البحوث والدراسات

تكافؤ الاختبارات الإلكترونية مع الاختبارات الورقية في قياس التحصيل الدراسي: دراسة تجريبية على طلبة كلية التربية بجامعة الكويت

د. فهد عبدالله الخزي*
د. محمد إبراهيم الزكري**

ملخص:

هدفت الدراسة إلى اختبار مدى التكافؤ بين الاختبارات الإلكترونية والاختبارات الورقية في قياس التحصيل الدراسي الجامعي، ومدى تأثير تعرض الطلبة للاختبارات الإلكترونية على اتجاهاتهم نحوها. وقد استخدم المنهج التجريبي حيث أعطي (٣١٦) طالباً وطالبة في كلية التربية بجامعة الكويت نسختين متماثلين من الاختبارات: ورقية وإلكترونية؛ وصاحب ذلك قياس اتجاه الطلبة، باستخدام استبانة نحو الاختبارات الإلكترونية قبل تعرضهم لها وبعده. وقد أظهرت النتائج: (أ) تكافؤ الاختبارات الإلكترونية والورقية في قياس التحصيل الدراسي للطلاب مع وجود فروق ذات دلالة إحصائية في الوقت اللازم لأداء الاختبار لصالح الاختبارات الإلكترونية. (ب) ارتفاع اتجاهات الطلاب نحو الاختبارات الإلكترونية بسبب تعرضهم لها. وقد أوصت الدراسة بتبني استخدام الاختبارات الإلكترونية في التعليم

* دكتوراه في تكنولوجيا التعليم والإحصاء التطبيقي، جامعة كولورادو الشمالية، الولايات المتحدة الأمريكية، عام ٢٠٠٢م، وأستاذ مساعد بقسم المناهج وطرق التدريس، كلية التربية، جامعة الكويت، دولة الكويت.

** دكتوراه في تكنولوجيا التعليم، جامعة ولاية كانساس، الولايات المتحدة الأمريكية، عام ٢٠٠٠م، وأستاذ مشارك بقسم التربية، جامعة الإمام محمد بن سعود، المملكة العربية السعودية.

الجامعي في التخصصات التربوية والإنسانية المشابهة، وبإجراء دراسات مماثلة في التخصصات الأخرى.

مقدمة:

للتقويم أهمية كبيرة وجوهرية في التعلُّم والتعليم؛ فعن طريقه تتحدد درجة تحقق الأهداف التعليمية؛ ومن ثم يمكن الحكم على أداء الطلبة وتحصيلهم العلمي ودرجة اكتسابهم للمهارات والمعارف والعلوم. ولا تتوقف أهمية التقويم عند ذلك، بل إنه يساعد على تعديل عمليات التعلُّم والتعليم وتطويرها. تشير Sutton (2004) إلى أن تحسين أساليب التقويم وتطويرها يعتبر واحداً من خمسة أهداف للمجتمع الأكاديمي. كما أن مؤسسات الاعتماد الأكاديمي على مستوى العالم – كرابطة الشمال المركزية (NCA) و (NCATE) – تعد التقييم واحداً من المعايير اللازمة للاعتماد، وقد قام الكثير من تلك المؤسسات بإصدار أدلة تهدف إلى مساعدة المؤسسات التربوية على تطوير وتحسين برامج التقييم الخاصة بها (NCATE, 2008).

وقد تزايد الاهتمام بشكل كبير – خلال العقدين السابقين – بحوسبة التعليم والاستفادة من تقنية المعلومات، في معظم جوانب العملية التعليمية – الإدارية والتدريسية – وعمليات التعلُّم والتواصل مع المجتمع والأهالي (Davidson, 2003). إلا أن هذا الاهتمام لم يطل عمليات التقويم والاختبارات بالدرجة الكافية على جميع المستويات التعليمية الأساسية والعليا. يشير إلى ذلك عددٌ من الباحثين مثل Davidson (2003) وعدد من الدراسات والتقريرات مثل الدراسة التي عرضها Bennett (1999) على الكونجرس الأمريكي، وتبيَّن أن أكبر مشكلة أمام التعليم الحديث هي أن أدوات التقويم التي نحاول أن نقيس بها مهارات اليوم هي نفسها أدوات الأمس.

ويؤكد Russell et al. (2003) أن استخدام الكمبيوتر لحوسبة الاختبارات في المدارس والجامعات الأمريكية كان غائباً حتى بداية هذا القرن. عندها بدأت الولايات المتحدة الأمريكية باستخدام الاختبارات الإلكترونية باطراد؛ كما أن دولاً أخرى مثل سنغافورة والنرويج بدأت بدراسة واستكشاف طرق لحوسبة أدوات

التقويم والقياس لطلبة المدارس العامة والجامعات (Ashton et al. 2003; Nugent, 2003). ومن المتوقع أن تزداد الدول وتزداد المحاولات لتحويل الاختبارات الورقية إلى إلكترونية، إلا أن النمو في هذا الاتجاه يظل بطيئاً عالمياً.

وعلى الرغم من نمو الإحساس بأهمية الاختبارات الإلكترونية وفائدتها المتزايدة، فإن كثيراً من المعلِّمين يميلون إلى تفضيل استخدام الاختبارات الورقية التقليدية، مشككين في تكافؤ الاختبارات الإلكترونية معها (Davidson, 2003).

ومن هنا، فإن دراسة كفاءة الاختبارات الإلكترونية غاية في الأهمية، وهو أمر تفرضه الفجوة الحاصلة بين التوجُّه المتسارع في استخدام الحاسوب في عمليات التعلُّم والتعليم من جهة، والبطء الشديد في اعتماد الاختبارات الإلكترونية من جهة أخرى.

مشكلة الدراسة:

شهدت السنوات العشرون الأخيرة تأثيراً واسعاً لتقنيات الحاسوب والمعلومات في جميع مراحل التعليم في دول العالم المتقدم، كالولايات المتحدة الأمريكية وبريطانيا واليابان وكوريا الجنوبية. وقد تمثل هذا التأثير بشكل رئيس في استخدام تقنيات الحاسوب والمعلومات كأدوات تدريس وتعلُّم ومصادر للبحث العلمي. وعلى الرغم من التأثير الهائل لهذه التقنيات في التعليم، فإن تأثيرها في الاختبارات ظل قليلاً (Davidson, 2003). وبعبارة أخرى لم يكن استخدام الاختبارات الإلكترونية لقياس التحصيل الدراسي بالمستوى نفسه لاستخدام الأدوات الإلكترونية في التدريس والتعلُّم والبحث العلمي.

لم يكن تأثير تقنيات الحاسوب والمعلومات في التعليم بجميع مراحله منحصراً في دول العالم المتقدم، بل وصل أيضاً إلى مؤسسات التعليم في الشرق الأوسط؛ حيث شهدنا في السنوات العشر الأخيرة استخدامات متنوعة لتقنيات الحاسوب والمعلومات في مدارس التعليم الأساسي ومؤسسات التعليم العالي. ولكن ظلَّت هذه الاستخدامات منحصرة في التدريس والتعلُّم، ولم يكن من ضمنها استخدام الأدوات الإلكترونية في قياس التحصيل الدراسي. وربما

يعود السبب في تدني استخدام الاختبارات الإلكترونية لقياس التحصيل العلمي إلى الظن في عدم مصداقيتها وموثوقيتها لدى التربويين.

حالياً، قد يتنازع استخدام الاختبارات الإلكترونية في التعليم في العالم العربي توجهان: الأول مندفع نحو استخدامها والآخر متوجس. ويرى أصحاب التوجه المندفع أنه سيكون لها الأثر نفسه في التدريس والتعلُّم، في حين يذهب أصحاب الرأي الآخر إلى أن استخدام الاختبارات الإلكترونية يمثل نوعاً من المجازفة التي لا داعي لها؛ لأنه لا يتوافر فيها المصداقية والموثوقية التي تتوافر في الاختبارات الورقية. والحقيقة أن كلا التوجهين لا يدعمهما مسوغ منطقي ولا أبحاث أو دراسات علمية قاطعة.

وهنا تكمن مشكلة الدراسة الحالية في أنها تختبر مدى تكافؤ الاختبارات الإلكترونية مع الاختبارات الورقية في قياس التحصيل الدراسي على المستوى الجامعي. ومن ثم قد تعطي مؤشرات حول إمكانية الوثوق في الاختبارات الإلكترونية من عدمها، سعياً إلى استخدامها واستغلال الجوانب الإيجابية التي توفرها، أو الاقتناع بعدم استخدامها استناداً إلى مسوغات علمية وليس إلى مجرد تكهنات وآراء شخصية.

