دراسة أثر بعض المتغيرات على أداء طلبة الصف الحادي عشر في مدارس دولة الكويت في الاختبارات الإلكترونية

ملخص

هدفت الدراسة التي طبقت على (521) طالبا وطالبة من طلبة الصف الحادي عشر في مدارس التعليم العام بدولة الكويت إلى اختبار أثر بعض المتغيرات (الجنس، التخصص، امتلاك جهاز حاسوب خاص، طبيعة المادة العلمية، والقدرة على المراجعة وتغيير الإجابات) على الأداء في الاختبارات الإلكترونية. وقد جمعت البيانات عن طريق ثلاثة اختبارات تحصيلية: اللغة العربية، واللغة الإنجليزية، إضافة إلى الإحصاء. وباستخدام مجموعة من الأساليب الإحصائية المناسبة لتحليل البيانات، أشارت النتائج إلى وجود فروق في الأداء في الاختبارات الإلكترونية تعزى لمتغيري طبيعة المادة العلمية، والقدرة على المراجعة وتغيير الإجابات. وقد خلصت الدراسة إلى مجموعة من التوصيات متعلقة بالاهتمام بحوسبة الاختبارات وتعميمها وتدريب المعلمين ومعدي الاختبارات عليها، وإجراء المزيد من الدراسات في هذا المجال.

د. فهد عبدالله الخزي
كلية التربية
جامعة الكويت
الكويت

Abstract

The current correlational descriptive study aims at examining the effect of some variables on the performance Of the computerized exam. Data were collected from (521) grade 11th students at Kuwaiti public schools, using three exams: Arabic, English, and Statistics. Analysis of data, using the appropriate statistical techniques, showed that there is a significant difference in students' performance in the computerized exam due to "exam subject" and "the ability to review and change

مقدمة

لعب تطور تكنولوجيا المعلومات والاتصالات، وخصوصا تقنية الحاسوب، دورا كبيرا في مجال التعليم خلال السنوات القليلة الماضية. ولا شك أن استخدام التكنولوجيا في التعليم يعد من أكثر الموضوعات التي شغلت، وتشغل، أذهان المتخصصين في مجال التربيـة والتعليم خلال العقود الثلاثة الماضيـة

© جامعة منتوري، قسنطينة، الجزائر 2011.

answers". The study concluded with relevant recommendations including the adaption of computerized tests, preparing and encouraging teachers and exam makers to use them, in addition to conducting more studies in this topic.

(Al-Gahtani, 2003). وإن كانت النزعات والاتجاهات في المجال التربوي تأتي وتذهب، فإن التكنولوجيا واستخدامها في العملية التعليمية يعد من الموضوعات التي ستبقى طافية على السطح لمدة ليست بالقصيرة (Sherritt & Basom, 1997). ولئن تنافست

المدارس والمؤسسات التعليمية في العالم بشكل عام – وفي وطننا العربي بشكل خاص - لتنال السبق في تبني التحول الإلكتروني في التعليم، إلا أن الاستخدامات ما زالت تعد تقليدية – نسبيا- وتتمثل في استخدام التكنولوجيا بوصفها عملية مساعدة في التعليم أو للمهام الإدارية كالطباعة وإعداد التقارير وعمل الكشوف وغيرها. وحتى على مستوى العالم وفي الدول التي خطت خطوات حثيثة في اتجاه استثمار التكنولوجيا في التعليم، لم يكن للاختبارات الالكترونية الأكاديمية في المدارس ومؤسسات التعليم العالي الدور المتوقع برغم المزايا العديدة التي تميزها (Davidson, 2003a). فاستخدام التكنولوجيا في التعليم كان يتوقف عادة عند نقطة تسبق عملية التقييم، أو كما يعبر Kilgore (2008) أن استخدام الحاسوب للتقييم كان خارج حدود التفكير بالنسبة للتربويين، وكأن التقييم يعد جزءا مستقلا عن العملية التعليمية. والحقيقة أنه يجب ألا يستثنى التقييم – كجزء لا يتجزأ من العملية التعليمية – من جهود الاستفادة من التطور التكنولوجي. حيث يمكن أن تسهم التكنولوجيا بشكل فاعل في تحسينه وتطويره. بل إنها من أهم وأبرز العوامل التي ستساعد في إعادة تشكيل مفهوم القياس وعلى إعادة النظر في أدوات القياس والتقويم (Bennett, 1999). ومن غير المنطقي أن يسعى التربويون إلى استكشاف طرق حديثة للتدريس معتمدة اعتمادا أساسيا على التكنولوجيا مع الإصرار والتشبث في ذات الوقت بالطرق التقليدية في قياس مخرجات ذلك التدريس. فقد ذكر Bennett (2001) أن استخدام أدوت الأمس لقياس مهارات اليوم يعد من أهم المشكلات التي تواجه التعليم الحديث. كما قرر (Morris, Brandsford, & Franks, 1977) أنه يجب أن تكون طرق التدريس وطرق القياس والتقييم يعملان بنفس الطريقة، فإن كانت الدراسة تقليدية، فيجب أن يكون الاختبار تقليديا، وإن كانت الكترونية فيجب أن يكون الاختبار الكترونيا.

مشكلة الدراسة:

تسعى وزارة التربية في دولة الكويت إلى تطوير نظم التعليم التقليدية للوصول إلى أنظمة من التعليم الحديث تسهم في بناء بيئات جديدة تحفز التعليم وتؤهل الأجيال لمواجهة المتغيرات الحديثة. وإيمانا منها بدورها في التعليم، تولي الوزارة اهتماما

خاصا لمواكبة التطور الهائل والسريع في مجال تكنولوجيا المعلومات والاتصالات. وقد خطت الوزارة خطوات واسعة في مجال حوسبة التعليم وقامت بإجراءات عملية كلفت الكثير من الوقت والجهد والمال؛ كعقد المؤتمرات المتخصصة، وتدريب أعضاء هيئة التدريس في مجال التكنولوجيا ، وتطوير الموقع الالكتروني، وإعداد المناهج الإلكترونية، وتجهيز الفصول الذكية، ومشروع "لابتوب لكل طالب"، وتوقيع الاتفاقيات مع الشركات المتخصصة كاتفاقية "شركاء في التعليم مع مايكروسوفت" و "مذكرة التفاهم مع شركة إتش بي" وغيرها. إلا أنه من المستغرب فعلا ألا يكون للتقييم الالكتروني – على أهميته - جزء من هذا الاهتمام.

وحتى على المستوى العالمي، وبرغم الاهتمام الواضح بتكنولوجيا المعلومات والاتصالات، وبرغم ما تتكلفه الحكومات من جهد ومال في استثمار التكنولوجيا في التعليم، يبدو واضحا أن التقييم لم ينل الحظ الوافي من ذلك الاستثمار (الخزي والزكري، قيد النشر). ولئن كانت الخطوات حثيثة في مجال حوسبة التعليم وبطيئة في مجال حوسبة التقييم، إلا أن انتشار التقييم الإلكتروني واعتماده قادم لا محالة (Davisdon, 2003b). فقد بدأت جهات عدة أكاديمية وغير أكاديمية حول العالم بتبني أنواع من الاختبارات الإلكترونية وتطبيقها على منسوبيها من طلاب ومتدربين، كما تخطط بعض الجهات الأكاديمية لاستبدال النسخ الورقية من الاختبارات بالنسخ الالكترونية خلال السنوات القليلة القادمة (Ashton, Schofield, & Woodger, 2003). ويرى بعض المهتمين أن الحوسبة هي المستقبل الحتمي للاختبارات. وقد بالغ البعض بتبني حوسبة الاختبارات، كالمدير التنفيذي لمؤسسة كامبردج للتقييم، Simon Lebus ، والذي يتوقع أن تختفي النسخ الورقية من الاختبارات خلال 10-15 سنة (Curtis, 2009). بل إن البعض قد ذهب لأبعد من ذلك مثل Isabel Nisbet المديرة السابقة لمكتب (Ofqual) والتي تقول إنّه (يجب) استبدال كافة الاختبارات الورقية بأخرى إلكترونية (Stewart, 2011).

وفي دولة الكويت، وفي ظل التوجه العالمي والنزعات التطويرية الملحوظة والخطط التكنولوجية الطموحة لوزارة التربية، يتوقع أن تظهر نماذج من الاختبارات الإلكترونية في مدارس التعليم العام كونها خبرا مفيدا لمبتدأ حوسبة التعليم. إلا أن التوجه لتبني طرق حديثة للتقييم - شأنها شأن تبني أي فكرة حديثة - لا يجب أن يتم باندفاع ودون دراسة متأنية لأثار وتداعيات ذلك التحول، ومدى تقبل الأفراد طلابا كانوا أو معلمين له (Roger, 2003). بل لابد من عمل الدراسات المناسبة للتأكد من أن التحول إلى الصيغ الإلكترونية من التقييم لا يؤثر على فاعلية وكفاءة أدوات القياس كأن تتأثر نتائجه بعوامل نفسية أو شخصية او أن يقيس متغيرات أخرى غير التحصيل

كقدرات الطالب الحاسوبية او درجة قلقه من الحاسوب أو غيرها. وعليه فلا بد من دراسة موضوع التقييم الإلكتروني واختبار العوامل التي – قد – تؤثر على أداء الطلبة في الاختبارات الإلكترونية بغية التأكد من أن درجاتهم فيها تعبر عن مستواهم الحقيقي وليس متأثرة بعوامل أخرى.

