.
٥. معامل ارتباط سبيرمان لحساب معامل الصدق التلازمي للاختبار في الطرق الثلاث لتقديم الاختبار.
٦. اختبار فريدمان لاختبار دلالة الفروق بين متوسطات رتب كل من:

- معاملات صعوبة المفردات المحسوبة في كل طريقة من طرق تقديم الاختبار.
- معاملات تمييز المفردات المحسوبة في كل طريقة من طرق تقديم الاختبار.

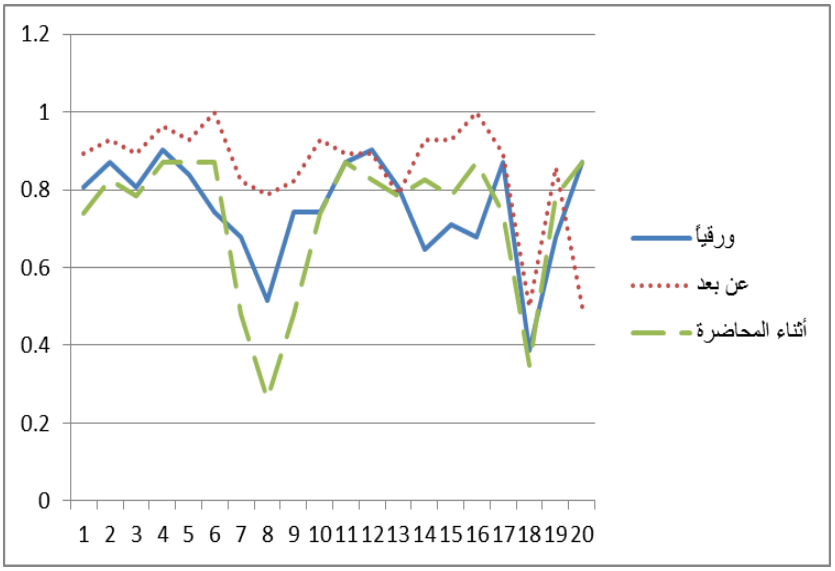
نتائج البحث:

الفرض الأول: "لا توجد فروق ذات دلالة إحصائية عند مستوى دلالة ($\alpha \leq$ ٠.٠٥) بين معاملات صعوبة مفردات الاختبار التحصيلي عند تقديمه إلكترونياً أثناء المحاضرة وتقديمه إلكترونياً عن بعد، وتقديمه ورقياً". لاختبار هذا الفرض تم حساب معاملات صعوبة المفردات عند تقديم الاختبار إلكترونياً أثناء المحاضرة، وعند تقديمه إلكترونياً عن بعد، وعند تقديمه ورقياً، كما هو موضح في جدول رقم (٣)، ثم تم حساب دلالة الفروق بين متوسطات رتب معاملات الصعوبة الناتجة باستخدام اختبار فريدمان Friedman Test، كما هو موضح في جدول رقم (٤).

جدول (٣)

معاملات الصعوبة تبعاً لطريقة تقديم الاختبار

ورقياً	إلكترونياً عن بعد	إلكترونياً أثناء المحاضرة	المفردة	ورقياً	إلكترونياً عن بعد	إلكترونياً أثناء المحاضرة	المفردة
0.87	0.89	0.87	Item 11	0.81	0.89	0.74	Item 1
0.9	0.89	0.83	Item 12	0.87	0.93	0.83	Item 2
0.81	0.79	0.78	Item 13	0.81	0.89	0.78	Item 3
0.65	0.93	0.83	Item 14	0.9	0.96	0.87	Item 4
0.71	0.93	0.78	Item 15	0.84	0.93	0.87	Item 5
0.68	1	0.87	Item 16	0.74	1	0.87	Item 6
0.87	0.89	0.74	Item 17	0.68	0.82	0.48	Item 7
0.39	0.5	0.35	Item 18	0.52	0.79	0.26	Item 8
0.68	0.86	0.78	Item 19	0.74	0.82	0.48	Item 9
0.87	0.5	0.87	Item 20	0.81	0.89	0.74	Item 10



شكل (١) توزيع معاملات الصعوبة عند تقديم الاختبار بالطرق الثلاث

وتراوحت معاملات الصعوبة بين (٠.٢٦) و (٠.٨٧) عند تقديمه إلكترونياً أثناء المحاضرة، وبين (0.5) و (١) عند تقديمه إلكترونياً عن بعد، وتراوحت بين (٠.٣٩) و (٠.٩٠) عند تقديمه ورقياً. ويتضح من الشكل رقم (١) أن معاملات الصعوبة كانت أعلى عند تقديم الاختبار إلكترونياً عن بعد.