أسئلة الدراسة:

تحديداً، تجيب الدراسة عن أربعة أسئلة، هي:

١ – هل الاختبارات الإلكترونية متكافئة في قياس التحصيل الدراسي الجامعي مع الاختبارات الورقية؟

٢ – هل هناك تكافؤ في الوقت المستغرق للإنجاز بين الاختبارات الإلكترونية والاختبارات الورقية؟

٣ – ما اتجاهات طلبة المرحلة الجامعية نحو الاختبارات الإلكترونية قبل التعرُّض لها؟

٤ – هل يؤثر تعرض طلبة المرحلة الجامعية للاختبارات الإلكترونية على اتجاهاتهم نحوها؟

المصطلحات الإجرائية:

أينما وردت في هذه الدراسة المصطلحات التالية فإن لها معاني إجرائية محددة، وذلك على النحو التالي:

- الاختبارات الإلكترونية (أو الاختبارات المحوسبة): يُقصد بها الاختبارات التي تتم من خلال الحاسوب، حيث يتم عرض بنود الاختبار على الشاشة، ويجيب الطالب عنها مباشرة باستخدام أدوات الإدخال (لوحة المفاتيح، الفأرة، شاشة اللمس).
- التحصيل الدراسي: الدرجة التي يحصل عليها الطالب في اختبار التحصيل سواء كان إلكترونياً أم ورقياً (مطبوعاً)، بغض النظر عن مستوى هذه الدرجة.
- الاتجاه نحو الاختبارات الإلكترونية: ميل نفسي لتقييم الاختبارات الإلكترونية بدرجة من التفضيل أو عدم التفضيل.

حدود الدراسة:

الحدود الزمانية: تم تطبيق الدراسة خلال الفصلين الأول والثاني من العام الجامعي (٢٠٠٩/ ٢٠١٠).

الحدود المكانية: تم تطبيق الدراسة في كلية التربية بجامعة الكويت.

الحدود البشرية: تم تطبيق الدراسة على طلبة كلية التربية (درجة البكالوريوس) المسجلين في مقرر الحاسوب في التربية.

أهداف الدراسة:

إن التوسع في استخدام تقنية المعلومات في التدريس من جهة، والإحجام عن استخدام القياس الإلكتروني من جهة أخرى، سوف يُعمِّق الفجوة الرقمية بين عمليات التدريس وعمليات التقويم؛ إذ من الأفضل عند تبني أدوات إلكترونية في التدريس أن نتبنى أدوات إلكترونية في القياس والتقويم (Morris et al., 1977). إلا أنه يجب التأكد من أن الأدوات الإلكترونية المزمع استخدامها ذات فاعلية ومصداقية لا تقل عن نظيراتها الورقية. وعليه فإن للدراسة الحالية هدفين، هما:

١ – تحديد كفاءة (Reliability) الاختبارات الإلكترونية مقارنةً بالاختبارات الورقية؛ وذلك عن طريق تعرُّف الفروق بين أداء الطلبة في الاختبارات المحوسبة مقابل أدائهم في الاختبارات الورقية؛ وعن طريق تعرُّف الوقت المستغرق في أداء الطلبة للامتحانين: الورقي والإلكتروني.

٢ – دراسة اتجاهات الطلبة نحو الاختبارات الإلكترونية، وإذا ما كان تعرضهم لها يؤثر في اتجاهاتهم نحوها.

أهمية الدراسة:

تنبع أهمية الدراسة من ناحيتين: الناحية الأولى أنها ستساعد التربويين في اتخاذ القرارات المناسبة حول تبنِّي استخدام الاختبارات الإلكترونية. ومن ثم فإنها ستسهم في معالجة الفجوة الحاصلة بين الاستخدام المتزايد للحاسوب في عمليات التعلُّم والتعليم من ناحية، وبين الإحجام الكبير عن استخدام الحاسوب في الاختبارات وأساليب التقويم من ناحية أخرى. وتنبع الأهمية الثانية في أنها ستكون من الدراسات الأساسية حول تحديد فاعلية الاختبارات الإلكترونية في المراحل الجامعية وفي التخصصات التربوية في العالم العربي؛ إذ إن الدراسة الحالية تبحث في موضوع لم يعط حقه من البحث، وخاصةً من قِبل الباحثين العرب. وعليه فإنها ستفتح المجال لدراسات مشابهة، وعلى مستويات تعليمية مختلفة وفي تخصصات تطبيقية وسلوكية أخرى.

الإطار النظري

نشأة الاختبارات الإلكترونية:

تاريخياً، بدأ استخدام الاختبارات الإلكترونية في أواخر الستينيات من القرن الماضي، من خلال وزارة الدفاع الأمريكية، التي اعتمدت على علماء النفس الإكلينيكيين الذين يُعدُّون رواد الاختبارات الإلكترونية المحوسبة (Russell et al., 2003). وخلال تلك الفترة، قام عدد من الباحثين بإجراء دراسات مقارنة بين الاختبارات الإلكترونية والورقية، من أجل تحديد كفاءة الاختبارات الإلكترونية وصدقها وثباتها في مقاييس الشخصية والذكاء (Morris et al., 1977).

أما استخدام الاختبارات الإلكترونية في مجال التعليم، فلم يكن حولها دراسات تُذكر قبل عام ١٩٨٦م. وبعد هذا التاريخ، ظهرت أبحاث ركَّزت على دراسة درجة التكافؤ بين الاختبارات الإلكترونية والورقية، وتحديد العوامل التي تؤثر على أداء المستخدمين للاختبارات الإلكترونية (Russell et al., 2003).

حالياً، يوجد ازدياد متسارع في استخدام الحاسوب لأغراض التقويم في مؤسسات التعليم العالي حول العالم (Akdemir & Oguz, 2008). كما أن هناك اتجاهات حكومية تسعى إلى استخدام الاختبارات المحوسبة في مراحل التعليم العام. من هذه الاتجاهات الاستراتيجية التي تبنَّتها الحكومة البريطانية لاستخدام الاختبارات المحوسبة في المرحلة الدراسية الثالثة (١٣ – ١٤ سنة)؛ بحلول عام ٢٠٠٨م (Ashton et al., 2003; Nugent, 2003).

خصائص الاختبارات الإلكترونية:

حدد عددٌ من الباحثين خصائص الاختبارات الإلكترونية. من هؤلاء الباحثين Ryan et al. (2000) الذين أشاروا إلى وجود خمس خصائص للاختبارات الإلكترونية هي: التفاعلية، والمرونة، وتوفير الوقت، والتغذية الراجعة الفورية، واختصار الموارد، والاحتفاظ بالسجلات. ويضيف Bennett, (2001) خصائص أخرى هي: سهولة مراجعة بنود الاختبار، وتوفير تكلفة الطباعة، وتقديم مثيرات ديناميكية بالصوت والصورة والرسوم المتحركة دون الحاجة لأجهزة أخرى، وإعطاء الدرجات للطلبة بشكل آني وفوري ودون تحيز، والحد من الحاجة لقيام المصححين بالقراءة والمراجعة، وتوزيع نتائج الاختبارات دون تكلفة الطباعة وإرسال التقريرات. ويتفق De Beer & Visser (1998) على أن الأمان وتوفير الوقت ودقة النتائج وسرعة التصحيح هي من المميزات الجوهرية للاختبارات الإلكترونية. ويؤكد ذلك أيضاً Schmit & Ryan (1993) بقولهما: إن من مزايا الاختبارات الإلكترونية سهولة الإجراء وآنية التصحيح والتقليل من أخطاء التصحيح في الاختبارات الورقية. ويضيف Wiechmann, & Ryan (2003) أن

الاختبارات الإلكترونية تُستخدم في كثير من الأحيان بغرض زيادة الاختبارات المعيارية، وتوفير الوقت اللازم للتصحيح، وتخفيض الكلف المادية.

يتضح مما سبق تركيز الباحثين على المميزات المتنوعة للاختبارات الإلكترونية، سواء ما يتعلق منها بالسرعة وتوفير الجهد والدقة والعدالة في التقييم، وما يتعلق منها بتوفير الكلف المادية، وثالثها ما يتعلق بالجوانب التصميمية كالمؤثرات الصوتية والمتحركة التي يمكن إضافتها إلى الاختبارات الإلكترونية.

عيوب الاختبارات الإلكترونية:

لا تقتصر الاختبارات الإلكترونية على المزايا، بل لها عدد من العيوب التي بينها بعض الباحثين. يرى Bennett (2001) أن من أهم عيوب الاختبارات الإلكترونية أن إعدادها يستهلك وقتاً طويلاً، لما تتطلبه من مهارات وتدريب لعمليات الإعداد والتنفيذ. ويضيف أن قياس المهارات العليا أمر صعب في الاختبارات المحوسبة كونها تعتمد – بشكل عام – على نوعية الاختبارات الموضوعية، كما أن إمكانية وجود أعطال في الأجهزة أمر وارد في أثناء الاختبار. يضاف إلى ذلك، أن الاختبارات الإلكترونية تتطلب مهارات وخبرة جيدة بتكنولوجيا المعلومات من قِبل الطالب والأستاذ، وهذا يستوجب تدريباً كافياً بعكس الاختبارات الورقية.

ويشير بعض الباحثين إلى عيوب جوهرية مختلفة؛ حيث ترى Harrison & Bergen (2000) أن من أهم عيوب الاختبارات الإلكترونية هو تدني القدرة على ضمان النزاهة الدراسية. أما Kearsley (2000) فيشير في الدراسة التي أجراها إلى أن أحد الجوانب التي تثير الجدل كثيراً حول الاختبارات الإلكترونية هو الغش؛ فمن المحتمل أن الذي قدم الاختبار على الشاشة شخص آخر غير الطالب. وتؤكد إحدى الدراسات أن من عيوب الاختبارات الإلكترونية عدم القدرة على مراجعة الأسئلة، وعدم القدرة على تغيير الإجابات، وعدم القدرة على تجاهل بعض الأسئلة، ومشكلات خاصة بدافعية الممتحنين؛ وهذا قد يؤثر على

الممتحنين ويفقد المصداقية في تساوي النتائج بين الاختبارات الورقية والإلكترونية (Wise & Plake, 1989).

