أثر سلوك مراجعة وتعديل المفردات على تقدير القدرة وكفاية القياس في الاختبارات التواؤمية المحوسبة

د./ أمين محمد صبري نور الدين
مدرس بقسم علم النفس التربوي
كلية التربية- جامعة عين شمس

ملخص الدراسة :

تعد أنظمة الاختبارات التواؤمية المحوسبة أحد أهم وأحدث تطبيقات نظرية الاستجابة للمفردة وتعتبر تحولا جديدا في القياس النفسي. إلا أن مدى إمكانية السماح للمفحوص بمراجعة إجاباته السابقة، وتعديلها، أو تغييرها بشكل أو بآخر من أكثر الإشكاليات السيكومترية في مجال الاختبار التواؤمي المحوسب. تهدف الدراسة الحالية إلى الكشف عن مدى جدوى تضمين خاصية مراجعة وتعديل المفردات في الاختبارات التواؤمية المحوسبة وتأثير ذلك على تقديرات القدرة للمفحوصين وعلى دقة وكفاية القياس. تم تطبيق الاختبار التواؤمي المحوسب للمصفوفات المتدرجة على عينة قوامها (٢٣٥) طالب وطالبة بكلية التربية من مختلف التخصصات. أسفرت نتائج الدراسة أن أكثر من نصف حجم عينة الدراسة قد استفاد من خاصية التعديل. وأن متوسط درجات الأفراد الذين راجعوا وغيروا إجاباتهم كان أعلى بدلالة من متوسط درجات الأفراد الآخرين الذين لم يراجعوا. غير أنه في المقابل انخفضت كفاية الاختبار بسبب التعديل بنسبة (٥%). وكان فقدان المعلومات بمقدار (١,٦٨). كما زاد سلوك التعديل من زمن الاختبار الأصلي بنسبة زيادة متوسطها (٣٣%). وبينت نتائج الدراسة أيضا أن سلوك المراجعة والتعديل لم يتأثر بالجنس، أو بمستوى محدد من القدرة، وأشارت نتائج الدراسة أنه كلما زادت قدرة المفحوص انخفضت التغييرات التي يجريها على إجاباته. وناقشت الدراسة بعض التضمينات السيكومترية المتعلقة بسلوك تعديل الإجابات في سياق الاختبار التواؤمي المحوسب.

أثر سلوك مراجعة وتعديل المفردات على تقدير القدرة وكفاية القياس في الاختبارات التواؤمية المحوسبة

د./ أمين محمد صبري نور الدين
مدرس بقسم علم النفس التربوي
كلية التربية- جامعة عين شمس

مقدمة :

يهدف القياس النفسي -فيما يهدف إليه- إلى تحسين عملية القياس، وتطوير أدواته، وصولا إلى توفير معلومات دقيقة حول الأفراد، والمساعدة على اتخاذ قرارات بشأنهم. وسواء كان اتخاذ القرار يتعلق بانتقاء أفراد بعينهم لمهنة ما، أو تسكين تلاميذ في صف دراسي معين، أو توجيه تعليمي مناسب لهم، أو غير ذلك؛ فإن للقياس النفسي تأثيرا بالغا وحيويا حين يستخدم لأغراض تربوية، أوتشخيصية، أو تقويمية، أو علاجية.

ولقد تطورت أساليب القياس النفسي في الآونة الأخيرة تطورا هائلا على المستوى النظري، حيث قدمت نظرية الاستجابة للمفردة ***Item Response Theory (IRT)*** حلولا مناسبة لمعالجة وتحليل المفردات، ومعادلة الدرجات، وتشبيك الاختبارات بعضها ببعض. وتمكنت هذه النظرية بفضل تطبيقاتها الواسعة من مواجهة الكثير من المشكلات التي كان يعاني منها القياس النفسي من قبل. فالنظرية تقدم نماذج لتقديرات صعوبة المفردات، وقوتها التمييزية، واحتمالية الإجابة الصحيحة لمفردة بعينها؛ اعتمادا على عينة مختبرة.

كما تطورت أساليب القياس النفسي أيضا على مستوى الممارسة مثل استخدام الحاسب في تقديم وعرض المفردات ومعالجتها وتحليلها وتصحيحها أيضا. وقد واكب هذا التطور المتلاحق للقياس النفسي طفرة هائلة في تقنيات الحاسبات سواء في مجال معدات الحاسبات ***Hardware*** مثل سعة حجم ذاكرة الحاسب، أو القرص الصلب، أو في مجال البرمجيات ***Software*** وأنظمة التشغيل خاصة مع بداية الثمانينيات من القرن الماضي، حيث أصبحت الأجهزة الآن في متناول كل فرد.

واستثمر علماء القياس -مسايرة لهذا التطور- التقنيات المعاصرة في تطبيقات أخرى حديثة لنظرية الاستجابة للمفردة. وكان من نتاج ذلك أن ظهر مصطلح الاختبارات التواؤمية المحوسبة ***Computerized Adaptive Testing (CAT)***. لقد انبثق مفهوم الاختبار التواؤمي المحوسب من خلال دمج نظرية الاستجابة للمفردة، وأسلوب القياس التواؤمي، مع تطبيق الاختبارات على الحاسب. ففيه يتم تطبيق المفردة، والإجابة عليها، ورصد الاستجابة، وتقدير سمة المفحوص، ثم

انتقاء مفردة جديدة للمفحوص وهكذا حتى يتحقق أحد مؤشرات دقة القياس المحددة سلفا؛ عندئذ يتوقف الاختبار. وتعد أنظمة الاختبارات التواؤمية المحوسبة أحد أهم وأحدث تطبيقات نظرية الاستجابة للمفردة؛ حيث يعتبر تحولا جديدا في القياس النفسي.

مشكلة الدراسة :

لقد أجريت مئات الدراسات في مجال الاختبار التواؤمي المحوسب؛ استهدفت معظمها فحص تأثير بعض مكونات الاختبارات التواؤمية المحوسبة على مؤشرات دقة القياس المختلفة، ثم دراسات قارنت بين أساليب انتقاء المفردة الرئيسية، أو طرح أساليب جديدة. علاوة على فحص الاختبار الموائم ذاتيا ***self adapted testing*** كأحد الأساليب المقيدة لانتقاء المفردة وتأثيره على كفاية الاختبار. كما سعت دراسات أخرى لاستخدام محكات مختلفة لإنهاء الاختبار. وفحصت بعض الدراسات خصائص أساليب تقدير القدرة في سياقات مختلفة من القياس. وهناك عدد غير قليل من الدراسات استهدفت التحقق من مدى تكافؤ الاختبارات المحوسبة في مقارنتها مع الاختبارات الورقية العادية.

والنتيجة العامة لتلك الدراسات (***Alkhader, Clark & Anderson, 1998; McBride & Martine, 1983; Vispoel, Wang & Bleiler, 1997***) هي أن للاختبار التواؤمي المحوسب معاملات ثبات وصدق مكافئة لمعاملات ثبات وصدق الاختبارات التقليدية الورقية، أو أعلى منها حتى مع خفض طول الاختبارات إلى حوالي (٥٠%)، بالرغم أنها تقيس نفس المتغيرات. أي أن الاختبار التواؤمي المحوسب لا يحتاج لأكثر من نصف عدد المفردات للوصول إلى نفس مستوى ثبات الاختبار إذا ما طبق بالطريقة الورقية العادية. وتراوح انخفاض زمن الاختبار من (٢٥%) إلى (٧٥%) من زمن الاختبار الأصلي دون فقدان دقة القياس. كما تم الوصول إلى تقديرات ثبات بلغت (٠,٨٠) عن طريق تطبيق (٩ مفردات) فقط (***Alkhader, et al., 1998, p. 205***).

ومع ذلك فإن للاختبارات الورقية خصائص قد لا تتوفر في الاختبارات التواؤمية المحوسبة. فإحدى القضايا المثيرة للجدل في مجال الاختبارات التواؤمية المحوسبة هي مدى إمكانية السماح للمفحوص بمراجعة إجاباته السابقة، وتعديلها، أو تغييرها بشكل أو بآخر ***Reviewing and changing answers***؛ حيث أن هذه الخاصية تختلف في تأثيرها عن الاختبارات الورقية. ففي اختبارات الورقة والقلم التقليدية؛ يطبق على المفحوص الاختبار بترتيب معين صممه واضع الاختبار. ومع ذلك؛ فإن الترتيب الحقيقي للاستجابة على تلك المفردات يحددها المفحوص بنفسه. فالمفحوص قد يتخطى مفردة أو مفردات ثم يعود إليها فيما بعد، وقد لا يرغب في تعديل إجاباته إلا بعد الانتهاء من الإجابة على جميع أسئلة الاختبار. وبصفة عامة تتسم

الاختبارات الورقية التقليدية بشيء من المرونة من جانب المفحوص في ترتيب الإجابة على أسئلة الاختبار.

لكن في الاختبار التواؤمي ***adaptive test*** فإنَّ المفحوص قد لا يكون مخيرا. ويكمن السبب في ذلك أن المفردات تنتقى "ديناميكيا" أثناء التطبيق. فالمفحوص الذي يجيب على المفردة بصورة صحيحة تقدم له المفردة التي غالبا ما تكون من مستوى أعلى في الصعوبة. والعكس يحدث أيضا؛ فعندما يخطئ المفحوص في الإجابة تنتقى له المفردة الأقل صعوبة، وهكذا تتواءم "***adapts***" المفردات مع قدرة المفحوص.

بناء على ذلك؛ فإنه عند توفير خاصية مراجعة المفحوص لاستجاباته وتعديلها بعد انتهائه من الإجابة على جميع المفردات في الاختبارات التواؤمية المحوسبة؛ فإن المفردة التي تلي المفردة المبدلة لن تكون هي المستهدفة، مما قد يؤدي إلى شيء ما من عدم الدقة، علاوة على أنه قد يزيد من المسافة بين تقدير قدرة المفحوص وصعوبة المفردات المنتقاة سلفا. إن تكرار المفردات المبدلة من الصحيحة إلى الخاطئة، وبالعكس من الخاطئة إلى الصحيحة يؤدي إلى انتقاء مفردات غير المستهدفة، أو لا تلائم مستوى المفحوص الحقيقي بعد التحسين. أي أن بعض الخصائص التي تمتاز بها اختبارات الورقة والقلم بصفة عامة يصعب توفيرها في الاختبارات التواؤمية.

لذا؛ فقد تساءلت الدراسات عن مدى إمكانية تزويد هذه الخاصية في الاختبارات التواؤمية المحوسبة. وإلى أي مدى يُنقص إضافة هذه الخاصية من كفاءة الاختبار ***efficiency*** من جانب المفحوص ويرفع من تقدير قدرة الأفراد. وهل السماح باستخدامها يمس عدالة الاختبار ***test fairness***؟ وهل يمكن أن يستغل بعض المفحوصين هذه الخاصية للحصول على درجات مرتفعة بغير استحقاق لها؟ لقد أثارت نتائج هذه الدراسات جدلا واسعا لدى الباحثين حول أهمية تضمين أو وجوب استبعاد هذا العنصر من الاختبارات التواؤمية المحوسبة لما يتعلق بها من تضمينات سيكومترية ومدى عدالة عملية القياس. (***Lunz, Bergstrom, & Wright, 1992, ; Vispoel, Wang, de la Torre, Bleiler, & Dings, 1992; Vispoel, Hendrickson, & Bleiler, 2000; Vispoel, 1998; Wise, Finney, Enders, Freeman, & Severance, 1999***). فهناك بعض الدراسات(***Gershon & Bergstrom, 1995; Stone & Lunz, 1994***) التي أكدت وجوب تضمين هذه الخاصية في الاختبارات التواؤمية المحوسبة حتى تكون مكافئة للاختبارات الورقية، بينما هناك دراسات أخرى (***Stocking, 1997; Welch & Frick, 1993***) ترى أهمية استبعاد هذا العنصر من باب الحفاظ على دقة قياس الاختبارات التواؤمية المحوسبة، وحتى لا يستغل بعض المفحوصين الخاصية لتضخيم درجاتهم النهائية.

وعلى ذلك تسعى الدراسة الحالية إلى محاولة الإجابة على السؤال التالي: ما تأثير خاصية مراجعة وتعديل الإجابات على الاختبارات التواؤمية المحوسبة؟ وما مدى تأثير السماح بذلك العنصر على تقديرات القدرة للمفحوصين وعلى دقة وكفاية القياس؟

أسئلة الدراسة:

بناء على ذلك؛ يتفرع من السؤال الرئيسي للدراسة الأسئلة التالية:

١- كم نسبة أفراد العينة الذين استفادوا من خاصية مراجعة وتعديل الإجابات في الاختبارات التواؤمية المحوسبة؟

٢- هل يوجد فرق دال إحصائيا بين متوسطات درجات الأفراد الذين راجعوا والذين عدّلوا إجاباتهم والذين لم يعدلوا أو يراجعوا إجاباتهم مطلقا؟

٣- إلى أي مدى يحسّن سلوك مراجعة وتعديل الإجابات من درجات المفحوصين في الاختبار؟

٤- إلى أي مدى يُخفّض تعديل المفحوص لاستجاباته من كفاية الاختبار التواؤمي المحوسب؟ وإلى أي مدى تتأثر دالة معلومات الاختبار بتغيير الإجابات؟

٥- إلى أي مدى يستغرق سلوك المراجعة والتعديل زمنا إضافيا على الوقت الأصلي للاختبار؟

٦- هل يختلف سلوك مراجعة وتعديل الإجابات في الاختبار التواؤمي المحوسب باختلاف الجنس؟

٧- هل يختلف سلوك مراجعة وتعديل الإجابات في الاختبار التواؤمي المحوسب باختلاف مستويات القدرة؟

٨- من الأكثر استفادة من خاصية مراجعة وتعديل الإجابات في الاختبار التواؤمي المحوسب هل هو منخفض القدرة أم متوسط القدرة أم مرتفع القدرة؟

٩- ما هو اتجاه تغيير الإجابات الأكثر تكرارا: هل هو من الصح إلى الخطأ، أم من الخطأ إلى الصح، أم من الخطأ إلى الخطأ؟

أهداف الدراسة :

- الكشف عن مدى استفادة الأفراد لخاصية مراجعة وتعديل المفردات في الاختبارات التواؤمية المحوسبة.
- دراسة تأثير سلوك تعديل وتبديل الإجابات على تقدير القدرة ودرجات المفحوصين.
- فحص مدى تأثير مراجعة وتعديل المفردات من جانب المفحوصين على كفاية الاختبار ومقدار المعلومات ودقة القياس.
- بحث مدى جدوى تضمين خاصية مراجعة وتعديل المفردات في الاختبارات التواؤمية المحوسبة.

تحديد المصطلحات الواردة في الدراسة:

مؤشرات دقة القياس ***Measurement Precision Indices*** مجموعة من الدلالات التي يحتكم إليها لبيان مدى إحكام أسلوب التقدير في قياس السمة المراد قياسها. ومؤشرات دقة القياس تنقسم إلى: مؤشرات عامة حيث تعطي قيمة مختصرة عن دقة التقدير للاختبار ككل مثل الكفاية. ومؤشرات المشروطة على القدرة حيث يتم حسابها عند كل نقطة من نقاط تدريج القدرة مثل الخطأ المعياري لتقدير القدرة، ودالة المعلومات. ***(Meijar & Nering, 1999, p. 191)***

الخطأ المعياري لتقدير القدرة ***Standard Error of Estimation***أحد مؤشرات دقة القياس المشروطة على القدرة. ويمثل هذا المؤشر الانحراف المعياري لتقديرات القدرة حول خط الانحدار عند مستوى ثابت من مستوى القدرة الحقيقي. ويحسب الخطأ المعياري عن طريق الجذر التربيعي لمقلوب دالة المعلومات. ***(Wang & Wang, 2001, p. 322)***

دالة المعلومات المفردة/الاختبار ***Item/Test Information Function*** يقصد بها مقدار ما تعطيه المفردة أو الاختبار من معلومات عند المستويات المختلفة لمتصل القدرة. وتوفر نظرية الاستجابة للمفردة حساب دالة المعلومات سواء للاختبار ككل أو لكل مفردة على حدة. ودالة المعلومات في نظرية الاستجابة للمفردة مشروطة على القدرة، ويتم حسابها عن طريق مقلوب تباين الخطأ المعياري لتقدير القدرة. ***(Stocking, 1984, p. 4)***

الكفاية ***Efficiency*** أحد مؤشرات دقة القياس العامة. ويشير هذا المؤشر إلى مقدار المعلومات التي يتم الحصول عليها من كل مفردة تطبق في الاختبار. فالاختبار الأقصى كفاية هو ذلك الاختبار الذي يعطي أفضل قيمة للمعلومات عن المفحوص بأقل عدد ممكن من المفردات. وفي حالة تثبيت عدد المفردات لكل مفحوص كمحك لإنهاء الاختبار كما هو الحال في الدراسة الحالية؛ فإن الاختبار الأفضل هو ذلك الاختبار الذي يعطي أعلى قيمة للمعلومات بأقل خطأ معياري متاح. ***(Wang & Vispoel, 1998)***

مراجعة وتغيير الإجابات: يشير مصطلحا "مراجعة المفردات" ***reviewing items*** و"تغيير الإجابات" ***changing initial answers*** إلى أسلوب في تطبيق الاختبارات يُسمح بمقتضاه أن يقوم المفحوص أو الطالب بمراجعة وتغيير إجاباته السابقة. وقد ورد مصطلح تغيير الإجابات بألفاظ مختلفة في الدراسات الأجنبية؛ غالبا ما تكون: ***Answer changing*** وأحيانا ***changing answers*** أو ***switching answers*** أو ***altering answers***. كما وصفت الإجابات الأصلية قبل التبديل إما ***original answers*** أو ***initial answers***، ويوصف هذا الإجراء أحيانا بأنه عملية ***process*** أو ممارسة ***practice*** أو سلوك***behavior*** ، والمصطلح الأخير هو الأكثر شيوعا في الدراسات السابقة ولذا سيلتزم به الباحث اتساقا مع تلك الدراسات.

الإطار النظري والدراسات السابقة :

أولا مراجعة وتغيير المفردات في الاختبارات الورقية:

عادة ما يواجه الطلاب مشكلة أثناء أدائهم الاختبارات الموضوعية عندما يرغبون في تصحيح بعض الإجابات التي اختاروها، فيشطبون إجابة، ويختارون بديلا آخر يرونه مناسبا.. وهكذا؛ لكنهم يصطدمون بتعليمات المعلمين والممتحنين بعدم تغيير إجاباتهم الأصلية، وأنه "ممنوع تغيير الإجابات" أو "لا تستغرق وقتا طويلا عند أي عبارة" أو "إعط الإجابة التي يبدو أنها تصف ما تشعر به في العادة" أو غيرها من العبارات مثل "***go with your first response***" أو "***stick with your first response***" أو "***first impressions***".

لقد ساد الاعتقاد قديما لدى واضعو الاختبارات الموضوعية على ضرورة الإبقاء على الإجابات الأصلية للمفحوصين، ولا يجب على متلقي الاختبارات ***test-takers*** تغيير إجاباتهم؛ لأن الإجابات الأولى والأصلية هي الأدق، وأي تغيير لها يكون أقل دقة. بينما الطلاب من جانب آخر يرغبون في تصحيح إجاباتهم وتعديلها بما يرونه مناسبا.

وقد كرس هذا الاعتقاد بعض مراقبي الاختبارات الذين يشددون على الطلاب بعدم تغيير الإجابات، فقد أقر أكثر من (٦٣%) من الممتحنين في الجامعات بأنهم يحذرون طلابهم بالفعل من تغيير إجاباتهم. (***Benjamin, Cavell, & Shallenberger, 1984, P. 133***)كما يشير (***Geiger, 1997b***) إلى أن (٧٣%) من عينة دراسته من الطلاب تلقوا بالفعل تحذيرات من معلميهم بعدم تغيير إجاباتهم في اختبارات الاختيار من متعدد. بالرغم أن هؤلاء الطلاب المحذَّرين مازالوا راغبين في مراجعة وتغيير إجاباتهم. كما كان مصدر هذا الاعتقاد أيضا الطلاب أنفسهم الذين يخشون من أن تغيير ومراجعة إجاباتهم ربما يُخفِّض من درجاتهم (***Benjamin, et al., 1984, P. 133***).

بناء على ما تقدم، أجريت عشرات الدراسات متسائلة عن جدوى هذه التعليمات؛ وهل تستند على بحوث ميدانية أم هي من قبيل الفهم العام***common sense*** والاعتقادات الشائعة؟ وهل إتباع تلك التعليمات يفوّت على الطلاب والمفحوصين فرصة لكسب المزيد من الدرجات في الاختبارات المختلفة؟ وما تأثير الفروق بين الجنسين، ومستويات القدرة، ومستويات قلق الامتحان، والأساليب المعرفية المختلفة، على سلوك مراجعة وتغيير المفردات لدى المفحوصين؟ وما تأثير كل من صعوبة المفردة وموضعها أو ترتيبها في الاختبار على تغيير المفردات؟ وما مدى إدراك الطلاب لسلوك تغيير الإجابات؟ لقد أجابت عن هذه الأسئلة وغيرها عدد من الدراسات ومنها ما يلي:

دراسة (***Jacobs, 1971***) هدفت إلى دراسة تأثير صعوبة المفردات وقدرة الأفراد على سلوك تغيير الإجابات. حيث تم تطبيق اختبار تحصيلي مكون من مفردات ذات ثلاثة مستويات من

الصعوبة من خلال شرائح معروضة متبوعة بنسخة مطبوعة من الاختبار. أظهرت النتائج أنه لا توجد علاقة بين مستوى قدرة الطلاب وتغيير الإجابات. بينما ارتبطت صعوبة المفردات بعدد المفردات التي تغيرت، وأيضا نوعية التغيير. فأقل نسبة تغييرات تمت مع المفردات الأسهل مع درجات أعلى على مفردات متوسطة الصعوبة. كما ظهرت علاقة عكسية بين اتجاه التغيير وصعوبة المفردات.