أهمية الدراسة:

بالرغم أن موضوع استخدام تكنولوجيا المعلومات والاتصالات في التعليم قد أشبع دراسة من قبل الباحثين التربويين، إلا أنه يبدو جليا ندرة الدراسات المتعلقة في مجال الاختبارات الإلكترونية وخصوصا في وطننا العربي. وتتجلى أهمية الدراسة الحالية في أنها تبحث في موضوع لم يعط حقه في البحث وبالتالي تقدم المقترحات المناسبة بناء على نتائجها. وتأتي هذه الدراسة في ظل غياب الدراسات العربية والمحلية للبحث في سياق الأدبيات – العالمية - المنشورة والتي أشارت إلى أن الاختبارات المحوسبة شأنها شأن الاختبارات الورقية تتأثر بعوامل فردية وأخرى بيئية قد تعزز من كفاءة استخدام هذا النوع من الاختبارات أو تعيق من حسن استخدامها. ويمكن للمسؤولين وراسمي السياسات التعليمية الاسترشاد بنتائج هذه الدراسة فيما يتعلق باستخدام التكنولوجيا للتقييم. كما يمكن للدراسة، في ظل النقص الواضح للدراسات – المحلية والعربية - في هذا الخصوص، أن تكون منطلقا لدراسات مماثلة وأن توسع مجالات البحث في مجال التقييم الإلكتروني.

أسئلة الدراسة:

في ضوء ما تقدم، فإن عددا من العوامل قد وجد أن لها أثرا على الأداء في الاختبارات المحوسبة. منها ما يتعلق بالطالب من عوامل ديموغرافية ونفسية ومنها ما يتعلق بخصائص الاختبار الالكتروني. وعليه، فإن الدراسة الحالية تحاول الإجابة عن السؤال الرئيس التالي:

هل يتأثر أداء الطلاب في الاختبار الالكتروني بأي من المتغيرات (الجنس، التخصص، امتلاك جهاز حاسوب خاص، طبيعة المادة العلمية، القدرة على المراجعة وتغيير الإجابات)؟

ويتفرع من هذا السؤال عدد من الأسئلة الفرعية:

- هل توجد فروق ذات دلالة إحصائية في أداء الطلاب في الاختبار الالكتروني تعزى لمتغير الجنس؟
- هل توجد فروق ذات دلالة إحصائية في أداء الطلاب في الاختبار الالكتروني تعزى لمتغير التخصص؟
- هل توجد فروق ذات دلالة إحصائية في أداء الطلاب في الاختبار الالكتروني تعزى لمتغير امتلاك جهاز حاسوب خاص؟
- هل توجد فروق ذات دلالة إحصائية في أداء الطلاب في الاختبار الالكتروني تعزى لمتغير طبيعة المادة العلمية؟

- هل توجد فروق ذات دلالة إحصائية في أداء الطلاب في الاختبار الالكتروني تعزى لمتغير القدرة على مراجعة وتغيير الإجابات؟

حدود الدراسة:

- الحدود البشرية: اقتصرت هذه الدراسة على طلبة الصف الحادي عشر بقسميه العلمي والأدبي في مدارس التعليم العام بدولة الكويت.
- الحدود الموضوعية: كما اقتصرت على بعض العوامل التي قد تؤثر على الأداء في الاختبارات الإلكترونية وهي الجنس، والتخصص، وامتلاك حاسوب خاص، وطبيعة المادة العلمية، والقدرة على المراجعة وتغيير الإجابات.
- الحدود الزمانية: طبقت هذه الدراسة خلال الفصل الثاني من العام الدراسي 2011/2010
- الحدود المكانية: طبقت هذه الدراسة في مجموعة من المدارس الثانوية في جميع المناطق التعليمية بدولة الكويت.

مصطلحات الدراسة:

أينما وردت المصطلحات التالية في هذه الدراسة فإن لها معان إجرائية محددة، وذلك كما يلي:

- الاختبارات التقليدية: هي الاختبارات الورقية المعروفة التي درج على استخدامها في المدارس ومؤسسات التعليم العالي.
- الاختبارات الإلكترونية (أو الاختبارات المحوسبة): يُقصد بها الاختبارات التي تتم من خلال الحاسوب، حيث يتم عرض بنود الاختبار على الشاشة ويجيب الطالب عليها مباشرة باستخدام أدوات الإدخال (لوحة المفاتيح، الفأرة، شاشة اللمس).
- تخصص الطالب: هو التحاق طالب الصف الحادي عشر بأحد القسمين العلمي أو الأدبي.
- امتلاك جهاز حاسوب خاص: هو امتلاك الطالب لجهاز حاسوب شخصي خاص به في المنزل أيا كان نوع هذا الجهاز (محمول– مكتبي- نقال).

الإطار النظري:

يتناول الباحث في هذا القسم مجموعة من الموضوعات التي تساعد على وضع تصور عن الاختبارات الالكترونية. ولا بد عند الحديث عن الاختبارات الالكترونية من التطرق إلى مفهوم الاختبارات الإلكترونية وخصائصها ونشأتها وتاريخ البحث فيها.

مفهوم الاختبارات الالكترونية

الاختبارات الإلكترونية أو المحوسبة هي نسخ مطابقة للاختبارات الورقية التقليدية إلا أنها تؤدى عن طريق الحاسوب، حيث يقوم الطالب بقراءة الأسئلة من الشاشة

مباشرة والإجابة عنها عن طريق أدوات الإدخال التقليدية (الماوس، لوحة المفاتيح، شاشة اللمس).

وتنقسم الاختبارات الإلكترونية إلى نوعين رئيسيين، (Wang & Shin, 2010) هما:

- الاختبارات الإلكترونية غير المتكيفة (Non Adaptive Test): وهي اختبارات يتم استعراض الأسئلة فيها بشكل خطي متطابق مع النسخ التقليدية من ناحية عدد الأسئلة وترتيبها وطريقة عرضها على الشاشة.
- الاختبارات الالكترونية المتكيفة (Computerized Adaptive Test): وهي الاختبارات التي تتكيف من ناحية الصعوبة والسهولة مع مستوى الممتحن، حيث أن درجة صعوبة السؤال المعروض تعتمد على إجابة السؤال الذي يسبقه.

مزايا الاختبارات الالكترونية

للاختبارات الالكترونية خصائص فريدة لا تتوافر في مكافئاتها من الاختبارات الورقية التقليدية. وقد تطرق الكثير من الباحثين إلى مزايا الاختبارات الإلكترونية. ويمكن تلخيص المزايا في ثلاثة محاور رئيسية:

- **خصائص الاختبار:**

إن إعداد الاختبارات الالكترونية أقل تكلفة على المدى الطويل من الاختبارات التقليدية (Roid, 1989; McDonald, 2002)، حيث يمكن بفضل (بنوك الأسئلة) إعادة عمل الاختبار مرات عديدة بطريقة آلية دون الحاجة إلى بذل الجهد والوقت في عملية الإعداد (Vispoel et al., 2001; Pomplun et al., 2002). كما أنّه يوفر الكثير من تكاليف الطباعة ونقل الأوراق (Bennett, 2001). كما أنه لا يتطلب تكاليف ولا مساحات تخزين لأوراق الاختبار. فالاختبارات ونتائجها تحفظ إلكترونيا في أقراص أو وحدات تخزين قد لا يتجاوز حجمها البوصة. ويرى الكثير من الباحثين أنّ الاختبارات الالكترونية أكثر صدقا لأنها تسمح بإعداد صيغ متعددة من الاختبار بما فيها الأسئلة التفاعلية التي تقيس مهارات حل المشكلات المعقدة والتي لا يمكن للاختبار التقليدي عملها (Wise & Plake, 1989; Mead & Drasgow, 1993; Wainer, 1993; Huff & Sireci, 2001). ويرى Bennett (2001) أن الاختبارات الإلكترونية تتميز بسهولة مراجعتها واكتشاف الأخطاء وكذلك تصحيح بنودها إن دعت الحاجة لذلك. ويذكر أيضا أن للاختبارات الإلكترونية القدرة على تقديم مثيرات ديناميكية متعددة الوسائط كالصوت والصورة والرسوم المتحركة وغيرها دون الحاجة إلى أجهزة إضافية خاصة.

- **إجراءات التطبيق:**

تتميز الاختبارات الإلكترونية بسهولة التطبيق وبالمرونة بالزمان والمكان. فكل ما يحتاجه المعلم هو توافر جهاز حاسوب يمكن للطالب أن يستخدمه لتقديم الاختبار (Zandvliet & Farragher, 1997; Roever, 2001). كما يمكن أيضا عمل نماذج متعددة وصيغ مختلفة من الاختبار في ذات الوقت (Olsen et al., 1989; Vispoel, 2000)، فمثلا قد يكون الطلاب الذين يقدمون الاختبار ذاته في الوقت ذاته قد حصلوا على نماذج مختلفة، وهو ما يعد ميزة أخرى فهو يحد من إمكانية الغش ونسخ الإجابات أثناء الاختبار حيث أنه يظهر لكل طالب سؤال مختلف عن زميله الذي يخبر بجانبه (Young et al., 1996; Vispoel, 2000; Huff & Sireci, 2001). وقد أظهرت بعض الدراسات (Spray et al. 1989; Mason et al., 1999; Bugbee & Bernt, 1990; Chin & Donn, 1991; Ogilvie et al. 2001) أن الطلاب يفضلون الاختبارات الالكترونية على الورقية نظرا لأنها تتطلب وقتا أقصر من الاختبار التقليدي. كما أن الممتحنين يفضلون الاختبارات الالكترونية المتكيفة (CAT) كونها تزودهم بأسئلة تتناسب مع مستواهم، لا تتطلب منهم إجابة أسئلة صعبة جدا أو سهلة جدا (Dunkel, 1997). وقد أظهرت دراسة (Bugbee & Bernt, 1990) أن فرصة الخطأ في تسجيل الإجابات أقل في النسخ الالكترونية من الورقية، فالطالب يقوم باختيار الإجابة على الشاشة مباشرة. وبالتالي، فكون الطالب على علم بالإجابة الصحيحة، ولكنه يقوم بتضليل الإجابة الخاطئة في ورقة الإجابة دون قصد أمر أقل احتمالية. أخيرا، أظهرت دراسة (Stowell & Bennett, 2010) أن العلاقة بين قلق الاختبار وبين الأداء في الاختبار أضعف في الاختبارات الالكترونية منها في الاختبارات الورقية. كما خلص Powers (2001) إلى أن الممتحنين يعانون قلق اختبار اقل من الذي يعايشونه عادة في الاختبارات الورقية.