باختصار يمكن أن يكون للاختبارات الإلكترونية كل المشكلات المحتملة للبيئات الاختبارية الورقية، يضاف إليها مشكلات البيئات الإلكترونية من أعطال محتملة، وغيرها من الجوانب السلبية، التي أضافتها البيئات الإلكترونية إلى العملية التعلُّمية.

فاعلية الاختبارات الإلكترونية:

يشير كثير من الباحثين إلى أن الاختبارات الإلكترونية لها فاعلية كبيرة في قياس التحصيل الدراسي للطلبة. من هؤلاء الباحثين (Woodfield, 2003) الذي يرى أن الاختبارات الورقية لا تظهر بوضوح النمو الدراسي للطلبة، ولا تقيس مستواهم قياساً دقيقاً، ولا تقدم رؤية للمعلِّم عن كيفية تعديل المنهج الدراسي سعياً لتحسين مخرجات التعلُّم.

ويؤكد Kearsley (2000) أن الاختبارات الإلكترونية أكثر فاعلية من التقليدية؛ فهي تساعد المعلِّمين على تقييم درجة استيعاب الطلبة ومشاركاتهم، بسبب أن استجابات الطلبة يمكن تسجيلها، بما في ذلك الشاشات والروابط والمفاتيح التي يضغطون عليها؛ مما يوفر قدراً كبيراً من المعلومات لتحليلها. إضافةً إلى ذلك، يمكن تجميع وتخزين جميع مشاركات الطالب في حلقات النقاش والمجموعات البريدية، وحفظ نتائج الاختبارات والواجبات في قواعد بيانات سجّله تلقائياً (Kearsley, 2000).

وهناك العديد من الباحثين الغربيين الذين قاموا بدراسة فاعلية الاختبارات الإلكترونية. معظم هذه الأبحاث كانت تقارن بين الاختبارات الإلكترونية والورقية من أجل الوصول إلى تحديد درجة فاعلية الاختبارات الإلكترونية. ويمكن تلخيص أهم نتائج هذه الأبحاث فيما يلي:

- وجود تكافؤ بين الاختبارات الورقية والإلكترونية في قياس التحصيل الدراسي (Bunderso et al., 1989; Hofer & Green, 1985; Kingsbury &

(Houser, 1993). ويرى Hofer & Green (1985) أن هناك شرطين يجب توافرهما لتحقيق هذا التكافؤ: أولهما ألا يكون هناك فروق في رتب الدرجات بين نتائج النوعين من الاختبار؛ وثانيهما ألا يكون هناك فروق بين متوسطات الدرجات في نتائج النوعين من الاختبار. ويؤكد Bookman (1989) أنه – بشكل عام – يمكن اعتبار الاختبارات الورقية نسخاً مباشرة لمحتوى وترتيب الاختبارات الورقية، والفرق الوحيد هو أن أداءها يكون على الحاسوب بدلاً من الورق.

- عند وجود اختلاف بين نتائج الاختبارات الورقية والإلكترونية فإن هذا الاختلاف يعود إلى عامل "أثر صيغة الاختبار" (Bunderson et al, 1989). ويشير بندرسون وآخرون في تحليلهم لعدد من الدراسات، إلى أن الأبحاث التي وجدت فروقاً لصالح الاختبارات الورقية كانت تسع دراسات، ولصالح المحوسبة كانت ثلاث دراسات، والتي لم تظهر فرقاً كانت إحدى عشرة دراسة، إلا أن الفروق – بشكل عام – بين نتائج الاختبارات كانت صغيرة جداً. كما أشاروا إلى أن النتائج التحليلية للدراسات تشير إلى أن نسبة الحصول على نتائج متماثلة بين الاختبارات الورقية والمحوسبة هي فقط ٥٠ ٪ (Bunderson et al, 1989). وبعد تحليلهم لنتائج بعض الدراسات التي قارنت بين النوعين من الاختبارات وجد Mead & Drasgow (1993) أن درجة الاختبارات الورقية أكبر قليلاً من درجات الاختبارات المحوسبة. ويشيران إلى أن هناك الكثير من الدلائل على أن الاختبار الورقي والاختبار الإلكتروني المطابق له لن يؤديا إلى النتيجة نفسها بسبب "أثر صيغة الاختبار". ويشير McDonald (2002) إلى جدل كبير في أنه عند تحويل الاختبارات إلى صيغة إلكترونية، ينتج ما يُسمى بالخبرة النوعية المتخلفة للاختبار.

- أثبت عدد من الدراسات التي تقيس اتجاهات الطلبة، أنه قد تكوَّن لديهم اتجاهات إيجابية نحو الاختبارات الإلكترونية، وأنهم يفضلونها على

الاختبارات الورقية (Barbera et al. 1995; Schmidt et al., 1978; Slack & Van Cura, 1968). وتشير دراسة Schmitt et al., (1993) إلى أن الطلبة يرون في الاختبار الإلكتروني أكثر موضوعية ودقة وأنه لا يتأثر بالأداء الشخصي للمصححين. كما تشير كثير من الدراسات إلى أن الطلبة الأصغر سناً أكثر إقبالاً على الاختبارات الإلكترونية من الكبار، والنساء أكثر إقبالاً من الرجال (Bajtelsmit, 1986).

- وجدت بعض الدراسات أن درجات الممتحنين على الحاسوب أفضل من درجاتهم في الاختبارات الورقية لكن لم يعرف السبب (Bajtelsmit, 1986). وعلى العكس من ذلك وجد Legg & Buhr (1992) أن حوسبة الاختبارات قد تؤدي إلى فروق في النتائج، مرجعها قلق الاختبار.
- يقل الوقت اللازم لأداء الاختبار الإلكتروني بنسبة ٢٥٪ – ٧٢ ٪ دون التأثير على القدرة القياسية للاختبار (Olsen, 1990; Weiss, 1985; Moreno et al., 1984). يؤكد ذلك أيضاً Austin (1991)؛ حيث يشير إلى أن الدراسات قد أظهرت أن الاختبارات المحوسبة والورقية لها النتائج نفسها مع توفير نحو نصف الوقت اللازم للاختبار.
- تتميز بعض أنواع الاختبارات الإلكترونية بالتكيُّف. وتعني الاختبارات الإلكترونية المتكيفة (Computerized Adaptive Tests) أنها تتكيف من حيث السهولة والصعوبة؛ بحيث تتوافق مع قدرات الطلبة المختبرين ومستوياتهم (Vispoel, 1993; De Beer & Visser, 1998). كانت بداية استخدام الاختبارات المتكيفة في الثمانينيات عندما ظهرت الحاسبات الشخصية، وانتشرت بشكل كبير (Austin, 1991). ويحتوي الاختبار المتكيف عادةً على عدد أقل من الأسئلة مقارنةً بالاختبارات غير المتكيفة؛ لأن الأسئلة القريبة لمستوى الطالب المختبر هي التي تظهر فقط؛ ومن ثم تؤدي إلى كفاءة عالية ودقة كبيرة في تقييم مستواه (Embertson, 1992; Tylor, 1982; Wainer, 1990).

- حوسبة الاختبارات تغير ديناميكيتها (Kingsbury & Houser, 1993).

الدراسات السابقة:

قام كثير من الباحثين بدراسات طُبقت في أمريكا وأوروبا ودول متقدمة أخرى، وهدفت إلى مقارنة الاختبارات الإلكترونية بالاختبارات الورقية في المراحل الدراسية المختلفة. وقد توصلت هذه الدراسات إما إلى وجود فروق ضئيلة لصالح الاختبارات الورقية أو فروق ضئيلة لصالح الاختبارات الإلكترونية، وإما إلى عدم وجود فروق بينهما. وفيما يلي استعراض لعدد من هذه الدراسات.

دراسة (Neuman & Baydoun, 1998):

كان الهدف من هذه الدراسة تعرُّف درجة التكافؤ بين الاختبارات الورقية والمحوسبة في المرحلة الجامعية. اشترك في الدراسة (٤١١) طالباً وطالبة، خضعوا لعشرة اختبارات قبلية وبعدية. وتوصلت الدراسة إلى عدم وجود فروق ذات دلالة إحصائية بين الاختبارات الورقية والإلكترونية، وخاصةً عند اتباع نفس التعليمات والظروف في الاختبارين.

دراسة (DeBeer & Visser, 1998):

أُجريت هذه الدراسة على (٦١٣) طالباً وطالبة في المرحلة الثانوية في جنوب إفريقيا، بهدف مقارنة نتائج الاختبارات الورقية والمحوسبة، من أجل تعرُّف التكافؤ بين الاختبارين في قياس التحصيل الدراسي. وتوصلت الدراسة إلى عدم وجود تكافؤ بين الاختبارين، وإلى أن الاختبار الورقي كان الأفضل في قياس التحصيل الدراسي للطلبة.