دراسة (*Mueller & Shwedel, 1975*) أجريت على عينة مؤلفة من (٤٧١) طالب جامعي، أظهرت الدراسة أن (٨٠%) من الطلاب غيروا على الأقل إجابة واحدة فقط، ونسبة التغيير لكل طالب كانت (٣,٧%). كما أسفرت النتائج أيضا أن (٦٥%) من الطلاب استفادوا من سلوك التغيير بل وارتفعت درجاتهم أيضا، ولا يوجد فرق بين الذكور والإناث في اتجاه التغيير. وعن مدى إدراك الطلاب لأهمية تغيير الإجابات؛ أكد أكثر من (٦٤%) من الطلاب أن التغيير يخفض من درجاتهم، و(٣٦%) منهم يعتقدون أن التغيير لن يغني من أمرهم شيئا.

دراسة (*Jackson, 1978*) فحصت تأثير موضع المفردات وصعوبة المفردة على سلوك تغيير المفردات، حيث وجدت الدراسة أن الطلاب يميلون إلى أن يغيروا آخر مفردات الاختبار أكثر من أوله، وأن الطلاب المنخفضين يغيرون المفردات الصعبة أكثر من المفردات الأسهل، وأن الطلاب يميلون إلى تغيير المفردات ذات القوة التمييزية الضعيفة.

كما أكدت أيضا دراسة (*Green, 1981*) التي أجريت على (٧٠ طالبا وطالبة) أن التغيير انصب على المفردات الأصعب سواء كان مرتفع أو منخفض القلق، وأن مرتفع القلق كان أكثر تغييرا للمفردات من منخفض القلق، ولا يوجد فرق في الاستفادة من تغيير المفردات بين مستويات القلق الثلاثة.

أما دراسة (*Skinner, 1983*) فقد وجدت أن نسبة الإجابات التي تغيرت من الخطأ للصح كانت (٥١%)، -وذلك على عينة من (٦٨ طالب وطالبة)- وأن الطالبات قد غيرن عددا من الإجابات ضعف ما غيره الطلاب.

وبحثت دراسة (*Matter, 1986*) سلوك مراجعة وتغيير الإجابات على اختبار ***Iowa Basic Skills*** وذلك على عينة من تلاميذ المرحلة الابتدائية. أسفرت النتائج أن جميع المفحوصين قد غيروا على الأقل إجابة واحدة فقط، وأن (٥٦%) من التغييرات كانت من الخطأ للصح، و(١٧%) من الصح للخطأ، و(٢٧%) من الخطأ للخطأ.

واستهدفت دراسة (*McMorris, DeMers, & Schwarz, 1987*) مدى تأثير توعية الطلاب (٩١ طالب وطالبة) بأهمية التعديلات ومراجعة وتغيير المفردات على تغيير المفردات بالفعل، وتمت مقارنة أدائهم بأداء عينة أخرى مماثلة لم تلق عليهم التعليمات، تبين أن إلقاء

التعليمات وتوعية الطلاب لم يؤثر كثيرا في نسبة تغيير المفردات. فمعظم الطلاب ارتفعت درجاتهم بالتغيير أكثر ممن انخفضت درجاتهم. وأن المفردات الأصعب هي الأكثر تغييرا من المفردات السهلة. وأن أسباب تغيير الإجابات هو المراجعة أو أخطاء كتابية أو إعادة القراءة أو التقاط تلميحات *clues* من أسئلة أخرى في نفس الاختبار.

أما دراسة (*Ramsey, Ramsey, & Barnes 1987*) التي أجريت على (٩٥ طالب وطالبة) فقد استهدفت فحص مدى إدراك الطلاب لأهمية تغيير المفردات. كشفت النتائج أن مرتفعي الثقة أحدثوا زيادة في الدرجات بسبب التغيير أكثر من منخفضي الثقة مثل دراسة (*Skinner, 1983*) كما كشفت النتائج أيضا أنه لا توجد فروق دالة بين مستويات القدرة في عدد المفردات التي تغيرت، ولا توجد فروق بين الجنسين في نسبة تغيير الإجابات.

كما أظهرت دراسة (*Shatz, & Best, 1987*) التي أجريت على (٦٥ طالب وطالبة) أن هناك (٨١%) من الطلاب أحدثوا تغييرات بمعدل (٤,٤ تغييرا) لكل فرد، كما كشفت نتائج الدراسة أن أهم أسباب تغيير الإجابات هو التخمين في الإجابات الخاطئة أما في تصحيح الإجابات فكان من أهم أسباب التغيير هو سوء قراءة السؤال أو التقاط تلميحات من أسئلة أخرى في نفس الاختبار.

أما دراسة (*Friedman & Cook, 1995*) التي أجريت على (٢٠٩ طالب وطالبة) فقد استهدفت التعرف على تأثير الأساليب المعرفية (التأمل- الاندفاع، والاستقلال- الاعتماد عن المجال الإدراكي) على تغيير الإجابات. وقد وجد الباحثان أن معظم الذين غيروا إجاباتهم قد زادت درجاتهم، كما أن هناك علاقة موجبة دالة بين زيادة الدرجات والأداء الكلي على الاختبار. غير أنهما لم يجدا أي ارتباط بين الأسلوب المعرفي وسلوك تغيير الإجابات.

كما كشفت نتائج دراسة (*Geiger, 1996*) التي أجريت على (٢٧٩ طالب وطالبة) أن نسبة الإجابات التي تغيرت من الخطأ للصح كانت (٥٦,٢%) وهناك (٧٢%) من الطلاب زادت درجاتهم بمراجعة إجاباتهم وتغييرها. وعلى خلاف النتائج التي أظهرتها الدراسات السابقة؛ فإن الطالبات قد غيرن من الإجابات بنسبة أعلى بدلالة من الطلاب، ولكن لم تزد درجاتهن بدلالة عن البنين. كما كان هناك (٧٦%) من الطلاب لديهم مفاهيم خاطئة عن تغيير الإجابات، وأن الطلاب الذين حصلوا على درجات مرتفعة هم فقط الذين يدركون مزايا التغيير.

أما دراسة (*Geiger, 1997a*) فقد هدفت إلى كشف العلاقة بين كل من سلوك مراجعة وتغيير المفردات واستشفاف الاختبار *testwiseness* ودرجة الأداء الكلي لأحد اختبارات الكليات، على عينة مكونة من (١٥٠ طالب جامعي). أسفرت النتائج عن أن (٩٣%) من الطلاب قاموا بتغيير إجاباتهم، وأن نسبة اتجاه التغيير من الخطأ للصح كان أعلى منه من الصح للخطأ أو من الخطأ للخطأ، وأن التغيير قد أدى بالفعل إلى زيادة دالة في درجاتهم. وكان هناك ارتباط موجب دال بين

نمط اتجاه التغيير من الخطأ للصح والدرجة الكلية للاختبار. غير أنه لم يكن هناك ارتباط دال بين درجة استشفاف الاختبار وسلوك تغيير الإجابات.

كما أسفرت نتائج دراسة (*Heidenberg, 2000*) التي شارك فيها (١٨١٩ طالب) أن (٧٧%) من الطلاب غيروا على الأقل إجابة واحدة فقط،و كان متوسط التغيير هو (٣ تغييرات) لكل طالب. وأن (٧٤%) من الطلاب الذين غيروا إجاباتهم زادت درجاتهم بسبب التغيير، أما النسبة المتبقية فإن درجاتهم إما أنها لم تتأثر أو انخفضت بالتغيير. وفيما يتعلق بالفروق بين الجنسين؛ أظهرت الدراسة أن البنات قد غيرن أعلى قليلا من البنين (٧٨% مقابل ٧٣%) وأن اتجاه التغيير (من الخطأ للصح) كان أعلى قليلا للبنات عن البنين (٦٦% مقابل ٦٣%) وكشفت نتائج الدراسة أن من أسباب تغيير الإجابات إما سوء فهم أو سوء قراءة أو خطأ أثناء الكتابة أو تخمين.

وقامت دراسة (*Ferguson, Kreiter, Peterson, Rowat, & Elliott, 2002*)بتطبيق اختبارات مكونة من (٨٠ مفردة) على (١٩٠ طالب وطالبة). وقد أسفرت النتائج عن أن (٤٩%) من الطلاب زادت درجاتهم بسبب التعديل. وأكدت الدراسة أنه كلما كانت المفردات صعبة كان احتمال تغييرها أكبر. وأن الطلاب المرتفعين كانوا أكثر استفادة من الطلاب المنخفضين.

وطبقت دراسة (*Al-Hamly, & Coombe, 2005*) مجموعة من اختبارات للغة الإنجليزية على (٢٨٦ طالب وطالبة) من جامعتي الكويت ودبي، فأظهرت النتائج أن معظم الطلاب (٦٧%) لجؤوا لتغيير إجاباتهم. من هؤلاء؛ (٥٧%) قد ارتفعت درجاتهم من أثر تغيير وتصحيح إجاباتهم، وأن نسبة تغيير الإجابات كانت (٢,٦٥%) لكل مفحوص. كما أظهرت النتائج أيضا أن هناك علاقة سالبة بين الدرجات النهائية وعدد التغييرات التي تمت؛ أي أن مرتفعي القدرة يميلون إلى عدم تغيير إجاباتهم. وأنه لا توجد فروق تذكر بين البنين والبنات في نسبة تغيير الإجابات، غير أن البنات كن أكثر من البنين في اتجاه الصحيح (من الخطأ إلى الصح). وأسفرت نتائج الدراسة أيضا أن معظم المفحوصين قد غيروا المفردات الأسهل وليس الأصعب، كما كان التغيير من الخطأ للخطأ على المفردات الأصعب وليس الأسهل.

وهدفت دراسة (*Higham & Gerrard 2005*) إلى تحديد دور ما وراء المعرفة في تغيير الإجابات على أسئلة معلومات عامة من نمط الاختيار من متعدد، وذلك على عينتين من (٢٤ و٤٨ من طلاب الجامعة). أظهرت النتائج أن إعطاء الفرصة للمفحوصين لمراجعة استجاباتهم يؤدي إلى زيادة درجاتهم؛ إلا أن النتائج أظهرت أيضا أنه بتحليل الأخطاء ليس كل الأخطاء متساوية أو قابلة للتصحيح.

أما دراسة (*Balance, 2006*) فقد أجرت معاملات ارتباطية بين خصائص المفردة

والدرجات الناتجة من تغيير المفردات، وذلك على (١١٣ سؤال) من نمط الاختيار من متعدد. أظهرت نتائج الدراسة أن مؤشرات الصعوبة لم ترتبط بدلالة مع درجات المفردات التي تغيرت. أما تمييز المفردات فقد كان له ارتباط منخفض مع تلك الدرجات خلافا لما اتفقت عليه الدراسات السابقة.

تعقيب على دراسات سلوك تغيير الإجابات في الإختبارات الورقية:

يتضح مما سبق عرضه؛ أنه قد أجريت عشرات الدراسات حول سلوك تغيير الإجابات ***answer-changing behavior*** في الاختبارات (الورقية) منذ أكثر من ٧٧ عاما حين أجرى ***Mathews, 1929*** (في ***Vispoel, 1998***) أول دراسة حول سلوك تغيير الإجابات ووجد أن (٨٦%) من الطلاب يعتقدون أن تغيير المفردات يؤدي إلى خفض درجاتهم في الاختبارات، وحتى دراسة (***Higham & Gerrard 2005***) التي أكدت أن إعطاء الفرصة للمفحوصين لمراجعة استجاباتهم يؤدي إلى زيادة درجاتهم.

وطبقا لنتائج البحوث في هذا المضمار؛ فقد أكدت النتائج أنه – وبخلاف الاعتقادات السائدة – لا ضرر مطلقا من تغيير الإجابات الأصلية التي وضعها المفحوص؛ بل يجب تشجيع الطلاب على ذلك. (***Al-Hamly, & Coombe, 2005; Stoffer, Davis, & Brown, 1977***).

كما اتفقت الدراسات أيضا أن معظم الطلاب يلجأون إلى سلوك تغيير الإجابات تراوح من (٦٧%) في دراسة (***Al-Hamly, & Coombe, 2005***) إلى (٩٣%) في دراسة (***Geiger, 1997a***). ومع ذلك لم تتغير إلا نسبة ضئيلة من الإجابات لكل مفحوص لم تتعد (٤,٤%) في دراسة (***Shatz, & Best, 1987***). وفي دراسة (***Friedman-Erickson, 1994***) التي تألفت عينتها من (٢٤٤طالب) لم تتجاوز الإجابات التي تغيرت (٣%) فقط.

كما كان معظم اتجاه التغيير من الخطأ إلى الصح، أي أن معظم الذين غيروا إجاباتهم قد صححوا أخطائهم أكثر من كونهم وقعوا في الأخطاء. حيث تراوحت نسبة التغيير في الاتجاه الصحيح من (٤٩,٥%) كما في دراسة (***Skinner, 1983***) إلى (٦٥%) كما في دراسة (***Stoffer, et al., 1977***). وفي دراسة (***Friedman-Erickson, 1994***) كانت نسبة اتجاه المفردات التي تغيرت من الخطأ إلى الصح (٥٦%).

كما أسفرت نتائج الدراسات أيضا أن معظم المفحوصين الذين غيروا إجاباتهم تحسنت درجاتهم بسبب التغيير. ففي دراسة(***Geiger, 1996; Stoffer, et al., 1977***) هناك (٧٢%) من الطلاب الذين غيروا إجاباتهم تحسنت درجاتهم. وفي دراسة (***Heidenberg, 2000***) فإن (٧٤%) من الطلاب الذين غيروا إجاباتهم زادت درجاتهم بسبب التغيير.

واختلفت نتائج الدراسات السابقة فيما يرتبط بمستويات قدرة الطلاب وعلاقتها بسلوك مراجعة

وتغيير المفردات؛ فبينما أشارت دراسة (*Jacobs, 1971*) أنه لا توجد علاقة بين مستوى قدرة الطلاب وتغيير الإجابات، ودراسة (*Ramsey, et al., 1987*) أنه لا توجد فروق دالة بين مستويات القدرة في عدد المفردات التي تغيرت؛ فإن دراستي (*Ferguson, et al., 2002; Mueller & Wasser, 1977*) أكدتا أن مرتفعي القدرة كانوا أكثر استفادة من تغيير المفردات من الطلاب المنخفضين. بينما دراسة (*Al-Hamly, & Coombe, 2005*) أشارت إلى أن مرتفعي القدرة يميلون إلى عدم تغيير إجاباتهم.

وهناك دراسات أخرى بحثت الفروق بين الجنسين على سلوك تغيير المفردات والتي لم تظهر في نتائجها فروقا دالة بين البنين والبنات في الاستفادة من سلوك التغيير (*Mueller & Wasser, 1977*) أو في اتجاه التغيير (*Mueller & Shwedel, 1975*) أو في نسبة تغيير الإجابات (*Al-Hamly, & Coombe, 2005; Ramsey, et al., 1987*). بينما وجدت دراسة (*Heidenberg, 2000*) أن هناك ارتفاع طفيف للإناث عن الذكور في نسبة التغيير (٧٨% مقابل ٧٣%)، وأن اتجاه التغيير (من الخطأ للصح) كان أعلى قليلا للبنات عن البنين (٦٦% مقابل ٦٣%). أما دراستا (*Geiger, 1996; Skinner, 1983*) فقد أكدتا أن الطالبات قد غيرن عددا من الإجابات بنسبة أعلى بدلالة من الطلاب.

بعض الدراسات فحصت سلوك تغيير المفردات في علاقته بمتغيرات الشخصية؛ ولم تكن ثم فروق دالة. فدراسة (*Stoffer, et al., 1977*) لم تجدا فرقا يذكر بين وجهتي الضبط الداخلي والخارجي وبين الاستفادة من تغيير المفردات. ودراسة (*Friedman & Cook, 1995*) لم تجد أي ارتباط بين الأسلوب المعرفي وسلوك تغيير الإجابات. أما دراسة (*Green, 1981*) فلم تجد فرقا دالا في الاستفادة من تغيير المفردات بين مستويات القلق الثلاثة. ولكن مرتفع القلق كان أكثر تغييرا للمفردات من منخفض القلق. ودراسة (*Ramsey, et al., 1987; Skinner, 1983*) أكدت أن مرتفعي الثقة أحدثوا زيادة في الدرجات بسبب التغيير أكثر من منخفضي الثقة.

وعن مدى إدراك الطلاب لأهمية تغيير الإجابات؛ تبين أن (٧٦%) من الطلاب لديهم مفاهيم خاطئة عن تغيير الإجابات (*Geiger, 1996*)، كما أكد أكثر من (٦٤%) من الطلاب أن التغيير يخفض من درجاتهم (*Mueller & Shwedel, 1975*)، حتى مع توعية الطلاب بأهمية هذه التغييرات لم يؤثر كثيرا في نسبة تغيير المفردات (*McMorris, et al., 1987*).

وعن أسباب تغيير المفردات بينت الدراسات أنها إما أخطاء كتابية أو أخطاء في قراءة السؤال أو التقاط تلميحات *clues* من أسئلة أخرى في نفس الاختبار أو التباس في الفهم أو مجرد تخمين. (*Heidenberg, 2000; McMorris, et al., 1987; Shatz, & Best, 1987*)

وفيما يتعلق بخصائص المفردات في ارتباطها بسلوك تغيير المفردات؛ فينك اتفاق بين

الدراسات (*Jacobs, 1971*) أن صعوبة المفردات مرتبطة بعدد المفردات التي تغيرت. ولكن باستثناء دراسة (*Al-Hamly, & Coombe, 2005*) التي أشارت إلى أن معظم المفحوصين قد غيروا المفردات الأسهل وليس الأصعب؛ فإن معظم الدراسات الأخرى (*Ferguson, et al., 2002; Green, 1981; Jackson, 1978; McMorris, et al., 1987*) أكدت أن المفردات الأصعب هي الأكثر تغييرا من المفردات السهلة. كما أسفرت دراسة (*Jackson, 1978*) أن الطلاب يميلون إلى أن يغيروا آخر مفردات الاختبار أكثر من أوله، وأن الطلاب يميلون إلى تغيير المفردات ذات القوة التمييزية الضعيفة؛ إلا أن دراسة (*Ballance, 2006*) أكدت أن مؤشرات تمييز المفردات التي تغيرت كانت لها معاملات ارتباط منخفضة الدرجات التي أحرزها الطلاب.

بالرغم من ثراء تلك النتائج؛ إلا أن التعرف على مدى التغيير الذي أحدثه المفحوصون على إجاباتهم في الاختبارات الورقية يكتنفه صعوبات جمة مقارنة بالاختبار التواؤمي المحوسب. فأساسا لا يمكن التعرف على تعديل الاستجابات أو تغييرها بصورة تامة في الاختبارات الورقية إلا من خلال الإجابات المشطوبة على ورقة الإجابة، أو من خلال التقارير الذاتية ***self reports*** من جانب الطلاب أنفسهم، وأحيانا ما تكون مضللة لأن الطلاب الذين غيروا إجاباتهم من الإجابة الخاطئة إلى الإجابة الصحيحة من الممكن ألا يتذكروا جيدا مثلما الحال بالنسبة الطلاب الذين غيروا إجاباتهم إلى الإجابات الخاطئة. بخلاف التطبيق المحوسب الذي يرصد اتجاه التعديل بكل سهولة ودقة وبحسب زمن عرض كل مفردة وغيرها من الخصائص المميزة التي يعرضها الباحث لاحقا.

ثانيا مراجعة وتغيير المفردات في الاختبارات التواؤمية المحوسبة:

تعتبر الاختبارات التواؤمية المحوسبة أحد أهم وأحدث تطبيقات نظرية الاستجابة للمفردة. ونظرية الاستجابة للمفردة هي أساس نظري جديد للقياس النفسي والتربوي لمعالجة المفردات وتصميم الاختبارات ظهرت بعد أن سادت نظرية الاختبار التقليدية[1] ***Classical test theory (CTT)*** طوال ما يزيد عن ٥٠ عاما مضت (والتي هي في الحقيقة مجموعة من الإجراءات التي تعالج بها المفردات مثل ثبات وصدق وتمييز المفردة بالطرق التقليدية).

ولهذه النظرية مجموعة من الفروض لابد من تحقيقها حتى يمكن استخدامها. كما تشترط مطابقة بيانات الاختبار لأحد النماذج الاستجابة للمفردة المحددة للنظرية. ولا تعتمد في تقدير درجة الأفراد على عدد الإجابات الصحيحة كما هو الحال في نظرية الاختبار التقليدية، إنما تعتمد على

---

[1] المتتبع لأدبيات القياس النفسي في هذا السياق يجد ثلاثة مصطلحات وصفت بها نظرية الاختبار التقليدية وهي: traditional/conventional/classical test theory.

نمط إجابة الفرد بناء على دالة رياضية مستندة على نظرية الاحتمالات تسمى دالة الاستجابة للمفردة.

وتستند الكثير من الاختبارات النفسية واختبارات التحصيل والاستعداد المقننة في الوقت الحالي على مبادئ تلك النظرية. ومن تلك الاختبارات اختبار الاستعداد المدرسي *Scholastic Aptitude Test (SAT)* واختبار كاليفورنيا للتحصيل *California Achievement Test (CAT)* واختبار استنفرد للتحصيل *Stanford Achievement Test* وبطارية وُدكُك-جنسن النفسية-التربوية *Woodcock-Johnson Psycho-Educational Battery (Hambleton, Zaal, & Pieters, 1994, p. 341)*.

وهناك اهتمام مواز في مصر بالاختبارات والمقاييس في سياق نظرية الاستجابة للمفردة (أمين نور الدين: ٢٠٠٢؛ أمينة كاظم: ١٩٨٨أ؛ أمينة كاظم ١٩٨٨ب؛ صلاح الدين علام: ١٩٨٥؛ محمد الشافعي ١٩٩٦) إلا أن هناك حاجة إلى مزيد من البحث والتقصي في هذا المضمار، وذلك لاستدراك التطورات الهائلة التي تعرض لها القياس النفسي في الغرب واللحاق بركب حركة القياس المعاصرة.