جدول (٤)

نتائج اختبار فريدمان للفروق بين معاملات الصعوبة

كا^{٢}	درجات الحرية	عدد المفردات	مستوى الدلالة
١٨.٩	٢	٢٠	٠.٠٠٠

وكانت قيمة كا^{٢} دالة عند مستوى (٠.٠١)، أي أنه توجد فروق ذات دلالة إحصائية بين معاملات صعوبة مفردات الاختبار التحصيلي عند تقديمه إلكترونياً أثناء المحاضرة وتقديمه إلكترونياً عن بعد، وتقديمه ورقياً، ولتحديد مصدر هذه الفروق تم استخدام اختبار ويلكوكسون للفروق بين الرتب Wilcoxon Signed Ranks Test.

جدول (٥)

نتائج المقارنات الثنائية للفروق بين معاملات الصعوبة

	إلكترونياً أثناء المحاضرة – إلكترونياً عن بعد	إلكترونياً أثناء المحاضرة – ورقياً	إلكترونياً عن بعد – ورقياً
Z	٣.١٧٥ –	١.٠٤٥ –	٢.٩٨٧
مستوى الدلالة	٠.٠٠٢	٠.٢٩٦	٠.٠٠٣

يتضح من الجدول السابق عدم وجود فروق ذات دلالة إحصائية بين معاملات صعوبة المفردات عند تقديم الاختبار إلكترونياً أثناء المحاضرة وعند تقديمه ورقياً، بينما هناك فروق ذات دلالة إحصائية عند مستوى دلالة ($\alpha \leq$ ٠.٠٥) بين معاملات صعوبة المفردات عند تقديم الاختبار إلكترونياً أثناء المحاضرة وعند تقديمه إلكترونياً عن بعد. وهناك فروق ذات دلالة إحصائية عند مستوى دلالة ($\alpha \leq$ ٠.٠٥) بين معاملات صعوبة المفردات عند تقديم الاختبار ورقياً وعند تقديمه إلكترونياً عن بعد. ومن ثم فقد كانت معاملات صعوبة المفردات أكبر عند تقديم الاختبار إلكترونياً عن بعد مقارنة بمعاملات الصعوبة عند تقديمه أثناء المحاضرة إلكترونياً أو ورقياً، ويمكن تفسير ذلك اعتماداً على توفر الرقابة على أداء الاختبار أثناء المحاضرة، ومن ذلك نستنتج أن أداء الاختبار التحصيلي عن بعد يترتب عليه ارتفاع في معاملات الصعوبة، أي زيادة نسبة الممتحنين الذين استطاعوا الإجابة عن المفردة إجابة صحيحة، وقد يرجع ذلك إلى احتمالين: الأول هو انخفاض الرقابة الذاتية لدى الممتحنين، مما يترتب عليه الاستعانة بمصادر

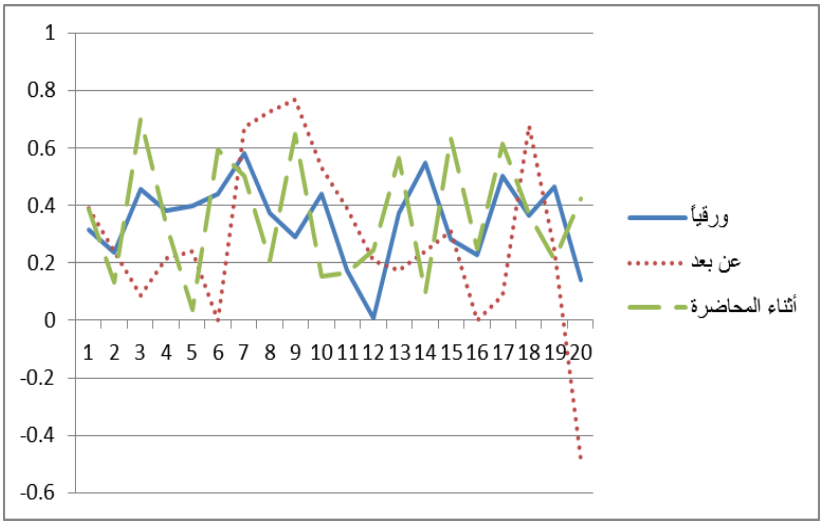
للتعلم أو الاستجابة في مجموعات على فقرات الاختبار. والثاني هو الفرصة المتاحة للممتحنين في التحكم في بيئة الاختبار، وانخفاض معدل قلق الاختبار.

