الفصل الخامس

الاختبارات التكيفية

Computerized-adaptive testing CAT

محتويات الفصل :

- مقدمة.
- مفهوم الاختبارات التكيفية.
- مسميات الاختبارات التكيفية.
- أنواع الاختبارات الحاسوبية.
- الفرق بين الاختبارات: CBT ، وIBT، وCAT.
- الفرق بين الاختبارات الخطية والاختبارات التكيفية.
- كيف تعمل الاختبارات التكيفية CAT.
- فوائد ومميزات الاختبارات التكيفية.
- مزايا الاختبارات التكيفية مقارنة بالاختبارات الخطية.
- عيوب ومعوقات الاختبارات التكيفية.
- مبررات اعتماد الاختبارات التكيفية في المدارس.
- مُؤيدي ومُعارضي الاختبارات التكيفية.

تابع محتويات الفصل :

- ❖ مبررات مُؤيـدي الاختبـارات الحاسـوبية التكيفية.
- ❖ مبررات مُعارضي الاختبـارات الحاسـوبية التكيفية.
- ❖ مكونات الاختبارات التكيفية.
- ❖ دراسات سابقة حول الاختبارات التكيفية.
- ❖ مراجع الفصل الخامس.

❖ مقدمة :

يتزايد الاهتمام ببناء وتطوير المقاييس والاختبارات العقلية والأدائية لقياس وتقييم قدرات الأفراد، وتحقيق معايير القياس العلمي الذي يضمن تقديرًا كميًّا صادقًا وبدرجة مقبولة من الدقة والموضوعية للدرجات التي تم اعتمادها كتقدير كمي لهذه القدرات، وقد واكب هذا الاهتمام تطور في البرمجيات ذات الصلة بتطبيقات النظرية الحديثة في القياس، والذي أدى إلى ظهور الاختبارات المحوسبة **(Murphy and Davidshofer, 1994)** التي توفر إمكانية التخزين والتطبيق للاختبارات.

وتختلف الاختبارات المحوسبة في طريقة تطبيقها، فقد يتم تقديم جميع الفقرات لجميع المفحوصين بنفس الترتيب، ويسمى ذلك الاختبار «**الاختبار الخطي المحوسب**»، أو أن تقدم للمفحوص الواحد الفقرات التي تتناسب مع مستواه فقط، ويطلق عليه «**الاختبار التكيفي المحوسب**» **(CAT) testing Computerized adaptive**؛ حيث يُعرض على المفحوص بعض الفقرات لتحديد قدرته المبدئية، وبناءً على قدرته المبدئية المقدرة من خلال دائه على تلك الفقرات تُقدم له فقرات أخرى لاحقة من بنك الأسئلة بما يتناسب مع التقدير المستمر لمستواه، ويعتمد اختيار الفقرة اللاحقة على استجابات المفحوص على الفقرات السابقة، وخصائص تلك الفقرات **(Pinter and Cisar, 2010 .(Cisar, Radosav, Markoski،**

ومن فوائد الاختبار التكيفي أنهُ أكثر مرونة من الخطي، ويوفر عدة صور للاختبار **(Cisar, et al, 2010)** فهو يقلص عدد الفقرات الضرورية للوصول إلى مستوى معين من الدقة في القياس، ويقل الوقت المطلوب للاختبار **(Davey, Stone and 2011)**، كما يقلل من احتمالية معاناة المفحوص من التعب والملل وتراجع مستوى الدافعية **(Huo, 2009, Magis and Raiche, 2011)** كما أنه

يقدم معلومات أكثر عند نهاية تنفيذ الاختبار أكثر من أي اختبار آخر، ويوفر معلومات كافية عن مستويات متوسطة من الصعوبة للمهارة أو القدرة، ويتطلب عددًا أقل من الفقرات للحصول على المستوى المطلوب من الدقة والثبات مقارنة بالاختبارات الخطية **(Vispoel, 1993)**.

وتكمن أهمية استخدام الاختبارات التكيفية المحوسبة في مزاياها عن الاختبارات التقليدية، حيث أن خوارزميات الحاسب تقوم بدور أخصائي الاختبارات والقياس النفسي، والاختبار يتكيف مع قدرة المفحوص؛ فهي تختلف تمامًا عن الاختبارات الخطية والعشوائية ليس بسبب اختلاف عدد وخصائص الفقرات التي يتلقاها كل مفحوص عن الآخر فقط، بل يمكن أن يتم تقدير القدرة عن طريق النظرية الحديثة في القياس ونماذجها المختلفة بعد الاستجابة عن كل فقرة إمَّا بطريقة الأرجحية العظمى **Maximum Likelihood** أو بالطرق البييزية **Bayes**، وتعتبر عملية اختيار طريقة التقدير الملائمة للقدرة ذات أهمية بالغة لأنها لا تؤثر فقط في النتيجة النهائية للاختبار، بل تؤثر في طريقة اختيار الفقرات التي يتم تطبيقها، وفي قاعدة إنهاء الاختبار.