دراسة (Clariana & Wallace, 2002):

طُبقت هذه الدراسة على (١٠٥) من طلاب وطالبات المرحلة الجامعية في إدارة الأعمال، بهدف دراسة الفرق في (٤) عوامل بين الاختبارات الورقية والإلكترونية. هذه العوامل هي المعرفة بالمحتوى، المعرفة بالحاسوب، التنافسية، والجنس. وقد تعرَّض الطلبة لنسختين من الاختبار نفسه، ولكن

أحدهما ورقي والآخر محوسب. أظهرت النتائج تفوقاً للطلبة الذين تعرَّضوا للاختبار المحوسب، ولم يكن هناك فرق في العوامل المتعلقة بالجنس والتنافسية والمعرفة بالحاسوب، فيما كان هناك فرق في عامل المعرفة بالمحتوى.

دراسة (Sim & Horton, 2005):

تم تطبيق هذه الدراسة على (٢٠) طالباً وطالبة في الصف الثالث الابتدائي في المملكة المتحدة؛ وكان الهدف منها تعرُّف الفروق في أداء الطلبة بين الاختبار الورقي والاختبار المحوسب، ولمعرفة اتجاهات الطلبة نحو الاختبار المحوسب. أظهرت الدراسة أن (٥٠٪) من الطلبة كان أداؤهم أفضل في الاختبار الورقي، فيما كان أداء (٢٥٪) منهم أفضل في الاختبار المحوسب، وكان أداء (٢٥٪) منهم متماثلاً في الاختبارين الورقي والإلكتروني. وأظهرت الدراسة أيضاً تفضيل الطلبة للاختبارات المحوسبة على الاختبارات الورقية.

دراسة (Lim et al., 2006):

أجريت الدراسة لمعرفة اتجاهات طلاب كلية الطب في جامعة سنغافورة نحو الاختبارات المحوسبة ومدى تفضيل الطلاب لها مقارنة بالاختبارات الورقية. وقد أظهرت نتائج الدراسة التي طبقت على (٢١٣) من طلبة السنة الأخيرة أن نسبة كبيرة (٧٩,٨٪) من الطلاب فضلوا النسخة المحوسبة من الاختبار الموضوعي، كما أن أكثر من نصف العينة فضلت النسخة المحوسبة من الاختبار المقالي (٥٤,٤٪). وأظهرت النتائج أيضاً أن استحسان الطلاب للاختبار المحوسب يعود لعدة أسباب، أهمها نقاوة الصور والأشكال في الاختبار المحوسب، والقدرة على التقدم بحسب رغبة الممتحن. وقد خرجت الدراسة بنتيجة أن الصيغة المحوسبة للاختبارات قد تكون أنسب للاختبارات الموضوعية أكثر من مناسبتها للاختبارات المقالية.

دراسة (Demirci, 2007):

بحثت الدراسة في اتجاهات الطلاب نحو التعيينات والمتطلبات (الواجبات) المحوسبة مقارنة بالواجبات الورقية، وقد طبقت الدراسة على (١٠٣) من طلاب

وطالبات الجامعة في تركيا، تم توزيعهم على مجموعتين، قامت إحداهما بعمل الواجبات عن طريق الحاسوب والأخرى بالطريقة التقليدية الورقية. وقد أظهرت النتائج اتجاهات إيجابية نحو الواجبات الإلكترونية، إلا أن الدرجات في الواجبات الورقية كانت أكبر من درجات الطلاب في الواجبات الإلكترونية.

دراسة (Akdemir & Oguz, 2008):

طبقت الدراسة على (٤٧) طالباً وطالبة من طلاب الجامعة في تركيا لمعرفة مدى تكافؤ الاختبارات الورقية والمحوسبة، ولمعرفة ما إذا كان هناك علاقة بين الجنس والأداء في الاختبارات الإلكترونية. وقد أظهرت نتائج الدراسة أنه لا توجد فروق في درجات الطلاب بين النوعين من الاختبارات. كما أن الدراسة لم تظهر أي علاقة بين الجنس وبين الأداء في الاختبارات الإلكترونية. وقد خلص الباحثان إلى نتيجة مفادها أن الصيغ الإلكترونية من الاختبارات ستكون هي المستقبل للاختبارات في تركيا وأوصيا بتعميمها على المجتمع التركي لما لها من مزايا.

دراسة (Stowell & Bennett, 2010):

اعتمدت الدراسة على فرضية أن تقديم الاختبارات بطريقة إلكترونية سوف يؤدي إلى خفض مستوى قلق الاختبار لدى المختبرين وسيؤدي – ومن ثم – إلى تحسين مستوى الأداء والحصول على درجات أعلى في الاختبار. وقد أجريت الدراسة على (٦٩) طالباً جامعياً أخضعوا لاختبارين: ورقي وإلكتروني. وقد أظهرت النتائج أن الطلاب الذين عادة ما يعانون قلق اختبار – في أثناء تأدية الاختبارات التقليدية – قد انخفض معدل القلق لديهم بشكل كبير أثناء أداء الاختبار الإلكتروني. وفي المقابل أظهرت الدراسة أن الطلاب الذين لا يعانون – عادة – قلق الاختبار في أثناء أداء الاختبارات الورقية قد ازداد معدل القلق لديهم في أثناء أداء الاختبار الإلكتروني. كما وجد الباحثان أن العلاقة بين قلق الاختبار وبين الأداء في الاختبار أضعف في الاختبارات الإلكترونية منها في الاختبارات الورقية.

التعليق على الدراسات السابقة:

يظهر من استعراض الأدبيات والدراسات السابقة وجود نمطية في منهجية هذه الدراسات. فقد تمثَّلت معظم هذه الدراسات في المقارنة بين الاختبارات الإلكترونية والورقية من أجل تعرُّف مدى تكافؤ الاختبارات الورقية مع الإلكترونية. وهذا يعني وجود ثقة متراكمة في الاختبارات الورقية، وأن الباحثين قد توجَّهوا للتأكد من وجود هذه الثقة في الاختبارات الإلكترونية. ويعني هذا أيضاً أن الدراسات ما زالت تدور حول التأكد من وجود درجة كافية من الثقة والمصداقية للاختبارات الإلكترونية مقارنة بالاختبارات الورقية؛ وأنه عند تأسيس القدر الكافي من هذه الثقة في الاختبارات الإلكترونية، فإنه يمكن التوجُّه إلى دراسات أكثر تنوعاً؛ بحيث تتناول تصميم الاختبارات الإلكترونية، وأساليب التفاعلية، والعوامل المؤثرة عليها، وغيرها من الموضوعات.

وتتفق الدراسة الحالية مع الدراسات السابقة في الهدف، في حين تختلف معها في البيئة. فمعظم تلك الدراسات تم تطبيقها في دول متقدمة مثل أمريكا وأوروبا. أما الباحثون العرب فلم يوجِّهوا اهتمامهم إلى دراسة الاختبارات الإلكترونية. ولذا تسعى الدراسة الحالية إلى اختبار درجة الثقة والمصداقية في الاختبارات الإلكترونية في العالم العربي، عن طريق مقارنتها بالاختبارات الورقية التي تتصف بالموثوقية لدى التربويين على المستويين العالمي والعربي. ومن ثم فإنها مهمة، حيث إنها تؤسس لدراسات تكون أكثر عمقاً تتعلق بموضوعات مختلفة في الاختبارات أو المقاييس الإلكترونية.

منهجية البحث:

للإجابة عن أسئلة الدراسة ولتحقيق أهدافها، تم استخدام المنهج التجريبي، وبتصميم يتناسب مع أسئلة الدراسة. حيث تم تعريض جميع الطلبة المشاركين في الدراسة لكلا الاختبارين: الورقي والإلكتروني، لقياس مدى التكافؤ بين درجتيهما. ويعرف هذا التصميم باسم **(common person study design)**؛ لأن كل فرد من أفراد العينة يتم اختباره مرتين (Wan, et al., 2009). وللتحكم في أثر

الترتيب والتعرض للاختبار على الأداء في الاختبار الثاني، يتم عادة في الدراسات المشابهة تقسيم العينة إلى قسمين؛ يأخذ أحدهما الاختبار الإلكتروني أولاً ويأخذ القسم الآخر الاختبار الورقي أولاً (Wan, et al., 2009)، وهو ما يطلق عليه مصطلح **(counter-balanced order)**.

كما تم تطبيق استبانة الاتجاهات نحو الاختبارات الإلكترونية قبل الاختبار وبعده لقياس الفرق بين الاتجاهات قبل والتعرض للاختبار الإلكتروني وبعده؛ لمعرفة أثر التعرض للاختبار الإلكتروني على الاتجاه نحوه.

مجتمع الدراسة والعينة:

يتكون مجتمع الدراسة من جميع طلبة كلية التربية بجامعة الكويت. وتكوَّنت العينة من (٣١٦) طالباً وطالبة مسجلين في (١٥) شعبة من مقرر الحاسوب في التربية خلال الفصلين – الأول (٨ شعب) والثاني (٧ شعب) – من العام الجامعي (٢٠٠٩/ ٢٠١٠م). وقد تم اختيار الطلبة بطريقة شبه عشوائية؛ بدأت أولاً بتحديد طلبة مقرر الحاسوب بطريقة مقصودة، ثم اختيار عينة الدراسة منهم بطريقة عنقودية عشوائية **(Cluster Sampling)**. وكان اختيار شعب مقررات الحاسوب – قصداً – لعدة أسباب، منها:

- أن هذه المادة تُدرّس في المختبر، ومن ثم فإن البيئة الإلكترونية المطلوبة كانت متوافرة.
- أن محتوى هذا المقرر موحد لجميع طلبة الكلية وأن الموضوعات المدرسة متفق عليها من قِبل الأساتذة القائمين على التدريس؛ مما يعني أن مستوى الطلبة في جميع الشعب متقارب. كما أنه يتوافر بنك من الأسئلة الموضوعية التي يمكن عمل اختبارات منها.
- أن الباحثين حاولا استبعاد أثر الخبرة بالحاسوب كمتغير قد يؤثر على أداء الطلاب في هذه الدراسة. وكون المبحوثين طلبة في مقرر الحاسوب فإن هذا قد يعني أن لديهم القدر الأدنى من المهارات اللازمة للتعامل مع الحاسوب.