الفرض الثاني: "لا توجد فروق ذات دلالة إحصائية عند مستوى دلالة ($\alpha \leq$ ٠.٠٥) بين معاملات تمييز صعوبة مفردات الاختبار التحصيلي عند تقديمه إلكترونياً أثناء المحاضرة وتقديمه إلكترونياً عن بعد، وتقديمه ورقياً". لاختبار هذا الفرض تم حساب معامل ارتباط بوبينت بايسيريال بين درجات الممتحنين على المفردة ودرجاتهم الكلية، كما هو موضح في جدول (٦)، ثم تم حساب دلالة الفروق بين معاملات التمييز باستخدام اختبار فريدمان Friedman Test، كما هو موضح في جدول رقم (٧).

جدول (٦)

معاملات التمييز تبعاً لطريقة تقديم الاختبار

المفردة	إلكترونياً أثناء المحاضرة	إلكترونياً عن بعد	ورقياً	المفردة	إلكترونياً أثناء المحاضرة	إلكترونياً عن بعد	ورقياً
Item 1	0.38	0.39	0.32	Item 11	0.16	0.39	0.17
Item 2	0.13	0.24	0.24	Item 12	0.25	0.21	0.01
Item 3	0.7	0.09	0.46	Item 13	0.56	0.18	0.37
Item 4	0.34	0.22	0.38	Item 14	0.09	0.24	0.55
Item 5	0.04	0.24	0.4	Item 15	0.63	0.31	0.28
Item 6	0.59	0	0.44	Item 16	0.25	0	0.23
Item 7	0.5	0.67	0.58	Item 17	0.62	0.09	0.5
Item 8	0.21	0.73	0.37	Item 18	0.37	0.68	0.37
Item 9	0.65	0.77	0.29	Item 19	0.21	0.25	0.46
Item 10	0.15	0.53	0.44	Item 20	0.42	-0.49	0.14



شكل (٢) توزيع معاملات التمييز عند تقديم الاختبار بالطرق الثلاث

وقد تراوحت معاملات التمييز بين (٠.٠٠٤) و (٠.٧) عند تقديم الاختبار إلكترونياً أثناء المحاضرة، وبين (-٠.٤٩) و (٠.٧٧) عند تقديمه إلكترونياً عن بعد، ومن (٠.٠٠١) إلى (٠.٥٨) عند تقديمه ورقياً.

جدول (٧)

نتائج اختبار فريدمان للفروق بين معاملات التمييز

كا^2	درجات الحرية	عدد المفردات	مستوى الدلالة
٠.١	٢	٢٠	٠.٩٥١

وكانت قيمة كا^2 غير دالة، إذ إن قيمة الخطأ من النوع الأول بلغت (٠.٩٥١)، لذا لا يمكن رفض الفرض الصفري، أي أنه لا توجد فروق ذات دلالة إحصائية بين معاملات تمييز صعوبة مفردات الاختبار التحصيلي عند تقديمه إلكترونياً أثناء المحاضرة وتقديمه إلكترونياً عن بعد وتقديمه ورقياً، ورغم عدم دلالة الفروق بين معاملات التمييز، يلاحظ من الشكل رقم (٢) أن المفردات (٣، ٥، ٦، ١١، ١٢، ١٦، ١٧، ٢٠) ميزت بطريقة مختلفة تماماً؛ نتيجة اختلاف طريقة تقديم الاختبار.

