مقالات تربوية

الاختبارات المحبوكة المحوسبة : تطوير لبنوك الأسئلة

◙ إعداد: ا.م.د/ محمد إبراهيم محمد محمد
أستاذ مساعد علم النفس التربوي (تخصص القياس والتقويم)-كلية التربية – جامعة المنيا

◙ المقدمة :

انتشرت تطبيقات الكمبيوتر لتشمل جميع مجالات الحياة ومن تلك المجالات بناء الاختبارات النفسية والتربوية حيث ظهر مفهوم " الاختبارات المحوسبة " Computer-based tests(CBT) التي تميزت بعدة فوائد منها : آلية عرض الفقرات على المتحنين ، وآلية التصحيح ، وتوفير الوقت والجهد ، والحفاظ على سرية الامتحانات ،ورغم تلك المميزات فقد أشارت العديد من الدراسات أن هذا النمط لا يحقق مستوي الكفاءة الاختبارية المقبولة في القياس وذلك لأنه لا يتم عرض المفردات بصورة تتناسب مع مستوي قدرة كل فرد ، فهو يعرض المفردات نفسها وبذات الترتيب على جميع المتحنين بصرف النظر عن الفروق الفردية بين المتحنين في قدراتهم . (Lilley,Barkert,Britton,2004,110)

وقد تطور مجال استخدام الكمبيوتر في بناء الاختبارات للاستفادة من إمكانياته الهائلة لذا فقد استخدم مفهوم " الاختبار المحبوك المحوسبة " Computer-Adaptive Test(CAT) والذي يطلق عليه " الاختبار المفصل او الاختبار التكيفى المحوسب" حيث يتم تفصيل وعرض مفردات لكل ممتحن تبعا لمستوي قدراتهم وبذلك يمتحن كل ممتحن بمفردات مختلفة عن الآخرين مراعاة للفروق الفردية بين المتحنين .

وترجع فكرة الاختبارات المحبوكة المحوسبة إلي أعمال (Binet,1980) من خلال اختبار الذكاء الذي يتكون من (١٢٩) اختبارا فرعيا يتم تطبيقها في زمن يمتد من ساعة إلي ساعة ونصف ، ورغم أنه من المستحيل تطبيق (١٢٩) اختبارا في زمن ساعة ونصف ، إلا أنه ابتكر طريقة تسمح بمراعاة الفروق الفردية بين الأفراد حيث يتم عرض مفردات لكل فرد تبعا لمستوي قدرته في المفردات الأولية ، ومن ثم يتعرض كل فرد لمفردات مختلفة عن الآخرين لاختلاف قدراتهم ،ويتم تحديد مستوي كل فرد من خلال مفهوم " العمر القاعدي " الذي يتحدد من الاختبارات الأولية التي يتعرض له كل فرد وفي ضوء العمر القاعدي يتعرض الفرد لمفردات مختلفة عن الأفراد الآخرين وذلك لاختلاف العمر القاعدي لكل فرد عن الآخرين ومن ثم يتم اختزال الـ(١٢٩) اختبار فرعي في عدد قليل من الاختبارات التي يمكن تطبيقها في زمن ساعة ونصف .(Linacre,2000,4)

ويعرف الاختبار المحبوك المحوسب على أنه : نظام لإدارة الاختبار يستخدم الكمبيوتر لانتقاء وعرض المفردات على المتحنين ويطلق عليه " تكيفي" لأن الكمبيوتر يختار لكل فرد مفردات مختلفة عن الآخرين تبعا لاختلاف خصائصهم ، وبذلك فهو يكيف الاختبار تبعا لخصائص المتحنين .(Sereci,2003,4) .

وبالتالي فإن هناك مميزات للاختبارات التكيفية المحوسبة عن الاختبارات التقليدية التي كانت تهتم بعرض جميع مفردات الاختبارات بنفس الطريقة على جميع المتحنين وفيما يلي جدول (١) يوضح الفرق بينهما.