- **نتائج الاختبار:**

تتميز الاختبارات الالكترونية بأنها آلية التصحيح، وهو ما يخفف على المعلمين أعباء جمع الأوراق وتصحيحها ومراجعتها ورصد درجاتها (Meijer & Nering,

1999; Niemeyer, 1999). كما أنه يقلل من أخطاء التصحيح والرصد، وخصوصا مع الاختبارات الطويلة والأعداد الكبيرة التي تجهد المعلمين (Pomplun et al., 2002). علاوة على ذلك يتم التصحيح آنيا وفور انتهاء الطالب من أداء الاختبار (Wise & Plake, 1989; Bugbee & Bernt, 1990). ويذكر بعض الباحثين (Niemeyer, 1999; Parshall & Balizet, 2001) أن من أهم مزايا الاختبارات الإلكترونية أن تصحيحها يكون موضوعيا وموحدا لجميع الممتحنين لأنه يتم إلكترونيا ولا يتأثر بشخصانية المقيمين. وتساعد الاختبارات الالكترونية في رسم صورة تفصيلية عن الطالب وتبين مواطن الضعف والقوة في المختبرين، حيث أنها يمكن أن تقدم تقريرا مفصلا يتجاوز عدد الإجابات الصحيحة والخاطئة (Vispoel et al., 2001; Pomplun et al., 2002). هذه المعلومات يمكن استخدامها من قبل المعلمين لتفريد التدريس، ويمكن استخدامها من قبل مصممي المناهج لتصميم مناهج خاصة تناسب بعض الفئات من الطلبة (Davidson, 2003a). وأخيرا، تمكن الاختبارات الالكترونية من الاحتفاظ بالسجلات والبيانات لمدة طويلة، ويمكن عن طريق تلك المعلومات عمل الدراسات والمقارنات ومراجعة السياسات بطريقة سهلة وسريعة لا تتطلب وقتا أو جهدا يذكر (Ryan et al., 2000).

عيوب الاختبارات الإلكترونية:

يثير عدد من الباحثين بعض الملاحظات والأسئلة حول الاختبارات الإلكترونية. ومع أن هذه الملاحظات لا يبدو أنها تعيق استخدام الاختبارات الالكترونية ولم تؤثر على قرار الطلبة في أداء الاختبارات بصورة إلكترونية (Bernt & Bugbee, 1990)، إلا أنه لا يمكن التغافل عن وجودها عند التفكير بالتحول إلى الاختبارات الإلكترونية. ويمكن تلخيص تلك الملاحظات في ثلاثة محاور رئيسة:

- **خصائص الاختبار:**

يرى بعض الباحثين أن صدق الاختبارات الالكترونية موضع تساؤل، حيث انه لا يمكن التحقق من أن الاختبار يقيس فعلا معلومات الطلاب (ما صمم لقياسه) وليس مهاراتهم في استخدام الحاسوب (Kobrin, 2000). كما أن القراءة من الشاشة مباشرة تتطلب عمليات عقلية مختلفة عن تلك التي تتطلبها القراءة التقليدية من الورقة مما قد يشكك في تكافؤ الاختبارات الورقية والالكترونية (Sawaki, 2001). ويرى (Ryan et al., 2000) أن من أهم عيوب الاختبارات الإلكترونية هو كونها – بشكل عام –

من النوع الموضوعي، ومن المعروف أن إعداد هذا النوع من الاختبارات يتطلب مهارة وتدريب وبالتالي تستهلك وقتا طويلا. كما أن قياس المهارات العليا أمر صعب في الاختبارات الموضوعية. ويضيف (Davidson, 2003a) أن من التحديات التي تواجه الاختبارات الإلكترونية وتحد من تقبل المعلمين لها هو أن المعلم يحتاج إلى تدريب خاص على إعداد التقييم الإلكتروني ومهارات تكنولوجيا المعلومات وإدارة الامتحانات. كما أن الأسئلة وبنوك المعلومات تحتاج إلى المراجعة والتغيير الدائم حتى لا تستخدم أكثر من اللازم (Ban et al., 2001). ويذكر (Meijer & Nering, 1999) أن من عيوب الاختبارات الإلكترونية هو موضوع الأمان والوصول غير المشروع من قبل بعض الهاكرز (Hackers) إلى قواعد البيانات والحصول على الأسئلة.

- **إجراءات التطبيق:**

لا بد من الإقرار بأن للاختبارات الالكترونية كل المشكلات المحتملة للاختبارات التقليدية يضاف إليها المشكلات المحتملة في البيئة الالكترونية كاحتمال انقطاع التيار الكهربائي أثناء الاختبار أو تعطل جهاز الحاسوب أو إمكانية إصابته بفيروسات أو غيرها (Bugbee & Burnet, 1990). إن تعطل الحاسوب أثناء أو فور الانتهاء من تأدية الاختبار هو أشبه ما يكون بضياع أوراق الإجابة قبل تصحيحها من قبل المعلمين. وهو ما يتطلب مراقبة أجهزة الحاسوب والبرامج بدقة لتجنب الأعطال أثناء الاختبار وتوفير الدعم الفني اللازم للتدخل حال حدوث أي مشكلات. كما أن عدم توافر أجهزة الحاسوب أو الاتصال الجيد بالإنترنت في بعض الأماكن قد يحد من إمكانية تطبيق الاختبارات الالكترونية (Wise & Plake, 1989). وقد ذكر بعض الباحثين (Bracey, 1990; Webster & Compeau, 1996; Kobrin, 2000; Davidson, 2003a) مجموعة من العوامل التي قد تؤدي – حال توافرها - إلى عدم موثوقية الاختبارات الالكتروني، منها علاقة قلق الحاسوب بالاختبار. فقد أظهرت بعض الدراسات أن درجات الطلاب قد تتأثر بمستوى قلقهم من استخدام الحاسوب (Russell et al., 2003). كما توصل Legg & Buhr (1992) إلى أن حوسبة الاختبارات قد تؤدي إلى فروق من النتائج مرجعها قلق الاختبار. بالإضافة إلى أن طريقة عرض الأسئلة على الشاشة، وعددها في الشاشة الواحدة، وترتيبها يودي إلى عدم تكافؤها مع الاختبارات التقليدية وبالتالي يثير التساؤل حول موثوقيتها. ويخلص Green et al. (1984) إلى أنه لا يمكن الجزم بأن درجة صعوبة السؤال لا تتأثر باختلاف الوسيط الذي يعرض عليه (ورقة أو شاشة حاسوب)، فبعض الأسئلة تبدوا

أسهل على الحاسوب وبعضها تبدوا أصعب. كما يقررون أن سبب ذلك غير معلوم ويحتاج إلى مزيد من البحث لاستكشافه.

- **نتائج الاختبار:**

من أهم العيوب التي ذكرها الباحثون هو احتمالية أن يعتمد المعلمون أكثر من اللازم على الآلة لتقييم الطلاب وتحديد مستواهم (McMinn et al, 1999). وتبرز المشكلة بشكل أكبر في الاختبارات التي لا تتطلب عددا كبيرا من الأسئلة، أو تلك التي لم تكن قد أعدت بشكل جيد وبالتالي فالأحكام والقرارات المترتبة عليها فيما يخص مستوى الطلاب قد يشوبها الكثير من الشكوك (Davidson, 2003a). وبشكل عام، يرى الباحثون التربويون أن الاعتماد على الآلة وحدها– الحاسوب- لتقيم الطلاب واتخاذ القرارات التعليمية بحقهم دون تضمين الجانب الإنساني- المتمثل في المعلم- أمر غير مقبول، حيث لا يمكن للآلة التعامل مع الجوانب النفسية والذهنية للطلاب.

إن التساؤلات التي تثار حول الاختبارات الإلكترونية والمتمثلة في العيوب آنفة الذكر قد تفسر التردد والبطء في تقبل المعلمين والتربويين للاختبارات الإلكترونية والتحول لها. إلا إنه يبدو جليا أن مزايا الاختبارات الالكترونية تفوق العيوب المحتملة لها خصوصا عند اتخاذ الإجراءات اللازمة لتجنب المشكلات المحتملة، وكذلك التدريب الجيد لمعدي الاختبار والممتحنين (Davison, 2003a).

تاريخ البحث في الاختبارات الالكترونية:

منذ بدايات تطوير أداوت القياس والاختبارات السيكولوجية المقننة، كانت الصيغ الورقية هي النوع السائد لهذه الاختبارات (Weiss & Kingsbury, 1984). ومع تطور التكنولوجيا، وبعد ظهور الحاسبات وانتشارها بشكل كبير، وبسبب انخفاض تكاليفها وزيادة كفاءتها، بدأت بعض النماذج من الاختبارات الالكترونية بالظهور والانتشار. ويمكن تقسيم تطور الاختبارات الالكترونية إلى فترتين رئيسيتين:

- **الفترة الأولى (1969-1985)**

في أواخر الستينات وبداية السبعينات، كانت وزارة الدفاع الأمريكية، وعلماء النفس الإكلينيكيين هم رواد تطوير الاختبارات المحوسبة. لأنها كما يرى الأخصائيين النفسيين، تزيد من كفاءة الاختبار وموضوعيته، ولا تتأثر بشخصانية المقيمين (Russell et al., 2003). أما في حقل التربية، فلم توجد دراسات تذكر قبل عام 1986. ويلاحظ في تلك الفترة أنه أجريت العديد من الدراسات لمقارنة النسخ التقليدية مع نظيرتها الالكترونية من مقاييس الشخصية ومقاييس الذكاء وغيرها. وقد كان الهدف

الأساسي من هذه الدراسات هو التأكد من تكافؤ هذين النوعين من الاختبارات وبالتالي صدق وثبات النسخ الالكترونية من تلك المقاييس (Russell et al., 2003).