ولا تقتصر تطبيقات نظرية الاستجابة للمفردة على تقنين الاختبارات فحسب؛ بل تتيح أيضا بناء بنوك الأسئلة التي يمكن من خلالها إنشاء صور متكافئة من الاختبارات. كما تقدم نظرية الاستجابة للمفردة تطبيقا آخر يستهدف إمكانية تشبيك ***linking*** عدد من الاختبارات التي تقيس متغيرا واحدا بحيث تصير مقياسا واحدا؛ بالرغم من تطبيقها على عينات متفاوتة خلافا للأساليب التقليدية للقياس، والتي تشترط أن يكون التقنين على نفس العينة. غير أن أهم وأحدث تطبيقات نظرية الاستجابة للمفردة في مجال القياس النفسي الحديث هي الاختبارات التواؤمية المحوسبة *(Meijer & Nering, 1999; Wang & Kolen 2001)*

والاختبار التواؤمي المحوسب هو ذلك الاختبار النفسي أو التحصيلي أو غير ذلك الذي يطبق على المفحوصين من خلال الحاسب. إلا أنه في ذلك النمط من الاختبارات لا يتلقى كل مفحوص نفس العدد من الاختبارات الذي يتلقاه الآخرون، وليس أيضا بنفس الترتيب كما هو الشائع في الاختبارات التقليدية الورقية؛ إنما يختلف عدد المفردات المطبقة وترتيبها من فرد لآخر؛ حسب قدرة كل فرد. أي أنه يتم "تفصيل"[٢] ***tailoring*** الاختبار لكل فرد على حدة؛ بحيث لا يتعرض للمفردات السهلة جدا، ولا الصعبة جداً بالنسبة له؛ بل تقدم له المفردات التي تتناسب فقط مع

---

[2] تجدر الإشارة إلى أن الاختبارات التواؤمية كانت تسمى من قبل بالاختبارات المفصلة tailoring tests.

مستوى قدرته بهدف الوصول إلى دقة أعلى في عملية القياس. ونظرا لأنه يتم تقديم الاختبار التواؤمي عن طريق الحاسب، لذا يوصف بأنه محوسب ***computerized***.

وتستند العملية التواؤمية ***adapting*** في الاختبار على نظرية الاستجابة للمفردة كأساس نظري حيث تقدم احتمالية الإجابة الصحيحة لكل مفردة على حدة؛ وذلك بناء على عينة مختبرة من قبل. وللاختبار التواؤمي المحوسب أركان محددة تتضمن اختيار نموذج استجابة للمفردة مناسب، وبناء مستودع المفردات، وأسلوب انتقاء المفردة، وأسلوب تقدير القدرة، وأسلوب قاعدة الإنهاء [٣].

وهناك عدد من المؤسسات والهيئات في كثير من الدول الغربية والولايات المتحدة الأمريكية اعتمدت الاختبار التواؤمي المحوسب في أنظمتها مثل مؤسسة الدفاع الأمريكية التي تبنت إصدار بطارية الاستعداد المهني للقوات المسلحة ***Armed Services Vocational Aptitude Battery (ASVAB)*** وجعلتها بطارية تواؤمية محوسبة. وتتألف البطارية من (١٠ اختبارات فرعية)؛ وقد اشتركت جامعة مينسوتا مع مؤسسات الدفاع الأمريكية في إجراء تجارب أولية حول الاختبار التواؤمي المحوسب لتلك البطارية منذ عام (١٩٧٧)(***Moreno, Segall, & Hetter 1997***). وكذلك مؤسسة الاختبارات التربوية ***Educational Testing Services (ETS)*** التي حولت عددا من الاختبارات الشهيرة إلى اختبارات تواؤمية محوسبة مثل "اختبار اللغة الإنجليزية كلغة أجنبية" ***Test of English as a Foreign Language (TOEFL)*** و ***Scholastic Assessment tests***، ***Graduate Record Examination (GRE)***.

كما قامت الزمالة الطبية الأمريكية بتحويل اختباراتها أيضا إلى تواؤمية محوسبة. وأصدرت بعض دور النشر الخاصة الصورة التواؤمية المحوسبة لاختباراتها مثل اختبارات الاستعدادات الفارقة ***Differential Aptitude Tests. (Meijar & Nering, 1999, p. 187; Wang & Vispoel, 1998, p. 109; Zwick, Thayer, & Wingersky, 1994a, p. 121)***.

ومن الجدير بالذكر في هذا الصدد أن شركة مايكروسوفت الشهيرة في أنظمة الحاسب قد تبنت مؤخرا نظام الاختبارات التواؤمية المحوسبة في منحها لشهادتها الرسمية المعروفة في مجال صيانة الشبكات والبرمجة وغيرها. ويشير (***Meijar & Nering, 1999, p. 187***) إلى أن هناك تجارب ناجحة في هولندا لاستخدام الاختبارات التواؤمية المحوسبة في القياس التربوي. ونتيجة لهذا التوسع الهائل في الاستخدام؛ فقد نشطت العديد من المنظمات والجمعيات العلمية والمجالس القومية في الولايات المتحدة الأمريكية منذ أكثر من عشرة سنوات لوضع معايير لتصميم وتطبيق

---

[3] لمزيد من التفاصيل حول هذا الموضوع انظر (أمين نور الدين، ٢٠٠٢) في قائمة المراجع، أو على موقع الباحث www.geocities.com/ameensabry

الاختبارات التربوية والنفسية التواؤمية المحوسبة (*Georgiadou, Triantafillou, & Economides, 2006*).

ويحرص علماء القياس النفسي على إنشاء اختبار تواؤمي محوسب مكافئ تماما للإصدار الورقي سواء في الخصائص السيكومترية أو في الجوانب التطبيقية. ومن تلك الجوانب التطبيقية إتاحة الفرصة للمفحوصين لمراجعة استجابتهم وتعديلها ما أمكن لهم ذلك، بمثلما هو قائم في الاختبارات الورقية، مما دفع العديد من الباحثين إلى القول بضرورة تضمين هذا الإجراء أيضا في الاختبارات التواؤمية المحوسبة (*Mills, & Stocking, 1996*).

فضلا عن أنّ التطبيق على الحاسب يفسح المجال لخصائص جديدة لم تكن متاحة من قبل في الاختبارات الورقية، ويحل مشكلات إجرائية لرصد التعديلات، وتحديد التغييرات على الاستجابات.

فمن الخصائص المميزة للتطبيق على الحاسب فيما يتعلق بتعديل استجابات الطلاب:

١- رصد تحركات المفحوص داخل الاختبار بدقة تامة وبكل سهولة.

٢- تحديد اتجاهات التعديل الثلاثة التقليدية (أي من الخطأ للصح، ومن الصح للخطأ، ومن الخطأ للخطأ) بدقة تامة.

٣- بل وتحديد اتجاه رابع جديد، وهو التراجع عن التعديل (أي اختيار المفحوص إجابة ما، ثم تعديلها باختيار إجابة أخرى، ثم التراجع عنها هذا التعديل، والإبقاء على الإجابة الأصلية)، مما يتيح الفرصة لدراسة بعض الأساليب المعرفية للمفحوصين.

٤- تمييز الطلاب الذين قاموا بمراجعة استجاباتهم السابقة واستعراضها فقط، حتى وإن لم يعدّلوا استجاباتهم، مع رصد زمن مراجعتهم لكل مفردة. وهذا يكاد يكون غير متاح في الاختبارات الورقية.

٥- قياس زمن عرض كل مفردة بجزء من الثانية، وذلك يفسح المجال لقياس زمن كمون الاستجابة.

بالرغم من تلك الخصائص الفريدة لرصد خاصية مراجعة المفردات وتغيير الإجابات في تطبيق الاختبارات على الحاسب؛ إلا أن السماح بتعديل المفردات في الاختبارات التواؤمية المحوسبة يكتنفه إشكاليات سيكومترية يتمثل في عدم الدقة حيث أن السماح للمفحوص بمراجعة وتبديل إجاباته من الصحيحة إلى الخاطئة، أو بالعكس من الخاطئة إلى الصحيحة يؤدي إلى انتقاء مفردات غير المستهدفة، أو لا تلائم مستوى المفحوص الحقيقي بعد التحسين فالمفردة التي تلي المفردة المبدّلة لن تكون هي المستهدفة، مما قد يؤدي إلى شيء ما من عدم الدقة، علاوة على أنه قد يزيد من المسافة بين تقدير قدرة المفحوص وصعوبة المفردات المنتقاة سلفا.

ولذلك يرى بعض علماء القياس النفسي أنه من الصعوبة بمكان تحقيق التكافؤ بين الاختبارات

الورقية ونظيراتها التواؤمية. ومع ذلك فالمفحوصون يرغبون في تضمين هذا الإجراء لتصحيح بعض أخطائهم، فالفرق الرئيسي بين اختبارات الورقة والقلم والاختبارات التواؤمية -من وجهة نظر المفحوصين- هو أنه لا يسمح لهم بالعودة لمراجعة استجاباتهم في الاختبارات التواؤمية المحوسبة (*Vispoel, Rocklin, & Wang, 1994; Vispoel, 1998; Wise, Roos, Plake, & Nebelsick-Gullett, 1994*).

ويرى بعض الباحثين أن صدق الاختبار التواؤمي المحوسب يمكن أن يزداد عندما يراجع المفحوص إجاباته، لأن ذلك يجعله يعيد التفكير في إجاباته، ويصحح إجابة السؤال الذي ربما قرأها خطأ أو فهمها خطأ (*Vispoel, 1998*). كما أن السماح بالمراجعة أيضا يقرب تقدير القدرة النهائي من قدرة المفحوص الحقيقية، ويجعلها أكثر دقة عندما يقوم بتصحيح أخطائه السابقة (*Wise, 1996; Vispoel, et al., 2000*).

بينما يرى باحثون آخرون أن كفاية الاختبار التواؤمي المحوسب يمكن أن تنخفض إذا سمح للمفحوصين بالمراجعة، مما يفقده الغرض الذي أنشئ من أجله، ولذلك فمراجعة وتعديل الاستجابات غير مسموح به في معظم الاختبارات التواؤمية المحوسبة (*Georgiadou, et al., 2006; stocking, 1997; Vispoel, Rocklin, Wang, & Bleiler, 1999; Wise, 1996*). بل إن السماح بالمراجعة ينتقص من صدق الاختبار التواؤمي المحوسب أيضا، لأن إتاحة الفرصة للمفحوص بالمراجعة يمكنه من استشفاف الاختبار.

فهو قد يجيب -متعمدا- إجابات خاطئة من أول مفردة إلى آخر مفردة، حتى تنتقى له دائما المفردات الأسهل تباعا. وفي نهاية الاختبار يقوم المفحوص -عن طريق استغلال خاصية تغيير الإجابات- بالعودة مرة أخرى إلى نفس تلك الأسئلة التي لن تكون صعبة بالنسبة له؛ فيعدل الإجابات ويختار الإجابات الصحيحة كلما أمكنه ذلك. ومن ثم يحصل على تقدير للقدرة أعلى مما ينبغي، ويسمى هذا الأسلوب باستراتيجية (*Kingsbury (Stocking, 1997*) واستراتيجية (*Wainer, 1993*) وهما أول من حذرا من هذا السلوك، وهي أحد أساليب استشفاف الاختبار *testwiseness or cheating*.

وبناء عليه؛ فقد لاقى تضمين مراجعة وتعديل الاستجابات في الاختبارات التواؤمية المحوسبة معارضة شديدة من جانب بعض علماء القياس النفسي لأنه يمكن أن يعرقل مسار العملية التواؤمية، ويؤخر زمن الاختبار، ويقلل من دقة القياس، ويزيد من احتمالية حصول المفحوص على درجات متضخمة. (*Wise, 1996*)

لذا؛ فقد تساءلت الدراسات عن مدى إمكانية تزويد هذه الخاصية في الاختبارات التواؤمية المحوسبة. وإلى أي مدى يُنقص إضافة هذه الخاصية من كفاءة الاختبار من جانب المفحوص،

ويرفع من تقدير قدرة الأفراد. وهل السماح باستخدامها يمس عدالة الاختبار؟ ومن تلك الدراسات ما يلي:

دراسة (*Lunz, et al., 1992*) التي انصبت على فحص مراجعة المفردات وتبديل الاستجابات وتأثير ذلك على تقديرات قدرة المفحوصين وكفاية الاختبار التواؤمي المحوسب. حيث تكونت عينة الدراسة من مجموعتين الأولى من (٢٢٠ مفحوص)؛ سمح لها بتعديل استجاباتها بعد انتهاء الاختبار، أما الثانية فتكونت من (٤٩٢ مفحوص)؛ لم يسمح لها بالإجراء السابق، وجميعهم من أحد المعاهد الطبية. كشفت الدراسة عن وجود ارتباط عال بين تقديرات القدرة للمجموعتين: المسموح لها بالتعديل والأخرى غير المسموح لها بلغ (٠,٩٨) أما متوسط كفاءة الاختبار فلم ينقص إلا (١%) فقط لمجموعة التعديل. وحوالي (٣٢%) من المفحوصين حسنوا من درجاتهم بعد التعديل، غير أنه لم يتغير موقفهم من النجاح أو الرسوب. ونظرا لأن أداء العينة المسموح لها بالتعديل قد تحسن مع تغير طفيف في الثبات والكفاءة؛ فقد أوصى الباحثون بضرورة تضمين إمكانية التعديل في الاختبار التواؤمي المحوسب.

واستهدفت دراسة (*Vispoel, et al., 1992*) فحص تأثيرات مراجعة وتغيير الإجابات على حجم الدرجات، وثباتها، وكفايتها، وصدقها التلازمي، من خلال تطبيق ثلاثة أنماط من الاختبارات: اختبار خطي محوسب، واختبار موائم ذاتيا[٤] *Self-adapted testing* واختبار تواؤمي محوسب على عينة قوامها (٩٧ طالبا) من الكليات مقسمين إلى ثلاث مجموعات. أظهرت الدراسة أن مراجعة المفحوص للإجابات حسّن من أداء الاختبار؛ إلا أنه أطال من مدته بمقدار حوالي (٢٧%)، كما زادت نسبة تباين الخطأ أيضا. وقد كان معامل الارتباط بين درجات المفحوصين قبل وبعد المراجعة (٠,٩٨) وخلصت الدراسة إلى تضمين هذا الإجراء في الاختبارات التواؤمية المحوسبة؛ حيث أنه هو المفضل للمفحوصين.

أما دراسة (*Stone & Lunz, 1994*)فقد هدفت إلى فحص تعديل المفحوص لاستجاباته، وتأثير ذلك على تقدير القدرة، ودقة الاختبار، ودالة معلومات الاختبار، والثقة في القرار *decision confidence* وموقف النجاح والرسوب *pass/fail status*. تكونت عينة الدراسة من مجموعتين (٢٠٨، ١٦٨ مفحوص)، من أحد المعاهد الطبية حيث طُبِّق على كل من المجموعتين اختبار قبلي لم يسمح فيه بتغيير استجاباتهم بعد انتهاء الاختبار؛ ثم أعيد بعد ذلك نفس

---

[4] الاختبار الموائم ذاتيا وهو الانتقاء الذاتي لمستوى الصعوبة من جانب الفرد في الاختبارات التواؤمية المحوسبة. حيث يقرر هو بنفسه المسار الذي يسلكه سواء الارتقاء إلى مستوى أصعب؛ أو اختيار مستوى آخر أسهل. ويهدف هذا الأسلوب إلى خفض قلق الاختبار عندما يتحكم الفرد بنفسه في موقف الاختبار.

الاختبار وسُمح لهم فيه بمراجعة المفردات وتعديلها. بلغ مقدار ارتباط تقديرات القدرة قبل وبعد التعديل (٠,٩٨) بالرغم من زيادة متوسط تقديرات القدرة بعد المراجعة والتعديل عن قبله بصورة طفيفة. أما الثقة في القرار فقد كانت أقل نوعا ما بعد التعديل. وقد أثَّر السماح بتبديل الإجابة في دقة الاختبار بصورة طفيفة.

وهدفت دراسة (***Lunz & Bergstrom, 1994***) التي أجريت على (٦٤٥ طالب) من أحد المعاهد الطبية إلى فحص بعض خصائص تطبيق الاختبار التواؤمي المحوسب متضمنا أربعة أنماط للتحكم في الإجابة على المفردات مقسمين إلى أربع مجموعات لعينة الاختبار: المجموعة الأولى "تخطي المفردات" (٣٣% من العينة) حيث يمكن للطالب أن يتخطى ***skip*** المفردة التي لا يريد الإجابة عليها؛ وعندئذ تظهر له مفردة أخرى من نفس محتوى المجال المستهدف قياسه. المجموعة الثانية "مراجعة وتعديل المفردات" (١٧% من العينة) حيث يسمح للطالب بعد الانتهاء من الاختبار أن يقوم بمراجعة إجاباته وتعديلها إن رغب في ذلك. المجموعة الثالثة "المفردات المحددة" (١٥% من العينة) حيث يقوم الطالب أثناء الإجابة على أسئلة الاختبار بتحديد المفردات التي يرغب في تغييرها وبعد الانتهاء من أسئلة الاختبار تظهر له هذه المفردات فقط التي يعدلها. المجموعة الرابعة (٣٥% من العينة) لم يسمح لها بالتعديل مطلقا. أظهرت الدراسة أن متوسط درجات المجموعة الأولى أعلى بدلالة من المجموعة الرابعة، بينما لم تكن هناك فروق دالة بين المجموعات الثلاثة الأولى. غير أن المجوعة الأولى حصلت على أقل معامل ثبات من المجموعات الأخرى.

أما دراسة (***Gershon & Bergstrom, 1995***) فقد فحصت إمكانية استشفاف المفحوص إجاباته من خلال استغلال خاصية المراجعة وتعديل الإجابات. فقام الباحثان بإنشاء مستودعين للمفردات أحدهما سهل (بمتوسط –٢,٠) والآخر شديد السهولة (بمتوسط –٤,٠). ثم قاما الباحثان بمحاكاة استجابات المفحوصين على أساس الإجابة بصورة خاطئة في البداية ثم تصحيح الإجابات بعد ذلك. توصلت الدراسة إلى أن الخطأ المعياري لتقدير القدرة لم يتأثر بأقصى درجة حصل عليها المفحوص حيث لم تتعد (١,٤٢) من الخطأ المعياري في أي من مستودعي المفردات. كذلك لم تتأثر تقديرات القدرة حيث لم تتعد أكثر من (٣,١٩) من وحدة اللوجيت مما يشير إلى أن تضمين خاصية المراجعة وتعديل الإجابات في الاختبارات التواؤمية المحوسبة لا يؤثر في تقديرات القدرة أو الأخطاء المعيارية للقياس.

وفحصت دراسة (***Ferrara, Frances, Gilmartin, Knott, Michaels, Pollack, Schuder, Vaeth, & Wise, 1996***) أسباب ودوافع تغيير الإجابات لأحد اختبارات الحساب التواؤمية المحوسبة لدى عينة صغيرة من (٢٩ تلميذا وتلميذة) في المرحلة الإعدادية. حيث وجدت

أن (٦٢%) من التلاميذ قاموا بتغيير إجاباتهم. وأن (٤٥%) من الذين غيروا إجاباتهم لم يجتازوا الاختبار المقدم لهم. وأنهم قد استغرقوا من (٢٠-٣٠ دقيقة) إضافية على الوقت الأصلي لإتمام الاختبار. وأن معظم اتجاه التغيير كان من الخطأ إلى الصح (٤٩% من الخطأ إلى الصح؛ ٢٦% من الصح إلى الخطأ؛ ٢٥% من الخطأ إلى الخطأ). وأظهرت الدراسة كذلك أن أحد أهم أسباب التغيير هو اكتشاف أخطاء أثناء إعادة الحساب مرة أخرى (٤٩%). غير أن كثيرا من التلاميذ لم يفصحوا عن الأسباب الحقيقية للتغيير.

وفي المقابل أجرت (*Stocking, 1997*) دراسة من خلال عينة محاكاة على عدد من نماذج المراجعة والتعديل. وقد وجدت الباحثة أن استخدام أسلوب الأرجحية القصوى في تقدير القدرة قد شوَّه النتائج بدرجة ملحوظة عند استثمار خاصية المراجعة والتعديل. وأن أفضل أسلوب توصلت إليه الدراسة هو إضافة عنصر المراجعة والتعديل، ولكن بصورة مقيدة عن طريق تقسيم المستودع إلى مجموعات صغيرة من المفردات المرتبطة بالمستودع الرئيسي. ودلت النتائج أن هذا الأسلوب حقق نتائج مقبولة سيكومتريا.

أما دراسة (*Vispoel, 1998*) فإنها هدفت إلى فحص أداء (٣٧٩ طالب جامعي) على كل من الاختبار التواؤمي المحوسب والاختبار الموائم ذاتيا من خلال السماح بتعديل الإجابات مقابل عدم السماح بذلك. وقد اتسقت نتائج الدراسة مع النتائج السابقة؛ فعندما تعطى للمفحوصين الفرصة لتغيير استجاباتهم؛ فإن معظمهم يفعلون ذلك، غير أنهم يغيرون بعض الإجابات فقط، وقد استفاد المفحوصون من هذه الخاصية كما تمثل من خلال ارتفاع درجاتهم على القدرة المقاسة مقارنة بدرجاتهم فيما لو لم يحدث تعديل. وكشفت الدراسة عن أن اتجاه التعديل كان من الخطأ للصح أكثر من التعديل من الصح للخطأ. غير أن السماح بالتعديل زاد من زمن الاختبار بنسبة تراوحت بين (٣٧-٦١%) مؤكدة بذلك على نتائج الدراسات السابقة. كما حصلت المجموعة التي سمح لها بالمراجعة والتعديل على خطأ معياري للقياس أعلى من المجموعة الأخرى بغير دلالة بخلاف الدراسات السابقة والتي كانت فيها فروق دالة.