جدول (١):المقارنة بين الاختبارات التقليدية والاختبارات المحبوكة المحوسبة

وجه المقارنة	الاختبارات التقليدية	الاختبارات المحبوكة المحوسبة
تركيب الاختبار	كل المشتركين يؤدون نفس الاختبار	يؤدى كل ممتحن اختبار مختلف عن الاخر
صعوبة الاختبار	تحدد في ضوء متوسط إجابات المشتركين	تحدد في ضوء إجابة كل فرد
طول الاختبار	طول الاختبار واحد لكل المشتركين يفضل الاختبارات الطويلة	طول الاختبار يختلف باختلاف المشتركين يفضل الاختبارات القصيرة
زمن الاختبار	زمن واحد لكل المشتركين	يختلف الزمن لكل مشترك
تنظيم الاختبار	يحتاج لوقت طويل	يحتاج لوقت قصير
فورية النتائج	لا	نعم

يلاحظ من الجدول أن انتقاء المفردات من بنك الأسئلة في الاختبارات المحبوكة المحوسبة يعتمد على قدرات المشتركين التي تم تحديدها من المفردات الأولي التي تم عرضها في بداية الدقائق الاولي من الاختبار لتحديد قدرات كل ممتحن وبالتالي يتم استدعاء مفردات تتناسب مع كل ممتحن حسب قدراته كما أن زمن الاختبار بالتالي يختلف باختلاف تلك المفردات التي تظهر لكل ممتحن نظرا لاختلاف طول الاختبار الذي يتناسب أيضا مع قدرات كل ممتحن على حده . (Tian et al , 2007)

◙ مبادئ الاختبارات المحبوكة الإليكترونية:

يقوم الاختبار المحبوك المحوسب على ثلاثة مبادئ هي :

- تقديم الأسئلة السهلة للطلاب المتفوقين يقلل من كفاءة الاختبار.
- تقديم الأسئلة الصعبة للطلاب الضعفاء يقلل من كفاءة الاختبار.
- تقديم أسئلة لكل طالب تتناسب مع مستوي قدرته حتى نهاية الاختبار.

(Laurier,1990,3-4)

يتضح من المبادئ السابقة أن كل اختبار يتكون من أسئلة سهلة وأخري صعبة وأن تقديم الأسئلة السهلة للطالب المتفوق لا يساعد في الحصول علي معلومات دقيقة عنه بل إنه يؤدي إلي الملل مما يؤثر على كفاءة الاختبار ، كما أن عرض الأسئلة الصعبة على الطالب الضعيف يؤدي إلي إحباطه مما يؤثر على كفاءة الاختبار ، ومن ثم فإن تقديم الأسئلة السهلة للطالب الضعيف والأسئلة الصعبة للطالب المتفوق يؤدي إلي زيادة كفاءة الاختبار والحصول على معلومات أكثر دقة عن المتعلم ، ويقلل من طول الاختبار مع زيادة كفاءته.

◙ خطوات بناء الاختبارات المحبوكة المحوسبة:

يتم بناء الاختبار المحبوك المحوسب من خلال إتباع الخطوات التالية :

- بناء مستودع مفردات Item Pool.
- اختيار أحد نماذج نظرية الاستجابة للمفردة الاختبارية المناسب.
- تحديد استراتيجية انتقاء وعرض العناقيد .
- اختيار أسلوب تقدير القدرة المناسب .
- تحديد قاعدة إنهاء الاختبار .

(Wainer,Kaplan,Lewis,1991,2-3)

◙ أولا- بناء مستودع المفردات :

إنه من الضروري لبناء الاختبارات المحبوكة المحوسبة تواجد بنك أسئلة يتضمن عدد ضخم من المفردات .(Wright and Bell, 1984) حيث يتضمن البنك تحديد نوع المفردة وخصائصها من حيث الصعوبة والتمييز والتخمين ، ويتضمن معلومات عن التدريج للمفردات والمحتوي التي تقيسه كل مفردة . حيث أن معظم الاختبارات التكيفية المحوسبة تعتمد في بناءها على شروط الاختبارات التقليدية من حيث الصعوبة والتمييز والتخمين ، إلا أنه عندما يتم استخدام تلك المفردات ضمن الاختبارات الكمبيوترية فإن مستويات الصعوبة والتمييز لبعض المفردات من الممكن أن تتغير بسبب أن هناك بعض المفردات تكون الإجابة عنها أسهل عندما تعرض على شاشة الكمبيوتر مثل المفردات التي تتناول الرسوم البيانية لذلك يتطلب نقل الاختبارات الورقية إلي الاختبارات الكمبيوترية عمل معادلة بين الصور الاختبارية الورقية والكمبيوترية. (Wright & Stone, 1979)