- **الفترة الثانية (1986 وما بعده):**

بدأت بعد هذا التاريخ تظهر بعض النماذج من الاختبارات الالكترونية في مجال التربية وبدأت المدارس ومؤسسات التعليم العالي في تقبل واستخدام هذا النوع من الاختبارات. وإذا كانت الدراسات في الفترة الأولى تركز على تأسيس التكافؤ بين الاختبارات الالكترونية ونظيراتها الورقية، فإنه بعد عام 1986 بدأ تركيز الباحثين على العوامل التي قد تؤثر على أداء المستخدمين للاختبارات الالكترونية وتحد من تكافؤ النوعين من الاختبارات.

الدراسات السابقة:

هناك العديد من الدراسات - الأجنبية - التي تناولت موضوع الاختبارات الإلكترونية من عدة جوانب. ففي حين ركزت مجموعة من الدراسات على مقارنة النوعين لتأسيس التكافؤ بينهما، ركزت أخرى على استكشاف العوامل التي قد تؤثر على التكافؤ بين النوعين من الاختبارات. وقد تم التركيز في هذه الدراسة على الدراسات التي تناولت العوامل المؤثرة على تكافؤ الاختبارات.

فدراسة (Parshall & Komrey, 1993) أثبتت وجود فروق في نتائج الاختبارات الورقية والمحوسبة تعزى لمتغيرات ديموغرافية كالجنس والعمر والخلفية العرقية. كما أن المعرفة بالكمبيوتر، والاتجاه نحو الاختبارات المحوسبة لم يكن لها أي اثر على الأداء في الاختبارات المحوسبة.

وفي دراسة أجراها (Vispoel, 2000) لمعرفة أثر تمكين الطلاب من مراجعة تغيير إجاباتهم أثناء الاختبارات الالكترونية، قام بمقارنة نتائج اختبار الكتروني يسمح بالمراجعة وتغيير الإجابات مع اختبار الكترونية لا يسمح بالمراجعة وتغيير الإجابات. وبرغم أن نتائج الطلاب في الاختبارين لم يكن بينها اختلافات ذات دلالة إحصائية، إلا أن الطلاب في المجموعتين كانوا يفضلون توافر هذه الخاصية خلال الاختبار. كما أن تحليل النتائج أظهر أن نسبة قليلة من الطلاب قامت بتغيير إجاباتها، وان نسبة التغيير للإجابات الصحيحة كان اكبر من نسبة التغيير إلى الإجابات الخاطئة بنسبة 2.25:1 مما أدى إلى تحسين أداء معظم الطلاب. كما أظهرت النتائج أن تمكين الطلاب من المراجعة أدى إلى تحسين الأداء للطلاب المتمكنين من المادة، وانه كان مفضلا أكثر من قبل الطلاب الذي يعانون من "قلق الاختبار". كما أظهرت النتائج وجود سلبية واحدة فقط من تمكين الطلاب من المراجعة وهي أن مدة أداء الاختبار قد زادت.

وفي الدراسة التي قام بها (Mason, Patry, & Berstein, 2001) وتضمنت (27) من طلبة الجامعة، أظهرت النتائج انه في حالة وجود حافز للطلاب لأخذ الاختبار (ككونه متطلبا في المقرر) وفي حالة توافر الخصائص الخاصة بالاختبار الورقي كمراجعة الإجابات وتخطي بعض الأسئلة وتغيير الإجابات، فانه الاختبارات الالكترونية ونظيراتها الورقية تكون متكافئة. وبالتالي فإنه لا توجد فروق في النتائج المتحصلة من نوعي الاختبار. وقد خرجت الدراسة بتوصيات تتعلق بضرورة الاهتمام بتوفير الخصائص المناسبة عند إعداد الاختبار الالكترونية.

أجرى (Clariana & Wallace, 2002) دراسة على (105) من طلاب الجامعة هدفت إلى مقارنة الاختبارات المحوسبة والورقية لمعرفة ما إذا كان هناك فروق في درجات الطلاب بين الاختبارات الالكترونية والورقية. وعلاقة بعض المتغيرات كالجنس والمعرفة بالكمبيوتر والمعرفة بالمادة والتنافس على الأداء في الاختبار. وقد أثبتت الدراسة أن أداء الطلاب في الاختبار المحوسب أفضل من الورقي. كما أثبتت نتائج المقارنة عدم تأثير الإلمام بالحاسب أو المنافسة أو الجنس على النتائج بين الاختبارين. أما الدراية والإلمام بمحتوى الاختبار فقد اظهر فارقا واضحا بين الأداء في الاختبارين. حيث كان أداء الطلاب ذوو المعرفة بالمادة العلمية أفضل في الاختبار الإلكتروني.

وفي دراسة (Sim & Horton, 2005) التي أجريت على 20 طالب في الصف الثالث الابتدائي في بريطانيا ، لمعرفة الفروق بين أداء الطلاب في الاختبار الورقي والمحوسب ، ولمعرفة اتجاهاتهم نحو الاختبار المحوسب. أظهرت النتائج أن 50% من الطلاب كان أداؤهم أفضل في الاختبار الورقي ، و25% كان أداؤهم في الاختبار المحوسب أفضل. كما أن 25% كان أداؤهم متماثل في الاختبارين. إلا انه لم تظهر أي فروق في أداء الطلاب بشكل عام في الاختبارين ، كما أظهرت عدم وجود فروق بين أداء الذكور والإناث. كذلك أظهرت النتائج تفضيل الطلاب للاختبارات المحسوبة على الاختبارات الورقية.

وفي دراسة (Akdemir & Oguz, 2008) التي طبقت على (47) طالبا وطالبة من طلاب الجامعة في تركيا لمعرفة مدى تكافؤ الاختبارات الورقية والمحسوبة، ولمعرفة ما إذا كان هناك علاقة بين الجنس وبين الأداء في الاختبارات الالكترونية. وقد أظهرت نتائج الدراسة انه 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 وأوصوا بتعميمها على المجتمع التركي لما لها من مزايا.

كما قام (Kilgore, 2008) بدراسة على طلبة الصفوف السادس والسابع والثامن بولاية متشيجان في الولايات المتحدة، لاستكشاف العوامل التي تؤثر على الاتجاهات والأداء في الاختبارات الإلكترونية. وقد درس الباحث عدة عوامل هي الجنس، والمستوى الاجتماعي، وتوافر أجهزة الحاسوب، والاتجاهات نحو الحاسوب. وقد أظهرت النتائج عدم وجود علاقة بين توافر أجهزة حاسوب في المنزل والأداء في الاختبارات الالكترونية. كما أن الاتجاهات نحو الحاسوب ليس لها أي تأثير على الأداء في الاختبارات الإلكترونية. كما أن الطلاب ذوي المستوى الاجتماعي المنخفض كانوا اضعف أداء في الاختبارات الإلكترونية من زملائهم ذوو المستوى الاجتماعي العالي. كما أن الذكور كانوا أكثر ثقة من الإناث أثناء أداء الاختبارات الالكترونية.

التعليق على الدراسات السابقة:

يتضح مما سبق الاهتمام الكبير الذي يوليه الباحثون التربويون لموضوع الاختبارات الالكترونية، وهو ما يدلل على أهمية الموضوع في المجال التربوي. ويلاحظ أن الباحثين قد تناولوا عوامل مختلفة بغية التحقق من أثرها – المحتمل – على الأداء في الاختبارات الإلكترونية ورغبة منهم للوصول إلى درجة من الموثوقية فيها وفي نتائجها. ويتضح – جليا - أن نتائج تلك الدراسات ليس بينها توافق، بحيث نجد أن هناك اختلافا بل وتعارضا في نتائج بعضها عن بعض. ففي حين تؤكد دراسة علاقة متغير ما بالأداء في الاختبارات الإلكترونية، نجد دراسة أخرى تنفي علاقة ذلك المتغير بالأداء. لكن الإطار العام لها يشير إلى حقيقة مؤداها أن الأداء في الاختبارات الالكترونية– إجمالا- ليس بمعزل عن عوامل فردية أو بيئية قد تؤثر فيه سلبا أو إيجابا.

منهجية الدراسة:

تسعى الدراسة الحالية لمعرفة علاقة الأداء في الاختبار الإلكتروني ببعض المتغيرات الديموغرافية. لذا فإن المنهج الوصفي الارتباطي، وهو أحد أنواع المنهج الوصفي (العساف، 2003) يعد المنهج الأنسب لهذه الدراسة.

عينة الدراسة

تكوَّنت العينة من (521) طالبا وطالبة من طلبة الصف الحادي عشر في مدارس التعليم العام بدولة الكويت. وقد تم اختيار الطلبة بطريقة طبقية عنقودية؛ حرص فيها

على تمثيل الجنسين في جميع مناطق الكويت التعليمية. وقد تم عشوائيا اختيار مدرستين ثانويتين من كل منطقة تعليمية (مدرسة للذكور ومدرسة للإناث)، ثم اختيار صفين من كل مدرسة (صف علمي وصف أدبي). ومن ثم تضمين جميع الطلبة في تلك الفصول. وقد تراوح عدد الطلاب في كل صف ما بين (19-23) طالبا. ويبين الجدول (1) توزيع أفراد العينة تبعا لبعض المتغيرات الديموغرافية.

جدول (1): توزيع أفراد العينة تبعا للمتغيرات الديموغرافية

المتغيرات	العدد		النسبة
الجنس	ذكور	236	45.3%
	إناث	285	54.7%
التخصص	أدبي	273	52.4%
	علمي	248	47.6%
امتلاك حاسوب في المنزل	نعم	441	84.6%
	لا	80	15.4%

أداة الدراسة

تم جمع بيانات الدراسة باستخدام عدة أداوت: (أ) الاختبار التحصيلي لمادة اللغة العربية؛ و(ب) الاختبار التحصيلي لمادة اللغة الإنجليزية؛ و(ج) الاختبار التحصيلي لمادة الإحصاء.