وافترضت دراسة (*Vispoel, et al., 1999*) أن أسلوب تقدير القدرة له دور في التقديرات المتحيزة إذا أجاب المفحوص إجابة خاطئة ثم استغل ميزة إمكانية التعديل فيصحح الاختبار ويحصل بذلك على درجات متحيزة أي مرتفعة بصورة زائفة. حيث فحصت الدراسة تأثير أسلوبي تقدير القدرة (أسلوب تقدير الأرجحية القصوى، والأسلوب البييزي) على حصول المفحوص على تقديرات متحيزة إذا استغل ميزة إمكانية التعديل. ومن خلال عينة محاكاة مؤلفة من (١٧٠٠ فرد)، وأخرى حقيقية مؤلفة من (٨٧ طالبا)، وجدت الدراسة أن تقدير القدرة بأسلوب الأرجحية القصوى يمكن أن يعطي تقديرات قدرة متحيزة إذا سلك المفحوص هذا المسلك حتى أن المفحوص قد يكون

تقدير قدرته الحقيقي من (٠,٤) ومع هذا؛ يحصل على تقدير مرتفع للقدرة يصل إلى (٤,٠) بخلاف أسلوب البعدي الأقصى الذي قد يحمي الدرجات إلى حد كبير من التضخم.

أما دراسة (*Wise, et al., 1999*) فقد حاولت أن تتحقق من أن مسلك استشفاف الإجابات من جانب المفحوص باستغلال خاصية تعديل الإجابات يعتمد على مدى تمييز المفحوص لصعوبة المفردات وتدرجها (أي يستطيع تمييز تدرج صعوبة المفردة فيجيب إجابة خاطئة حتى تظهر له مفردة أسهل. وبناء على هذا؛ أجريت الدراسة ثلاثة تجارب مختلفة على عينات واقعية من طلاب الجامعة، وعينات محاكاة مشتقة من أداء سابق. دلت نتائج الدراسة على أن (٧٣%) من المفحوصين يمكنهم تمييز صعوبة المفردات بصورة جيدة، ومن ثم يكون لديهم فرصة كبيرة في تغيير إجاباتهم بصورة مزيفة؛ من خلال اكتشاف تدرج صعوبة المفردة ومن الممكن عندئذ تغيير إجاباتهم.

وقام (*Vispoel, et al., 2000*) بدراسة استهدفت فحص تأثير إتاحة درجات متفاوتة من تعديل المفردات على زمن الاختبار، وتقدير القدرة، والأخطاء المعيارية، وعدد الإجابات الصحيحة. فقام الباحثون بتقسيم أفراد العينة الجامعية البالغ عددها (٢٤٢ طالب) إلى أربع مجموعات: أتيح للمجموعة الأولى تعديل المفردات في نهاية الاختبار بصورة مطلقة، بينما لم يسمح للمجموعة الأخرى بذلك التعديل، أما المجموعتين الأخريين فقد سمح لهما بالتعديل بصورة مقيدة حيث يتم تقسيم مستودع المفردات إلى وحدات أو مجموعات ***blocks*** تحوي كل منها (٥ أو ١٠ مفردات) حيث يسمح للفرد بالتعديل داخل المجموعة فقط فإذا ما انتقل إلى المجموعة الأخرى لا يسمح له بالتعديل في المجموعة السابقة.

بينت نتائج الدراسة أنه لا توجد فروق دالة بين المجموعات الأربعة سواء في تقدير القدرة أو في خطأ القياس أو في زمن المستغرق. وأن عدد الإجابات الصحيحة قد زاد بصورة طفيفة بعد التعديل، وأن اتجاه التغيير كان من الخطأ للصح أكثر منه من الصح للخطأ.

واقترحت دراسة (*Papanastasiou, 2002*) إجراءا جديدا يسمح بمقتضاه للمفحوص بمراجعة وتغيير المفردات التي يراها خاطئة (تم تحديد خمس مفردات للتعديل فقط)، وإعادة تنظيم وتخطي مفردات الاختبار، ثم إعادة تقدير القدرة ومن ثم تزداد الدرجات دقة عن ذي قبل بدون السماح للمفحوصين باستشفاف الاختبار. أظهرت النتائج التي أجريت على (٢٦٠٠٠ فرد) محاكى ***simulator***أن هذا الإجراء فعال في خفض التحيز لتقديرات القدرة، حيث انخفض التحيز بعد تنفيذ هذا الإجراء بنسبة (١٥,٦%) مقارنة بمتوسط قيمة التحيز بعد أسلوب المراجعة العادي، كما انخفض التحيز عند معظم مستويات القدرة باستثناء المستويات الطرفية، ولكن انخفض الثبات بصورة طفيفة بعد تنفيذ هذا الإجراء. فقد كان بعد المراجعة (٠,٨١١) ولكن بعد تنفيذ الإجراء كان

(٠,٨٠٦). وتؤكد الباحثة أن هذا الإجراء لا يؤثر في تقديرات القدرة كما لا يحتاج إلى زمن إضافي للأفراد للمراجعة.

وامتدادا لنتائج الدراسات السابقة؛ وبخلاف ما توصل إليه (*Wise, et al., 1999*)؛ أجريت دراسة (*Vispoel, Clough, Bleiler, Hendrickson, & Ihrig, 2002*) على عينة واقعية – وليست محاكاة– لمفحوصين تعلموا كيف يستخدمون استراتيجيات استشفاف مفردات الاختبار مع وضع بعض المراجعة المقيدة لذلك ***restricted review*** مقارنة بمجموعات أخرى راجعوا وغيروا من الاختبار بدون أي قيود على عملية المراجعة. قسمت عينة الدراسة إلى مجموعات تستخدم كل منها حالة من حالات استراتجيات استشفاف الاختبارات؛ ثلاثة منهم ذات عدد محدد من المفردات، والثلاثة الآخرين ذات عدد مقيد من التعديلات، علاوة على حالة سابعة يسمح للمفحوصين فيها بالمراجعة والتغيير دون أية قيود على ذلك. أثبتت الدراسة أن مجموعة المراجعة غير المقيدة كانت أعلى في تقديرات القدرة من المجموعات التي استخدمت معها القيود على المراجعة لتلافي الاستشفاف. مما يؤكد مرة أخرى على إتاحة الفرصة للمفحوصين بالمراجعة والتعديل بصورته الطبيعية يؤدي إلى نتائج أفضل.

كما قام (*Vispoel, Clough, & Bleiler 2005*) بدراسة مماثلة لدراسته السابقة، وتأكيدا على افتراض أن استخدام المفحوصين لاستراتيجيات استشفاف الاختبار يعتمد على كيفية عمل الاختبار التواؤمي المحوسب أكثر من كونه تمييز المفحوص لصعوبة المفردات. أجريت الدراسة على (٢١٨ طالب جامعي) مقسمين إلى سبع مجموعات؛ ست مجموعات منها تم تعليمهم كيف يستخدموا استراتيجيات متنوعة لاستشفاف الاختبار للحصول على أفضل نتيجة ممكنة. أظهرت الدراسة أن ارتفاع أو انخفاض قدرة المفحوص يعتمد على تقديره لصعوبة المفردات، واتساق أسلوب انتقاء المفردات، وعندما يكون غير متأكد من صحة إجاباته. كما أكدت الدراسة ما سبق أن توصلت إليه الدراسات السابقة أن أعلى تقدير للقدرة حازت عليه مجموعة المراجعة غير المقيدة.

تعقيب على دراسات سلوك تغيير الإجابات في الاختبارات التواؤمية المحوسبة:

لقد كان الهدف من هذه الدراسات هو فحص تأثير تعديل المفحوص لاستجاباته على الخصائص السيكومترية المختلفة في مجال الاختبار التواؤمي المحوسب. وتضمنت هذه الخصائص السيكومترية المستهدفة للدراسة: تقديرات قدرة المفحوصين، وكفاءة الدرجات، وصدقها التلازمي، ودقة الاختبار، والثقة في القرار، وموقف النجاح والرسوب. وكان من أهم المحكات في هذه الدراسات: متوسط تقديرات القدرة، وعدد المفردات المراجعة والمستبدلة، واتجاه التعديل (من الصح إلى الخطأ أو من الخطأ إلى الصح أو من الخطأ إلى الخطأ) وتجدر الإشارة إلى أن التعديل كان بعد انتهاء الاختبار وليس أثناءه.

وبالرغم من اختلاف الدراسات السابقة في إجراءاتها التجريبية؛ إلا أنها اتفقت فيما بينها في نتائجها: حيث تمكن حوالي ثلث عينة المفحوصين من تحسين درجاتهم بعد التعديل. وبلغ معامل الارتباط بين درجات المفحوصين مع إجراء التعديل وبدونه (٠,٩٨) في معظم الدراسات. كما لم يتغير موقفهم من النجاح/الرسوب.

وفيما يتعلق باتجاه التصحيح للمفحوصين الذين قاموا بتغيير إجاباتهم؛ أشارت الدراسات (*Ferrara, et al., 1996; Vispoel, 1998*) أن معظم اتجاه التغيير كان من الخطأ إلى الصح أكثر من التغيير من الصح للخطأ. أفادت الدراسات في هذا الإطار أيضا أن معظم المفحوصين رغبوا في تضمين هذه الميزة حتى يتمكنوا من تعديل استجاباتهم. (٨٧%) في دراسة (*Vispoel, et al., 1992*) و(٨٥%) في دراسة (*Vispoel, 1998*). فتوفير هذه الخاصية يجعل موقف الاختبار أقل ضغطا لاسيما لدى المفحوصين ذوي قلق الاختبار المرتفع. بالإضافة إلى هذا؛ فإن بعض المفحوصين يعتبر توفير تلك الخاصية "حق مكتسب" على حد تعبير (*Vispoel, 1998, p. 338*). وفي دراسة (*Baghi, Gabrys, & Ferrara, 1991, p. 9*) أعرب الطلاب الذين طبق عليهم كلا من التطبيقين التواؤمي المحوسب والاختبار الورقي عن ضيقهم من عدم تضمين أي وسيلة لمراجعة إجاباتهم لاسيما تعديلها.

ومن اللافت للنظر أن النتائج السابقة كانت متسقة إلى حد بعيد مع النتائج التي أجريت باستخدام اختبارات الورقة والقلم التقليدية طوال سبعة وسبعين عاما الماضية. مما دفع العديد من الباحثين إلى القول بضرورة تضمين هذا الإجراء أيضا في الاختبارات التواؤمية المحوسبة. ولكن في المقابل؛ انخفضت دقة الاختبار للمجموعة التي أجرت التعديل، وتأثرت كفاءة الاختبار بالتبعية بنسبة (١%) على الأقل. وزاد زمن الاختبار من (٢٧%) في دراسة (*Vispoel, et al., 1992*) إلى حوالي (٦١%) في دراسة (*Vispoel, 1998*).

وهناك اتجاه آخر للدراسات التي أجريت في سياق الاختبارات التواؤمية المحوسبة وهو تضمين عنصر مراجعة وتعديل المفردات مع تقييد المراجعة (*Papanastasiou, 2002; Stocking, 1997*) ، إلا أن هذا التقييد لم يكن فعالا إلا مع المفحوصين الذين يسعون بالفعل لاستشفاف الاختبارات. (*Vispoel, et al., 2002*) وأن وضع للمراجعة والتعديل هو أن تكون بصورتها الطبيعية كما أشار إلى ذلك (*Vispoel, et al., 2005*) دون قيود على أسلوب انتقاء المفردات كما اقترح (*Papanastasiou, 2002; Bowles, & Pommerich 2001*) أو على تقسيم مستودع المفردات كما طرحت ذلك (*Stocking, 1997*).

لقد أثارت هذه النتائج جدلا واسعا لدى الباحثين حول وجوب تضمين أو استبعاد هذا العنصر من الاختبارات التواؤمية المحوسبة لما يتعلق بها من تضمينات سيكومترية ومدى عدالة

الاختبارات. ففريق يرى ضرورة السماح للمفحوص بالمراجعة والتعديل مطلقا أسوة بالاختبارات الورقية(*Gershon & Bergstrom, 1995; Stone & Lunz, 1994; Vispoel, et al., 2005*) ، وفريق آخر يرى وجوب استبعادها أو تقييدها بشكل أو بآخر(*stocking, 1997; Wainer, 1993; Wise, 1996*).

فمع التوسع في استخدام الإختبار التواؤمي المحوسب، وتحول العديد من الاختبارات الورقية الى اختبارات تواؤمية محوسبة، ومع تزايد الرغبة في تضمين عنصر مراجعة وتعديل الإجابات، وما أثار ذلك من قضايا جديدة تتعلق فعالية وكفاية الاختبارات التواؤمية وتأثير ذلك الإجراء على تقديرات القدرة ودقة القياس؛ تتساءل الدراسة الحالية عن جدوى توفير خاصية مراجعة وتعديل الإجابات في الاختبارات التواؤمية المحوسبة؟ وما تأثير ذلك على تقديرات القدرة للمفحوصين وعلى دقة وكفاية القياس؟ لا سيما وأن هذه الدراسات لم تهتم بفحص تأثير الفروق بين الجنسين على سلوك التعديل، وكذلك الفروق بين مستويات القدرة، كما أنها لم تدرس بصورة دقيقة تأثير هذين المتغيرين على اتجاه التصحيح.

المنهج والإجراءات

تهدف الدراسة الحالية إلى الكشف عن مدى استفادة الأفراد لخاصية مراجعة وتعديل المفردات في الاختبارات التواؤمية المحوسبة، وتأثير ذلك على تقدير قدرة المفحوصين ودرجاتهم في الاختبار. ومدى تأثر كفاية الاختبار ومقدار المعلومات ودقة القياس بهذا الإجراء. وجدوى تضمين خاصية هذه الخاصية في الاختبارات التواؤمية المحوسبة. ولتحقيق هذا الهدف اتخذ الباحث الإجراءات التالية:

١- اختيار عينة الدراسة

تكونت عينة الدراسة من (٢٣٥) طالب وطالبة من طلاب الفرقة الأولى والثالثة بكلية التربية جامعة عين شمس من تخصصات: لغة عربية ولغة إنجليزية وأحياء وتاريخ وجغرافيا وفلسفة وفيزياء. وقد تراوحت أعمار عينة الدراسة من (١٦ – ٢٢ سنة) بمتوسط مقداره (١٨ عاما) وانحراف معياري قدره (١,٠٥ سنة). ويوضح الجدول رقم (١) الأعداد والنسب المئوية لعينة الدراسة.

٢- أدوات الدراسة:

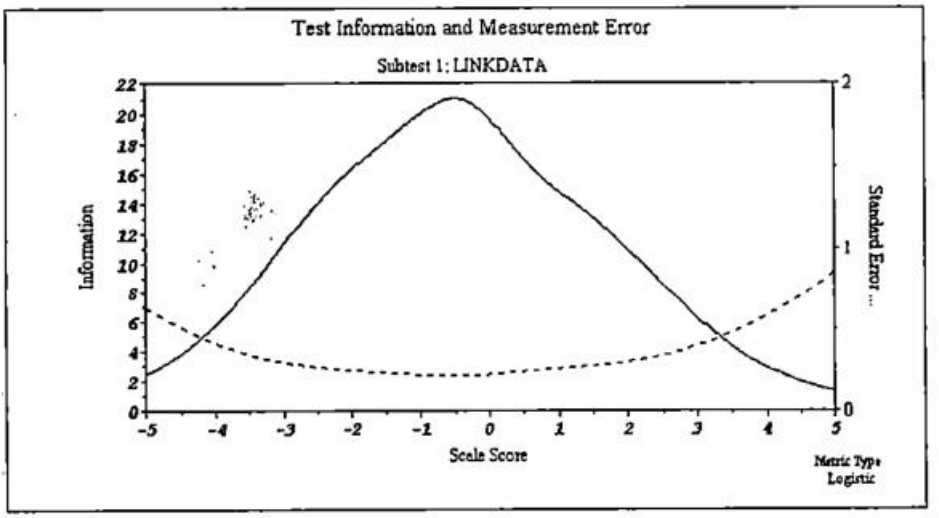
استعان الباحث -لتحقيق أهداف الدراسة-بالإصدار التواؤمي المحوسب لاختبار المصفوفات المتدرجة (أمين نور الدين، ٢٠٠٢) من خلال برنامج ***FastTEST Pro (Assessment Systems Corporation, 1999)***. والمشتق أساسا من النسخ الثلاثة الورقية لاختبار المصفوفات: المصفوفات المتدرجة الملونة ***The Coloured Progressive Matrices Test***،

والمصفوفات المتدرجة العادية *The Standard Progressive Matrices Test* ، والمصفوفات المتدرجة المتقدمة *The Advanced Progressive Matrices Test*.

الجدول رقم (١) الأعداد والنسب المئوية لعينة الدراسة.

الجنس / التخصص	طالبات	طلاب	المجموع (%)
الأدبي	١٢٠	٥٩	١٧٩ (٧٦,٢%)
العلمي	٥١	٥	٥٦ (٢٣,٨%)
المجموع (%)	١٧١ (٧٢,٨%)	٦٤ (٢٧,٢%)	٢٣٥

ويتألف الاختبار التواؤمي المحوسب للمصفوفات المتدرجة من (١٣٣ مفردة) داخل بنك المفردات *item bank* ويعتبر هذا العدد مناسبا تماما في الاختبارات التواؤمية المحوسبة في الاستجابات الثنائية كما أشار لذلك (*Stocking, 1984, p. 11; Urry, 1977, p. 184*). ويعرض شكل رقم (١) دالة معلومات الاختبار التي تظهر قيمة عالية من المعلومات التي يقدمها الاختبار خاصة عند منتصف القدرة، ويقابلها دقة عالية للاختبار، أو خطأ معياري لتقدير القدرة منخفض للغاية.



شكل رقم (١) دالة معلومات الاختبار التواؤمي المحوسب للمصفوفات المتدرجة

كما بلغت قيمة أقصى معلومات (*maximum information (15.33*) عند مستوى قدرة (١,٣٧) وهذا يعني أن بنك المفردات يمكنه أن يعطي أفضل دقة للقياس عند الأفراد ذوي مستوى القدرة المرتفع عن المتوسط بمقدار (١,٣٧) وهو يقابل خطأ معياري لتقدير القدرة ما قيمته

(٠,٢٦). وبلغ متوسط دوال المعلومات للمفردات (١٢,٩٩) ومعامل الثبات (٠,٩٥) كما يتضح من الجدول رقم (٢). ويدل اقتراب متوسط تقدير القدرة -في ذات الجدول- من الصفر (٠,٠٤٥) والانحراف المعياري من (١,٠٥٥) على اعتدالية توزيع القدرة وعدم التوائها.

الجدول رقم (٢) الخصائص السيكومترية للاختبار التواؤمي المحوسب للمصفوفات المتدرجة

الخصائص	متوسط المعلومات	الثبات	القدرة	
			م	ع
القيمة	١٢,٩٩	٠,٩٤٧	٠,٠٤٥	١,٠٥٥

توصيف أداء الاختبار التواؤمي المحوسب للمصفوفات المتدرجة:

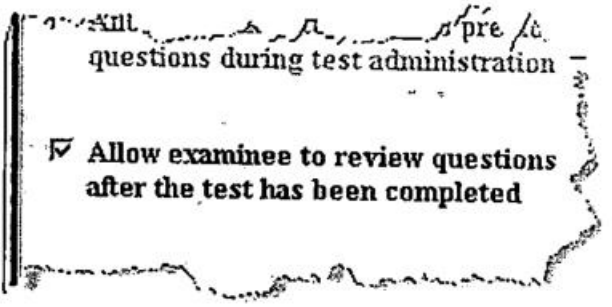
أسلوب تقدير القدرة: يتيح برنامج ***FastTEST Pro*** إجراء تقدير القدرة إما بأسلوب تقدير الأرجحية القصوى ***maximum likelihood estimation (MLE)***، أو بأسلوب البعدي الأقصى ***maximum a posteriori (MAP)***، وقد اختار الباحث أسلوب البعدي الأقصى لتقدير القدرة نظرا لتميزه -باعتباره أحد الأساليب البيزية- على أسلوب الأرجحية القصوى في أنه يعطي خطأ معياريا أقل وكفاءة أكبر من الأرجحية القصوى، كما أشارت إلى ذلك دراسات السابقة (***Wang & Vispoel, 1998***).

أسلوب انتقاء المفردة: استخدم الباحث أسلوب المعلومات القصوى ***Maximum Information Method*** لانتقاء المفردة، وهو الأسلوب الوحيد الذي يتيحه البرنامج.

محك إنهاء الاختبار: اختار الباحث محك إنهاء الاختبار التواؤمي بعدد محدد من المفردات؛ حيث ينتهي الاختبار عند وصول المفحوص إلى (٣٠ مفردة). وبتثبيت قيمة عدد المفردات التي يتلقاها كل طالب في الاختبار التواؤمي المحوسب؛ يتفاوت الخطأ المعياري لكل طالب. فقد أكد (***McBride, Wetzel & Hetter, 1997***) أن (٣٠ مفردة) في الاختبارات التواؤمية المحوسبة مناسبة تماما لتقدير قدرة المفحوصين. كما أظهرت دراسة (أمين نور الدين، ٢٠٠٢) أن (٣٠ مفردة) في الاختبار التواؤمي المحوسب للمصفوفات المتدرجة كافية تماما لتحقيق دقة في القياس حيث كانت دالة المعلومات ذات متوسط مقداره (١٨,٥٥) أي معامل ثبات يعادل (٠,٩٥) أو خطأ معياري يوازي (٠,٢٣).

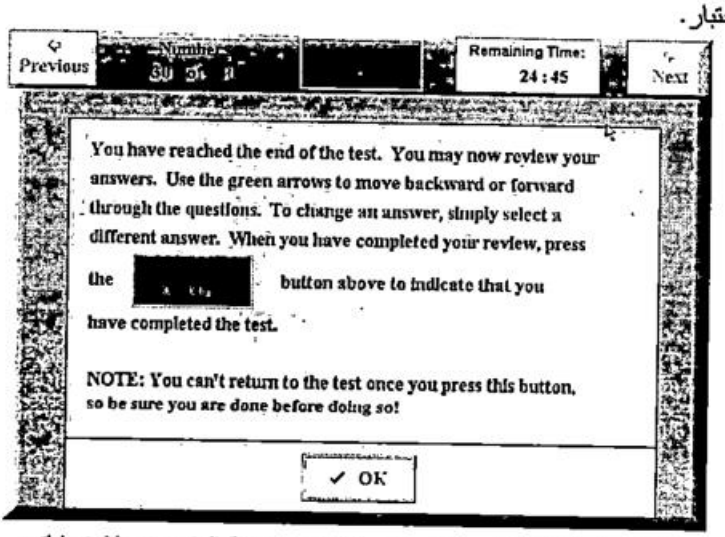
اختيار عنصر مراجعة وتعديل المفردات:

أشر الباحث في إعدادات البرنامج على السماح للمفحوص بمراجعة وتغيير إجاباته حسبما يرغب، وذلك لكي يترك للمفحوصين الفرصة لمن يريد تعديل استجاباته كما يظهر من الشكل رقم (٢).