يعد المستودع بنك للمفردات يتكون من مجموعة من المفردات التي يتم إدارتها في الاختبار وبارامترات تلك المفردات ، وبمعني آخر إنه مجموعة من المفردات الاختبارية المناسبة التي تم ترميزها بعدة طرق (التخصص ، المستوي الدراسي ، الأهداف ، خصائص المفردة). (Gronlund,1998,130)

وتتحقق جودة مستودع المفردات طبقا لمحكين أساسيين هما :

- أن يكون العدد الكلي لمفرداته كافي لتجهيز اختبار غني بالمعلومات ومفيد أثناء الجلسة الاختبارية .
- أن تتضمن مفردات المستودع الخصائص المفيدة في تقديم معلومات قيمة عن المتحنين . وذلك لأن جودة المستودع تؤدي إلي جودة الاختبار فلا يمكن تحقيقها إذا كان المستودع محدد وغير كاف ومفرداته رديئة الجودة .(Flaugher,2000,38)

وتوجد عدة نقاط يجب أخذها في الاعتبار عند بناء مستودع المفردات

◙ حجم مستودع المفردات :

بالنسبة لحجم مستودع المفردات يجب ألا يقل عن ١٠٠ مفردة بحيث يسمح للاختبار المحبوك المحوسب إجراء القياسات بدقة أفضل من الاختبارات الورقية (Urry, 1977.) لذلك لا يمكن أن يكون هناك مستودع مفردات يقل حجمه عن ١٠٠ مفردة كما يسمح للمستودع بتكوين عدد من الاختبارات القصيرة التي تختلف باختلاف قدرات المتحنين .

◙ تحديد أبعاد مستودع المفردات .

يتم تحديد الأبعاد من خلال دراسة أبعاد الاستجابات التي قام بها الطلاب على الامتحانات لذا يحتاج القائم على الاختبارات المحبوكة المحوسبة لقياس أبعاد الاستجابة للوقوف على إذا كانت تلك الاستجابات أحادية البعد أو متعددة الأبعاد ، ولا تكفي نظرية الاستجابة للمفردة لتحديد الأبعاد أو حسن المطابقة بل هناك حاجة لاستخدام احصاءات أخري مثل : التحليل العاملي الاستكشافي ، والتحليل العاملي التوكيدي ، وبالتالي يتم تحديد أحادية البعد من خلال اجراء التحليل العاملي الاستكشافي (Bejar, 1980)، ويفيد تحديد أبعاد المستودع في تحديد نماذج الاستجابة للمفردة المناسبة لتدريج المستودع إذا كانت أحادية البعد يستخدم معها نماذج الاستجابة للمفردة أحادية البعد ، وإن كانت متعددة الأبعاد نستخدم نماذج الاستجابة للمفردات متعددة الأبعاد .

◙ المفردات التي تحذف والخاضعة للمراجعة .

تعد من المشكلات التي تؤخذ في الحسبان عند بناء الاختبارات التكيفية تحديد المفردات الغير مناسبة (المعيبة) حيث خصائص المفردات المعيبة باختلاف نماذج الاستجابة للمفردة المستخدمة ، ولذا فإن المفردات ذات المطابقة الضعيفة poor item يجب أن تحذف من مستودع المفردات فور اكتشاف أنها معيبة حتي لا يؤثر ذلك على كفاءة القياس .

ويحتاج مستودع المفردات إلي صيانة مستمرة للتأكد من دقة القياسات المستخدمة على المفردات حتي لا يتم حذف المفردات دون الدقة في القياس ، وبالتالي فقد يتم إعادة التطبيق للمفردات التي يبدو أنها معيبة للتأكد من ذلك كما أن عملية المراجعة تفيد في أن بعض النماذج المستخدمة قد يقترح حذف عدد كبير جدا من المفردات وبالتالي فهذا يستدعي إعادة النظر في أبعاد الاختبار (البعدية dimensionality) والنموذج المستخدم في التدريج .