الفرض الثالث: "لا تختلف قيمة معامل ثبات الاتساق الداخلي للاختبار التحصيلي عند تقديمه إلكترونياً أثناء المحاضرة وتقديمه إلكترونياً عن بعد، وتقديمه ورقياً". ولاختبار هذا الفرض تم حساب معامل الثبات باستخدام كل من طريقة ألفا كرونباخ والتجزئة النصفية، وتصحيح المعامل باستخدام معادلة سبيرمان براون، ويوضح جدول (٨) قيم معاملات الثبات المحسوبة.

جدول (٨)

معامل الثبات باستخدام طريقة ألفا كرونباخ تبعاً لطريقة تقديم الاختبار

طريقة تقديم الاختبار	إلكترونياً أثناء المحاضرة	إلكترونياً عن بعد	ورقياً
قيمة ألفا كرونباخ	٠.٦٦٩	٠.٤٤٢	٠.٦٣٨
قيمة معامل ثبات التجزئة النصفية	٠.٤٧١	٠.١٦٤	٠.٥٧٠
عدد الممتحنين	٢٣	٢٩	٣١

يتضح من الجدول السابق أن أقصى قيمة لألفا كرونباخ (٠.٦٦٩) تم الحصول عليها عند تقديم الاختبار إلكترونياً أثناء المحاضرة، وأن أدنى قيمة لألفا كرونباخ (٠.٤٤٢) تم الحصول عليها عند تقديم الاختبار إلكترونياً عن بعد، وأن قيمة ألفا كرونباخ عند تقديم الاختبار ورقياً بلغت (٠.٦٣٨)، وهي قريبة جداً من قيمتها في حالة تقديم الاختبار إلكترونياً أثناء المحاضرة، وأن أقصى قيمة لمعامل ثبات التجزئة النصفية (٠.٥٧٠) تم الحصول عليها عند تقديم الاختبار ورقياً، وأن أدنى قيمة لمعامل ثبات التجزئة النصفية (٠.١٦٤) تم الحصول عليها عند تقديم الاختبار إلكترونياً عن بعد، وأن قيمة معامل ثبات التجزئة النصفية عند تقديم الاختبار إلكترونياً أثناء المحاضرة (٠.٤٧١) كانت قريبة من قيمتها في حالة تقديم الاختبار ورقياً.

ويمكن تفسير انخفاض ثبات الاختبار عند تقديمه إلكترونياً عن بعد في ضوء معاملات الصعوبة؛ حيث تؤثر صعوبة المفردات في قيم معاملات الثبات، فيزيد معامل الثبات كلما اقتربت قيم معاملات الصعوبة من (٠.٥)، وعندما يتكون الاختبار من مجموعة مفردات متطرفة الصعوبة يقل تباين درجات الممتحنين على الاختبار، مما ينعكس سلباً على معامل الثبات، فعندما تكون مجموعة المفردات ذات معامل صعوبة مرتفع ينجح معظم الممتحنين في الإجابة عنها، وعندما تكون مجموعة المفردات ذات معامل صعوبة منخفض يفشل معظم الممتحنين في الإجابة عنها، وفي كلتا الحالتين لا ينجح الاختبار في الكشف عن الفروق بين الأفراد في السمة المقاسة، ويقل تباين الدرجات، وبالتالي ينخفض معامل الثبات.

كما أن معامل ثبات التجزئة النصفية كان متشابهاً عند تقديم الاختبار ورقياً وعند تقديمه إلكترونياً أثناء المحاضرة، وقد ترجع الزيادة الطفيفة في قيمة معامل ثبات التجزئة النصفية عند تقديم الاختبار ورقياً إلى زيادة عدد الممتحنين الذين أخذوا الاختبار ورقياً، فكلما زاد حجم العينة ارتفع معامل الثبات.

الفرض الرابع "لا تختلف قيمة معامل الصدق التلازمي للاختبار التحصيلي عند تقديمه إلكترونياً أثناء المحاضرة وتقديمه إلكترونياً عن بعد، وتقديمه ورقياً". لاختبار هذا الفرض تم حساب معامل ارتباط سبيرمان بين درجات المجموعات على الاختبار الفصلي ودرجاتهم على الاختبار النهائي، حيث تم استخدام الاختبار النهائي كمحك. وذلك نظراً لعدم تحقق شرط اعتدالية توزيع الدرجات؛ حيث كان توزيع الدرجات سالب الالتواء، وهذا يتسق مع نتائج الاختبارات التحصيلية في البرامج المعتمدة؛ حيث يتوقع أن يحقق غالبية المتعلمين نواتج التعلم. ويوضح جدول (١٠) معاملات الصدق التلازمي للاختبار المقابلة للطرق الثلاث لتقديمه، كما يعبر عنها معامل ارتباط سبيرمان.