الشكل رقم (٢) لقطة من إعدادات الاختبار للسماح للمفحوصين بالمراجعة

وتجدر الإشارة أن التعديل لا يتم إلا بعد انتهاء المفحوص من الاختبار، وليس أثناءه. فبعد الانتهاء من الاختبار تظهر رسالة للمفحوص -كما في الشكل رقم (٣)- إن كان يود القيام بالمراجعة والتعديل؛ وبناء على ذلك تظهر المفردات التي استجاب عليها مرة أخرى من المفردة الأخيرة حتى الأولى؛ فإذا قرر المفحوص الانتهاء من المراجعة يضغط على الزر المحدد وينتهي الاختبار تماما. وإن كان المفحوص لا يريد مراجعة استجاباته وتعديلها فإنه يضغط الزر المحدد فينتهي به الاختبار.

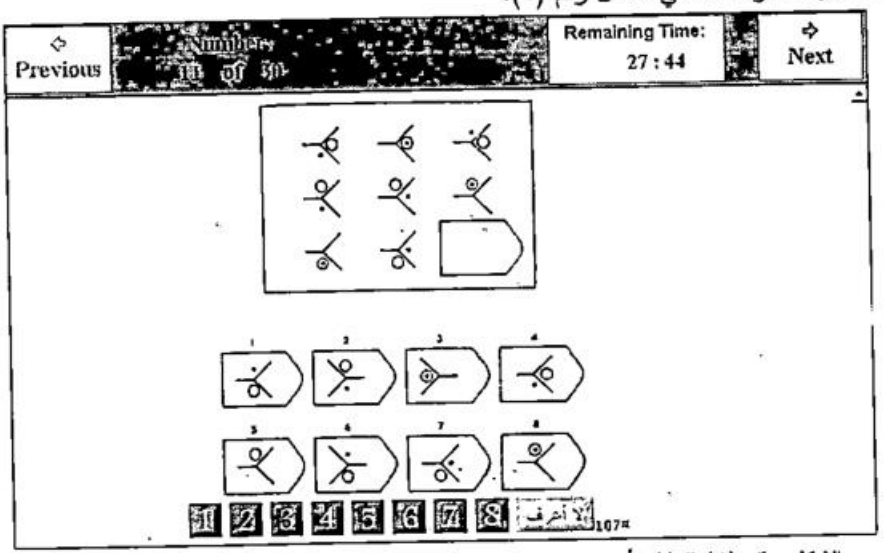


الشكل رقم (٣) لقطة من البرنامج لإنهاء مراجعة المفحوص لاستجاباته

٣- الإجراءات:

تم تطبيق الاختبار التواؤمي المحوسب للمصفوفات المتدرجة في الفصل الأول من العام الدراسي ٢٠٠٢/٢٠٠٣ داخل معمل الحاسب بقسم علم النفس التربوي بكلية التربية جامعة عين شمس من خلال أجهزة مناسبة للتطبيق. وقد بدأ تطبيق الاختبار عن طريق جلوس كل طالب أو طالبة أمام الحاسب. وقام الباحث بإلقاء التعليمات المطلوبة لتطبيق الاختبار المحوسب. وعُرضت

أمامهم أمثلة لتطبيق الاختبار، وكيفية الإجابة عليها. ولا يتطلب الإجابة على الاختبار أكثر من استخدام الفأرة (*mouse*) للإشارة إلى رقم الإجابة التي يراها الطالب صحيحة، ثم الضغط على زر *Next* فقط. فيقوم البرنامج باختيار المفردة التالية بناء على أدائه على المفردة السابقة، ثم يختار له مفردة تالية بناء على استجاباته على المفردات السابقة، وهكذا.. وذلك استنادا إلى نظرية الاستجابة للمفردة كما في الشكل رقم (٤).



الشكل رقم (٤) لقطة لأحد مفردات الاختبار التواؤمي المحوسب للمصفوفات المتدرجة

٤- التحليل الإحصائي:

قام الباحث بتفريغ (٢٣٥ ملف) -من أجهزة الحاسب التي تم التطبيق عليها- والتي تمثل أداءات جميع أفراد العينة. فعند بدء تطبيق الاختبار التواؤمي المحوسب؛ يقوم البرنامج بإنشاء ملف خاص لكل فرد، حيث يسجل جميع الخطوات التي أداها، فيحسب القدرة المقدرة له والخطأ المعياري[٥] بعد كل مفردة حتى نهاية الاختبار. وكذلك الزمن الذي استغرقه الفرد في أداء كل مفردة، وما إذا كان المفحوص قد عدّل استجاباته -بعد انتهاء الاختبار- أم أنهى الاختبار مباشرة ولم يشأ مراجعته. والوقت الذي استغرقه المفحوص في سلوك التعديل والمراجعة (إن كان قد قام بالفعل بالتعديل). ويعرض الشكل رقم (٥) لقطة للجزء النهائي من ملف أحد أفراد العينة.

---

[5] تجدر الإشارة إلى أن الخطأ المعياري في الاختبارات التواؤمية المحوسبة هو الخطأ المعياري لتقدير القدرة (Standard Error of Estimation (SEE) وليس الخطأ المعياري للقياس (Standard Error of Measurement (SEM.

```
NA CT=22:5:51.720 RT=953
SI ID=RAVEN|028 CT=22:5:52.100 RT=953
RS RN=6 AC=N CT=22:8:52.750 RT=782
EI ID=RAVEN|028 FR=6 AC=N CR=5 NR=9 SI=Y BM=-0.367 BE=0.272 CT=22:8:54.840 RT=780
NA CT=22:8:54.840 RT=780
SI ID=RAVEN|034 CT=22:8:55.220 RT=780
RS RN=6 AC=N CT=22:10:51.390 RT=670
EI ID=RAVEN|034 FR=6 AC=N CR=7 NR=9 SI=Y BM=-0.41 BE=0.271 CT=22:10:53.260 RT=669
NA CT=22:10:53.260 RT=669
ET NA=30 NC=13 BM=-0.41 BE=0.271 CT=22:10:53.310 RT=669
SR CT=22:11:1.330 RT=669
ER NA=30 NC=13 BM=-0.41 BE=0.271 CT=22:11:6.550 RT=665
```

نهاية الاختبار ..لا توجد مراجعة

الشكل رقم (٥) لقطة لجزء من نهاية ملف لأحد أفراد العينة

ونظرا لأن إجراءات الدراسة الحالية للاختبار التواؤمي المحوسب للمصفوفات المتدرجة تسمح للمفحوصين بإمكانية تعديل إجاباتهم ومراجعتها بعد انتهاء الاختبار إذا رغبوا في ذلك؛ فقد انقسمت العينة تلقائيا (٢٣٥ طالب وطالبة) أثناء إجراء الاختبار التواؤمي المحوسب إلى ثلاث مجموعات:

المجموعة الأولى هي مجموعة عدم المراجعة ***no-review group*** وهي تلك المجموعة التي فضلت بعد انتهاء الاختبار عدم تعديل إجاباتها أو حتى مراجعتها، فلم تستفد من هذه الخاصية بالرغم أنها ممنوحة لهم. من ثم؛ فإن تقديرات القدرة ***ability estimates*** والخطأ المعياري لتقدير القدرة ***standard error of estimation*** لكل منهم ظل كما هو.

المجموعة الثانية وهي مجموعة المراجعة ***reviewing group*** وهم الطلاب الذين فضلوا عقب انتهاء الاختبار أن يراجعوا إجاباتهم مفردة تلو الأخرى، ويدققوا فيها، أي استعرضوها فقط، ولم يغيروا أي شيء في إجاباتهم. وعلى ذلك؛ فإن تقدير القدرة والخطأ المعياري لتقدير القدرة لكل منهم ظل أيضا كما هو مثل المجموعة الأولى غير أنهم استغرقوا زمنا إضافيا في المراجعة.

المجموعة الثالثة وهي مجموعة التعديل ***changing group*** وهم الطلاب الذين قاموا بعد انتهاء الاختبار بمراجعة إجاباتهم بل وتعديلها أيضا كيفما يشاءون، وذلك بغرض تحسين إجاباتهم -كما يرونها- وتدارك ما سبق أن فاتهم من أخطاء. ومن ثم؛ اختلفت تقديرات القدرة وكذلك الخطأ المعياري لهذه المجموعة من نهاية الاختبار الأصلي بالنسبة لهم إلى نهاية الاختبار بعد التعديل والتصحيح الذي أجروه لاستجاباتهم، كما استغرقوا زمنا في ذلك أيضا.

وعلى ذلك؛ تم استخلاص جميع هذه المعلومات من ملفات أفراد العينة وإخضاعها للتحليل الإحصائي. علاوة على حصر لعدد الإجابات التي تم تغييرها وتحديد ما هو اتجاه التغيير هل من الصح إلى الخطأ أو من الخطأ إلى الصح أم من الخطأ إلى الخطأ. وقد تم ذلك يدويا حيث أن البرنامج لا يستخلص هذه البيانات.

ومن خلال الخطأ المعياري لتقدير القدرة؛ قام الباحث بحساب دالة المعلومات وقيمة معامل الثبات لكل فرد بناء على أن دالة المعلومات = مربع مقلوب الخطأ المعياري وأن معامل الثبات =

١- (الخطأ المعياري)[٢] (*Urry, 1977, p. 183*). وعلى ذلك ففقدان المعلومات كأحد المؤشرات يتم حسابها عن طريق الفرق بين قيمة دالة المعلومات قبل وبعد تعديل المفحوص لمفردات الاختبار، أما كفاية الاختبار فيتم حسابها عن طريق قيمة الخطأ المعياري؛ فالاختبار الأكثر كفاية هو الذي يحصل على أقل قيمة في الخطأ المعياري.

نتائج الدراسة :

أسفرت نتائج الدراسة بالنسبة للاختبار التواؤمي المحوسب للمصفوفات المتدرجة على حصول أفراد العينة إجمالا على تقديرات للقدرة بمتوسط قدره (١,٣٥) بوحدة اللوجيت وانحراف معياري مقداره (٠,٩٥). كما كان متوسط الخطأ المعياري لتقدير القدرة (٠,٢٥). ويعرض الجدول رقم (٣) الخصائص الإحصائية للاختبار على عينة الدراسة الحالية.

الجدول رقم (٣) الخصائص الإحصائية لأداء عينة الدراسة على الاختبار

الإحصاءات / الخصائص	م	ع	المدى	
			من	إلى
تقدير القدرة	١,٣٥	٠,٩٥	-١,٧٠	٣,٦٧
الخطأ المعياري	٠,٢٥	٠,٠٣٣	٠,٢٠١	٠,٤١٦
معامل الثبات	٠,٩٤	٠,٠١٨	٠,٩٦	٠,٨٣

ولم يختلف متوسطا تقديرات القدرة للبنين عن البنات اختلافا دالا في عينة الدراسة الحالية حيث كان متوسط تقدير القدرة للبنين هو (١,٣٢) بانحراف معياري مقداره (١,٠٢) بينما متوسط تقدير القدرة للبنات هو (١,٣٦) بانحراف معياري مقداره (٠,٩٣). ويوضح الجدول رقم (٤) دلالة الفرق بين المتوسطين باستخدام اختبار "ت" للعينيتين المستقلتين.

الجدول رقم (٤) اختبار "ت" لدلالة الفرق بين متوسطي تقديرات القدرة للجنسين

الإحصاءات / الجنس	ن	م	ع	قيمة ت	د. ح.	الدلالة
الطلاب	٦٤	١,٣٢	١,٠٢	-٠,٣١٦	٢٣٣	غير دالة
الطالبات	١٧١	١,٣٦	٠,٩٣			

ينص السؤال الأول على:

كم نسبة أفراد العينة الذين استفادوا من خاصية مراجعة وتعديل الإجابات في الاختبارات التواؤمية المحوسبة؟

بينت نتائج الدراسة أن مجموعة عدم المراجعة وهي تلك المجموعة التي فضلت بعد انتهاء

الاختبار عدم تعديل إجاباتها أو حتى مراجعته، كان عددها (٩٥) طالبا وطالبة بنسبة (٤٠,٤%) من إجمالي العينة. منهم (٢٥%) من الطلاب و(٧٥%) من الطالبات.

أما مجموعة المراجعة وهم الطلاب الذين فضلوا عقب انتهاء الاختبار أن يراجعوا إجاباتهم دون أن يغيروا أي منها؛ فقد كان عددها (٤٣) طالبا وطالبة بنسبة (١٨,٣%)، من هؤلاء؛ (٤٤%) من الطلاب و(٥٦%) من الطالبات. أما مجموعة التعديل وهم الطلاب الذين قاموا بعد انتهاء الاختبار بتغيير إجاباتهم فقد تألفت من (٩٧) طالبا وطالبة بنسبة (٤١,٣%)، منهم (٢٢%) من الطلاب و(٧٨%) من الإناث. ويعرض الجدول رقم (٥) الأعداد ونسب الطلاب للمجموعات الثلاثة.

الجدول رقم (٥) أعداد ونسب الطلاب للمجموعات الثلاثة حسب سلوك المراجعة والتعديل

المجموعة / الجنس	التعديل	المراجعة	عدم المراجعة	الإجمالي
الطلاب	٢١ (٣٢,٨%)	١٩ (٢٩,٧%)	٢٤(٣٧,٥%)	٦٤ (١٠٠%)
الطالبات	٧٦ (٤٤,٤%)	٢٤ (١٤%)	٧١ (٤١,٥%)	١٧١ (١٠٠%)
الإجمالي	٩٧ (٤١,٣%)	٤٣ (١٨,٣%)	٩٥ (٤٠,٤%)	٢٣٥ (١٠٠%)

يتضح مما سبق أن (١٤٠طالب وطالبة) أي أكثر من نصف حجم عينة الدراسة (٦٠%) قد استفاد من خاصية التعديل سواء عدلوا استجاباتهم أم لا. وأن نسبة عدد الطلاب الذين غيروا إجاباتهم بالفعل بلغ (٤١,٣%) من إجمالي حجم العينة.

ينص السؤال الثاني على:

هل يوجد فرق دال إحصائيا بين متوسطات درجات الأفراد الذين راجعوا والذين عدّلوا إجاباتهم والذين لم يعدلوا أو يراجعوا إجاباتهم مطلقا؟

للإجابة على هذا السؤال؛ قام الباحث بمقارنة كل من تقديرات القدرة وعدد الإجابات الصحيحة لمجموعة التعديل مع مجموعتي المراجعة وعدم التعديل. ففيما يتعلق بتقديرات القدرة؛ أظهرت النتائج أن مجموعة التعديل حصلت على تقديرات للقدرة بمتوسط قدره (١,٥٣) باللوجيت، وانحراف معياري قدره (٠,٨٢). بينما مجموعة المراجعة أحرزت تقديرات للقدرة بمتوسط (١,٢٨) بانحراف معياري مقداره (٠,٨٤). أما مجموعة عدم المراجعة فقد حققت تقديرات للقدرة ذات متوسط (١,٢٠) وانحراف معياري قدره (١,١٠). ويبين الجدول رقم (٦) الإحصاءات الوصفية لتقديرات القدرة للمجموعات الثلاثة. وبحساب دلالة الفروق بين متوسطات تقديرات القدرة تبين أن هناك فرق دال بين المتوسطات الثلاثة كما يوضح ذلك الجدول رقم (٧)

الجدول رقم (٦) الإحصاءات الوصفية لتقديرات القدرة للمجموعات الثلاثة

الإحصاءات / المجموعة	ن	م	ع
عدم المراجعة	٩٥	١,٢٠	١,١٠
المراجعة	٤٣	١,٢٨	٠,٨٤
التعديل	٩٧	١,٥٣	٠,٨٢
الإجمالي	٢٣٥	١,٣٥	٠,٩٥

الجدول رقم (٧) تحليل التباين الأحادي لمتوسطات تقديرات القدرة بين المجموعات الثلاثة

مصدر التباين	مجموع المربعات	درجات الحرية	متوسط المربعات	ف	الدلالة
بين المجموعات	٥,٣١٥	٢	٢,٦٥٧	٢,٩٦٩	٠,٠٥
داخل المجموعات	٢٠٧,٦٦٠	٢٣٢	٠,٨٩٥		
الكلي	٢١٢,٩٧٥	٢٣٤			

وبإجراء التحليل البعدي ***Post Hoc*** باختبار ***Tukey*** لحساب أي من المتوسطات أعلى بدلالة من المتوسطات الأخرى؛ تبين أن متوسط تقدير القدرة لمجموعة التعديل كان أعلى بدلالة من متوسط مجموعة عدم المراجعة، ولم يكن أعلى بدلالة من مجموعة المراجعة، كما لم يكن هناك فرق دال بين مجموعتي المراجعة وعدم المراجعة. ويعرض الجدول رقم (٨) المقارنات الثنائية لفروق المتوسطات.

الجدول رقم (٨) المقارنات الثنائية لفروق متوسطات تقديرات القدرة بين المجموعات الثلاثة

	التعديل	المراجعة	عدم المراجعة
التعديل		٠,٢٥	*٠,٣٢
المراجعة			٠,٠٩
عدم المراجعة			

كما قام الباحث كذلك بمقارنة متوسطات عدد الإجابات الصحيحة لمجموعة التعديل (بعد تعديل إجاباتهم)، مع المجموعتين الأخريين؛ فأظهرت النتائج أن هناك فرقا دالا بين متوسطات المجموعات الثلاثة. ويعرض الجدول رقم (٩) الإحصاءات الوصفية لعدد الإجابات الصحيحة، كما

يوضح الجدول رقم (١٠) دلالة الفروق بين متوسطات عدد الإجابات الصحيحة للمجموعات الثلاثة.

الجدول رقم (٩) الإحصاءات الوصفية لعدد الإجابات الصحيحة للمجموعات الثلاثة

الإحصاءات / المجموعة	ن	م	ع
عدم المراجعة	٩٥	٢٠,٤٣	٢,٧٤
المراجعة	٤٣	٢٠,٦٥	٢,١٤
التعديل	٩٧	٢٣,٦١	٢,٨٩
الإجمالي	٢٣٥	٢١,٧٨	٣,١٠

الجدول رقم (١٠) تحليل التباين الأحادي بين متوسطات عدد الإجابات الصحيحة للمجموعات الثلاثة

مصدر التباين	مجموع المربعات	درجات الحرية	متوسط المربعات	ف	الدلالة
بين المجموعات	٥٥١,٦٤٦	٢	٢٧٥,٨٧	٣٧,٧٣	٠,٠٠٠١
داخل المجموعات	١٦٩٦,١٩	٢٣٢	٧,٣١		
الكلي	٢٢٤٧,٩٣	٢٣٤			

ولحساب أي من المتوسطات أعلى بدلالة من المتوسطات الأخرى؛ أُجري التحليل البعدي ***Post Hoc*** باختبار ***Tukey*** الذي يشير إلى أن متوسط عدد الإجابات الصحيحة لمجموعة التعديل كان أعلى بدلالة من متوسطي مجموعة المراجعة وعدم المراجعة، ولم يكن هناك فرق دال بين مجموعة المراجعة ومجموعة عدم المراجعة. ويعرض الجدول رقم (١١) المقارنات الثنائية لفروق المتوسطات.

الجدول رقم (١١) المقارنات الثنائية لفروق المتوسطات بين المجموعات الثلاثة

	التعديل	المراجعة	عدم المراجعة
التعديل		٢,٩٦*	٣,١٨*
المراجعة			٠,٢٢
عدم المراجعة			

وللتأكد -مرة أخرى- من أن التحسن الذي طرأ على درجات مجموعة التعديل كان بفعل سلوك المراجعة وتعديل الإجابات؛ قام الباحث بمقارنة درجاتهم (تقديرات القدرة وعدد الإجابات

الصحيحة) قبل إجراء المراجعة وتغيير إجاباتهم[٦] مع المجموعتين الأخريين بدرجاتهم الأصلية، ويبين الجدول رقم (١٢) الإحصاءات الوصفية لتقدير القدرة وعدد الإجابات الصحيحة للمجموعات الثلاثة. وبإجراء تحليل التباين الأحادي بين المتوسطات الثلاثة لتقدير القدرة تبين أنه لا توجد فروق دالة، كذلك لا توجد فروق دالة بين المجموعات الثلاثة في متوسطات عدد الإجابات الصحيحة كما يوضح ذلك الجدولان رقما (١٣)، (١٤)

الجدول رقم (١٢) الإحصاءات الوصفية للمجموعات الثلاثة قبل (سلوك التعديل)

الإحصاءات / المجموعة	ن	تقدير القدرة		الإجابات الصحيحة	
		م	ع	م	ع
عدم المراجعة	٩٥	١,٢٠	١,١٠	٢٠,٤٣	٢,٧٤
المراجعة	٤٣	١,٢٨	٠,٨٤	٢٠,٦٥	٢,١٤
التعديل	٩٧	١,٢١	٠,٨٩	٢١,٠٤	٢,٣٩
الإجمالي	٢٣٥	١,٢٢	٠,٩٧	٢٠,٧٢	٢,٥٠

الجدول رقم (١٣) تحليل التباين الأحادي لمتوسطات تقدير القدرة بين المجموعات الثلاثة

مصدر التباين	مجموع المربعات	درجات الحرية	متوسط المربعات	ف	الدلالة
بين المجموعات	٠,١٩	٢	٠,٠٩٤	٠,١٠	غير دالة
داخل المجموعات	٢٢٠,٠٨	٢٣٢	٠,٩٥		
الكلي	٢٢٠,٢٧	٢٣٤			

الجدول رقم (١٤) تحليل التباين الأحادي لمتوسطات عدد الإجابات الصحيحة بين المجموعات الثلاثة

مصدر التباين	مجموع المربعات	درجات الحرية	متوسط المربعات	ف	الدلالة
بين المجموعات	١٨,١١	٢	٩,٠٦	١,٤٥	غير دالة
داخل المجموعات	١٤٤٤,٩١	٢٣٢	٩,٢٣		
الكلي	١٤٦٣,٠٢	٢٣٤			

[٦] تجدر الإشارة إلى أن البرنامج الذي يطبق على المفحوصين يحتفظ بدرجاتهم الأصلية، علاوة على درجاتهم أيضا بعد المراجعة والتعديل.