◙ إضافة مفردات لمستودع المفردات :

بعد إجراء عمليات التدريج لمستودع المفردات وتحديد الاختبارات المحبوكة المحوسبة ، توجد حاجة لإضافة مفردات جديدة للمستودع والتي تعد مشكلة أن يتم إضافة مفردات جديدة على المفردات القديمة التي تم تدريجها وتدويرها ، وبالتالي فإن عملية تدريج المفردات الجديدة وإدخالها في الاختبارات المحبوكة تحتاج إلي دقة شديدة ، وهناك توجهان في هذا المجال ، الأول : أن يتم تقديم المفردات الجديدة والقديمة معا لنفس الأشخاص ليتم تدريجها معا وبالتالي فإن التدريج القديم للمفردات سوف يتعدل بالتدريج الجديد الذي تم من خلال الدمج بين المفردات القديمة والجديدة ، وهذه الطريقة لا يتم استخدامها في الاختبارات التكيفية بسبب أن أفراد مختلفين يتناولون مفردات مختلفة في نفس الاختبار التكيفي مما يؤدي إلي الحصول على بيانات كثيرة عن الطبيعي الذي يتم الحصول عليه من المفردات القديمة حيث أن وضع المفردات الجديدة مع المفردات القديمة يؤدي إلي إطالة الاختبار . والسبب الثاني : إن التحول إلي أداة القياس عملية جماعية حساسة جدا للمفردات المستخدمة لذلك لايقترح استخدام تلك الطريقة .

الطريقة الثانية أن المفردات الجديدة يتم تدريجها على نفس الأفراد الذين لديهم نفس مستوى القدرة التي تقيسه تلك المفردات كما قدر من المفردات القديمة حيث أن لكل فرد مستوى قدرة ثابت (Ingebo, 1997) ومن مميزاتها انها مرتبطة مباشرة بالدقة في تقدير السمات وعدد المفردات المتضمنة في الاختبار التكيفي .

◙ صيانة إتساق المقياس

إذا كان المقياس يستخدم منذ عدة سنوات فإن المجتمع الأصل الذي تم بناء المقياس عليه يكون قد تغير ، وبالتالي فمن المحتمل أن يكون هناك تغير في خضائصه ، ونتيجة لذلك فإن المتوسطات والنسب المئوية وغيرها من الإحصاءات المستخدمة تكون قد تغيرت مع مرور السنين ، لذلك فإن الاختبارات التي تم إعدادها تحتاج لإعادة التدريج وبالتالي فإن اتساق المقياس يكون مهددا في ظل الظروف التي يتم فيها التطبيق مثل : إجراء التعديلات ، الوقت المحدد ، العينة المناسبة ، والظروف البيئية الأخري ، كل تلك الظروف ممكن أن تؤدي إلي عدم الاستقرار في الدراسة التعديلية draft لذلك يجب التحكم والضبط في تلك العوامل إذا أردنا بناء مقياس أكثر استقرار عبر الفترات الزمنية الطويلة وهذا يستدعي جهود كبيرة لإجراء قياسات ضخمة للاختبار في مواقع مختلفة وعلى عينات متعددة ، وكما تشير معظم نماذج الاستجابة للمفردة أن المفردات يتم قياسها تحت اختبارات القوة دون التأثر بعامل السرعة . وبالتالي فإن تغير زمن الاختبار يعد مؤشرا على أنه يقيس السرعة حيث أن المفردات التي تظهر في بداية الاختبار تكون أسهل من المفردات نفسها عندما تظهر في نهاية الاختبار هذه المشكلة تستمر في السنوات المقبلة ، وبالتالي فإنه يبدو أن هناك صعوبة كبيرة في التدريج كجزء من إجراءات الاختبار وكنتيجة للنمو الطولي وتغير التقديرات تصبح التدريجات مشكوك فيها ، أسوأ ما في هذه الظاهرة أنه لا يمكن تصحيحها من خلال التعامل مع المفردات المحذوفة ، حيث أنها تزال متأثرة بالسرعة في الاستجابة . وتظهر تلك المشكلة أيضا عند محاولة تدريج المفردات الجديدة مع القديمة ولحل تلك المشكلة يجب ١- تنوع موضع المفردات المشتركة في الاختبار عبر الزمن ٢- استخدام اختبارات غير محددة بزمن مما يسهل على القضاء على مشكلة السرعة التي تؤثر على تقديرات الاختبار .