- وصف الأداة

تم الاستعانة بالاختبارات الموحدة لمواد اللغة العربية واللغة الانجليزية والإحصاء للصف الحادي عشر والمعتمدة من قبل التوجيه في وزارة التربية. حيث قام الباحث بالاستعانة بمشرفين في التخصصات الثلاث لاختيار مجموعة من الاسئلة الموضوعية للمواد المختارة من اختبارات الفترة الأولى والثانية للفصل الدراسي الأول للعام الدراسي 2011/2010. وقد تم عمل ثلاثة اختبارات (اختبار لكل مادة) يحتوي كلا منها على 20 سؤالا (10 أسئلة من أسئلة الاختيار من متعدد و 10 أسئلة من أسئلة الصواب والخطأ). ويقوم الطلبة بالإجابة عنها بتظليل الإجابة الصحيحة على ورقة الإجابة. وقد قام الباحث – بعد التأكد من صدق وثبات الاختبارات - بعمل نسختين إلكترونيتين (نموذج أ ، ب) مطابقتين للنسخة الورقية من الاختبار تحتوي على ذات الأسئلة بنفس الترتيب، يقوم الطالب بالإجابة عن الأسئلة فيها عن طريق اختيار الإجابة الصحيحة بواسطة فأرة الكمبيوتر، حيث:

- في نموذج (أ)، تعرض الأسئلة بالكامل في شاشة واحدة ويتمكن الطالب فيها من المراجعة وتخطي الأسئلة والعودة لها لاحقاً، وتغيير الإجابة في أي وقت خلال زمن الاختبار.
- في نموذج (ب)، تعرض الأسئلة سؤالا واحدة في الشاشة الواحدة، ولا يظهر السؤال التالي إلا بعد إجابة واعتماد السؤال الذي يسبقه. ولا يمكن للطالب أن يقوم بالمراجعة أو تخطي الأسئلة والعودة لها لاحقاً، ولا تغيير الإجابات بعد اعتمادها والذهاب للسؤال التالي.

وقد وضعت النسخ الالكترونية من الاختبار على موقع الباحث للتعليم الإلكتروني على الانترنت (ithraa.com).

- صدق الأداة

أعتمد الباحث على واحد من أشهر طرق قياس الصدق وأكثرها انتشارا وهو الصدق الظاهري (Face Validity) ويعني إلى أي مدى تبدو الأداة ملائمة ومناسبة لقياس ما يراد قياسه (Walsh & Betz, 1995). ولقياس الصدق الظاهري، يتم عادة عرض الأداء على مجموعة من المختصين ليقوموا بتقدير مدى تمثيل البنود للصفة المراد قياسها (Ebel, 1972). وحيث أن الاختبارات قد تم عملها بواسطة مشرفين متخصصين في المواد المذكورة، وحيث أن الاسئلة تم اختيارها بمعرفة المشرفين من نسخ موحدة ومعتمدة من قبل التوجيه في وزارة التربية، فقد اطمأن الباحث من تحقق درجة صدق مريحة للأدوات. وقد عرضت الاختبارات – زيادة في التأكد – على عدد من معلمي المواد المختصين في بعض المدارس الثانوية وقد ابدوا موافقتهم على محتواها.

- ثبات الأداة

تم حساب الثبات باستخدام طريقة التقسيم النصفي (Split Half)، وهي من الطرق شائعة الاستخدام خصوصا في الظروف التي يتعذر فيها تطبيق الاختبار أكثر من مرة كضيق الوقت او قلة الموارد المادية (Cohen & Swerdlik, 2001). ويتم في هذه الطريقة تطبيق الاختبار مرة واحدة على مجموعة من الأشخاص ثم يتم تقسيم الاسئلة إلى نصفين بطريقة معينة، أشهرها ما يطلق عليه (Odd–Even reliability) وفيه توضع الاسئلة الفردية في نصف والزوجية في النصف الاخر (Kaplan &

Saccuzzo, 2001)، ومن ثم يتم حساب معامل الارتباط (r) ما بين درجات النصفين.

وقد تم تطبيق الاختبارات على صفين مختلفين (أدبي وعلمي) في احدى المدارس الثانوية إناث خلال الأسبوع الأول من الفصل الدراسي الثاني للعام الجامعي 2011/2010. وكان مجموع الطلبة 41 طالبة. وقد تم تقسيم كل اختبار إلى نصفين وحساب معامل الارتباط (r) بين درجتي كل نصف. ويبين الجدول (2) معامل الارتباط بين نصفي كل اختبار.

جدول (2): معامل الارتباط (r) بين نصفي كل اختبار

الاختبار	معامل الارتباط (r)
اللغة العربية	0.77
اللغة الإنجليزية	0.79
الإحصاء	0.76

وبالرجوع إلى الإرشادات التي اقترحها Cohen (1988) والتي تبين أن قيمة الارتباط تعد صغيرة إذا قلت عن |0.29|، ومتوسطة ما بين |0.30| و |0.49|، وكبيرة إذا تجاوزت |0.50|، يتضح أن قيمة معامل الارتباط عالية جدا تعطي الثقة في الأداة وتمكن من الاعتماد عليها في إجراء الدراسة.

المعالجة الإحصائية

تم في هذه الدراسة استخدام مجموعة من الأساليب والاختبارات الإحصائية للتعامل مع البيانات:

- الإحصاء الوصفي (descriptive statistics) والتكرارات والنسب المئوية (frequencies & percentage) لأغراض وصفية.
- معامل الارتباط (r) للتأسيس لثبات الأدوات (الاختبارات التحصيلية).
- تحليل التباين المصاحب الأحادي (ANCOVA) للإجابة عن أسئلة الدراسة.

الاختبار البعدي (Bonferroni) للمقارنات البعدية في حال وجود خلافات بين المجموعات

إجراءات الدراسة

قام الباحث مع نهاية الفصل الدراسي الأول للعام الدراسي 2011/2010 بالاتفاق مع مشرفي ومدرسي المدارس المختارة بتحديد موعد للتطبيق، وقد حرص الباحث على ان يكون التطبيق خلال الشهر الأول من بدء الفصل الثاني. وقد قام جميع المعلمين مشكورين بتطبيق الاختبار القلبي – الورقي – خلال الاسبوع الأول من الفصل الدراسي المحدد. وبعد مرور أسبوعين على التطبيق الأول، قام المعلمون بإخضاع الطلبة للنسخة الإلكترونية من الاختبار في مختبر المدرسة حيث تم تزويد كل طالب بكلمة مرور للدخول إلى النسخة الإلكترونية في الموقع. وقد تم توزيع طلبة كل فصل بطريقة عشوائية - إلكترونية - على النموذجين الإلكترونيين (أ، ب).

نتائج الدراسة

للإجابة عن أسئلة الدراسة تم استخدام الأساليب الإحصائية لاختبار الفروق بين المجموعات المستقلة. وحيث أنه من المحتمل أن يكون المتغير التابع (الأداء في الاختبار الإلكتروني) قد تأثر ببعض المتغيرات المصاحبة (covariate)، فلا بد هنا من استخدام إجراء إحصائي يسمح باكتشاف الاختلافات بين المجموعات مع التخلص – إحصائيا- من اثر المتغيرات المصاحبة (Pallant, 2001). لذا فقد تم استخدام اختبار تحليل التباين المصاحب أحادي الاتجاه (ANCOVA) للإجابة عن الأسئلة. وقد تم إجراء بعض عمليات التحقق المبدئية للتأكد من عدم مخالفة الفرضيات اللازمة لاستخدام هذا الاختبار. وقد تم التعامل مع درجة التحصيل في الاختبار الورقي كمتغير مصاحبة (covariate) بغية التخلص من أثره على الأداء في الاختبار الإلكتروني. وفيما يلي استعراض للنتائج:

إجابة السؤال الأول: هل توجد فروق ذات دلالة إحصائية في أداء الطلاب في الاختبار الالكتروني تعزى لمتغير الجنس؟للإجابة عن هذا السؤال، تم إجراء اختبار تحليل التباين المصاحب أحادي الاتجاه (ANCOVA) ما بين الذكور والإناث لإيجاد الفروق في الأداء في الاختبار الإلكتروني. ويبين جدول (3) نتائج هذا التحليل:

جدول (3): نتيجة اختبار تحليل التباين أحادي الاتجاه (ANCOVA) لإيجاد الفروق في الأداء في الاختبار الإلكتروني تبعا لمتغير الجنس.

الجنس	العدد	المتوسط	الانحراف المعياري	قيمة F	درجات الحرية	Sig	Partial Eta^2
ذكور	236	68.531	17.245	0.121	1,518	0.728	0.000

				17.55	68.725	285	إناث

يتضح من جدول (3) أنه لا توجد فروق ذات دلالة إحصائية بين درجات الذكور والإناث في الاختبار الإلكتروني [F(1,518)=0.121, p=0.728, partial eta^2= 0.000]. ويؤكد ذلك أيضا حجم التأثير والمتمثل في قيمة إيتا تربيع الجزئية (partial eta^2=0.000) والتي تدل على أن حجم التأثير لمتغير الجنس يعد معدوما وفقا للمعايير التي وضعها Cohen (1988). حيث حدد الإرشادات التالية لتفسير قوة حجم التأثير: (0.01= تأثير ضئيل، 0.06= تأثير متوسط، 0.14= تأثير كبير). وهذا يعني أن الجنس ليس له تأثير على الأداء في الاختبار الإلكتروني.