ينص السؤال الثالث على:

إلى أي مدى يحسّن سلوك مراجعة وتعديل الإجابات من درجات المفحوصين في الاختبار؟

أظهرت النتائج أن المفحوصين الذين قاموا بمراجعة وتعديل إجاباتهم (مجموعة التعديل) قد زادت تقديرات قدرتهم عمًا قبل المراجعة؛ وأن هذه الزيادة كانت دالة عند مستوى دلالة أقل من (٠,٠٠١). حيث كان متوسط تقديرات القدرة قبل التعديل (١,٢١) باللوجيت، بينما كان بعد التعديل (١,٥٣).

أي أن المفحوصين الذين قاموا بتغيير إجاباتهم قد تحسن تقديرات القدرة لديهم بصورة كبيرة. وأن متوسط التحسن الذي طرأ كان بمقدار (٠,٣٢)، وأن (٩٢%) من مجموعة التعديل تحسنت درجاتهم بسب التعديل وبينما (٨%) فقط من نفس المجموعة هم الذين انخفضت درجاتهم بسبب التعديل.

كما أسفرت النتائج أيضا أنّ هؤلاء المفحوصين قد أحرزوا عددا من الإجابات الصحيحة بعد التعديل بمتوسط (٢٣,٦١) أعلى بدلالة من متوسط عدد الإجابات الصحيحة قبل التعديل (٢١,٠٤). كما كانت العلاقة الارتباطية بين تقديرات القدرة قبل التعديل وبعدها بلغت (٠,٩٤) بمستوى دلالة أقل من (٠,٠٠١).

ويعرض الجدول رقم (١٥) اختبار "ت" للعينتين المرتبطتين لدلالة الفرق بين متوسطي تقديرات القدرة قبل وبعد التعديل. كما يعرض الجدول رقم (١٦) اختبار "ت" للعينتين المرتبطتين لدلالة الفرق بين متوسطي عدد الإجابات الصحيحة.

الجدول رقم (١٥) اختبار "ت" لدلالة الفرق بين متوسطي تقدير القدرة قبل وبعد التعديل

المتغير / الإحصاءات	ن	م	ع	قيمة ت	د. ح.	الدلالة
قبل المراجعة	٩٧	١,٢١	٠,٨٩	-١٠,٣٩	٩٦	٠,٠٠١
بعد المراجعة		١,٥٣	٠,٨٢			

جدول (١٦) اختبار "ت" لدلالة الفرق بين متوسطي عدد الإجابات الصحيحة قبل وبعد التعديل

المتغير / الإحصاءات	ن	م	ع	قيمة ت	د. ح.	الدلالة
قبل المراجعة	٩٧	٢١,٠٤	٢,٣٩	-١١,٤٥	٩٦	٠,٠٠١
بعد المراجعة		٢٣,٦١	٢,٨٩			

ينص السؤال الرابع على:

إلى أي مدى يُخفض تعديل المفحوص لاستجاباته من كفاية الاختبار التواؤمي المحوسب؟ وإلى أي مدى تتأثر دالة معلومات الاختبار بتغيير الإجابات؟

للإجابة على هذا السؤال؛ قام الباحث باستخلاص قيم الخطأ المعياري لكل فرد في مجموعة التعديل (٩٧ طالب وطالبة) من ملفات عينة الدراسة مباشرة قبل إجراء التعديل وبعد الانتهاء من التعديل. فكشفت النتائج أن المفحوصين بعد انتهاء الاختبار الأصلي قد حصلوا على خطأ معياري لتقدير القدرة بمتوسط (٠,٢٣٨٥) وبعد أن قام المفحوصون بمراجعة إجاباتهم وتغييرها؛ حصلوا على خطأ معياري بمتوسط (٠,٢٥٥١). وبحساب نسبة الفقدان في كفاية الاختبار؛ تبين أن متوسط نسبة الكفاية كان (٠,٩٥)، أي أن كفاية الاختبار انخفضت بسبب التعديل بنسبة (٥%).

وللإجابة على الشق الثاني من السؤال والمتعلق بمقدار المعلومات المفقود؛ قام الباحث بحساب قيمة دالة المعلومات عن طريق مقلوب تباين الخطأ المعياري لتقدير القدرة. فأسفرت النتائج أن متوسط مقدار المعلومات في الاختبار الأصلي قبل التعديل كان (١٧,٩٨) وبعد التعديل أصبح (١٦,٣٠). أي أن فقدان المعلومات كان بمقدار (١,٦٨). ويعرض الجدول رقم (١٧) الإحصاءات الوصفية لتقديرات القدرة وكفاية الاختبار ومقدار المعلومات لمجموعة التعديل.

الجدول رقم (١٧) الإحصاءات الوصفية لمجموعة التعديل

الإحصاءات / المتغير	م	ع	المدى	
			من	إلى
تقدير القدرة (باللوجيت):				
قبل المراجعة	١,٢١	٠,٨٩	-٠,٨١٦	٣,٣٩
بعد المراجعة	١,٥٣	٠,٨٢	-٠,٤١	٣,٦٧
الفرق بين(بعد - قبل) المراجعة	٠,٣٢	٠,٣٠		
نسبة كفاية الاختبار:				
الخطأ المعياري (القبلي)	٠,٢٤	٠,٠٢	٠,١٩٩	٠,٢٨٨
الخطأ المعياري (البعدي)	٠,٢٦	٠,٠٤	٠,٢٠١	٠,٤١٦
المعلومات:				
المعلومات قبل التعديل	١٨	٣,١١	١٢,٠٦	٢٥,٢٥
المعلومات بعد التعديل	١٦,٣٠	٤,١٢	٥,٧٧٩	٢٤,٧٥٢
الفرق	-١,٦٨	٢,٨٤		

ينص السؤال الخامس على:

إلى أي مدى يستغرق سلوك المراجعة والتعديل زمنا إضافيا على الوقت الأصلي للاختبار؟

بينت نتائج الدراسة أن إجراء الاختبار الأصلي بالنسبة لجميع أفراد العينة استغرق زمنا بمتوسط قدره (٢٨ دقيقة و٣٣ ثانية). حيث استغرقت مجموعة عدم المراجعة زمنا بمتوسط قدره (٢٩ دقيقة و١٢ ثانية)، بينما استغرقت مجموعة المراجعة زمنا بمتوسط قدره (٢٧ دقيقة و٤٠ ثانية)، أما مجموعة التعديل فقد استغرقت زمنا بمتوسط مقداره (٢٨ دقيقة و١٨ ثانية). ويعرض الجدول رقم (١٨) الإحصاءات الوصفية لمقدار الزمن الذي استغرقته المجموعات الثلاثة.

الجدول رقم (١٨) الإحصاءات الوصفية لمقدار الزمن المستغرق للمجموعات الثلاثة

الإحصاءات / المجموعة	ن	زمن الاختبار الأصلي[٧]	
		م	ع
عدم المراجعة	٩٥	٢٩:١٢	١٠:٠٢
المراجعة	٤٣	٢٧:٤٠	٠٩:١٨
التعديل	٩٧	٢٨:١٨	٠٨:٤٦
إجمالي العينة	٢٣٥	٢٨:٣٣	٠٩:٢٢

وبحساب دلالة الفروق بين المتوسطات؛ تبين أنه لا توجد فروق دالة بين المجموعات الثلاثة في متوسطات الزمن الذي استغرقته في الاختبار الأصلي (قبل سلوك المراجعة لمجموعة التعديل) كما يوضح ذلك الجدول رقم (١٩)

الجدول رقم (١٩) تحليل التباين الأحادي بين متوسطات الزمن للمجموعات الثلاثة

مصدر التباين	مجموع المربعات	درجات الحرية	متوسط المربعات	ف	الدلالة
بين المجموعات	٢٨٦٨٦١,٨٤٩	٢	١٤٣٤٣٠,٩٢	٠,٤٥١	غير دالة
داخل المجموعات	٧٣٨٥٥٢١٧,٧١	٢٣٢	٣١٨٣٤١,٤٦		
الكلي	٧٤١٤٢٠٧٩,٥٥	٢٣٤			

[٧] يشير الرقمان على اليسار إلى الزمن بالدقائق، بينما يشير الرقمان على اليمين إلى الزمن بالثواني.

وأسفرت النتائج أيضا عن أن إجراء المراجعة والتعديل قد استغرق زمنا إضافيا بمتوسط مقداره (٧ دقائق و٥٥ ثانية) أي بنسبة زيادة متوسطها (٣٣%) فوق زمن الاختبار الأصلي، تراوح مدته من دقيقة واحدة إلى خمس وعشرين دقيقة.

ينص السؤال السادس على:

هل يختلف سلوك مراجعة وتعديل الإجابات في الاختبار التواؤمي المحوسب باختلاف الجنس؟

أسفرت نتائج الدراسة عن أن الذين رغبوا في تعديل إجاباتهم أو حتى فقط مراجعتها كانوا بنسبة (٦٢,٥%) من الذكور؛ بينما كن بنسبة (٥٨,٥%) من الإناث. بينما الذين لم يفضلوا مراجعة إجاباتهم كانوا بنسبة (٣٧,٥%) من الذكور؛ و(٤١,٥%) من الإناث. ويوضح الجدول رقم (٢٠) أعداد ونسب الطلاب والطالبات وفقا لسلوك المراجعة والتعديل. وبحساب دلالة الفروق بين الذكور والإناث؛ كشفت نتائج الدراسة عن أن قيمة كا^2 = (٠,٣١٣) عند درجات حرية = (١). وهي غير دالة أي أنه لم تكن هناك فروق دالة بين الطلاب والطالبات في سلوك المراجعة والتعديل، أو أن سلوك المراجعة والتعديل لم يتأثر بالجنس.

الجدول رقم (٢٠) أعداد ونسب الذكور والإناث وفقا لسلوك المراجعة والتعديل

الجنس / المجموعة	الذكور	الإناث	المجموع
مجموعة عدم المراجعة	٢٤ (٣٧,٥%)	٧١ (٤١,٥%)	٩٥ (٤٠,٤%)
مجموعتا المراجعة والتعديل معا	٤٠ (٦٢,٥%)	١٠٠ (٥٨,٥%)	١٤٠ (٥٩,٦%)
المجموع	٦٤ (١٠٠%)	١٧١ (١٠٠%)	٢٣٥ (١٠٠%)

ينص السؤال السابع على:

هل يختلف سلوك مراجعة وتعديل الإجابات في الاختبار التواؤمي المحوسب باختلاف مستويات القدرة؟

للإجابة على هذا السؤال؛ تم تقسيم تقديرات القدرة إلى ثلاثة مستويات: المنخفض من (٠,٨٣٦) فأقل، والمرتفع من (١,٦٧٦) فأعلى، ومابين ذلك المستوى المتوسط، حيث كانت قيمة الوسيط (١,٣١٦). ومن خلال جدول الاقتران (٢×٣) كما يظهره الجدول رقم (٢١) تبين أن الطلاب الذين رغبوا في إجراء التعديل أو حتى المراجعة كانوا بنسبة (٣٥%) في المستوى المنخفض من القدرة، و(٣٠,٧%) من المستوى المتوسط، و(٣٤,٣%) من المستوى المرتفع من القدرة. وبالتحقق من دلالة الفروق بين النسب للمستويات الثلاثة وسلوك التعديل؛ تبين أن قيمة كا^2

= (١,٣٣٤) عند درجات حرية (٢) وكانت غير دالة أيضا. أي أن سلوك المراجعة والتعديل لم يتأثر بمستوى محدد من القدرة.

الجدول رقم (٢١) أعداد ونسب الطلاب وفقا لمستويات القدرة وسلوك المراجعة والتعديل

المستوى / المجموعة	المنخفض	المتوسط	المرتفع	المجموع
مجموعة عدم المراجعة	٢٩ (٣٠,٥%)	٣٦ (٣٧,٩%)	٣٠ (٣١,٦%)	٩٥ (١٠٠%)
مجموعتا المراجعة والتعديل معا	٤٩ (٣٥%)	٤٣ (٣٠,٧%)	٤٨ (٣٤,٣%)	١٤٠ (١٠٠%)
المجموع	٧٨ (٣٣,٢%)	٧٩ (٣٣,٦%)	٧٨ (٣٣,٢%)	٢٣٥ (١٠٠%)

ينص السؤال الثامن على:

من الأكثر استفادة من خاصية مراجعة وتعديل الإجابات في الاختبار التواؤمي المحوسب هل هو منخفض القدرة أم متوسط القدرة أم مرتفع القدرة؟

للإجابة على هذا السؤال؛ تم تقسيم مستويات القدرة القبلية (أي قبل سلوك التعديل) وذلك لمجموعة التعديل إلى ثلاثة مستويات، المستوى المنخفض من (٠,٦٦) فأقل، والمستوى المرتفع من (١,٦٦) فأعلى؛ ومابين ذلك المستوى المتوسط، حيث كانت قيمة الوسيط (١,٢٥). كما تم تقسيم مستويات القدرة البعدية (أي بعد سلوك التعديل) لنفس المجموعة إلى ثلاثة مستويات، المستوى المنخفض من (١,٠٧) فأقل، والمستوى المرتفع من (١,٩٣) فأعلى، ومابين ذلك المستوى المتوسط، حيث كانت قيمة الوسيط (١,٥٧).

ومن خلال جدول الاقتران (٣×٣) كما يعرضه الجدول رقم (٢٢) الذي يعرض لأعداد ونسب الطلاب وفقا لمستويات القدرة قبل وبعد التعديل؛ يتضح أن (٦ طلاب) انتقلوا من المستوى المنخفض في القدرة القبلية إلى مستوى المتوسط في القدرة البعدية بعد سلوك التعديل بنسبة (١٨,٨%). بينما هناك في المقابل (٤ طلاب) بنسبة (١٢,٥%) هم الذين انخفض مستواهم من المرتفع إلى المتوسط. بينما في المستوى المتوسط؛ فإن هناك حوالي (١٨%) من الطلاب انخفض مستواهم بعد سلوك التعديل وحوالي (١٢%) ارتفع مستواهم. أما المستوى المرتفع من القدرة فحوالي (٨٦%) من الطلاب في ذات المستوى حافظ على مستواه بعد التعديل. وقد أظهرت نتائج الدراسة أنه كانت هناك فروق دالة عند مستوى دلالة أقل من (٠,٠٠١) بين النسب المختلفة حيث كانت قيمة كا^2 = (٩٧,٩٧) عند درجات حرية (٤). أي أن الذي استفاد أكثر من سلوك التعديل هو المستوى المرتفع من القدرة عن المستوى المنخفض والمتوسط.

الجدول رقم (٢٢) أعداد ونسب الطلاب وفقا لمستويات القدرة ولسلوك المراجعة والتعديل

البعدي / القبلي	المنخفض	المتوسط	المرتفع	المجموع
المنخفض	٢٦ (٨١,٣%)	٦ (١٨,٨%)	------	٣٢ (١٠٠%)
المتوسط	٦ (١٨,٢%)	٢٣ (٦٩,٧%)	٤ (١٢,١%)	٣٣ (١٠٠%)
المرتفع	------	٤ (١٢,٥%)	٢٨ (٨٧,٥%)	٣٢ (١٠٠%)
المجموع	٣٢ (٣٣%)	٣٣ (٣٤%)	٣٢ (٣٣%)	٩٧ (١٠٠%)

ينص السؤال التاسع على:

ما هو اتجاه تغيير الإجابات الأكثر تكرارا: هل هو من الصح إلى الخطأ، أم من الخطأ إلى الصح، أم من الخطأ إلى الخطأ؟

للإجابة على هذا السؤال؛ قام الباحث بتفريغ بيانات مجموعة التغيير من عينة الدراسة يدويا لحصر عدد واتجاه ونسبة تغيير الإجابات، حيث أسفرت نتائج الدراسة أن عدد المفردات التي تغيرت كانت (٣٨٥) مفردة من واقع (٢٩١٠ مفردة) بنسبة (١٣%) من إجمالي عدد المفردات التي تم تطبيقها (أي ٣٠ مفردة × ٩٧ مفحوصا)؛ وبمعدل (٤ تغييرات) تقريبا لكل مفحوص، حيث تراوحت عدد التغييرات لكل مفحوص من تغيير واحد إلى ١٣ تغييرا.

وقد كشفت نتائج الدراسة أن معظم اتجاه التغيير كان من الخطأ للصح بنسبة (٧١%)، وأن نسبة اتجاه التغيير من الصح للخطأ كانت (٧%)، بينما اتجاه التغيير من الخطأ للخطأ كان بنسبة (٢٢%).

وفيما يتعلق بتأثير الجنس على عدد التغييرات واتجاه تغيير الإجابات؛ فقد أظهرت نتائج الدراسة أيضا أنه بالرغم أن متوسط عدد التغييرات التي أجرينها الطالبات (٤,٢١ تغييرا) كان أعلى مما أجراها الطلاب (٣,١٠ تغييرا)؛ إلا أن الفرق بين المتوسطين لم يكن ذا دلالة كما يدل على ذلك الجدول رقم (٢٣). أي أنه لا توجد فروق دالة بين الجنسين في متوسط عدد التغييرات.

الجدول رقم (٢٣) اختبار "ت" لدلالة الفرق بين متوسطي عدد التغييرات للذكور والإناث

الجنس / الإحصاءات	ن	م	ع	قيمة ت	د. ح.	الدلالة
الطلاب	٢١	٣,١٠	٢,٣٩	-١,٧٨	٩٥	غير دالة
الطالبات	٧٦	٤,٢١	٢,٥٨			

أما فيما يتعلق باتجاه تغيير الإجابات؛ أشارت نتائج الدراسة أن الطالبات قد غيرن من الخطأ للصح وكذلك من الصح إلى الخطأ بنسبة أعلى من الطلاب، بينما أحرز الطلاب نسبة أعلى من الطالبات في اتجاه التغيير من الخطأ للخطأ، غير أن هذه النسب لم تكن دالة حيث كانت قيمة كا^2= (١,٠٣) عند درجات حرية (٢). ويعرض الجدول رقم (٢٤) أعداد ونسب اتجاهات التغيير وفقا للجنس.

الجدول رقم (٢٤) أعداد ونسب اتجاهات التغيير وفقا للجنس

اتجاه التغيير / الجنس	من الخطأ للصح	من الصح للخطأ	من الخطأ للخطأ	عدد التغييرات
الطلاب	٤٤ (٦٨%)	٤ (٦%)	١٧ (٢٦%)	٦٥ (١٠٠%)
الطالبات	٢٣٠ (٧٢%)	٢٤ (٨%)	٦٦ (٢١%)	٣٢٠ (١٠٠%)
الإجمالي	٢٧٤ (٧١%)	٢٨ (٧%)	٨٣ (٢٢%)	٣٨٥ (١٠٠%)

وفيما يتعلق بتأثير مستويات القدرة على اتجاه التغيير؛ فقد أسفرت نتائج الدراسة أن المستوى المنخفض من القدرة قد أجرى تغييرات بمعدل (٥ تغييرات) لكل طالب، بينما المستوى المتوسط قد أجرى تغييرات بمعدل (٤ تغييرات) لكل طالب، والمستوى المرتفع أجرى تغييرات بمعدل (٣ تغييرات) لكل طالب. كما أظهرت نتائج الدراسة أيضا أن مرتفعي القدرة كانوا النسبة الأعلى فيما يتعلق باتجاه التغيير من الخطأ للصح، يليهم المتوسطين ثم المنخفضين. ونفس النتيجة تقريبا بالنسبة لاتجاه التغيير من الصح للخطأ، أما فيما يتعلق باتجاه التغيير من الخطأ للخطأ؛ فإن النتيجة عكسية حيث كان يمثل المنخفضين أعلى نسبة من التغييرات، يليهم المتوسطين، ثم المرتفعين. ويعرض الجدول رقم (٢٥) عدد ونسب اتجاهات تغيير الإجابات وفقا لمستويات القدرة. غير أنه أشارت نتائج الدراسة أيضا أنه لم تكن هذه النسب ذات فروق دالة بينها حيث كانت قيمة كا^2 = (٥,٢٣) عند درجات حرية (٤).

الجدول رقم (٢٥) الأعداد ونسب اتجاهات التغيير وفقا لمستويات القدرة

اتجاه التغيير / مستوى القدرة	من الخطأ للصح	من الصح للخطأ	من الخطأ للخطأ	عدد التغييرات
المنخفض	١٠٠ (٦٨%)	٨ (٥%)	٤٠ (٢٧%)	١٤٨ (١٠٠%)
المتوسط	١٠٥ (٧٣%)	١٢ (٨%)	٢٨ (١٩%)	١٤٥ (١٠٠%)
المرتفع	٦٩ (٧٥%)	٨ (٩%)	١٥ (١٦%)	٩٢ (١٠٠%)
الإجمالي	٢٧٤ (٧١%)	٢٨ (٧%)	٨٣ (٢٢%)	٣٨٥ (١٠٠%)

كما أظهرت نتائج الدراسة أيضا أن ثمة علاقة ارتباطية سالبة بدلالة بين تقديرات القدرة وعدد تغييرات الإجابات التي تمت (-٠,٢٩*)، أي أنه كلما زادت قدرة المفحوص انخفضت التغييرات التي يجريها على إجاباته. وتأكيدا لتلك النتيجة؛ تمت مقارنة متوسط عدد التغييرات التي تمت وفقا لمستويات القدرة الثلاثة السابقة، وبينت النتائج أن هناك فرقا دالا بين متوسطات عدد التغييرات التي أجريت كما يوضح ذلك الجدولان رقم (٢٦) و(٢٧)

الجدول رقم (٢٦) الإحصاءات الوصفية لعدد التغييرات وفقا لمستويات القدرة

الإحصاءات / المستوى	ن	م	ع
المنخفض	٣٢	٤,٦٣	٢,٩٣
المتوسط	٣٣	٤,٣٩	٢,٦٠
المرتفع	٣٢	٢,٨٨	١,٧٤
الإجمالي	٩٧	٣,٩٧	٢,٥٧

الجدول رقم (٢٧) تحليل التباين الأحادي لمتوسطات عدد التغييرات لمستويات القدرة

مصدر التباين	مجموع المربعات	درجات الحرية	متوسط المربعات	ف	الدلالة
بين المجموعات	٥٨,٠٣	٢	٢٩,٠١	٤,٧٤	٠,٠١
داخل المجموعات	٥٧٤,٨٨	٩٤	٦,١٢		
الكلي	٦٣٢,٩١	٩٦			

وبإجراء التحليل البعدي ***Post Hoc*** باختبار ***Tukey*** لحساب أي من المتوسطات أقل بدلالة من المتوسطات الأخرى؛ تبين أن متوسط عدد التغييرات للمستوى المرتفع من القدرة كان أقل بدلالة من متوسطي المستوى المنخفض والمتوسط من القدرة، ولم يكن هناك فرق دال بين المستويين الأخيرين من القدرة. أي أن المستوى المرتفع من القدرة أجرى عددا من التغييرات أقل بدلالة مما أجراها المستويين المتوسط والمنخفض. ويعرض الجدول رقم (٢٨) المقارنات الثنائية لفروق المتوسطات.