◙ إستخدام مستودعات مفردات متعددة :

إن الاختبارات التكيفية المحوسبة التي يبني عليها قرارات خطيرة يجب أن تتمتع بدرجة عالية من السرية ولتحقيق ذلك يتم استخدام مستودعات مفردات متعددة وهذا ييسر الحفاظ على الاختبارات إذا سرق مستودع المفردات أو كان في خطر ، حيث أن استخدام مستودعات مفردات متعددة يقلل من حجم المستودع وبالتالي يساعد على تذكر مكونات المستودع إذا سرق أو كان في خطر ، وييسر حساب احتمالات عرض المفردات على المتحنين . (Wise and Kingsbury , 2000)

◙ ثانيا - إختيار أحد نماذج نظرية الإستجابة للمفردة الإختبارية المناسب:

يعد عنصر اختيار النموذج المناسب ذو قيمة لأن الباحث يعتمد عليه في تدريج مفردات المستودع طبقا للبارامتر (الصعوبة ، التمييز ، التخمين) الذي يعتمد عليه النموذج ، فالنموذج الأحادي يعتمد على بارامتر الصعوبة فقط ويتضمن ذلك نماذج راش

اللوغاريتمي، في حين يعتمد النموذج الثنائي على بارامتري (الصعوبة والتمييز) ، بينما يعتمد النموذج الثلاثي على ثلاثة بارامترات (الصعوبة ، التمييز ، التخمين) .

وتتمثل عملية الانتقاء للنماذج الاستجابة للمفردة بعوامل عدة منها حجم العينة فإذا كان معلم يريد بناء اختبار علي عينة صغيرة فإن المناسب له استخدام نموذج راش أحادي البارامتر أما إذا كانت العينات كبيرة فهنا من الممكن استخدام النماذج ثنائية أو ثلاثية المعلم (Lord, 1983)، وبالنسبة لبنوك الأسئلة من الممكن أن يتعدد فيه استخدام النماذج فقد يتضمن بنك أسئلة بعض الاختبارات المدرجة بالنماذج الأحادية البارامتر ، واختبارات أخري متدرجة بنماذج ثنائية أو ثلاثية البارامتر كما قد يحتوي على مفردات ثنائية التدريج ومفردات أخري متعددة التدريج .

والقضية الثانية تتعلق باستخدام النماذج السابقة معا للعمل في الاختبارات التكيفية المحوسبة نفسها ،حيث أن الاختبارات التكيفية تختار المفردات اعتمادا على تقديرات البارامترات ، وتقدير تلك البارامترات أيضا يحتاج إلي الدقة ، و مراعاة أن تلك الدقة في الوقت ذاته لا بد أن توضح خصائص المفردة وأن تلك المفردات يتم تدويرها في الاختبارات التكيفية وبما أن البنك تم بناءه بنماذج متعددة فلابد أن يراعي ذلك عند تحديد البارامترات (Wise and Kingsbury , 2000)

◙ ثالثا- تحديد إستراتيجية إنتقاء وعرض العناقيد Exposure Control Strategies:

تعد تلك الاستراتيجيات إجابات تستخدم للتحكم في انتقاء وعرض المفردات والعناقيد بصورة تتناسب مع قدرات الطلاب بحيث تساعد على عرض جميع المفردات التي في المستودع مما يحافظ على سرية الاختبار ويراعي الفروق الفردية .

وتهدف تلك الاستراتيجيات لتحقيق الأهداف التالية :

- تحقق المفردات المنتقاة درجة عالية من الدقة في القياس من خلال انتقاء المفردات التي تعطي أقصي معلومات عن الطالب.
- المحافظة على سرية الاختبار من خلال عرض مفردات مختلفة لكل طالب.
- تحقق التوازن من خلال التأكد أن كل طالب تعرض لكل المحتويات التي يقيسها الاختبار. (Davis,Dodd,2003,33).

ويتم تصنيف استراتيجيات انتقاء المفردات والتحكم في العرض فيما يلي:

- الاستراتيجيات العشوائية .
- استراتيجيات الانتقاء المشروط .
- الاستراتيجيات الطبقية .
- الاستراتيجيات التكاملية .
- استراتيجيات الاختبار التكيفي متعدد المراحل .