ويتطلب الاختبار التكيفي المحوسب بناء بنك للأسئلة كخطوة أولى، لأنه يعتمد على مجموعة كبيرة نسبيًّا من الفقرات ذات معالم **Parameters** معروفة. وترى إمبرتسون ورايس **(Embretson and Reiase, 2000)** أن القياس الدقيق يتطلب احتواء هذا البنك على عدد كافٍ من الفقرات، وذات صعوبة موزعة بشكل جيد على متصل السمة.

ويُعرّف كل من ميلمان وآرثر **(Millman and Arther, 1984)** ذلك البنك بأنه "مجموعة كبيرة من الأسئلة التي يسهل استخدامها، وأن المعلومات عن كل فقرة مفهرسة وموصوفة بصورة كاملة، ويسهل انتقائها في أي اختبار بمواصفات محددة، ومن تلك المعلومات مثلاً: الإجابة الصحيحة، ومعالم الصعوبة، والتمييز، والتخمين. وأن 30 فقرة بهذه المواصفات يمكن أن تكون كافية وملائمة لإعداد

اختبار تكيفي **(Dodd, Koch and De Ayla, 1989; and Dodd, 1998 Chen, Hou)**

أمَّا الخطوة التالية فهي تحديد أسلوب اختيار الفقرة، ومنها أسلوب مطابقة الصعوبة **(bi)** مع المهارة أو القدرة المقدّرة ففي هذا الأسلوب يتم اختيار الفقرة ذات الصعوبة الأقرب للمهارة أو القدرة المقدّرة، وبعد أن يستجيب المفحوص على الفقرة أو الفقرات الأولية لتقدير القدرة يتم اختيار الفقرة ذات المعلومات الأكبر عند مستوى المهارة أو القدرة المقدّرة للمفحوص، ثم تقدر هذه القدرة من جديد وهكذا ..، حتى يتم إنهاء الاختبار. ويتمتع هذا الأسلوب بأنه يوفر أقصى معلومات بأقل خطأ في القياس عند مستوى القدرة المستهدفة. **(Kingsbury and Zara, 1989، Vispoel, 1993)**

وهناك عدة طرق لتقدير المهارة أو القدرة **methods Ability estimation** أبرزها طريقة الأرجحية العظمى **Maximum Likelihood Estimation (MLE)**، والطرق البييزية **Bayesian Methods** التي تتضمن طريقة التقدير البعدي المتوقع **Expected A Posterior (EAP)** وطريقة التقدير البعدي المُعظم **Maximum A Posterior (MAP)** إلاَّ أنها تختلف في دقة التقدير الذي يؤثر بالتالي في كيفية إنهاء الاختبار. وتُعد الأرجحية العظمى هي الأكثر استخدامًا، وتعتمد في تقدير المهارة أو القدرة على نمط إجابة المفحوص لمجموعة فقرات تأخذ واحدًا أو صفرًا.

أمَّا طريقة التقدير البعدي المُعظم **(MAP)** فهي تستخدم معلومات سابقة عن توزيع المهارة أو القدرة، ويتم افتراض هذا التوزيع بناء على معلومات سابقة، والتوزيع الأكثر استخدامًا هو التوزيع الطبيعي المعياري. والمشكلة في هذا الأسلوب أن التقدير يكون متحيزًا عندما يكون عدد الفقرات أقل من ٢٠ **(Embreston and Reiase, 2000).**

وأمَّا الخطوة الأهم في الاختبارات التكيفية فهي تحديد قاعدة الإنهاء للاختبار، حيث ينتهي الاختبار التكيفي المحوسب عادةً عند تطبيق عدد محدد من الفقرات **Fixed Length** بدرجة مقبولة من الدقة لكل مفحوص، ووفقًا لهذه القاعدة يأخذ كل مفحوص مجموعة فقرات مختلفة عن المفحوصين الآخرين وبنفس العدد. كما ينتهي تقديم الاختبار عند الوصول لأدنى خطأ معياري **Minimum Standard error** محدد مسبقا. وقد أشارت إمبرستون وريس **(Emperston and Reiase, 2000)** أنه في الغالب يستخدم الخطأ المعياري ٠.٢٥ كحد أقصى، وينتهي عند الوصول لأدنى قيمة للمعلومات، حيث يستمر تطبيق الفقرات طالما أن هناك إمكانية للحصول على مقدار أكبر من المعلومات عن المفحوص، وعندما يصل الاختبار إلى أدنى زيادة في المعلومات المصاحبة لتقدير القدرة يتوقف الاختبار