الجدول رقم (٢٨) المقارنات الثنائية لفروق المتوسطات لعدد التغييرات بين مستويات القدرة

	منخفض	متوسط	مرتفع
منخفض		٠,٢٣	*١,٧٥
متوسط			*١,٥٢
مرتفع			

مناقشة النتائج وتفسيرها

لقد تحولت العديد من الاختبارات الورقية إلى اختبارات تواؤمية محوسبة، واتسعت رقعة استخدامها في الفترة الأخيرة، ونشطت مئات الدراسات المتعلقة بتحسين مكونات الاختبار التواؤمي المحوسب سواء من حيث الحجم المناسب لمستودع المفردات، إلى الاستراتيجية الأفضل لانتقاء المفردات، إلى أدق أسلوب لتقدير القدرة، وأنسب طريقة لإنهاء الاختبار. وبصورة موازية أثيرت العديد من القضايا حول التكافؤ بين الاختبارات الورقية والاختبارات التواؤمية، ومن ذلك مراجعة المفحوص وتعديله للاستجابات. فيرى فريق من الباحثين أهمية السماح للمفحوص بمراجعة إجاباته السابقة وتعديلها إذا رغب في ذلك حتى تكون الاختبارات التواؤمية مكافئة تماما للاختبارات الورقية(*Gershon & Bergstrom, 1995; Stone & Lunz, 1994*)؛ بينما يشدد فريق آخر (*Stocking, 1997; Welch & Frick, 1993*) على ضرورة استبعاد هذا العنصر من الاختبارات التواؤمية المحوسبة لكي لا يقلل من فعالية وكفاءة عملية القياس، وللحفاظ على دقة قياس الاختبارات التواؤمية المحوسبة، وحتى لا يستغل بعض المفحوصين الخاصية لتضخيم درجاتهم النهائية.

وأجريت بناء على ذلك العديد من الدراسات (*Ferrara et al., 1996 ;Vispoel, et al., 2002; Wise, et al., 1999*) كما طرح أيضا عدد من المقترحات مثل السماح بالمراجعة مع تقييدها (*Papanastasiou, 2002; Vispoel, et al., 2000*)، إلا أن هناك دراسات (*Vispoel, et al., 2005*) أجريت مؤخرا دعت -مرة أخرى- إلى السماح للمفحوص بالمراجعة والتعديل بصورة طبيعية دونما أية قيود على عملية التعديل.

وبناء على ذلك استهدفت الدراسة الحالية الكشف عن جدوى توفير خاصية مراجعة وتعديل الإجابات في الاختبارات التواؤمية المحوسبة، وتأثير ذلك على تقديرات القدرة للمفحوصين وعلى دقة وكفاية القياس.

أسفرت نتائج الدراسة الحالية عن حصول أفراد عينة الدراسة على متوسط قدره (١,٣٥) بوحدة اللوجيت، مما يدل على ارتفاع تقدير قدرة الطلاب عامة، كما كانت قيمة متوسط الخطأ المعياري لتقدير القدرة (٠,٢٥) وهذه القيمة تعادل معامل ثبات بما قيمته (٠,٩٤) وهو معامل ثبات مرتفع يدل على دقة قياس هذا الاختبار. ويشير عدم وجود فروق دالة بين متوسطي تقديرات القدرة للبنين والبنات إلى تجانس المجموعتين.

أشارت نتائج الدراسة الحالية إلى أن (٦٠%) من أفراد العينة قد استفاد من خاصية التعديل سواء عدلوا استجاباتهم أم مجرد راجعوا إجابتهم. وأن نسبة الطلاب الذين غيروا إجاباتهم بالفعل بلغت (٤١,٣%) من إجمالي حجم العينة. وتتفق هذه النتيجة مع ما قررته نتائج الدراسات السابقة

مثل (*Ferrara, et al., 1996; Lunz & Bergstrom, 1994; Stone & Lunz, 1994; Vispoel, 1998*) وغيرها التي أجريت على الاختبارات التواؤمية المحوسبة. وكذلك التي أجريت على الاختبارات الورقية مثل (*Al-Hamly, & Coombe, 2005; Geiger, 1997a; Heidenberg, 2000; Mueller & Shwedel, 1975; Shatz, & Best, 1987*) فيما يتعلق بمراجعة المفحوص لاستجابته وتعديلها والتي خلصت إلى أنه عندما تعطى للمفحوصين الفرصة؛ فإن معظمهم يسعون لتغيير إجاباتهم بشكل أو بآخر.

وأشارت نتائج الدراسة الحالية أيضا إلى أن الطلاب الذين عدّلوا إجاباتهم قد ارتفعت درجاتهم عن الذين لم يعدلوا أو راجعوا فقط؛ سواء كان ارتفاع درجاتهم متمثل في عدد الإجابات الصحيحة، أو في تقديرات القدرة. وتأكيدا على هذه النتيجة -وهي أن التحسن الدال الذي طرأ على درجات مجموعة التعديل كان بفعل سلوك المراجعة والتعديل- فقد قام الباحث بمقارنة درجاتهم الأصلية مع درجات مع المجموعتين الأخريين (أي مجموعتي المراجعة وعدم المراجعة)، ولم تكن هناك فروق دالة بين متوسطات تقدير القدرة أو عدد الإجابات الصحيحة، مما يؤكد أن سلوك مراجعة وتعديل يحسّن بالفعل من درجات الطلاب.

كما أكدت نتائج الدراسة أيضا أن المفحوصين الذين قاموا بتغيير إجاباتهم قد تحسنت درجاتهم بصورة كبيرة ليس فقط مقارنة بالمفحوصين الذين لم يحدثوا أي تغييرات على إجاباتهم؛ بل أي مقارنة بدرجاتهم ذاتها قبل سلوك التعديل، وأن هذا التحسن كان دالا أي أن متوسط تقديرات القدرة بعد التعديل كان أعلى بدلالة عن متوسط تقديرات القدرة قبل التعديل، ونفس النتيجة كانت لعدد الإجابات الصحيحة التي أحرزوها، حيث كانت بعد التعديل أيضا أعلى بدلالة عما قبل التعديل حتى بلغت عشرة إجابات صحيحة لدى أحد الطلاب.

أي أن الذي فضّل مراجعة وتعديل إجاباته كان أحسن بدلالة ممن لم يراجع مطلقا. وأن إتاحة الفرصة للمفحوصين والسماح لهم لتعديل إجاباتهم يحسّن من تقديراتهم للقدرة. مما يؤكد حاجة المفحوصين إلى المراجعة، وإدراكهم أن هذا الإجراء يزيد من درجاتهم ولا ينقصها. وتتسق هذه النتيجة مع النتائج التي أكدتها الدراسات السابقة (*Al-Hamly, & Coombe, 2005; Ferguson, et al., 2002; Geiger, 1996; Heidenberg, 2000; Higham & Gerrard 2005; Lunz, et al., 1992; Mueller & Shwedel, 1975; Stone & Lunz, 1994Vispoel, 1998*) وتدحض المقولات الشائعة أن تغيير الإجابات سيؤدي إلى الوقوع في الأخطاء وانخفاض الدرجات.

أشارت النتائج أيضا إلى أن (٣٠%) من مجموعتي المراجعة والتعديل قد قاموا بمراجعة درجاتهم بشكل أو بآخر، وهو ما أشار إليه الباحث في موضع سابق من الدراسة[٨] أن الاختبارات على الحاسب تمتاز برصد تحركات المفحوص، وتمييز من قام بالمراجعة فقط دون تعديل أو تغيير، وتحديد هؤلاء الطلاب يكون شديد الصعوبة في الاختبارات الورقية. ولكن يبقى السؤال قائما عن الأسباب التي أحجمت هؤلاء الطلاب عن تغيير إجاباتهم، واكتفوا بالمراجعة فقط خاصة أن نسبتهم (١٨%) من إجمالي حجم العينة الكلية، حيث تحتاج هذه النتيجة إلى مزيد من الدراسات حول الأسلوب المعرفي الذي يقف وراء هذا السلوك.

بينت نتائج الدراسة أن نسبة اتجاه التغيير من الخطأ للصح كانت (٧١%)، أي أن معظم الذين غيروا إجاباتهم قد صححوا أخطائهم أكثر من كونهم وقعوا في الأخطاء، وذلك بمعدل (٤ تغييرات) لكل مفحوص. وتتفق هذه النتيجة مع ما توصلت إليه دراسات (***Ferrara, et al., 1996; Geiger, 1997a; Vispoel, 1998; Vispoel, et al., 2000***) والتي أكدت أن معظم اتجاه التغيير كان في الاتجاه الصحيح. كما كانت نسبة اتجاه التغيير في هذا الاتجاه للدراسة الحالية (٧١%) متفوقة في تلك النتيجة على نتائج دراسات (***Friedman-Erickson, 1994; Skinner, 1983; Stoffer, et al., 1977***) والتي تراوحت فيها نسبة التغيير من الخطأ إلى الصح من (٤٩,٥%) إلى (٦٥%).

لقد أشارت الدراسة الحالية أيضا أنه لم تكن هناك فروق دالة بين الطلاب والطالبات فيما يتعلق بسلوك المراجعة والتعديل، أي أن متغير الجنس لا يرتبط بتعديل الاستجابات. ولا يرتبط كذلك بعدد التغييرات التي تمت؛ بالرغم من أن الطالبات قد أجرت تغييرات بمتوسط أعلى قليلا من متوسط التغييرات التي أجراها الطلاب.

وتتفق هذه النتيجة مع نتائج دراسات (***Al-Hamly, & Coombe, 2005; Heidenberg, 2000; Ramsey, et al., 1987***) التي بينت أنه إما أنه لا توجد فروق بين الجنسين في نسبة تغيير الإجابات، أو أن البنات قد غيرن أعلى قليلا من البنين. وتختلف هذه النتيجة من جانب آخر مع نتائج دراستي (***Geiger, 1996; Skinner, 1983***) من أن الطالبات قد غيرن من الإجابات بنسبة أعلى بدلالة من الطلاب أو ضعف ما غيره الطلاب.

وفيما يتعلق بتأثير الجنس على اتجاه تغيير الإجابات؛ أشارت نتائج الدراسة أن الطالبات قد غيرن من الخطأ للصح وكذلك من الصح إلى الخطأ بنسبة أعلى بغير دلالة من الطلاب. وتتفق هذه النتائج مع نتائج بعض الدراسات السابقة حيث أكدت أنه إما لم تكن هناك فروق دالة بين الجنسين في اتجاه التغيير، أو أن الإناث كن أعلى قليلا من الذكور بغير دلالة في اتجاه التغيير من الخطأ

---

[8] ص (١٧)

للصح. (*Al-Hamly, & Coombe, 2005; Heidenberg, 2000; Mueller & Shwedel, 1975*)

والحقيقة أنه سواء ظهرت فروق دالة، أم لم تظهر بين الجنسين في سلوك التعديل أو نسبة التغيير أو متوسط عدد التغييرات أو في اتجاه التغيير من الخطأ للصح؛ إلا أنه من الواضح أن الطالبات تعدل من الإجابات أعلى من الطلاب سواء بدلالة أم بغير دلالة، وتحتاج هذه النتيجة إلى مزيد من الدراسات لتسلط الضوء على تأثير الجنسين في سلوك تعديل الإجابات.

لقد أكدت نتائج الدراسة الحالية أيضا أن سلوك المراجعة والتعديل لم يتأثر بمستوى محدد من القدرة، أي أن مستويات القدرة الثلاثة المختلفة قد لجئت إلى سلوك التعديل بنسب متساوية تقريبا، متفقة في ذلك مع نتائج الدراسات السابقة مثل دراسة (*Jacobs, 1971*) التي لم تجد علاقة بين مستوى قدرة الطلاب وتغيير الإجابات. ويدل هذا على أن الأفراد الذين في المستوى المرتفع من القدرة يرغبون في سلوك تعديل الإجابات بنفس القدر تقريبا للأفراد الذين من المستوى المنخفض من القدرة.

إلا أن الذي استفاد من سلوك المراجعة والتعديل هم الأفراد مرتفعو القدرة أكثر من منخفضي أو متوسطي القدرة. وتتفق تلك النتيجة مع دراستي (*Ferguson, et al., 2002; Mueller & Wasser, 1977*) التي أكدتا أن مرتفعي القدرة كانوا أكثر استفادة من تغيير المفردات من الطلاب المنخفضين. غير أن اللافت للانتباه في هذا السياق أنه بالرغم من ذلك؛ فإن الطلاب مرتفعي القدرة -في هذه الدراسة- أجروا عددا من التغييرات بمتوسط أقل بدلالة من الطلاب متوسطي القدرة أو منخفضي القدرة، وتختلف هذه النتيجة جزئيا مع ما أشارت إليه دراستي (*Al-Hamly, & Coombe, 2005; Ramsey, et al., 1987*) أنه لا توجد فروق دالة بين مستويات القدرة في عدد المفردات التي تغيرت، أو أن مرتفعي القدرة يميلون إلى عدم تغيير إجاباتهم. كما أن هذه النتيجة أيضا تلفت النظر إلى خصائص مرتفعي القدرة فيما يتعلق بسلوك التعديل، فدرجاتهم قد ارتفعت بعد إجراء التعديل بعدد أقل من التغييرات.

أما فيما يتعلق باتجاه التغيير فقد كشفت نتائج الدراسة أيضا عن أنه لم يكن هناك فرق دال في اتجاه التغيير بين مستويات القدرة؛ حيث أن مرتفعي القدرة كانوا النسبة الأعلى (بغير دلالة) فيما يتعلق باتجاه التغيير من الخطأ للصح، يليهم المتوسطين ثم المنخفضين. ولكن النتيجة الأخرى اللافتة للانتباه في هذا الشأن هو عدد التغييرات التي أجروها الطلاب ارتبطت سلبا مع مستوى القدرة، أي كلما زادت قدرة المفحوص انخفضت التغييرات التي يجريها على إجاباته. وأن الأفراد مرتفعي القدرة أجروا تغييرات أقل بدلالة من الأفراد متوسطي ومنخفضي القدرة.

غير أنه في المقابل؛ انخفضت كفاية الاختبار بسبب التعديل بنسبة (٥%)، كما أن فقدان

المعلومات كان بمقدار (١,٦٨)، وزاد زمن التعديل بمتوسط نسبته (٣٣%) عن زمن الاختبار الأصلي متميزة في ذلك عن دراستي (*Ferrara, et al., 1996; Vispoel, 1998*) الذي زاد فيه سلوك التعديل من زمن الاختبار بنسبة تراوحت بين (٣٧-٦١%). كما تتسق هذه النتائج مع ما قررته الدراسات السابقة الأخرى في مجال الاختبارات التواؤمية المحوسبة أن مراجعة المفحوص لاستجاباته ينقص من كفاية القياس، ويرفع قيمة الخطأ المعياري لتقدير القدرة، ويخفض من دقة الاختبار، ويزيد من زمن الاختبار (*Lunz, et al., 1992; Stone & Lunz, 1994; Vispoel, 1998*). غير أنه من الملاحظ في هذه الدراسة أن الخطأ المعياري لتقدير القدرة بعد التعديل لم يزد إلا زيادة طفيفة بلغت في أعلى فرق بين قبل التعديل وبعده قيمة (-٠,٠١٤)، وهي تعتبر قيمة طفيفة نسبيا عن الدراسات الأخرى، كما بلغ متوسط الفرق كذلك (-٠,٠١٧).

أما فيما يتعلق بقيمة معامل الثبات فقد كان قبل سلوك التعديل (٠,٩٤٣) بينما انخفضت قيمة معامل الثبات إلى (٠,٩٣٣) بعد التعديل. أي أن انخفاض قيمة معامل الثبات كان بمقدار (٠,٠١) فقط متفوقة بذلك على دراسة (*Papanastasiou, 2002*) التي كانت قيمة معامل الثبات للاختبار التواؤمي المحوسب (٠,٨١١) وبعد إجراء تعديل الإجابات كان (٠,٨٠٦).

تثير هذه النتائج بعض التضمينات السيكومترية المتعلقة بسلوك مراجعة وتعديل المفردات في الاختبارات التواؤمية المحوسبة. ففيما يرتبط بدقة القياس؛ أظهرت نتائج معظم الدراسات أن مراجعة المفحوص لاستجاباته يحسّن من درجته على الاختبار. وقد دعمت هذه النتائج نتائج أخرى سابقة أجريت باستخدام اختبارات الورقة والقلم منذ أواخر العشرينيات كما أشير سلفا.

أما في الاختبارات التواؤمية المحوسبة فإن تعديل المفحوص لاستجاباته من أكثر الإشكاليات السيكومترية في مجال الاختبارات التواؤمية المحوسبة؛ ففي الاختبار التواؤمي المحوسب ربما يكون من غير الممكن أن يتخطى المفحوص مفردة ما ثم يرجع إليها فيما بعد. هذا بسبب أن صعوبة المفردات المتتابعة متوقفة على مستوى صعوبة المفردات المعروضة سابقا. فإذا كان عدد المفردات المبدلة كثيرا؛ وفي نفس الاتجاه أي من الصح إلى الخطأ أو من الخطأ إلى الصح؛ فإن تقدير القدرة النهائي سوف يتغير بصورة جوهرية.

بل إنه من الممكن أن تندرج هذه العملية في عموم ما يسمى باستشفاف الاختبارات " ***test wiseness***" مما يؤدي إلى تكشف الإجابات (***Stone & Lunz, 1994, p. 221***)؛ فلن تعطي المفردات عندئذ أقصى معلومات عن تقدير القدرة للمفحوص. لأنه عندما تبدل الاستجابة على مفردة أجيب عليه سلفا بصورة صحيحة أو خاطئة؛ فإن تقدير القدرة والخطأ المعياري يبدلان أيضا، بالرغم أن صعوبة المفردة التالية المطبقة تبقى كما هي (***Lunz, et al., 1992, p. 38***).

ومن منظور آخر؛ تشير (*Lunz , et al., 1992, p. 39*) إلى أن الهدف من الاختبار ليس مجرد استدراج المفحوص للإبقاء على الأخطاء التي وقع فيها بدعوى "السلامة السيكومترية!" *psychometric integrity* ولكن لقياس قدراتهم بكل دقة ما أمكن ذلك. مؤكدين في ذلك أن التلاميذ معتادون منذ صغرهم على مراجعة إجاباتهم بعد الانتهاء من الاختبار لحذف أخطائهم وتعديلها لتحسين درجاتهم. كما أن (*Stone & Lunz, 1994, p. 212*) يقران بأن السماح للمفحوص بالمراجعة وتبديل الإجابات ربما تؤثر على دقة القياس سلبا، غير أن الاهتمام يجب أن ينصب على قدرة المفحوص في حقيقتها؛ وليس مجرد الاستجابات التي أدخلها.

وعلى صعيد آخر تؤكد (*Stocking, 1996, p. 4*) أنه من الصعب، بل ربما من المستحيل وضع أساليب تسمح للمفحوص بالتحكم في الترتيب الحقيقي للأسئلة التي سيجيب عنها. بل وأكثر من هذا؛ فإن بعض الإستراتيجيات الجيدة التي يتبعها المفحوص للتحكم في ترتيب الاستجابات في الاختبارات الورقية ربما تكون غير مناسبة في السياق الجديد للاختبارات التواؤمية المحوسبة.

ويرى (*Vispoel, 1998*) أن السماح بتغيير الاستجابات بعد المراجعة يمكن أن يزيد من صدق الاختبار إذا كانت تلك التغييرات تصحح الخطأ من خلال إعادة قراءة السؤال مرة أخرى لتصحيح سوء فهم أو لبس في الإدراك. ومن ثم يمكن لمراجعة المفردة أن تجعل الدرجات أكثر صدقا لأن الدرجات عندئذ سوف تمثل بصورة أدق مستوى المهارة لدى المفحوص عند الانتهاء من الاختبار.

أما فيما يتعلق بعدالة الاختبارات؛ ترى (*Stocking, 1997*) أن السماح للمفحوص بالمراجعة والتعديل في بيئة الاختبار التواؤمي المحوسب يمكن أن تتاح للمفحوص الفرصة لإنشاء اختبار سهل بطريقة غير ملائمة. حيث أشارت إلى استراتيجيات استشفاف الاختبار ***testwiseness*** التي تبدو واضحة للمفحوص الذي يستهدف تحقيق أعلى درجة ممكنة؛ بغض النظر عن النموذج السيكومتري المستند عليه الاختبار، وبغض النظر أيضا عن أسلوب تقدير مستوى قدرة المفحوص. لأن تطبيق أي اختبار غير ملائم يؤدي إلى أخطاء أكبر في تقدير مستوى القدرة. ويستفيد من ذلك المفحوصون من ذوي المستوى المتوسط أو المنخفض بالتحديد.