جدول (١٠)

معامل الصدق التلازمي للاختبار تبعاً لطريقة تقديمه

طريقة تقديم الاختبار	إلكترونياً أثناء المحاضرة	إلكترونياً عن بعد	ورقياً
معامل الصدق التلازمي	٠.٧٩١**	٠.٢٤٧	٠.٦٧٠**
مستوى الدلالة	٠.٠٠٠	٠.١٩٧	٠.٠٠٠
عدد الممتحنين	٢٣	٢٩	٣١

يتضح من الجدول السابق أن معامل الصدق التلازمي للاختبار كان دالاً إحصائياً عند مستوى دلالة (٠.٠١) عند تقديم الاختبار إلكترونياً أثناء المحاضرة أو ورقياً، مع الأفضلية عند تقديمه إلكترونياً أثناء المحاضرة (٠.٧٩١)، بينما كان معامل الصدق التلازمي للاختبار ضعيفاً وغير دال عند تقديمه إلكترونياً عن بعد (٠.٢٤٧)، مما يعد مؤشراً على تدني مصداقية الاختبارات التحصيلية عند تقديمها إلكترونياً عن بعد، وتتسق تلك النتيجة مع ارتفاع معاملات صعوبة المفردات حال تقديم الاختبار إلكترونياً عن بعد، مما يدل على أن تقييد أداء الاختبار بعامل الوقت وتوزيع كل من المفردات والبدائل بشكل عشوائي وحظر الرجوع للمفردات لا يكفي لتحقيق مصداقية الاختبار التحصيلي عند تقديمه إلكترونياً عن بعد.

مناقشة النتائج والتوصيات:

تتسق نتائج البحث الحالي في الجزء الخاص بتقديم الاختبار إلكترونياً أثناء المحاضرة وورقياً مع نتائج دراسة Retnawati (2010) ودراسة المومني (٢٠١٢) اللتين أشارتا إلى تشابه تقديرات الصدق والثبات للاختبارات الورقية والإلكترونية. بينما تتعارض مع دراسة حجازي والشريف (٢٠١٦) فيما يتعلق بوجود فرق ذي دلالة إحصائية بين معاملي الصدق المحكي للاختبار لصالح الصورة الورقية، وتتفق معها فيما يتعلق بعدم دلالة الفروق بين معاملات تمييز مفردات الاختبار عند تقديمه ورقياً أو إلكترونياً. ولم تتناول أي دراسة سابقة – في حدود علم الباحثة – الخصائص السيكومترية للاختبارات الإلكترونية عند تقديمها عن بعد.

ويوصي البحث المعلمين باستخدام الاختبارات الإلكترونية، والاستفادة من مزاياها، على أن يتم تقديم الاختبار في الصف، مع ضبط بيئة الاختبار بشكل مقنن بالنسبة لجميع الممتحنين، كما يوصي بعدم الاعتماد على نتائج الاختبارات الإلكترونية عند تقديمها عن بعد. ولأن هذه الاختبارات لا غنى عنها في برامج التعلم عن بعد فإن البحث ينشد من مطوري تقنيات الاختبارات الإلكترونية إضافة مزيد من خيارات ضبط أداء الاختبار الإلكتروني عند تقديمه عن بعد، كأن يرتبط دخول الممتحن بالبصمة الشخصية، أو أن يقوم جهاز الحاسوب الذي يقدَّم عليه الاختبار بتسجيل صورة مقدم الاختبار. كما يوصي البحث الحالي بإجراء مزيد من الأبحاث التي تتناول نفس المتغيرات عند استخدام أنواع مختلفة من المفردات، وعند قياس مستويات مختلفة من الأهداف.