ويُبيِّن الجدول (١) توزع أفراد العينة بحسب متغير الجنس والفصل الدراسي

الجدول (١)

أفراد عينة الدراسة موزعين بحسب المتغيرات الديموغرافية

العدد	الجنس	الفصل الأول	الفصل الثاني	المجموع	النسبة	المجموع الكلي
الشعب	ذكور	١	١	٢	١٣,٣٣ ٪	١٥
	إناث	٧	٦	١٣	٨٦,٦٦ ٪	
الطلاب	ذكور	١٨	١٤	٣٢	١٠,١ ٪	٣١٦
	إناث	١٤٤	١٤٠	٢٨٤	٨٩,٩ ٪	

يتبيَّن من الجدول (١) أن عدد الطلاب الذكور المشتركين في الدراسة كان (٣٢) طالباً، بنسبة (١٠,١ ٪) من أفراد العينة، وعدد الطالبات كان (٢٨٤) طالبة، بنسبة (٨٩,٩ ٪)؛ أي أن النسبة الكبرى من المشتركين في البحث هي من الإناث. والدراسة الحالية لا تقارن مستوى الأداء بين الإناث والذكور، ولذا فإن الفارق في نسبة الذكور إلى الإناث ليس له تأثير.

أداة الدراسة:

جمعت بيانات الدراسة عن طريق أداتين؛ استبانة الاتجاهات نحو الاختبارات الالكترونية واختبار تحصيلي. وفيما يلي تفصيل الأدوات المستخدمة في الدراسة:

الأداة الأولى – الاستبانة:

خُصِّصت هذه الأداة لقياس اتجاهات الطلبة نحو الاختبارات الإلكترونية قبل تعرُّضهم لهذا النوع من الاختبارات وبعده.

وصف الأداة:

قام الباحثان، في ضوء الأدبيات ذات العلاقة، بتصميم استبانة لقياس اتجاهات الطلاب نحو الاختبارات الإلكترونية. تكوَّنت الأداة من (١٠) بنود تتعلق بالاختبارات الإلكترونية، يقابلها استجابات بصورة مقياس ليكرت الخماسي (Likert Scale) (أوافق بشدة – أوافق – غير متأكد – أعارض – أعارض بشدة).

وقد طُلب من المبحوثين تحديد مدى موافقتهم على تلك البنود، وذلك باختيار الاستجابة المناسبة من ذلك المقياس. وقد تنوعت بنود الاستبانة ما بين السلبية والإيجابية؛ فعند موافقة الطلاب على البنود الإيجابية، فهو دليل على اتجاه إيجابي نحو الاختبارات الإلكترونية. أما في حال موافقتهم على البنود السلبية، فهذا دليل على أن الاتجاه كان سلبياً.

صدق الأداة وثباتها:

لتحديد صدق الاستبانة، تم الاعتماد على اختبار الصدق الظاهري لها. وبما أن أفضل طريقة للتثبت من الصدق الظاهري هو عرضها على عدد من المختصين ليقوموا بتقدير مدى تمثيل البنود للصفة المراد قياسها (Ebel, 1972)، فقد تم عرض الاستبانة بصورتها الأولية على عدد من المختصين في مجال القياس والتقويم ومجال تكنولوجيا التعليم. قام هؤلاء المختصون بإبداء بعض الملاحظات التي تم على ضوئها إجراء التعديلات المقترحة من حذف وتعديل وإضافة لبعض البنود. وبعد إجراء هذه التعديلات بصورتها النهائية، تم عرض الاستبانة على المختصين مرة أخرى لأخذ الموافقة النهائية عليها لتصبح صالحة لقياس ما صممت لقياسه. بعد ذلك طُبِّقت الاستبانة بصورتها النهائية على عينة استطلاعية مكونة من (٣٧) طالباً وطالبة مسجلين في شعبتين من خارج العينة، وحُسب الثبات باستخدام معامل الاتساق الداخلي (ألفا لكرونباخ)، وقد كانت قيمة معامل ثبات الأداة (٠,٨٤٣)، وهي قيمة عالية تعطي الثقة في استخدام الأداة.

الأداة الثانية – الاختبار التحصيلي:

صُممت هذه الأداة لقياس درجة التكافؤ بين الاختبار الإلكتروني والاختبار الورقي في قياس التحصيل الدراسي، من أجل التأكد من مصداقية الاختبارات الإلكترونية وفاعليتها.

وصف الأداة:

قام الباحثان باختيار (٦٠) سؤالاً من بنك الأسئلة للمادة، تغطي منهج مقرر الحاسوب حتى وقت تقديم الاختبار. وقم تم بناء اختبار تحصيلي مكون

من (٤٠) سؤالاً من نوع الاختيار من متعدد (لكل سؤال ٤ إجابات)، و (٢٠) سؤالاً من أسئلة الصواب والخطأ. وأعد الباحثان نسختين من الاختبار شبه متطابقتين: ورقية وإلكترونية؛ حيث يحتوي الاختبار في كلتا النسختين على الأسئلة ذاتها مع تغيير في ترتيبها بين النسختين. ويقوم الطلبة في النسخة الورقية بالإجابة عن الاختبار بالطريقة التقليدية (بالورق والقلم)، وتتم الإجابة في النسخة الإلكترونية عن طريق اختيار الإجابة الصحيحة بوساطة فأرة الكمبيوتر. وفي كلتا النسختين، يمكن المراجعة وتخطي الأسئلة والعودة إليها لاحقاً، وتغيير الإجابة في أي وقت خلال زمن الاختبار.

صدق الأداة وثباتها:

قام مجموعة من أساتذة مقرر الحاسوب بمراجعة أسئلة الاختبار وإبداء الموافقة على محتواه. كما تم تطبيقه – بنسخته الورقية – على شعبتين من شعب الحاسوب تحويان (٣٧) طالبة (غير مشمولة بعينة الدراسة). وقد تم حساب الثبات باستخدام طريقة (الاختبار/ إعادة الاختبار)، وأعيد تطبيق الاختبار على الشعبتين بعد (٣) أسابيع من تطبيقه للمرة الأولى. ومن المعروف علمياً أن فترة الاختبار وإعادة الاختبار هي ١– ٣ أسابيع كما يقرر (Walsh & Betz, 1995)، وكانت قيمة معامل ثبات الأداة هي (٠,٩٣٢) وهي قيمة عالية تعطي الثقة في استخدام الأداة.

إجراءات الدراسة:

بعد تحديد الشعب الدراسية، قام الباحثان بالاتفاق مع أساتذة المقرر بتحديد موعد للتطبيق، واتفق على أن يكون في بداية الأسبوع الدراسي السابع في كل من الفصلين الدراسيين الأول والثاني من العام الجامعي ٢٠٠٩/ ٢٠١٠م.

وطبقت أدوات الدراسة في كل شعبة على حدة؛ في البداية وزعت الاستبانة على طلبة المقرر قبل الاختبار. وبعد تسلمها معبأة قسم طلبة المقرر إلى قسمين: قسم يؤدي الاختبار بالطريقة التقليدية (الورقية)، وقسم يؤدي الاختبار بالطريقة الإلكترونية (على جهاز الحاسوب). وبعد ٣ أسابيع من أداء الاختبار

الأول، أعيد تطبيق الاختبار مع تبديل القسمين، حيث يقوم الطلاب الذين أدوا الاختبار الأول بطريقة تقليدية بأدائه على الحاسوب، والطلبة الذين أدوا الاختبار الأول إلكترونياً بأدائه بطريقة ورقية. مع ملاحظة أن الطلاب لم يتلقوا أي معلومات عن أدائهم في الاختبار الأول حتى لا يؤثر ذلك على أدائهم في الاختبار الثاني. وقد حسبت المدة اللازمة لأداء الاختبار الورقي عن طريق ساعة إيقاف يتم تشغيلها فور بدء الاختبار. أما المدة اللازمة لأداء الاختبار الإلكتروني فقد كانت تحسب أوتوماتيكياً، حيث يقوم جهاز الحاسوب بحساب المدة من الضغط على زر البداية حتى الضغط على زر التسليم. ويتم تصحيح الاختبار الورقي يدوياً، كما يتم تصحيح الاختبار الإلكتروني إلكترونياً. ويقوم الاختبار الإلكتروني بإرسال تقرير كامل للأستاذ عن أداء كل طالب يبيِّن وقت بداية الاختبار ووقت الانتهاء منه، والإجابات الخاطئة والصحيحة، والدرجة النهائية.

وبعد الانتهاء من أداء الاختبار الثاني، وزعت الاستبانة مرة أخرى لمقارنة استجابات الطلاب بعد تعرضهم للاختبار الإلكتروني.