إجابة السؤال الثاني: هل توجد فروق ذات دلالة إحصائية في أداء الطلاب في الاختبار الالكتروني تعزى لمتغير التخصص؟

للإجابة عن هذا السؤال، تم إجراء اختبار تحليل التباين المصاحب أحادي الاتجاه (ANCOVA) ما بين التخصص العلمي والأدبي لإيجاد الفروق في الأداء في الاختبار الإلكتروني. ويبين جدول (4) نتائج هذا التحليل:

جدول (4): نتيجة اختبار تحليل التباين أحادي الاتجاه (ANCOVA) لإيجاد الفروق في الأداء في الاختبار الإلكتروني تبعا لمتغير التخصص

الجنس	العدد	المتوسط	الانحراف المعياري	قيمة F	درجات الحرية	Sig	Partial Eta^2
أدبي	273	68.67	18.02	0.015	1,518	0.902	0.000
علمي	248	68.601	16.69				

يتضح من جدول (4) أنه لا توجد فروق ذات دلالة إحصائية بين درجات التخصصات العلمية والأدبية في الاختبار الإلكتروني [F(1,518)= 0.015,

p=0.902, partial eta^2= 0.000]. ويؤكد ذلك أيضا حجم التأثير والمتمثل في قيمة إيتا تربيع الجزئية (partial eta^2=0.000) والتي تدل على أن حجم التأثير لمتغير التخصص يعد معدوما وفقا للمعايير التي وضعها Cohen (1988). وهذا يعني تخصص الطالب ليس له تأثير على الأداء في الاختبار الإلكتروني.

إجابة السؤال الثالث: هل توجد فروق ذات دلالة إحصائية في أداء الطلاب في الاختبار الالكتروني تعزى لمتغير امتلاك حاسوب خاص؟

للإجابة عن هذا السؤال، تم إجراء اختبار تحليل التباين المصاحب أحادي الاتجاه (ANCOVA) ما بين الطلاب الذين يمتلكون جهاز حاسوب خاص والطلاب الذين لا يمتلكون جهازا خاصا وذلك لإيجاد الفروق في الأداء في الاختبار الإلكتروني. ويبين جدول (5) نتائج هذا التحليل:

جدول (5): نتيجة اختبار تحليل التباين أحادي الاتجاه (ANCOVA) لإيجاد الفروق في الأداء في الاختبار الإلكتروني تبعا لمتغير امتلاك جهاز حاسوب.

امتلاك حاسوب	العدد	المتوسط	الانحراف المعياري	قيمة F	درجات الحرية	Sig	Partial Eta^2
نعم	441	68.622	17.417	0.017	1,518	0.897	0.000
لا	80	68.722	17.01				

يتضح من جدول (5) أنه لا توجد فروق ذات دلالة إحصائية بين درجات الطلاب الذين يملكون جهاز حاسوب خاص والذين لا يملكون جهازا خاصا بهم [F(1,518)= 0.017, p=0.897, partial eta^2= 0.000]. ويؤكد ذلك أيضا حجم التأثير والمتمثل في قيمة إيتا تربيع الجزئية (partial eta^2=0.000) والتي تدل على أن حجم التأثير لمتغير امتلاك جهاز حاسوب خاص يعد معدوما وفقا للمعايير التي وضعها Cohen (1988).

إجابة السؤال الرابع: هل توجد فروق ذات دلالة إحصائية في أداء الطلاب في الاختبار الالكتروني تعزى لمتغير طبيعة المادة العلمية؟

للإجابة عن هذا السؤال، تم إجراء اختبار تحليل التباين المصاحب أحادي الاتجاه (ANCOVA) ما بين درجات الطلاب في المواد المختلفة لإيجاد الفروق في الأداء في الاختبار الإلكتروني. ويبين جدول (6) نتائج هذا التحليل:

جدول (6): نتيجة اختبار تحليل التباين أحادي الاتجاه (ANCOVA) لإيجاد الفروق في الأداء في الاختبار الإلكتروني تبعا لمتغير المادة الدراسية.

المادة	العدد	المتوسط	الانحراف المعياري	قيمة F	درجات الحرية	Sig	Partial Eta^2
اللغة العربية	521	70.779	19.244	20.786	2,1559	0.000	0.026
اللغة الإنجليزية	521	69.358	21.045				
الإحصاء	521	65.774	18.782				

يتضح من جدول (6) أن هناك فروقا ذات دلالة إحصائية بين درجات الطلاب في المواد المختلفة [$F(2,1559)= 20.786$, $p=0.000$, partial $eta^2= 0.026$]. وبعمل المقارنات البعدية (post hoc)، أظهرت النتائج نتائج اختبار (Bonferroni) أن أداء الطلاب في مادة الإحصاء يختلف عن أداءهم في مادتي اللغة العربية واللغة الإنجليزية. في حين لم يختلف الأداء في مادتي اللغة العربية واللغة الإنجليزية عن بعضهما. حيث كان الأداء في مادة الإحصاء أضعف من الأداء في مادتي اللغة العربية واللغة الإنجليزية.

إجابة السؤال الخامس: هل توجد فروق ذات دلالة إحصائية في أداء الطلاب في الاختبار الالكتروني تعزى لمتغير القدرة على مراجعة وتغيير الإجابات؟

للإجابة عن هذا السؤال، تم إجراء اختبار تحليل التباين المصاحب أحادي الاتجاه (ANCOVA) ما بين نتائج النموذجين (أ) و (ب) من الاختبار الإلكتروني وذلك لإيجاد الفروق في الأداء بين الطلاب في الاختبار الإلكتروني. ويبين جدول (7) نتائج هذا التحليل:

جدول (7): نتيجة اختبار تحليل التباين أحادي الاتجاه (ANCOVA) لإيجاد الفروق في الأداء في الاختبار الإلكتروني تبعا لمتغير القدرة على المراجعة وتغيير الاجابات.

القدرة على المراجعة وتغيير الاجابات	العدد	المتوسط	الانحراف المعياري	قيمة F	درجات الحرية	Sig	Partial Eta^2
نعم	796	70.04	19.72	63.996	1,1560	0.000	0.039
لا	767	67.18	19.79				

يتضح من جدول (7) أنه توجد فروق ذات دلالة إحصائية بين درجات الطلاب الذين اخذوا نموذج (أ) الذي يمكن من المراجعة وتغيير الاجابات، وبين الطلاب الذين أخذوا نموذج (ب) الذي لا يمكن من المراجعة وتغيير الإجابات [$F(1,1560)$= 63.996, $p=0.000$, partial eta^2= 0.039]. وبمراجعة المتوسطات الحسابية لدرجات الطلاب في النموذجين، يتبين أن متوسط أداء الطلاب في نموذج (أ) (م=70.04) كان أكبر من متوسط الأداء في النموذج (ب) (م=67.17). مما يدل على أن الطلبة الذين كان لهم القدرة على المراجعة وتغيير الإجابات كان أداؤهم أفضل من الطلاب الذين لم يتمكنوا من المراجعة وتغيير الإجابات.

تحليل النتائج

- الجنس: أظهرت نتائج الدراسة الحالية أنه لا توجد فروق ذات دلالة إحصائية في درجات طلاب الصف الحادي عشر بين الذكور والإناث في الاختبار الإلكتروني. وتتفق نتائج الدراسة مع نتائج دراسة كل من (Clariana & Wallace, 2002) و (Sim & Horton, 2005) و (Akdemir & Oguz, 2008) والتي لم تظهر أي علاقة بين الجنس وبين الأداء في الاختبارات الإلكترونية. كما أنها تتعارض مع ما توصلت إليه دراسة (Parshall & Komrey, 1993) والتي أثبتت وجود فروق في الأداء تعزى لمتغير الجنس، ودراسة (Kilgore, 2008) التي أظهرت أداء أفضل للذكور وثقة أكبر في الاختبار الالكتروني من الإناث. ويلاحظ من نتائج الدراسة الحالية والدراسات السابقة أن عامل الجنس يعد من العوامل الذي لم يتم الفصل في علاقتها بالأداء في الاختبارات الإلكترونية. وهو ما يتطلب دراسات اعمق للوقوف على العلاقة الحقيقية بين الجنس والاداء والاسباب التي ادت لوجود الفروق.

- تخصص الطالب: أظهرت نتائج الدراسة الحالية أنه لا توجد فروق ذات دلالة إحصائية بين درجات التخصصات العلمية والأدبية في الاختبار الإلكتروني. وقد يكون

السبب في ذلك عدم وجود اختلافات عملية متعلقة بالحاسوب بين التخصصين. فالدراسة وان اختلفت في نوعية المواد الا أنها – للتخصصين - في طبيعتها لا تتطلب استخداما مختلفا لتطبيقات الحاسوب. بعكس طلاب الكليات والتخصصات المختلفة على المستوى الجامعي حيث يعد الحاسوب واستخدامه أساسيا في التخصصات العليمة. مما يؤدي إلى ألفة أكبر واستخداما أكثر للحاسوب وبالتالي قد يؤدي إلى اختلافات في الأداء عند تقديم الاختبار إلكترونيا. أن طلاب المرحلة الثانوية لا يختلفون في التعرض للحاسوب فالمواد في القسمين – العلمي والأدبي – لا تعتمد على الحاسوب ، وكلا التقسمين يأخذون مقرر الحاسوب ذاته.

- امتلاك جهاز حاسوب: أظهرت نتائج الدراسة الحالية أنه لا توجد فروق ذات دلالة إحصائية بين درجات الطلاب الذين يملكون جهاز حاسوب خاص والذين لا يملكون جهازا خاصا بهم. وتتوافق نتائج هذه الدراسة مع ما توصل إليه (Kilgore, 2008) من أنه لا توجد علاقة بين توافر أجهزة حاسوب خاصة في المنزل وبين الأداء في الاختبارات الالكترونية.