المراجع:

١. آل جديع، مفلح بن قبلان بن بجاد (٢٠١٧). اتجاهات أعضاء هيئة التدريس نحو إجراء الاختبارات الإلكترونية ومعوقات تطبيقها بجامعة تبوك. *المجلة الدولية التربوية المتخصصة*، ٦ (٢)، ٧٧-٨٧.

٢. آل ملوذ، حصة محمد و الشربيني، غادة حمزة (٢٠١٥). معايير جودة الاختبارات الإلكترونية من وجهة نظر أعضاء هيئة التدريس والطالبات في جامعة الملك خالد. *المجلة الدولية التربوية المتخصصة*، ٤ (٤)، ٢٥-٤٢.

٣. بدوي، محمد محمد عبد الهادي (٢٠١٤). فعالية برنامج مقترح في التعليم الإلكتروني لتنمية مهارات تصميم الاختبارات الإلكترونية. *المجلة الدولية التربوية المتخصصة*، ٣ (٥)، ١٤٦-١٧٦.

٤. البطاح، عماد رمضان (٢٠٠٧). *أثر طريقة اختيار فقرات الاختبار في خصائصه السيكومترية* (رسالة ماجستير غير منشورة). جامعة مؤته، الكرك – الأردن.

٥. البلوي، سالم بن عبد الرحمن (٢٠١٣). التحقق من فاعلية برنامج اختباري محوسب في العملية الاختبارية. *مجلة القراءة والمعرفة –مصر*، (١٣٨)، ١٩٧-٢١٤.

٦. التخاينة، بهجت حمد عفنان و أبو موسى، مفيد (٢٠٠٩). أثر استخدام استراتيجية التقويم التكويني المحوسب في تحصيل طلبة التربية في الجامعة العربية المفتوحة واتجاهاتهم نحوها. *مجلة دراسات في المناهج وطرق التدريس – مصر*، (١٤٥)، ١١٦-١٣٢.

٧. الحجاج، أيمن عودة (٢٠١٣). *أثر ترتيب الفقرات حسب خصائصها السيكومترية على أداء الأفراد والخصائص السيكومترية للاختبار (رسالة ماجستير غير منشورة)*. جامعة مؤتة، الكرك –الأردن.

٨. حجازي، تغريد عبد الرحمن محمد، و الشريف، نايف (٢٠١٧). الخصائص السيكومترية لصورتي اختبار في الرياضيات للمرحلة المتوسطة: دراسة مقارنة. *مجلة الدراسات التربوية والنفسية – سلطنة عمان*، ١١ (١)، ١٧٨-١٩٠.

٩. الخزي، فهد عبد الله (٢٠١٠). أثر قلق الاختبار وبعض المتغيرات الديموغرافية على أداء طلبة جامعة الكويت في الاختبارات الإلكترونية: دراسة وصفية ارتباطية.

٧ *مجلة جامعة صنعاء للعلوم التربوية والنفسية –اليمن*، (١)، ٢١٩-٢٧٠.

١٠. الخزي، فهد عبد الله و الزكري، محمد بن إبراهيم (٢٠١١). تكافؤ الاختبارات الإلكترونية مع الاختبارات الورقية في قياس التحصيل الدراسي: دراسة تجريبية على طلبة كلية التربية بجامعة الكويت. *مجلة دراسات الخليج والجزيرة العربية (الكويت)*، (١٤٣)، ١٦٧-١٩٨.

١١. الذنيبات، سجى أحمد (٢٠١٣). *أثر اختلاف نمط الاختبار على الخصائص السيكومترية للاختبار* (رسالة ماجستير غير منشورة). جامعة مؤتة، تكريت –الأردن.

١٢. شعيب، إيمان محمد مكرم مهني (٢٠١٤). أثر برنامج تدريبي مقترح لإكساب أعضاء هيئة التدريس بجامعة حائل مهارات بناء الاختبارات الإلكترونية بنظام إدارة التعلم بلاك بورد Blackboard. *مجلة دراسات عربية في التربية وعلم النفس – السعودية*، (٥٣)، ١٧٩-٢٠١.

١٣. عودة، أحمد سليمان (٢٠١٠): القياس والتقويم في العملية التدريسية، دار الأمل، إربد.