المعالجة الإحصائية:

إضافةً إلى التكرارات والنسب، التي استخدمت لأغراض وصفية، قام الباحثان باستخدام الاختبار التائي للقياس المتكرر (Paired Sample T-test) لاختبار الفروق بين درجتي الاختبارين: الورقي والإلكتروني، والوقت المستغرق لأدائهما، والاتجاهات نحو الاختبارات الإلكترونية قبل تعرُّض الطلبة لها وبعده.

عرض نتائج الدراسة ومناقشتها:

توصلت الدراسة إلى عدد من النتائج، نعرضها فيما يلي:

نتائج السؤال الأول:

للإجابة عن السؤال الأول للدراسة "**هل الاختبارات الإلكترونية متكافئة في قياس التحصيل الدراسي الجامعي مع الاختبارات الورقية؟**" تم استخدام الاختبار التائي للقياس المتكرر **(Paired Samples T-test)** بمستوى دلالة (٠,٠٥). يُبيِّن الجدول (٢) نتائج التحليل الإحصائي لهذا الاختبار.

الجدول (٢)
نتائج الاختبار التائي للقياس المتكرر لدرجات الطلبة في الاختبار الورقي والإلكتروني

صيغة الاختبار	العدد	أدنى درجة	أعلى درجة	المتوسط	الانحراف المعياري	قيمة t	درجات الحرية	P-Value
ورقي	٣١٦	١٠	٢٧	٢٠,١٤	٤,٣٠٣٩٦	١,١٩١	٣١٥	,٢٣٤
إلكتروني	٣١٦	١١	٢٨	٢٠,٣٤	٤,٢٤٢١٥			

وبالنظر إلى الجدول (٢) يُلاحظ أن متوسط درجات الطلاب في الاختبار الورقي كانت (٢٠,١٤)، في حين أن متوسط درجاتهم في الاختبار الإلكتروني كانت (٢٠,٣٤). ويلاحظ من نتائج التحليل (t = 1.191, P-Value = 0.234) أنه لا توجد فروق ذات دلالة إحصائية بين درجات الطلاب في الاختبارين الورقي والمحوسب عند مستوى دلالة (٠,٠٥). وهذا يعني أن الاختبارات الإلكترونية متكافئة مع الاختبارات الورقية في قياس التحصيل العلمي للطلبة.

وتتفق هذه النتيجة مع دراستي (Neuman & Baydoun, 1998) و(Akdemir & Oguz, 2008) اللتين أظهرتا عدم وجود فروق ذات دلالة بين درجات الطلاب في الاختبارات الورقية والمحوسبة. إلا أنها تعارضت مع نتائج دراسة (DeBeer & Visser, 1998) التي أظهرت فروقاً لصالح الاختبارات الورقية عند مقارنة نتائجها بالاختبارات المحوسبة. وقد تعارض ذلك أيضاً مع ما أكده (Clariana, Wallace, 2002) من أن الاختبارات الورقية والإلكترونية لن تؤديا إلى النتيجة نفسها حتى في حال تطابقها.

نتائج السؤال الثاني:

للإجابة عن سؤال الدراسة الثاني "**هل هناك تكافؤ في الوقت المستغرق للإنجاز بين الاختبارات الإلكترونية والاختبارات الورقية؟**" تم استخدام الاختبار التائي للقياس المتكرر **(Paired Samples T-test)** بمستوى دلالة (٠,٠٥). ويُبيِّن الجدول (٣) نتائج التحليل الإحصائي لهذا الاختبار.

الجدول (٣)
نتائج الاختبار التائي للقياس المتكرر للوقت اللازم لأداء الاختبارين الورقي والإلكتروني

صيغة الاختبار	العدد	أدنى مدة	أعلى مدة	المتوسط	الانحراف المعياري	قيمة t	درجات الحرية	P-Value
ورقي	٣١٦	١٢٠٠	٣٩٦٠	١٩٠٢	٥٣١,٠٧	١٢,٣٩٣	٣١٥	٠,٠٠٠
إلكتروني	٣١٦	٩٧٠	٤٠١٢	١٤٧٤	٤٠٤,٧٨			

يتضح من الجدول (٣) أن متوسط الوقت اللازم لأداء الاختبار الورقي كان (١٩٠٢) ثانية؛ أي ما يعادل (٣١,٧ دقيقة) في حين أن متوسط الوقت اللازم لأداء الاختبار الإلكتروني كان (١٤٧٤) ثانية؛ أي ما يعادل (٢٤,٥٧ دقيقة). ويلاحظ من نتائج التحليل (t = 12.393, P-value = 0.000) أن هناك فروقاً ذات دلالة إحصائية بين الوقت اللازم لأداء الاختبارين الورقي والإلكتروني؛ حيث إن المدة اللازمة لأداء الاختبار الإلكتروني كانت أقل من المدة اللازمة لأداء الاختبار الورقي.

وتتفق هذه النتيجة مع ما توصل إليه كل من (Olsen, 1990; Weiss, 1985; Moreno et al. 1984; Austin, 1991) من أن الوقت اللازم لأداء الاختبار الإلكتروني يقل بنسبة تراوح بين (٢٥٪-٧٢٪). ويؤكد هذه النتيجة أيضاً ما ذكره (De Beer & Visser, 1998; Ryan et al., 2000) من أن توفير الوقت يعد من الخصائص التي تميز الاختبار الإلكتروني.

نتائج السؤال الثالث:

للإجابة عن السؤال الثالث "**ما اتجاهات طلبة المرحلة الجامعية نحو الاختبارات الإلكترونية قبل التعرُّض لها؟** تم تحليل البيانات من استبانة الاتجاهات نحو الاختبارات الإلكترونية. وقد تكوَّنت الاستبانة من (١٠) بنود تتعلق بآراء الطلبة حول الاختبارات الإلكترونية؛ بحيث يحدد المبحوثون درجة موافقتهم على كل منها باختيار الاستجابة المناسبة على مقياس ليكرد الخماسي. وقد حسبت المتوسطات والانحرافات المعيارية لاستجابات الطلبة على بنود الاستبانة. ويبيِّن الجدول (٤) نتائج هذا التحليل. يُلاحظ – في البنود الإيجابية – أن متوسط استجابة من (١– ٢,٥) يدل على عدم الموافقة، ومن ثم يدل على

اتجاه سلبي نحو الاختبار الإلكتروني. كما أن متوسط استجابة من (٢,٥١ – ٣,٥٠) يدل على موافقة متوسطة ومن ثم اتجاه محايد. أما متوسط استجابة (٣,٥١ – ٥)، فهو دليل على الاتجاه الإيجابي. وفي البنود السلبية فإن متوسط الموافقة العالي يدل على اتجاه سلبي، والمتوسط المنخفض يدل على اتجاه إيجابي. وقد تم إعادة ترميز البنود السلبية ليسهل عرضها في الجدول (٤).

الجدول (٤)
متوسط موافقة الطلاب على بنود استبانة الاتجاه نحو الاختبار الإلكتروني قبل التعرُّض له

البند	متوسط درجة الموافقة	إعادة ترميز البنود السلبية	الانحراف المعياري
الاختبارات الإلكترونية أسهل من الاختبارات الورقية	٢,٢٣١		٠,٨٦٢١
نتائج الاختبارات الإلكترونية لا تعبر عن المستوى الحقيقي للمختبرين	٤,٥٥٤	١,٤٤٦	١,٠٢٦٨
الاختبارات الإلكترونية مناسبة لجميع المراحل	٢,١٣		١,٢١٨٥
للاختبارات الإلكترونية عيوب كثيرة	٣,٧٩٤	٢,٠٩٨	٠,٩٤٢٢
الاختبارات الإلكترونية مناسبة لجميع المواد	٢,١٧٧		٠,٨٢٨٢
قد يزداد قلق الاختبار لدي عند أداء الاختبار إلكترونياً	٤,٥٧٣	١,٤٢٧	٠,٩٧١٥
الوقت اللازم لأداء الاختبار الإلكتروني يساوي الوقت اللازم لأداء الاختبار الورقي	٢,٣٨		١,٠٦٦٧
أفضل أداء جميع اختباراتي المستقبلية بطريقة إلكترونية	١,٩٩١		٠,٧١٩٣
الاختبارات الورقية أكثر صدقاً في قياس مستوى الطلاب	٤,٠٦	١,٩٣٧	١,٥٤٧
أنصح بتعميم الاختبارات الإلكترونية لتحل محل الاختبارات الورقية	١,٨١٣		٠,٨٧٦٥
متوسط اتجاه عام الطلبة على الاختبارات الإلكترونية	١,٩٦٣		,٣٢٠٤٠

يُلاحظ من الجدول (٤) أن الاتجاهات نحو الاختبارات الإلكترونية قبل التعرُّض لها كانت سلبية بشكل عام (١,٩٦٣ على مقياس خماسي). وقد يعود السبب في ذلك إلى أن الطلاب لم يختبروا إلكترونياً من قبل، بل إن غالبيتهم (٨١,٩٦٪) لم يسمعوا بالاختبارات الإلكترونية؛ فهم يجهلون ماهيتها، ومن ثم

فلديهم تصور غير واقعي عنها. يؤكد ذلك حصول البند المتعلق بقلق الاختبار والبند المتعلق بتعبير نتائج الاختبار عن مستوى الطلاب الحقيقي على أقل متوسط موافقة (١,٤٢٧ و ١,٤٤٦ على التوالي). وهذا يعكس تخوفاً من الاختبارات الإلكترونية.