(Geogriadou,Triantafillou,Economides,2007,7)

تتضمن كل إستراتيجية من الاستراتيجيات السابقة العديد من الاستراتيجيات الفرعية وفيما يلي تعريف بتلك الاستراتيجيات :

◙ الاستراتيجيات العشوائية :

تشير إلي أن الطلاب يتميزون بعد عرض عدة مفردات عليهم يتم انتقاؤها عشوائيا بطريقة تتناسب معهم ، حيث يتم تحديد عدد من المفردات أو العناقيد ليتم عرضها مبدئيا لتحديد مستوى الطالب ثم يتم انتقاء المفردات أو العناقيد التي تتلاءم مع مستوي القدرة الذي تم تحديده من المفردات أو العناقيد المبدئية . (Davis,Dodd,2003,338)

◙ استراتيجيات الانتقاء المشروط :

يتم تحديد قيمة مسبقة كحد أقصي لعرض المفردات أو العناقيد ويتم في ضوء تلك القيمة انتقاء وعرض المفردات والعناقيد حتى تصل إلي نسبة العرض لكل مفردة أو عنقود إلي تلك القيمة فيتم استبدالها بمفردات أو عناقيد أخري مما يساعد على الحفاظ على سرية الاختبار واستخدام جميع المفردات والعناقيد التي في المستودع بدلا من الاقتصار على بعض المفردات أو العناقيد . (Geogriadou,Triantafillou,Economides,2007,11)

◙ الاستراتيجيات الطبقية :

يتم ترتيب مفردات الاختبار في طبقات تبعا للبارامتر المستخدم (الصعوبة ، التمييز ، التخمين) حيث يتم ترتيب المفردات في مستويات (متفوق ، متوسط ، ضعيف) ثم يتم انتقاء المفردات من كل تلك الطبقات بحيث لا تتعدى القيمة القصوى لمعدل العرض .
(Geogriadou,Triantafillou,Economides,2007, 16)

◙ الاستراتيجيات التكاملية :

اتجه الباحثون إلي التحكم في العرض من خلال دمج أكثر من استراتيجية من الاستراتيجيات السابقة معا بدلا من الاعتماد على استراتيجية واحدة .
(Geogriadou,Triantafillou,Economides,2007,19)

◙ استراتيجيات الاختبار المحبوك متعدد المراحل .

وهي استراتيجيات تتناسب مع الاختبارات متعددة المراحل مثل : الاختبارات العنقودية حيث يتم بناء الاختبار متعدد المراحل في البداية من خلال اختبارات الورقة والقلم ثم تطبيقه للحصول على مواصفات وبارامترات المفردات والعناقيد المتضمنة بالاختبار.
(Geogriadou,Triantafillou,Economides,2007,21)

حيث تتضمن تلك الاستراتيجيات التعامل مع عناقيد المفردات حيث يبدأ الاختبار بعرض عدد من العناقيد التي تتناسب مع مستوي المتحنين لتحديد مستوي كل ممتحن وتحديد مساره في الامتحان . (Davis,Dodd,2003,7-8)

وقد أوصت العديد من الدراسات باستخدام تلك الاستراتيجيات لانتقاء العناقيد والتحكم في العرض منها دراسة :(Lord,1971,1974,)
(Leucht,2003, Armstrong ,Edmond,2004)

◙ رابعا- إختيار أسلوب تقدير القدرة المناسب :

يعتمد تقدير القدرة على أسلوبين أساسيين هما :

◙ تقدير الأرجحية القصوي :[Maximum Likelihood Estimation]

وهذا أسلوب إحصائي يتناسب مع النماذج أحادية البعد ، وذلك للحصول على تدريج لبارامترات القدرة من خلال إجابات الطلاب . (صلاح الدين محمود علام،٢٠٠٥، ١٠٠)

◙ الأساليب البيزية :[Bayesian Techniques]

وهي أكثر الأساليب التي تناسب نموذج الاختبارات العنقودية وخاصة النموذج البيزي العام للاختبارات العنقودية Ageneral Bayesian Model of Testlets .
(Wang,Bradlow,Wainer,2002,Wang,Wilson,2005)

◙ خامسا - تحديد قاعدة إنهاء الإختبار :

يعد اختيار قاعدة إنهاء الاختبار الخطوة الأخيرة في بناء الاختبارات التكيفية المحوسبة ويتم تحديد تلك القاعدة في ضوء المحكات التالية :