إن امتلاك جهاز حاسوب خاص قد يعني – في الغالب – استخداما أكثر وألفة أكبر مع الحاسوب. وهي – أي الاستخدام المتكرر والألفة مع الحاسوب – من العوامل المؤثرة إيجابا في الأداء في الاختبارات الإلكترونية (Ward, Hooper, & Hannafin, 1989). وقد وجدت بعض الدراسات أن الأشخاص قلقلي الخبرة مع الحاسوب يكون أداؤهم في الاختبار الالكتروني أسوأ من نظرائهم ذوي الخبرة الكبيرة (Russell et al., 2003). ووجد Clariana & Wallace (2002) أن من أهم المتغيرات التي كان لها اثر على الأداء هي المعرفة بالكمبيوتر. وقد خلصا إلى أن الطلاب الأقل خبرة بالكمبيوتر، لن يكون أداؤهم جيدا كما يمكن أن يكون في البيئة التقليدية. كما وجد Watson (2001) أن الطلاب الذين يستخدمون الكمبيوتر بشكل متكرر يمكن أن يستفيدوا أكثر من البرامج المحوسبة. وقد خلص إلى أن المعرفة والإلمام بالكمبيوتر هي أهم عامل ممكن أن يؤثر على الأداء في الاختبارات الالكترونية، خاصة إن لم يكن هناك إلمام بالمادة العلمية أو مع الطلاب ذوي المستويات الضعيفة.

إلا أن نتائج هذه الدراسة لم تؤد إلى نتيجة مماثلة، فقد أظهرت نتائجها أن لا فروق في الأداء بين من يمتلكون جهازا خاصا ومن لا يمتلكون. ويمكن تفسير ذلك بأن الاستخدام المتكرر والألفة مع الحاسوب لا يتحقق فقط عند امتلاك جهاز حاسوب خاص. فيمكن للطلاب استخدام الحاسوب بشكل متكرر في المنزل أو في بيت قريب أو عند صديق أو في المقاهي الخاصة أو في مختبرات المدرسة.

- طبيعة المادة العلمية: أظهرت نتائج الدراسة الحالية أن هناك فروقا ذات دلالة إحصائية بين درجات الطلاب في المواد المختلفة. وقد بينت النتائج أن الأداء في المواد

النظرية (اللغة العربية واللغة الإنجليزية) كان أفضل من الأداء في المادة التطبيقية (الإحصاء). وقد يكون السبب في ذلك أن المواد التطبيقية تحتاج إلى مساحة للحل لا تكون متوافرة في البيئات الالكترونية. وقد أظهرت عدة دراسات نتائج تدعم في مجملها نتائج الدراسة الحالية وتفسر ما توصلت إليه من فروق في الأداء بين المواد المختلفة. فقد يكون السبب في الاختلاف ما بين المواد النظرية والتطبيقية ما توصلت إليه دراسة (Weldon et al., 1985) من أن العرض على الحاسوب يقلل قدرة الفرد على حل المشكلات، وبالتالي فحل المسائل الرياضية – كما في مادة الإحصاء – يكون أصعب في الاختبارات الإلكترونية. كما وجد Kruk & Muter (1984) أن القراءة من الشاشة تأخذ وقتا أطول من القراءة من المواد المطبوعة على ورق، ووجد Wright & Lickorish (1983) انه من الأصعب اكتشاف الأخطاء من الشاشة مباشرة. وهو ما قد يؤثر على أداء الطلاب في الاختبارات الالكترونية للمواد التطبيقية. وقد تكون القراءة من الشاشة مباشرة أكثر إجهادا للطلاب من القراءة التقليدية كما أشار (Mourant et al.,1981). وهي في المسائل الرياضية أكثر إجهادا.

إن الاهتمام بتصميم الاختبارات الالكترونية له أهمية كبيرة في جودة الاختبارات. حيث يجب الاهتمام بطريقة عرض الأسئلة على الشاشة، والتحدي عند عرض المسائل الحسابية المعقدة التي تتطلب من المختبر مساحة إضافية للحل.

- القدرة على المراجعة وتغيير الإجابات: أظهرت نتائج الدراسة الحالية أنه توجد فروق ذات دلالة إحصائية بين درجات الطلاب الذين اخذوا نموذج (أ) الذي يمكن من المراجعة وتغيير الإجابات، وبين الطلاب الذين أخذوا نموذج (ب) الذي لا يمكن من المراجعة وتغيير الإجابات. وقد كان الفروق لصالح النموذج (أ). وهو ما يتعاض مع دراسة (Vispoel, 2003) التي أظهرت عدم وجود اختلافات ذات دلالة إحصائية بين الاختبار الذي يسمح بالمراجعة وتغيير الإجابات والذي لا يسمح. إلا أن دراسات أخرى عدة أظهرت دلائل على أن عدم القدرة على مراجعة وتغيير الإجابات لها اثر سلبي على أداء الطلاب في الاختبارات الالكترونية (Wise & Plake, 1989). فقد وجد (Vispoel et al., 1992) على سبيل المثال أن تمكين المختبر من مراجعة الإجابات وتغييرها قد حسن من أداء المختبرين، وقد تم تفضليه بشدة من قبل المختبرين. وقد خلص إلى نتيجة مفادها أن الاختبارات الالكترونية التي تمكن من المراجعة والتي لا تمكن من المراجعة قد لا تؤدي إلى نفس النتائج. كما لاحظ (Wise & Plake, 1989) وجود ثلاثة خصائص أساسية في الاختبار الورقي يجب أن توضع في الاعتبار عند تصميم الاختبارات الالكترونية وهي: (أ) إعطاء المختبر الفرصة لتخطي سؤال ما وإجابته في وقت لاحق؛ (ب) الفرصة لمراجعة الأسئلة التي تمت الإجابة

عليها سابقا؛ و(جـ) القدرة على تغيير الإجابات. أما (Lunz et al., 1992) فيرى أن إعطاء الطلاب فرصة لمراجعة إجاباتهم قد تحسن من أدائهم في الاختبار الالكتروني. وقد أظهرت نتائج دراسته أن (32%) من الطلاب استطاعوا أن يحسنوا من إجاباتهم في الاختبار عندما أعطوا الفرصة لمراجعة إجاباتهم وتغييرها.

ومع أن حرمان المختبر من تغيير الإجابة في الاختبارات الالكترونية لم يدرس بالشكل الكافي والوافي خلال السنوات الماضية، إلا أن الكثير من الدراسات قد أجريت على تأثير تغيير الإجابة في الاختبارات الورقية، وقد لخصها (Mueller & Waaser, 1977) في أن المختبر سيحصل على درجات أكثر من التي سيخسرها عندما يسمح له بتغيير الإجابات. ويمكن القول - قياسا على الاختبارات الورقية - أن القدرة على تغيير الإجابات خاصية مهمة لها اثر ايجابي على أداء المختبرين في الاختبارات الالكترونية.

ولتلخيص ما تقدم، يلاحظ أن من بين جميع العوامل التي تم بحثها في هذه الدراسة، فإن عاملي طبيعة المادة العلمية والقدرة على المراجعة وتغيير الإجابات هي العوامل المؤثرة على الأداء في الاختبارات الإلكترونية. مما يعني أنه يجب وضعها في عين الاعتبار عند الرغبة للتحول إلى الاختبارات الإلكترونية. فلا بد من توافر الخصائص المتوافرة في الاختبارات التقليدية كالقدرة على المراجعة وتغيير الاجابات. كما يجب الحذر عند الرغبة في التحول الإلكتروني، فربما كان ذلك لا يناسب جميع المواد وخصوصا التي تحتاج إلى مساحة للحل.

توصيات الدراسة

في ضوء نتائج الدراسة يوصي الباحث بالآتي :

1- أولاء اهتمام أكبر بحوسبة أدوات القياس والاختبارات وإدراجها ضمن الخطط التكنولوجية لوزارة التربية.

2- تعميم الاختبارات الإلكترونية واستحثاث المعلمين على الاستفادة منها نظرا لكفاءتها ومزاياها المتعددة.

3- التدريب الجيد للمعلمين ومعدي الاختبارات وذلك لتجنب المشكلات المحتملة للاختبارات الإلكترونية.

4- الاهتمام عند تصميم الاختبارات الالكترونية بالخصائص التي يجب توافرها في الاختبارات كالقدرة على المراجعة والتخطي وتغيير الإجابات.

5- البدء في التحول الإلكتروني بالمواد النظرية والتي لا تتطلب عمليات معقدة أو مساحة إضافية للحل.

6- التركيز على العوامل المؤثرة على الأداء في الاختبارات الإلكترونية وإيجاد الطرق لمعالجتها والتخلص من آثارها.

7- إجراء دراسات متعمقة في بعض العوامل المؤثر على الأداء والتي لم تفصل الدراسات بآثارها كعامل الجنس.

8- إجراء المزيد من الدراسات المتعلقة بالاختبارات الإلكترونية عامة كونها من الموضوعات التي لم تعط حقها من البحث في البيئة العربية والمحلية.

9- إجراء دراسات على فئات ومستويات أخرى للبحث في العوامل المؤثرة في الأداء في الاختبارات الإلكترونية.

المراجع

أولا: المراجع العربية

- الخزي، فهد و الزكري، محمد. (قيد النشر). تكافؤ الاختبارات الإلكترونية مع الورقية في قياس التحصيل الدراسي: دراسة تجريبية على طلبة كلية التربية بجامعة الكويت. مجلة دراسات الخليج والجزيرة العربية.

- العساف، صالح (2003). المدخل إلى البحث في العلوم السلوكية. الرياض: مكتبة العبيكان

ثانيا: المراجع الأجنبية

Akdemir, O., & Oguz, A. (2008). Computer-based testing: An alternative for the assessment of Turkish undergraduate students. *Computers & Education, 51*(3), 1198-1204.

Al-Gahtani, S. S. (2003). Computer technology adoption in Saudi Arabia: Correlates of perceived innovation attributes. *Information Technology for Development, 10*(1), 57-69.

Ashton, H., Schofield, D., & Woodger, S. (2003). Pilot summative web assessment in secondary education. Paper presented at the 7th International Assisted Assessment Conference, Loughborough.

Ban, J., Hanson, B., Wang, T., Yi, Q., & Harris, D. (2001). A comparative study of on-line pretest item-calibration/scaling methods in computerized adaptive testing. *Journal of Educational Measurement*, *38*(3), 191-212.