نتائج السؤال الرابع:

للإجابة عن السؤال الرابع "**هل يؤثر تعرض طلبة المرحلة الجامعية للاختبارات الإلكترونية على اتجاهاتهم نحوها؟**" تم حساب متوسط اتجاهات الطلاب على بنود استبانة الاتجاهات نحو الاختبارات الإلكترونية بعد انتهائهم من الاختبار الإلكتروني. ويبيِّن الجدول (٥) نتائج هذا التحليل.

الجدول (٥)
متوسط موافقة الطلبة على بنود استبانة الاتجاه نحو الاختبار الإلكتروني بعد التعرُّض له

البند	المتوسط	إعادة ترميز البنود السلبية	الانحراف المعيار
الاختبارات الإلكترونية أسهل من الاختبارات الورقية	٣,٢٧٨٥		١,١٣٤٥٨
نتائج الاختبارات الإلكترونية لا تعبر عن المستوى الحقيقي للمختبرين	٢,٨٨٢٩	٣,١٠٤٤	١,٧١٠٤٣
الاختبارات الإلكترونية مناسبة لجميع المراحل	٢,٩٨٧٣		١,٢٩٧٠٧
للاختبارات الإلكترونية عيوب كثيرة	٣,٢٧٢٢	٢,٧٢٧٨	,٧٧٣٧٧
الاختبارات الإلكترونية مناسبة لجميع المواد	٣,٣٣٥٤		١,٢٧٧٩٥
قد يزداد قلق الاختبار لدي عند أداء الاختبار إلكترونياً	٢,٩٧١٥	٣,٠٢٨٥	١,٩٠١٣٣
الوقت اللازم لأداء الاختبار الإلكتروني يساوي الوقت اللازم لأداء الاختبار الورقي	٣,٥٣١٦		١,٢٩٣٠٦
أفضل أداء جميع اختباراتي المستقبلية بطريقة إلكترونية	٣,٧٣٤٢		١,١٢٦٤١
الاختبارات الورقية أكثر صدقاً في قياس مستوى الطلاب	٣,٣٦٣٩	٢,٦٢٦٦	١,٨٠١١٨
أنصح بتعميم الاختبارات الإلكترونية لتحل محل الاختبارات الورقية	٣,٩٢٤١		,٩٥٩٧٩
متوسط اتجاه عام الطلبة على الاختبارات الإلكترونية	٣,٣٢٨٠		,٥٢٤٠

يُلاحظ من الجدول (٥) أن الاتجاهات نحو الاختبارات الإلكترونية بعد التعرُّض لها كانت إيجابية بشكل عام (٣,٣٢٨ على مقياس خماسي). وقد يعود السبب في ذلك إلى أن الطلاب لم يجدوا مسوغاً للقلق والخوف من الاختبار الإلكتروني بعد تجربتهم له. ويُلاحظ أيضاً، أن بعض البنود (مثل البند المتعلق بصدق الاختبارات الإلكترونية مقارنة بالورقية ٢,٦٢٦) قد حصلت على موافقة قليلة، وإن كانت أفضل من السابق. كما أن بند خلو الاختبارات الإلكترونية من العيوب وبند مناسبة الاختبارات لجميع المراحل حصلا على موافقة منخفضة أيضاً، على الرغم من أنها أفضل من الاستجابة السابقة نسبياً. وقد يعود السبب في ذلك إلى أن الطلاب وإن تحسَّنت اتجاهاتهم نحو الاختبارات الإلكترونية، لا يعني من وجهة نظرهم خلوها من العيوب، أو أنه يمكن أن تُطبق في مراحل دراسية أخرى كالمدارس الابتدائية والمتوسطة والثانوية.

ولمعرفة إذا ما كان هناك تغير في اتجاهات الطلاب نحو الاختبارات الإلكترونية بعد التعرض للاختبار الإلكتروني، تم استخدام الاختبار التائي للقياس المتكرر **(Paired Sample T-test)** لمقارنة متوسط اتجاهات الطلاب قبل الاختبار وبعده. يبيِّن الجدول (٦) نتائج هذا التحليل.

الجدول (٦)

نتائج الاختبار التائي للقياس المتكرر للاتجاهات نحو الاختبار الإلكتروني قبل التعرُّض له وبعده

اتجاهات الطلاب	المتوسط	الانحراف المعياري	قيمة t	درجات الحرية	P-Value
قبل الاختبار	١,٩٦٣	٠,٢٠٤	٣٦,٠٣١	٣١٥	٠,٠٠٠
بعد الاختبار	٣,٣٢٨	٠,٥٢٤			

يتضح من الجدول (٦) أن متوسط اتجاهات الطلاب نحو الاختبار الإلكتروني قبل التعرُّض له كان (١,٩٦٣)، في حين أن متوسط الاتجاهات بعد أخذ الاختبار كان (٣,٣٢٨). ويلاحظ من نتائج التحليل (t = 36.031, P-value = 0.000) أن هناك فروقاً ذات دلالة إحصائية بين اتجاهات الطلاب نحو الاختبارات الإلكترونية قبل

التعرُّض للاختبار الإلكتروني وبعده، حيث إن الاتجاهات نحو الاختبار الإلكتروني قد تحسَّنت بشكل ملحوظ بعد التعرُّض له.

وتتفق نتائج هذه الدراسة مع دراسة كل من (Barbera et al, 1995; Schmidt et al, 1978; Slack & Van Cura, 1968; Demirci, 2007) التي أثبتت أن الطلاب قد تكونت لديهم اتجاهات إيجابية نحو الاختبارات الإلكترونية بعد التعرض لها. كما تتفق مع دراسة (Sim & Horton, 2005) التي أظهرت تفضيل الطلبة للاختبارات المحوسبة على الورقية.

مستخلص النتائج:

هدفت هذه الدراسة إلى تحديد كفاءة الاختبارات الإلكترونية مقارنةً بالاختبارات الورقية، عن طريق تعرُّف الفروق في أداء الطلبة في الاختبارات المحوسبة والورقية، وعن طريق تعرُّف الوقت المستغرق من قِبل الطلبة في أداء الاختبارين. كما هدفت الدراسة إلى تعرُّف الأثر الذي يُحدثه تعرُّض الطلبة للاختبارات الإلكترونية في رفع اتجاههم الإيجابي نحوها.

وقد توصلت الدراسة إلى تكافؤ الاختبارات الإلكترونية والورقية؛ حيث لم توجد فروق ذات دلالة إحصائية بين درجات الطلبة في الاختبارين. إلا أن الدراسة أثبتت فاعلية الاختبار الإلكتروني في تقليص الوقت المستغرق لأداء الاختبار. كما أظهرت الدراسة وجود اتجاه إيجابي نحو الاختبارات الإلكترونية راجع إلى تعرُّض الطلبة لها.

ويلاحظ أن هناك توافقاً بين نتائج هذه الدراسة ونتائج معظم الدراسات السابقة التي تم مناقشتها في الجزء الخاص باستعراض الأدبيات والدراسات السابقة (Neuman & Baydoun, 1998; Akdemir & Oguz, 2008; De Beer & Visser, 1998; Demirci, 2007; Sim & Horton, 2005)، على الرغم من أن الدراسات السابقة تم تطبيقها في بيئات تعليمية مختلفة عن البيئة التعليمية العربية.

وباستعراض الدراسات السابقة يتضح أن معظمها تبنَّى نمطاً تجريبياً موحداً، يتمثَّل في مقارنة الاختبارات المحوسبة مع الاختبارات الورقية. وهذا يدل

على وجود ثقة متراكمة في الاختبارات الورقية؛ ومن ثم ذهب الباحثون إلى مقارنة نمط في الاختبارات (أي المحوسبة) غير معروف الثقة، بنمط آخر (ورقي) اكتسب ثقة تاريخية. وما يبرر التوجه إلى الاختبارات المحوسبة – على الرغم من الثقة المتراكمة في الاختبارات الورقية – وجود مميزات أثبتتها هذه الدراسة والدراسات السابقة تجعل التفكير في تبنِّي الاختبارات المحوسبة أمراً ضرورياً. تكمن هذه المميزات في تقليص الوقت المستغرق لأداء الاختبار في البيئة الإلكترونية، وهو ما أثبتته الدراسة. ومن ثم تمنح الطلبة مزيداً من الوقت لمراجعة إجاباتهم، وتوفر لهم الفرص في تحقيق درجات أعلى؛ حيث كان هناك شكوى وتذمر على مدة الوقت من أن ضيق أوقات الاختبارات كانت عائقاً أمام الطلبة في المراجعة وتمحيص الإجابات. والأمر الآخر الجدير بالمناقشة، هو أنه كلما تحوَّلت البيئات التعليمية إلى استخدام تقنية الحاسوب والمعلومات في التدريس والتعلُّم فإن من الأجدى أن تتناسب البيئات الاختبارية مع البيئات التعليمية والتدريسية.