١٤. غريب، أحمد محمود فخري (٢٠١٤). أثر اختلاف أدوات التشارك بالفصول الافتراضية على إكساب مهارات تصميم وإنتاج الاختبارات الإلكترونية. مجلة *تكنولوجيا التعليم* –مصر، ٢٤ (١)، ١٤١-١٨٨.

١٥. المصري، محمد عبد المجيد (٢٠٠٢). الاختبارات المكيفة: أنموذج معاصر في قياس السمات والخصائص النفسية والتربوية. *مجلة البحوث والدراسات التربوية –اليمن*، (١٦)، ٩٦-١١٢.

١٦. معان، عماد ثابت (٢٠١٢). أثر استخدام الاختبارات الإلكترونية التشعبية في التدريبات الرياضية على حل المسائل الرياضية وتخفيف القلق الرياضي لدى تلاميذ التعليم الإعدادي بسوهاج. *المجلة التربوية* –مصر، (٣١) ، ٤٣-٦٥.

١٧. مندور، إيناس محمد الحسيني (٢٠١٣). أثر برنامج تدريبي لطلاب الدراسات العليا بكلية التربية في تصميم الاختبارات الإلكترونية وفقاً لمعايير الجودة المقترحة. *مجلة دراسات تربوية واجتماعية* –مصر، ١٩ (٢)، ٣٩١-٤٦٠.

١٨. المومني، صبحة أحمد (٢٠١٢). *أثر طريقة تقديم اختبار اللغة الإنجليزية لدى طلبة الصف الخامس الأساسي على تقديرات القدرة وخصائص الاختبار وخصائص فقراته* السيكومترية (رسالة ماجستير غير منشورة). جامعة اليرموك، إربد.

١٩. هنداوي، أسامة سعيد علي (٢٠١٠). أثر ثلاثة تصميمات لأنماط الاستجابة على الاختبارات الإلكترونية على معدل الأداء الفوري والمؤجل لطلاب الجامعة في الاختبار. *مجلة العلوم التربوية* –مصر، ١٨ (٣)، ١٠٢-١٤٣.

20. Anakwe, B. (2008). Comparison of student performance in paper-based versus computer-based testing. *Journal of Education for Business*, 84 (1), 13-17.

21. Flowers, C., Kim, D., Lewis, P., & Davis, V. C. (2011). A comparison of computer-based testing and pencil-and-paper testing for students with a read-aloud accommodation. *Journal of Special Education Technology,* 26 (1), 1-12.

22. Hansen, E. G., & Mislevy, R. J. (2005). Accessibility of Computer-Based Testing for Individuals with Disablities and English Language Learners within a Validity Framework. In S. Howell, & M. Hricko (Eds.), *Online Assessment and Measurement: Foundations and Challenges, (214-262).*

23. Maguire, K. A., Smith, D. A., Brallier, S. A., & Palm, L. J. (2010). Computer-based testing: A comparison of computer-based and paper-and-pencil assessment. *Academy of Educational Leadership Journal*, 14 (4), 117.

24. Ozalp-Yaman, S., & Cagiltay, N. E. (2010). Paper-based versus computer-based testing in engineering education. *Paper presented at Education Engineering Conference: (EDUCON)*. 1631-1637.

25. Piaw, C. Y. (2012). Replacing paper-based testing with computer-based testing in assessment: Are we doing wrong*? Procedia - Social and Behavioral Sciences*, 64, 655-664.

26. Prisacari, A. A., & Danielson, J. (2017). Computer-based versus paper-based testing: Investigating testing mode with cognitive load and scratch paper use*. Computers in Human Behavior,* 77, 1-10.

27. Retnawati, H. (2015). The Comparison of Accuracy Scores on the Paper and Pencil Testing vs. Computer-Based Testing. *TOJET: The Turkish Online Journal of Educational Technology*, 14 (4), 135-142.

28. Todd, C. & Bradely, C. (2001). Evaluating the design and development of psychological scales (Eds.), *Handbook of Psychology and Diabetes: A Guide to Psychological Measurement in diabetes research and practice, (pp.15-42)*. New York: psychological press.

29. Wang, S., Jiao, H., Young, M. J., Brooks, T., & Olson, J. (2008). Comparability of computer-based and paper-and-pencil testing in K–12 reading assessments: A meta-analysis of testing mode effects*. Educational and Psychological Measurement*, 68 (1), 5-24.