وتؤكد (*Stocking, 1997*) أنه إذا استفاد جميع المفحوصين من خاصية التحكم في المفردات الحقيقية المطبقة كما في الاختبارات الخطية؛ فإن الاختبار التواؤمي المحوسب يصبح عندئذ غير عادل لأي من المفحوصين. لأنه سيقوم كل منهم بانتقاء مفردات من اختيارهم الخاص؛ بل من الممكن أن يكون أداة سيئة للقياس. أما إذا استفاد بعض المفحوصين من هذه الخاصية فإن الاختبار التواؤمي المحوسب لا يصير عادلا لأولئك المفحوصين الذين لم يستفيدوا من هذه الاستراتيجية.

لقد أجريت بعض الدراسات للتحقق من القضية السابقة، عن طريق التحكم في عدد المفردات

المبدلة أو إجراء مراجعة مقيدة أو انتقاء مفردات ليست بأسلوب المعلومات القصوى ولكن بأسلوب المعلومات المحددة (*Bowles, & Pommerich 2001*)، غير أنها حصلت على نتائج متباينة. كما أنها اعتمدت هذه الدراسات على بيانات محاكاة وتحتاج إلى إعادة التطبيق على بيانات حقيقية.

فبينما يرى البعض أنه من الضروري تضمين عنصر مراجعة المفحوص للمفردات وتعديلها بعد انتهاء الاختبار لأنه يحسن أداءه، كما اتضح ذلك من الدراسات السابقة (*Lunz, et al., 1992*) وحتى تكون الاختبارات التواؤمية المحوسبة مكافئة للاختبارات الخطية (*Vispoel, 1998*). ولأن المفحوصين يفضلون هذه الخاصية (*Stone & Lunz, 1994; Vispoel, et al., 1992*).

يؤكد البعض الآخر على ضرورة استبعاد هذا العنصر لأنها تخفض من الكفاءة. فخاصية تعديل الاستجابات تزيد من طول الاختبار، وتزيد أيضا من الخطأ المعياري لتقدير القدرة، ولأن تضمين هذه الخاصية لا يحقق العدالة المطلوبة في الاختبار. ومن المحتمل أن يستغل بعض المفحوصين الخاصية بتضخيم درجاتهم النهائية (*Stocking, 1997*). وأن بعض المفردات المطبقة قد تصير معينات ***clues*** إلى الإجابات الصحيحة لمفردات أخرى تالية. فإذا تعرف المفحوص على هذه المعينات قد يستخدم مراجعة المفردات لتغيير إجاباته من خلال هذه المعينات دون أن يتعرف عليها بقدرته الشخصية (*Wise, et al., 1999*). وأن تضمين هذا العنصر ينتهك بالتبعية الفرض الثاني لنظرية الاستجابة على المفردة المتعلق بالاستقلال المركزي (*Welch & Frick, 1993, p. 48*).

بينما يدافع (*Gershon & Bergstrom, 1995, p. 14*) بشدة عن هذه الخاصية مؤكدين على وجوب تضمينها في الاختبارات التواؤمية المحوسبة حيث أنه ".. لا يجب حرمان المفحوصين من الحصول على درجات أعلى -كما ظهر من الدراسات السابقة- لمجرد أن هناك مفحوصا أو آخر يحاول تزييف إجاباته" مؤكدين إلى أنه من السهل كشف زيف الإجابات.

لقد بينت نتائج الدراسة الحالية أن أكثر من نصف عدد عينة الدراسة رغبوا في المراجعة والتعديل، والذين راجعوا تحسنت لديهم تقديرات القدرة بصورة دالة عن الأفراد الذين لم يراجعوا، وكذلك عن درجاتهم الأصلية قبل المراجعة. ويرغب في سلوك المراجعة والتعديل الأفراد المرتفعين والمنخفضين على حد سواء؛ وإن كان أكثر ما يستفيد منه هم الطلاب مرتفعي القدرة. كما خاصية مراجعة وتعديل الإجابات لم تؤثر كثيرا على دقة القياس فقيمة دالة المعلومات انخفضت بمتوسط مقداره (١,٦٨) فقط. ونقصت كفاية الاختبار بسبب التعديل بنسبة (٥%). ولا خوف من استشفاف الاختبار؛ فمتوسط عدد التغييرات التي تمت لم تزد عن (٤ تغييرات) لكل طالب. بناء على ذلك؛ خلصت الدراسة إلى جدوى بل وأهمية تضمين عنصر مراجعة وتعديل المفحوص لإجاباته في أنظمة الاختبارات التواؤمية المحوسبة.

أوضحت النتائج السابقة أن هناك الحاجة إلى مزيد من الدراسات حول مراجعة المفحوص لإجاباته وتعديلها في الاختبار التواؤمي المحوسب، فيما يتعلق بإمكانية التعديل أثناء أداء الاختبار وليس بعده، كما أنه من الضروري دراسة هذا العنصر مع متغيرات أخرى مثل الأسلوب المعرفي، وإجراء مقابلة للتقصي من المفحوص عن سبب التغيير، والاستراتيجيات المعرفية في علاقتها بسلوك المراجعة والتعديل، وكذلك المزيد من الدراسات حول تأثير متغيرات الفروق الفردية مثل موضع التحكم، وفعالية الذات، ومفهوم الذات والتغذية المرتدة، وقلق الاختبار، وتحمل الفشل، والمخاطرة، على كفاءة ودقة القياس. هناك الحاجة لمزيد من الدراسات للإجابة على هذه التساؤلات وغيرها.

وبالرغم من أن النتائج التي توصل إليها البحث الحالي تتفق إلى حد كبير مع نتائج الدراسات السابقة بل وامتداد لها، حيث درست أيضا تأثير مستويات القدرة والجنس على مراجعة وتعديل المفردات وكذلك على اتجاه تغيير الإجابات؛ إلا أنه من الممكن التوصل إلى نتائج أدق إذا كان الاختبار الذي طبقه الطلاب يتعلق نواحي تحصيلية لديهم. فمجرد إدراك بعض الطلاب أن ما يؤدونه هو "مجرد اختبار ذكاء!"، وأنه لن يترتب عليه أي درجات بالنسبة لهم، كفيل بأن ينقص من كفاءة الاختبار في حد ذاته، ولن يعطوا أفضل ما لديهم، بخلاف لو طُبق عليهم اختبار تحصيلي تواؤمي محوسب فربما يكون لديهم الدافع أقوى للحصول على مزيد من الدرجات والتي تنعكس على مستواهم الدراسي وعدد الطلاب الذين يرغبون في المرجعة والتعديل ربما يزيد عما هو واقع الآن.

التوصيات:

في إطار ما توصلت إليه هذه الدراسة من نتائج وحتى يمكن الاستفادة المثلى منها؛ ثمة مجموعة من التوصيات التي يؤمل أن تؤخذ بعين الاعتبار يمكن إيجازها فيما يلي:

١- تطوير اختبارات تحصيلية تواؤمية محوسبة تتعلق بمقررات دراسية لدى الطلاب لإجراء مزيد من الدراسات حول سلوك مراجعة وتعديل المفردات بصورة أكثر واقعية.

٢- إجراء المزيد من الدراسات حول تأثير متغير الجنس (الذكور والإناث) على سلوك تعديل المفردات، حيث أن هذا المتغير لم يتضح بصورة قاطعة في الدراسات خاصة في الاختبارات التواؤمية المحوسبة.

٣- هناك ندرة في الدراسات حول تأثير الأسلوب المعرفي على سلوك التعديل خاصة أن هذا المتغير لم يأخذ حقه من الدراسة في هذا السياق.

٤- لازالت هناك الحاجة إلى مزيد من الدراسات حول بعض دور بعض متغيرات الشخصية مثل تأثير موضع التحكم، ووجهة الضبط، ومفهوم الذات، وفعالية الذات، والمخاطرة، وتحمل الفشل، على سلوك المراجعة والتعديل في الاختبارات التواؤمية المحوسبة.

٥- يرى الباحث أن متغير قلق الاختبار ربما يعد أكثر المتغيرات تأثيرا على سلوك تعديل الإجابات بل وتأثرا به أيضا، خاصة في الاختبارات التواؤمية المحوسبة، لذا توصي الدراسة بضرورة دراسته في هذا السياق في بحوث لاحقة.

٦- إجراء المزيد من الدراسات على بيانات حقيقية وليست محاكاة فيما طرح من آليات التحكم في عدد المفردات المبدلة، أو أساليب المراجعة المقيدة أو انتقاء مفردات ليست بأسلوب المعلومات القصوى، ولكن بأسلوب المعلومات المحددة.

٧- يحتاج الميدان إلى دراسة سلوك تعديل المفردات للاختبارات الموائمة ذاتيا حيث أن هذا النمط من الاختبارات يسلط الضوء على جوانب أخرى من سلوك التعديل غير التقليدية.

المراجـــع

١- أمين محمد صبري نور الدين (٢٠٠٢): فاعلية استخدام الاختبـار المـوائم باسـتخدام الحاسب في تقدير قدرة الأفراد وتحديد الخصائص السـيكومترية للمقيـاس، رسالة دكتوراه، كلية التربية- جامعة عين شمس.

٢- أمينة محمد كاظم (١٩٨٨ أ): دراسة نظرية نقدية حول القياس الموضـوعي للسـلوك (نموذج راش). سلسلة الكتب المتخصصة، مؤسسة الكويت للتقدم العلمي،

٣- أمينة محمد كاظم (١٩٨٨ ب): استخدام نموذج راش في بناء اختبار تحصيلي في علـم النفس وتحقيق التفسير الموضوعي للنتـائج. مطبوعـات جامعـة الكويـت، الكويت.

٤- صلاح الدين علام (١٩٨٥): تحليل بيانات الاختبارات العقلية باسـتخدام نمـوذج راش اللوغاريتمي الاحتمالي (دراسة تجريبية)، المجلة العربية للعلـوم الإنسـانية، العدد ٢٧: مجلد ٧، ص ص ١٠٠-١٢٢.

٥- محمد منصور الشافعي (١٩٩٦): أثر طرق معادلة درجات الاختبار وضوابط اختيـار العينة على تدريج بنوك بنك الأسئلة باستخدام نموذج راش، رسالة دكتـوراه، كلية التربية - جامعة المنصورة.

6- Al-Hamly, M. & Coombe, C. (2005). To change or not to change: investigating the value of MCQ answer changing for Gulf Arab students. Language testing, 22, 509-531.

7- Alkhader, O., Clark, D., & Anderson, N. (1998). Equivalence and predictive validity of paper-and-pencil and computerized adaptive formats of the Differential Aptitude Tests. Journal of occupational and organizational psychology, 713, 205-218.

8- Assessment Systems Corporation. (1999). The FastTest Professional Testing System. Version 1.5. [computer program]. St. Paul MN: Author.

9- Baghi, H. Gabrys, R., & Ferrara, S. (1991, April). Applications of computer-adaptive testing in Maryland. Paper presented at the annual meeting of the American Education Research Association, Chicago, IL.

10- Ballance, C. (2006). Item characteristics and answer-changing behaviors. Psychological Reports, 98, 205-208.

11- Benjamin, L. T., Cavell, T. A., & Shallenberger, W. R. (1984). Staying with initial answers on objective tests: Is it a myth? Teaching of Psychology, 11, 133-141.

12- Bowles, R. & Pommerich, M. (2001, April). An examination of item review on a CAT using the specific information item selection algorithm. Paper presented at the annual meeting of the National Council of Measurement in Education, Seattle, WA.

13- Ferguson, K. , Kreiter, C., Peterson, M. Rowat, J. & Elliott, S. (2002). Is that your final answer? Relationship of changed answers to overall performance on a computer-based medical school course examination. Teaching and learning in medicine, 14, 20-23.

14- Ferrara. S., Frances, A., Gilmartin, D., Knott, T., Michaels, H., Pollack, J., Schuder, T., Vaeth, R., & Wise, S. (1996, April). A qualitative study of the information examinees consider during item review on a computer-adaptive test. Paper presented at the annual meeting of the National Council on Measurement in Education, New York City, NY.

15- Friedman, S. J., & Cook, G. L. (1995). Is an examinee's cognitive style related to the impact of answer changing on multiple-choice tests? The Journal of Experimental Education, 63, 199-213.

16- Friedman-Erickson, S., (1994, June). To Change or Not To Change: The Multiple Choice Dilemma. Paper presented at the Annual Institute of the American Psychological Society on the Teaching of Psychology, Washington, DC.

17- Geiger, M. (1996). On the benefits of changing multiple-choice answers: Student perception and performance. Education, 117, 108-119.

18- Geiger, M. (1997a). An examination of the relationship between answer changing, testwiseness, and examination performance. Journal of Experimental Education, 66, 49-60.

19- Geiger, M. (1997b). Educators' warnings about changing examination answers: effects on student perceptions and performance. College Student Journal, 31, 429-433.

20- Georgiadou, E. Triantafillou, E. and Economides, A. (2006). Evaluation parameters for computer-adaptive testing. British Journal of Educational Technology, 37, 261-278.

21- Gershon, R. & Bergstrom, B. (1995, April). Does cheating on CAT pay: Not. Paper presented at the annual meeting of the American Education Research Association, San Francisco, CA.

22- Green, K. (1981, April). Item-response changes on multiple-choice tests as a function of test anxiety. Paper presented at the Annual Meeting of the American Educational Research Association, Los Angeles CA.

23- Hambleton, R., Zaal, J. & Pieters, Jo P. M. (1994). Computerized adaptive testing: Theory, applications, and standards. In R. K. Hambleton, & J. N. Zaal. (eds.), Advances in educational and psychological testing: theory and applications. (pp. 341-366) Boston: Kluwer academic.

24- Heidenberg, A. J. (2000). Answer changing: A conditional argument. College student journal, 34, 440-452.

25- Higham, P. & Gerrard, C. (2005). Not all errors are created equal: metacognition and changing answers on multiple-choice tests. Canadian journal of experimental psychology, 59, 28-34.

26- Jackson, P. (1978). Answer changing on objective tests. Journal of Educational Research, 71, 313-315.

27- Jacobs, S. (1971, February). An experimental analysis of answer-changing behavior on objective tests. Paper presented at the Annual Meeting of the American Educational Research Association, New York, NY.

28- Lunz, M., & Bergstrom, B. (1994). An empirical study of computerized adaptive test administration formats. Journal of Educational Measurement, 31; 251-263.

29- 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.

29- Matter, M. (1986, April). Eenie, Meenie, Minie, Mo[9]--Change This Answer--Yes or No? Paper presented at the Annual Meeting of the National Council on Measurement in Education, San Francisco, CA.

30- McBride, J., & Martine, J. (1983). Reliability and validity of adaptive ability tests in a military setting. In D. J. Weiss (Ed.), New horizons in testing. (Pp. 223-236). New York: Academic Press.

31- McBride, J. R., Wetzel, C. D., & Hetter, R. D. (1997). Preliminary psychometric reach for CAT-ASVAB: Selecting an adaptive testing strategy. In W. A. Sands, B. K. Waters, & J. R.

[9] مقاطع صوتية مسجوعة لإحدى أغنيات الأطفال ليس لها معنى محدد على غرار "حادي بادي" وتتمته هو : Eenie Meenie Minie Mo :: catch a tiger by the toe

McBride (Eds.), Computerized adaptive testing: From inquiry to operation (PP.83-95). Washington DC: American Psychological Association.

32- McMorris, R. F., DeMers, L. P., & Schwarz, S. P. (1987). Attitudes, behaviors, and reasons for changing responses following answer-changing instruction. Journal of Educational Measurement, 24, 131-143.

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

34- Mills, C. & Stocking, M. (1996). Practical issues in large-scale computerized adaptive testing. Applied Psychological Measurement, 9, 287-304.

35- Moreno, K. Segall, D. & Hetter, R. (1997). The use of computerized adaptive testing in the military. In R. F. Dillon. (ed.), Handbook of testing.(pp. 204-219) London: Greenwood press.

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

37- Mueller, D., & Shwedel, L. (1975). Some correlates of net gain resulting from answer changing of objective achievement test items. Journal of Educational Measurement, 12, 251-254.

38- Papanastasiou, E. (2002, April). A "rearrangement procedure" for scoring adaptive tests with review options. Paper presented at the Annual Meeting of the National Council on Measurement in Education, New Orleans, LA.

39- Ramsey, P. H., Ramsey, P. P., & Barnes, M. J. (1987). Effects of student confidence and item difficulty on test score gains due to answer changing. Teaching of Psychology, 14, 206-210.

40- Shatz, M. A., & Best, J. B. (1987). Students' reasons for changing answers on objective tests. Teaching of Psychology, 14, 241-242.

41- Skinner, N. F. (1983). Switching answers on multiple-choice questions: Shrewdness or Shibboleth? Teaching of Psychology, 10, 220-222.

42- Stocking, M. (1984). Two Simulated Feasibility Study in Computerized Adaptive Testing. (Research Rep. 84-15). Princeton NJ: Educational Testing Service.

43- Stocking, M. (1997). Revising item responses in computerized adaptive tests: A comparison of three models. Applied Psychological Measurement, 21, 129-142.

44- Stoffer, G., Davis, K., & Brown, J. (1977). The consequence of changing initial answers on objective tests: a stable effect and a stable misconception. Journal of educational research, 70, 272-277.

45- Stone, G. & Lunz, M. (1994). The effect of review on the psychometric characteristics of computerized adaptive tests. Applied Measurement in Education, 7, 211-222.

46- Vispoel, W. (1998). Reviewing and changing answers on computer-adaptive and self-adaptive vocabulary tests. Journal of Educational Measurement, 35, 328-347.

47- Vispoel, W. , Clough, S., & Bleiler, T. (2005). A Closer Look at Using Judgments of Item Difficulty to Change Answers on Computerized Adaptive Tests. Journal of Educational Measurement, 42, 331–350.

48- Vispoel, W. P., Clough, S. J., Bleiler, T., Hendrickson, A. B., & Ihrig, D. (2002). Can examinees use judgments of item difficulty to improve proficiency estimates on computerized adaptive vocabulary tests? Journal of Educational Measurement, 39, 311–330.

49- Vispoel, W., Hendrickson, A., & Bleiler, T. (2000). Limiting answer review and change on computerized adaptive vocabulary tests: Psychometric and attitude results. Journal of Educational Measurement,37, 21-38.

50- Vispoel, W., Rocklin, T., & Wang, T. (1994). Individual differences and test administration procedures: A comparison of fixed-item, computerized-adaptive and self-adapted testing. Applied Measurement in Education. 7, 53-79.

51- Vispoel, W., Rocklin, T., Wang, T. & Bleiler, T. (1999). Can examinees use a review option to obtain positively biased ability estimates on a computerized adaptive testing?. Journal of Educational Measurement,36,141-157.

52- Vispoel, W., Wang, T., & Bleiler, T. (1997). Computerized adaptive and fixed-item testing of music listening skill: A comparison of efficiency, precision and concurrent validity. Journal of Educational Measurement, 34, 34-63.

53- Vispoel, W., Wang, T., de la Torre, R., Bleiler, T., & Dings, J. (1992, April). How review options, administration Modes, and anxiety influence scores on computerized vocabulary tests. Paper presented at the meeting of the National Council on Measurement in Education, San Francisco.

54- Urry, V. (1977). Tailored testing A successful application of latent trait theory. Journal of Educational Measurement, 14, 181-196.

55- 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.

56- Wang, T., & Kolen, M. (2001). Evaluating comparability in computerized adaptive testing: Issues, criteria and an example. Journal of Educational Measurement, 38, 19-49.

57- Wang, T. & Vispoel, W. (1998). Properties of ability estimation methods in computerized adaptive testing. Journal of Educational Measurement, 35, 109-135.

58- Wang, S. & Wang, T. (2001). Precision of Warm's weighted likelihood estimates for a polytomous model in computerized adaptive testing. Applied Psychological Measurement, 25, 317-331.

59- Welch, R. Frick, T. (1993). Computerized adaptive testing in instructional settings. Educational-Technology, Research & development, 41, 47-62.

60- Wise, S. (1996, April). A critical analysis of the arguments for and against item review in computerized adaptive testing. Paper presented at the Annual Meeting of the National Council on Measurement in Education, New York, NY.

61- Wise, S., Finney, S., Enders, C., Freeman, S., & Severance, D. (1999). Examinee judgments of changes in item difficulty: Implications for item review in computerized adaptive testing. Applied Measurement in Education, 12, 185-198.

62- Wise, S., Roos, L., Plake, B., & Nebelsick-Gullett, L. (1994). The relationship between examinee anxiety and preference for self-adapted testing. Applied Measurement in Education, 7, 81-91.

63- Zwick, R., Thayer, D. & Wingersky, M. (1994a). A simulation study of methods for assessing differential item functioning in computerized adaptive tests. Applied Psychological Measurement, 18, 121-140.

THE EFFECT OF REVIEW AND ANSWER-CHANGING BEHAVIORS ON THE ABILITY ESTIMATING AND EFFICIENCY OF COMPUTERIZED ADAPTIVE TESTS

Dr. Amin Mohammad Sabry Nour EL-Deen
Lecturer in Dept. of Educational Psychology
Faculty of Education- Ain Shams University

Abstract:

One of the most interesting and potentially advantageous methods for test developers and examinees is computerized adaptive testing. The algorithms for item selection usually depend on item response theory. However, there is a great debate in the measurement community regarding whether examinees should be provided an opportunity to review, and possibly change, their answers on a computerized adaptive testing. The purpose of this study was to examine the effects of reviewing and changing answers on the magnitude of ability estimates, efficiency, and measurement precision of computerized adaptive testing. The study participants were (235) undergraduates from different departments in faculty of education taking computerized Adaptive Raven's Matrices Test. The results of this study are in agreement with the results of previous studies, which report more than half of the sample size benefited from review option. There was a significant net improvement in ability estimates and number of correct answers when items were changed. however, the average efficiency of the test was decreased by (5%) after changing and the average information index loss about (1.68). The average amount time spent on changing was increasing (33%) over than original time of the test. The results Revealed no significant effect on review and answer-changing behavior was found for ability levels or gender. As ability level increase, number of answer changes decreases. At last, There were some psychometric implications about answer-changing behavior of computerized adaptive testing have been discussed.