- استخدام جميع المفردات في المستودع : وتصلح تلك القاعدة مع المستودعات الصغيرة حيث يتوقف الاختبار عندما يستنفذ جميع المفردات في المستودع.
- الوصول للحد الأقصي المحدد لطول الاختبار: يتم تحديد قاعدة الإنهاء من خلال تحديد عدد محدد من المفردات الاختبارية التي يتعرض لها كل ممتحن.
- إذا تم قياس القدرة بدرجة عالية من الدقة: يتم تحديد قاعدة الإنهاء إذا تم قياس قدرة الطالب بدرجة عالية من الدقة وتقاس دقة التقدير من خلال انخفاض الخطأ المعياري في القياس ، والتي تقدر بـ(٢لوجيت) وهي الدرجة المثالية لقيمة الخطأ المعياري في القياس والتي يتوقف الاختبار عندما يصل إلي هذه القيمة .
- إذا تم قياس القدرة بدرجة كافية من خلال محك النجاح/الفشل: ويطلق على محك النجاح / الفشلPass/Fail Criterion محك مستوى الثقة حيث يتوقف الاختبار إذا تم الوصول إلي تطبيق نسبة ٨٥٪ من مفردات الاختبار ، حيث يتطلب في قاعدة الثقة المعتمدة على دقة قرار النجاح / الفشل أن يكون الخطأ المعياري أقل من (٢ لوجيت) وهو ما يقابل مستوى ثقة يتراوح من (٩٠ – ٩٥ ٪) .
- عندما يلاحظ الممتحن أن الاختبار به أعطال: ينتهي الاختبار عندما يلاحظ الممتحن وجود عطل مثل : تكرار عرض نفس الأسئلة على نفس الفرد ، أو أن يكون العرض بطيئا تارة أو سريعا تارة ، فهنا على الطالب تبليغ مشرف الاختبار لإنهاء الاختبار وتأجيل الامتحان. (Linacre,2000,11-12)

ولا يمكن إنهاء الاختبار إذا كان عدد قليل جدا من المفردات تم تطبيقه على الممتحن حيث أن الممتحن لا يشعر بأنه تم قياس قدرته في الاختبار قياسا كافيا إلا إذا أجاب على عدد من المفردات يتراوح بين (١٠ – ٢٠) مفردة ، كما لا يمكن إنهاء الاختبار إلا إذا تم التأكد من أن جميع الموضوعات التي يقيسها الاختبار قد تم تناولها حيث يتطلب قياس قدرة الفرد أن تعرض عليه مفردات من كل الموضوعات التي يهدف الاختبار في الأساس إلي قياسها.(Tian et al , 2007)

◙ مميزات الاختبارات المحبوكة المحوسبة :

عرض (Rudner 1998.) مميزات الاختبارات المحبوكة المحوسبة في النقاط التالية :

- بصفة عامة يتميز الاختبار التكيفي المحوسب بالمرونة .
- درجة المستوي والاختبار ككل يتم تقديرها فوريا .
- لا توجد جاجة لورقة الإجابة في إدارة الاختبار مما يقلل من أخطاء القياس .
- الاختبارات فردية لذلك لا يحتاج الممتحن لانتظار زميله أن ينتهي من الإجابة حتي ينتقل إلي المستوي الثاني وهذا يقلل من قلق الاختبار .
- زيادة سرية الامتحان حيث أنه لا توجد نسخ ورقية من الامتحانات .
- يتيح الاختبار التكيفي استخدام أزمنه وأشكال من المفردات متعددة مما يؤدي إلي تغطية أنماط متعددة من المفردات .
- يقلل الاختبار التكيفي طول الاختبار بنسبة ٥٠٪ مع الاحتفاظ بمعاملات الثبات.
- يؤدي تقليل زمن الاختبار وطوله إلي تقليل التعب الذي يؤثر على إجابات الممتحنين.
- الاختبار التكيفي يقدم تقديرات لكل مستويات القدرات للممتحنين إلا أن الاختبارات التقليدية تعطي متوسط إجابات الممتحنين .

◙ المراجـــــع

◙ أولاً- المراجع العربية :

- صلاح الدين علام (٢٠٠٥) :نماذج الاستجابة للمفردة الاختبارية أحادية البعد ومتعددة الأبعاد وتطبيقاتها في القياس النفسي والتربوي ، القاهرة : دار الفكر العربي.