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. Retrieved from http://epaa.asu.edu/

Bernt, F. & Bugbee, A. (1990). Factors influencing student resistance to computer administered testing. *Journal of Research on Computing in Education*, *22*(1), 265-276.

Bracey, G. (1990). Computerized testing: A possible alternative to paper & pencil? *Electronic Learning*, *9*(5), 16-17.

Bugbee, A. & Bernt, F. (1990). Testing by computer: findings in six years of use 1982-1988. *Journal of Research on Computing in Education, 23*(1), 87-100.

Chin, C. & Donn, J. (1991). Effects of computer-based tests on the achievement, anxiety, and attitudes of grade 10 science students. *Educational and Psychological Measurement, 51*(3), 735-745.

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.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Erlbaum.

Cohen, R. & Swerdlik, M. (2001). *Psychological Testing and Assessment: An Introduction To Tests and Measurement* (5th ed.). New York, NY: McGraw-Hill

Curtis, P. (2009, July 12). Computerised testing likely to replace traditional exams, says head of board. *The Guardian*. Retrieved November 06, 2010, from http://www.guardian.co.uk

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.

Davidson, P. (2003b). The Equivalence of Paper-Based and Computer-Based Tests. *IATEFL Testing, Evaluation and Assessment SIG Newsletter*, August.

Dunkel, P. (1997). Computer-Adaptive Testing of Listening Comprehension: A Blueprint for CAT Development. *The Language Teacher Online*, 21(10). Retrieved from http://jalt-publications.org

Ebel, R. (1972) *Essentials Of Educational Measurement*. Englewood Cliff, NJ: Prentice Hall.

Green, B., Bock R., Humphreys, L., Linn, R., & Reckase, M. (1984). Technical guidelines for assessing computerized adaptive tests. *Journal of Educational Measurement 21*, 347–360.

Haas, C. & Hayes, J. (1986). What did I just say? Reading problems in writing with the machine. *Research in the Teaching of English, 20*(1), 22-35.

Huff, K. & Sireci, S. (2001). Validity issues in computer-based testing. *Educational Measurement: Issues and Practices, 20(*3), 16-25.

Kaplan, R. & Saccuzzo, D. (2001). Psychological *Testing: Principle, Applications and Issues* (5th Ed). Belmont, CA: Wadsworth

Kilgore, J. (2009). *Exploring the factors that influence attitudes and achievement when students take computerized tests* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3342471)

Kobrin, J. (2000). An Investigation of the Cognitive Equivalence of Computerized and Paper-and-Pencil Reading Comprehension Test Items. Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, Louisiana. (ED 442 836).

Kruk, R. & Muter, P. (1984). Reading continuous test on video screens. *Human Factors, 26*, 339-345.

Legg, S. & Buhr, D. (1992). Computerized adaptive testing with different groups. *Educational Measuremen: Issues and Practice, 11,* 23-27.

Lunz, M., Bergstrom, B. & Wright, B. (1992). The effect of review on student ability and test efficiency for computerized adaptive tests. *Applied Psychological Measurement, 16*, 33-40.

Mason, B., Patry, M., & Bernstein, D. (2001). An examination of the equivalence between non-adaptive computer-based and traditional testing. *Journal of Educational Computing Research, 24*(1), 29-39.

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.

McKee, L., & Levinson, E. (1990). A review of the computerized version of the Self-Directed Search. *Career Development Quarterly, 38*(4), 325-333.

McMinn, M., Ellens, B., & Soref, E. (1999). Ethical perspectives and practice behaviors involving computer-based test interpretation. *Psychological Assessment, 6*(1), 71-77.

Mead, A. & Drasgow, F. (1993). Equivalence of computerized and paper-and-pencil cognitive ability tests: A meta-analysis. *Psychological Bulletin 114,* 449-458.

Meijer, R. & Nering, M. (1999). Computerized adaptive testing: Overview and introduction. *Applied Psychological Measurement*, *23*(3), 187-194.

Mourant, R., Lakshmanan, R, & Chantadisai, R. (1981). Visual fatigue and cathode ray tube display terminals. *Human Factors, 23*(5), 529-540.

Morris, C., Brandsford, J., & Franks, J. (1977). Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning and Verbal Behavior. 16*, 519-533

Mueller, J. & Wasser, V. (1977). Implications of changing answers on objective test items. *Journal of Educational Measurement, 14*, 9-13.

Niemeyer, C. (1999). A computerized final exam for a library skills course. *Reference Services Review*, *27*(1), 90-106.

Ogilvie, R., Trusk, T., & Blue, A. (1999). Student's attitudes towards computer testing in a basic science course. *Medical Education, 33*, 828-831.

Olsen, J., Maynes, D., Slawson, D., & Ho, K. (1989). Comparison of paper-administered, computer-administered and computerized adaptive achievement test. *Journal of Educational Computing Research, 5,* 311-326.

Pallant, J. (2001). SPSS survival manual. Suffolk, UK: St. Edmundsbury Press.

Parshall, C, & Kromrey, J. (1993). Computer testing versus paper-and-pencil: An analysis of examinee characteristics associated with mode effect. A paper presented at the Annual Meeting of the American Education Research Association, Atlanta, GA, April (Educational Resources Document Reproduction Service (ERIC) # ED363272).

Parshall, C. & Balizet, S. (2001). Audio computer-based tests (CBTs): An initial framework for the use of sound in computerized tests. *Educational Measurement :Issues and Practice, 20* (2), 5-15.

Parshall, C., Spray, J., Davey, T., & Kalohn, J. (2002). *Practical considerations in computer-based testing*. New York, NY: Springer.

Pomplun, M., Frey, S., & Becker, D. (2002). The score equivalence of paper-and-pencil and computerized versions of a speeded test of reading comprehension. *Educational and Psychological Measurement*, *62*(2), 337-354.

Powers, D. (2001). Test anxiety and test performance: Comparing paper-based and computer-adaptive versions of the graduate record examinations (GRE) general test. *Journal of Educational Computing Research, 24*(3), 249-273.

Roever, C. (2001). Web-based language testing. *Language Learning and Technology, 5*(2), 84-94.

Rogers, E. M. (2003). *Diffusion of innovations* (5th ed.). New York, NY: Free Press.

Roid, G. (1989). Item writing and item banking by microcomputer: An update. *Educational Measurement: Issues and Practice*, *8*(3), 17-20.

Russell, M., Goldberg, A., & O'Connor, 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., & Patel, D. (2000). *The Virtual University: The internet and resource-based learning*. London, UK: Kogan Page.

Sawaki, Y. (2001). Comparability of conventional and computerized tests of reading in a second language. *Language Learning and Technology, 5*(2), 38-59.

Sherritt, C., & Basom, M. (1997). *Using the internet for higher education*. Retrieved from ERIC database. (ED 407 546).

Spray, J., Ackerman, T., Reckase, M., & Carlson, J. (1989). Effect of the Medium of Item Presentation on Examinee Performance & Item Characteristics. *Journal of Educational Measurement, 26*(3), 261-271.

Stewart, W. (2011, February 25). Pen and paper must go, says Ofqual head. *TES Connect.* Retrieved Retrieved March 01, 2010, from http://www.tes.co.uk

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.

Vispoel, W. (2000). Computerized versus paper-and-pencil assessment of self-concept: Score comparability and respondent preference. *Measurement and Evaluation in Counseling and Development, 33*(2), 130-143.

Vispoel, W., Boo, J., & Bleiler, T. (2001). Computerized and paper-and-pencil versions of the Rosenberg self-esteem scale: A comparison of psychometric features and respondent preferences. *Educational and Psychological Measurement, 61*, 461-474.

Vispoel, W., Wang, T., de la Torre, R., Bleiler, T. & Dings, J. (1992).*How review options, administration mode and anxiety influence scores on computerized vocabulary tests*. Paper presented at the Meeting of the National Council on Measurement in Education, San Francisco (ERIC Document Reproduction service, No. TM018547).

Wainer, H. (1993). Some practical considerations when converting a linearly administered test to an adaptive format. *Educational Measurement: Issues and Practice*, 12(1), 15-20.

Wang, H. & Shin, C. (2010). Comparability of Computerized Adaptive and Paper-Pencil Tests. Test, Measurement & Research Services. Retrieved from http://www.pearsonassessments.com/

Ward, T., Hooper, S., & Hannafin, K.. (1989). The effect of computerized tests on the performance and attitudes of college students. *Journal of Educational Computing Research, 5,* 327-333.

Watson, B. (2001). Key factors affecting conceptual gains from CAL materials. *British Journal of Educational Technology, 32*(5), 587-593.

Webster, J. & Compeau, D. (1996). Computer-assisted versus paper-and-pencil administration of questionnaires. *Behavior Research Methods, Instruments, and Computers, 28*(4), 567-576.

Weiss, D. & Kingsbury, G. (1984). Application of computerized adaptive testing to educational problems. *Journal of Educational Measurement, 21*, 361-375.

Weldon, L., Mills, C., Koved, L., & Schneiderman, B. (1985). The structure of information in outline and paper technical manuals. In R. W. Sweeney (Ed.). Proceedings of the Human Factors Society 29th Annual Meeting: Vol. 2. Human Factors Society (pp. 1110-1113). Santa Monica, CA.

Wilson, R. (2001). HTML E-mail: test font readability study. Retrieved from http://www.wilsonweb.com/wmt6/html-email-fonts.htm

Wise, S., and Plake, B. (1989). Research on the effects of administering tests via computers. *Educational Measurement: Issues and Practice 8*(3), 5-10.

Wright, P. & Lickorish, A. (1983).Proofreading tests on screen and paper. *Behavior and Information Technology, 2*, 227-235.

Young, R., Shermis, M., Brutten, S., & Perkins, K. (1996). From Conventional to Computer-Adaptive Testing of ESL Reading Comprehension. *System, 24*(1), 23-40.

Zandvliet, D. & Farragher, P. (1997). A Comparison of Computer-Administered and Written Tests. *Journal of Research on Computing in Education, 29*(4), 423-438.