لم تكن نتائج هذه الدراسة نشازاً عن نتائج الدراسات السابقة التي تم تطبيقها في أوروبا وأمريكا وغيرها. ومن ثم فإن نتائج هذه الدراسة تشير إلى وجود درجة كبيرة في الثقة في الاختبارات الإلكترونية؛ ما يعني أن بإمكان التربويين في التعليم العالي في العالم العربي الثقة في تبنِّي الاختبارات الإلكترونية؛ نظراً لفاعليتها مقارنة بالاختبارات الورقية، ولما تتميز به من توفير الوقت المستغرق لأداء الاختبار.

توصيات الدراسة:

نظراً لثبوت فاعلية الاختبارات الإلكترونية، فإن الباحثين يوصيان باستخدامها في التعليم العالي في العالم العربي، وخاصة في التخصصات التربوية والتخصصات ذات التشابه مثل العلوم الإنسانية. كما يوصي الباحثان بإجراء دراسات علمية للتأكد من فاعلية الاختيارات الإلكترونية المقالية، التي تقيس مهارات تفكير عليا لا يمكن قياسها في الاختبارات الموضوعية. إضافة لذلك، يوصي الباحثان بإجراء دراسات مماثلة في التخصصات التطبيقية العلمية في التعليم العالي وفي جميع التخصصات في التعليم الأساسي: الابتدائي والمتوسط والثانوي.

المراجع

- Akdemir, O., & Oguz, A. (2008), Computer-Based Testing: An Alternative for the Assessment of Turkish Undergraduate Students. *Computers & Education,* 51(3), 1198-1204.
- Ashton, H., Schofield, D., & Woodger, S. (2003), Pilot summative web assessment in secondary education. Paper presented at the 7th International Assisted Assessment Conference. Loughborough.
- Austin, G. (1991), A Review of Computer Adaptive Testing. [Online]. Available at: http://www.marces.org/mdarch/pdf/M032042.pdf.
- Bajtelsmit, J. (1986), Students likely to be affected by an elimination of written exams: Profile difference of written vs. EOD registrants. *Research and Evaluation Memo,* 86/4-5. Bry Mawr, PA: The American College.
- Barbera, K., Ryan, A., Burris Desmarais, L., & Dyer, P. (1995), Multimedia employment tests: *Effects of attitudes and experiences on validity*. Unpublished manuscripts.
- Bennett, R. (1999), Using new technology to improve assessment. *Educational Measurement: Issues and Practice,* 18(3), 5-12.
- Bennett, R. (2001), How the Internet Will Help Large-Scale Assessment Reinvent Itself. *Education Policy Analysis Archive,* 9(5), 1-23. Also Available [online]: http://epaa.asu.edu/ojs/article/viewFile/334/460.
- Bookman, H. (1989), No more paper and pencil. *Personnel Administrator*, (March), 26-29.
- Bunderson, C., Inouye, D., & Olsen, J. (1989), The four generation of computerized educational measurement. in Linn, R.(ed.). Educational measurement. American Council on Education, Washington, DC, 367-407.
- Clariana, R. & Wallace, P. (2002), Paper-based versus computer-based assessment: key factors associated with the test mode effect. *British Journal of Educational Technology*, 33(5) 593-602.

- Davidson, P. (2003), Why Technology Has Had Only A Minimal Impact On Testing In Education. *Educational Technology Conference Proceedings, Sultan Qabus University, Oman,* 65-79.
- De Beer, M. & Visser, D. (1998), Comparability of the paper-and-pencil and computerized adaptive version of the general scholastic aptitude test (GSAT) senior. *South African Journal of Psychology,* 28(1), 21-27
- Demirci, N. (2007), University students' perceptions of web-based vs. paper-based homework in a general physics course. *Eurasia Journal of Mathematics, Science & Technology Education,* 3(1), 29-34.
- Ebel, R. (1972), *Essentials Of Educational Measurement*. New Jersey: Prentice Hall.
- Embertson, S, (1992), Computerized adaptive testing: its potential substantive contributions to psychological research and assessment. *Current Directions in Psychological Science,* 1, 129-131.
- Harrison, N. & Bergen, C. (2000), Some Design Strategies for Developing an Online Course. *Educational Technology,* 40 (1), 57-60
- Hofer, P. & Green, B. (1985), The challenge competence and creativity in computerized psychological testing. *The Journal of Consulting and Clinical Psychology,* 53, 826-838
- Kearsley, G. (2000), *Online Education: Learning and Teaching in Cyberspace*. Belmont, CA: Wadsworth.
- Kingsbury, G. & Houser, R. (1993), Assessing the utility of item response models: computerized adaptive testing. *Educational Measurement. Issues and practice*, 12, 21-27.
- Legg, S. & Buhr, D. (1992), Computerized adaptive testing with different groups. *Educational Measurement. Issues and Practice,* 11, 23-27.
- Lim, E., Ong, B., Wilder-Smith, E., & Seet, R. (2006), Computer-based versus pen-and-paper testing: students' perception. *Annals of the Academy of Medicine*. 35(9), 599-603.

- McDonald, A. (2002), The impact of individual differences on the equivalence of computer-based and paper-and-pencil educational assessment. *Computers & Education,* 39(4), 299-312
- Mead, A. & Drasgow, F. (1993), Equivalence of computerized and paper-and-pencil cognitive ability tests: a meta-analysis. *Psychological Bulletin 114,* 449-458.
- Moreno, K., Wetzel, C., McBride, J. & Weiss, D. (1984), Relationship between corresponding Armed Services Vocational Aptitude Battery (ASVAB). *Applied Psychological Measurement, 8,* 155-163.
- Morris, C., Brandsford, J., & Franks, J. (1977), Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning and Verbal Behavior*. 16, 519-533
- Neuman, G. & Baydoun, R. (1998), Computerization of Paper-and-pencil Tests: When Ate they equivalent?. *Applied Psychological Measurement, 22*(1), 71-83.
- NCATE. (2008), Professional standards Accreditation of Teacher Preparation Institutions. [Online]. Available at: http://ncate.org/documents/standards/NCATE%20Standards%202008.pdf
- Nugent, G. (2003), On-line multimedia assessment for K-4 students. *Proceedings of the World Conference on Educational Multimedia, Hypermedia and Telecommunications,* Hawaii, USA, 1051-1057.
- Olsen, J. (1990), Applying computerized adaptive testing in schools. *Measurement and Evaluation in Counseling and Development,* 23, 31-38.
- Russell, M., Goldberg, A., & OConnor, K. (2003), Computer-based testing and validity: A look back into the future. *Assessment in Education: Principles, Policy & Practice,* 10(3), 279-293.
- Ryan, S, Scott, B, Freeman, H and Patel, D. (2000), *The Virtual University: The Internet and Resource-Based Learning*. London: Kogan Page.
- Schmidt, F., Urry, V., & Gugel, J. (1978), *Educational and Psychological Measurement 38*, 265-273.

- Schmit, M. & Ryan, A. (1993), Test-taking disposition: A missing link? *Journal of Applied Psychology, 77*, 629-637.
- Schmitt, N., Gilliland, S., Landis, R., & Devine, D. (1993), Computer-based testing applied to selection of clerical applicant. *Personnel Psychology, 46*, 149-165.
- Sim, G. & Horton, M. (2005), Performance and attitude of children in computer based versus paper based testing. [Online]. Available at: http://www.uclan.ac.uk/facs/destech/compute/staff/read/Publish/ChiCi/references/performance_and_attitude.pdf
- Slack, W. & Van Cura, L. (1968), Patient reaction to computer-based medical interviewing. *Computers and Biomedical Research, 1,* 527-531.
- Stowell, J. & Bennett, D. (2010), Effects of Online Testing on Student Exam Performance and Test Anxiety. *Journal of Educational Computing Research. 42*(2), 161-171.
- Sutton, R. E. (2004), Teaching under high-stakes testing: Dilemmas and decisions of a teacher educator. *Journal of Teacher Education, 55*(5), 463-475.
- Taylor, T. (1982), Computerized testing. *South African Journal of Psychology, 13*, 23-31.
- Vispoel, P. (1993), Computerized adaptive and fixed-item versions of the Item Vocabulary subtest. *Educational and Psychological Measurement, 53*(3), 779-788
- Wainer, H. (1990), *Computerized adaptive testing: a primer*. Hillsdale, N.J.: Erlbaum.
- Walsh, W., & Betz, N. E. (1995), *Tests and assessment* (3rd ed.). Englewood Cliffs, NJ: Prentice Hall.
- Wan, L., Keng, L., McClarty, K., & Davis, L. (2009), Methods of Comparability Studies for Computerized and Paper-Based Tests. *Test, Measurement & research services, 10*. [Online]. Available at: http://www.pearsonassessments.com/NR/rdonlyres/7955109E-6724-4EC8-A25F-375FBE360D5F/0/Bulletin_10.pdf
- Weiss, D. (1995), Adaptive testing by computer. *Journal of Consulting and Clinical Psychology, 53,* 774-789.

- Wiechmann, D. & Ryan, A. (2003), Reactions to computerized testing in selection contests. *International Journal of Selection and Assessment 11*(2/3) 215-229.
- Wise, S., and B. Plake. (1989), Research on the Effects of Administering Tests via Computers. *Educational Measurement: Issues and Practice 8(3)*, 5-10.
- Woodfield, K. (2003), Getting on board with online testing. *THE Journal, 30*(6), 32-37. [Online]. Available at: http://thejournal.com/Articles/2003/01/01/Getting-On-Board-With-Online-Testing.aspx?Page=1.