◙ ثانيا- المراجع الأجنبية :

- Armstrong , R.,Edmond, J.(2004). Astudy of multiple stage adaptive test designs, Paper presented at the annual meeting of National Council of Measurement in Education (NCME) , San Diego, CA.
- Bejar, I. I. (1980). A procedure for investigating the unidimensionality of achievement tests based on item parameter estimates. Journal of Educational Measurement, 17,283-296.
- Davis , L.L.,Dodd,B.G.(2003). An Examination of Testlet scoring and item Exposure constraints in the verbal reasoning selection of MCAT, Applied Psychological Measurement ,27(3),335-356.
- Flaugher, R.(2000). Item pool . In H. Wainer(Eds.) Computerized adaptive testing :A primer (2nd ed.) (pp.37-59) .Mahwah , NH: Lawrence Elbawm Associates.
- Georgriadou,E.,Triantafillou,E.,Economides,A.(2007).A Review of item exposure control strategies for computerized adaptive testing developed from 1983 to 2005 , Journal of Technology learning and assessment , 5(8),available at: www.jtle.org.
- Ingebo, G. S. (1997). Probability in the measure of achievement. Chicago, IL: MESA Press.
- Laurier , M.(1990). What We can do with computerized adaptive testing … and What we cannot do , paper presented at the annual meeting of Regional language center seminar , available at : www.eric.ed.gov. Ed 322 729.

- Leucht,R.M.(2003). Exposure control using adaptive multistage item bundle , Paper presented at the annual meeting of National Council of Measurement in Education (NCME) , Chicago, TL.
- Lilley , M.,Barkert,T., Britton, C. (2004). The development Evaluation of Software Prototype for computer-adaptive testing , Computer&Education, 43,109- 123 , available at : www.sciencedirect.com.
- Linacre (2000). Computer – adaptive testing : Amethology whose time hascome , MESA Memorandum no.69, published in S.Chae,U.Kang, E.Jean,J.M.Linacre, Development of computerized middle school achievement tests (in Korean), Seoul,South Korean: Komesa press.
- Lord, F.M.(1971). A theoretical study of Two stage testing . Psychometrika, 36,227-241.
- Lord, F.M.(1974). Individualized testing and item characteristic curve theory , Contemporary Development in Mathematical Psychology , 2(pp.106-126), San Franciso , CA: W.H. Freeman.
- Lord, F. M. (1983). Small n justifies the Rasch model. In D. J. Weiss (Ed.), New horizons in testing: Latent trait test theory and computerized adaptive testing.New York: Academic Press.
- Rudner, L. (1998) An On-line, Interactive, Computer Adaptive Testing Mini-Tutorial. ERIC Clearinghouse on Assessment and Evaluation.
- Sereci,S.G.(2003).Computerized adaptive testing :An introducation in : Measuring up: assessment issues for teachers, Counsetors, and administrators : see CG032608, available at : www.eric.ed.gov.Ed 480083.
- TIAN Jian-quan, MIAO Dan-min, ZHU Xia, GONG Jing-jing (2007). An Introduction to the Computerized Adaptive Testing . US-China Education Review, ISSN1548-6613, USA . Volume 4, No.1 (Serial No.26) 72- 81
- Urry, V. (1977). Tailored testing: A successful application of latent trait theory. Journal of Educational Measurement, 14, 181-196.
- Wainer, H.,Kaplan,B.,Lewis,C.(1991).Some empirical guideline for building testlets, Educational Testing Service , available at : www.eric.ed.gov.
- Wang ,Bradlow, Wainer ,H.(2002). A general Bayesian for Testlets : Theory and Applications, Educational Testing Service , Princeton,NJ 08541.
- Wang , Wilson,(2005). The Rasch Testlet Model, Applied Psychological Measurement , 29(2),126-149.
- Wise, S.L and Kingsbury , G (2000). Practical Issues in Developing and Maintaining a Computerized Adaptive Testing Program . Psicológica , 21,135-155.
- Wright, B.D. & Bell, S.R. (1984) Item Banks: What, Why, How. Journal of Educational Measurement, (21): 331-345.
- Wright, B.D. & Stone, M.H. (1979) Best Test Design. Chicago: Mesa Press,.