ومن حيث عدد الفقرات الداخلة في الاختبار كوجه من أوجه المقارنة بين الاختبارات الخطية والاختبارات التكيفية فقد أشار وارد **(Ward, 1984)** أن هذه الاختبارات تحتاج بشكلٍ عام إلى عدد من الفقرات تقل بنسبة ٥٠ ٪ إلى ٦٠ ٪ عن الاختبارات الخطية عند المستوى نفسه من الدقة. كما أشار لِندن وبشلي **(Linden and Pushley, 2003)** إلى أن الاختبارات التكيفية المحوسبة ذات الطول المتغير التي تتوقف عند الوصول لخطأ معياري معين قد خفضت عدد الفقرات التي تطبق بمقدار يصل إلى حوالي ٩١ ٪. والاختبارات ذات الطول الثابت التي تتوقف عند تطبيق عدد محدد من الفقرات قد خفضت عددها إلى حوالي ٨٩ ٪.

❖ مفهوم الاختبارات التكيفية :

صمّمت اختبارات الكمبيوتر التكيفية لضبط مستوى الصعوبة القائمة على الردود المقدمة لتتناسب مع المعرفة والقدرة على اخذ الاختبار. فإذا كان الطالب يعطي إجابة خاطئة، فإن السؤال التالي سيكون سهلاً، وإذا كان الطالب يجيب بشكل صحيح، فإن السؤال التالي سيكون أكثر صعوبة، تعتبر على حافة الرائدة في مجال تكنولوجيا التقييم والاختبارات الكمبيوتر التكيفية حيث كونها تمثل محاولة لقياس قدرات الطلاب الفردية على نحو أدق، مع تجنب بعض القضايا غالبًا ما ترتبط مع طبيعة «**مقاس واحد يناسب الجميع**» من توحيد الاختبارات للطلاب، والاختبار التكيفي يقدم دورة اختبار أقصر مع عدد أقل من الأسئلة، حيث يتم عرض فقط تلك الأسئلة التي تعتبر مناسبة للطالب. ومن ناحية أخرى، تطورت هذه الاختبارات لخلق مجموعة أكبر من بنود الاختبار حيث أن اختبارات النظم لديها ما يكفي من الأسئلة لتتناسب مع قدرات متنوعة من جميع الطلاب في الامتحان، وعادة ما تدار أكثر الأشكال الحالية للاختبار الحاسوبي التكيفي على شبكة الإنترنت، ولأن التطبيق محوسب، يمكن للمدرسين والطلاب الحصول على نتائج الاختبار بسرعة أكثر عن ما يجرى في اختبارات الورقة والقلم.

فاختبارات الكمبيوتر التكيفية يمكن استخدامها لمجموعة متنوعة وواسعة من الأغراض، بما في ذلك على نطاق واسع، مثل الاختبارات عالية المخاطر. والتقويم التكويني، الذي يقدم للمعلمين ردود الفعل في العملية التعليمية للوقوف على مدى تقدم الطلاب والتي يمكن من خلالها تحديد الاستخدامات لتعديل التقنيات التعليمية، والاختبارات التحصيلية، والتي يستخدمها المعلمون لتحديد ما تعلمه الطلاب في نهاية الوحدة، أو في نهاية الفصل الدراسي، أو السنة. كما أنها تستخدم لتحديد الطلاب الذين قد يحتاجون إلى الدعم الأكاديمي المتخصص في مهارة معينة أو مجال الموضوع، أو مقرر أو مادة معينة.

وقد عرَّف احمد عودة وعمر عبيدات الاختبار التكيفي المحوسب بأنه: هو الاختبار الذي يُفصل لكل مفحوص على حدة، وذلك بعرض الفقرات التي تتناسب مع مستوى قدرته، مما يمكن من تقدير أدق للقدرة بأقل عدد من الفقرات. (أحمد عودة وعمر عبيدات، 2013م)

في حين أن الاختبار الخطي: هو الاختبار الذي يتم فيه تطبيق جميع فقراته على جميع المفحوصين بالعدد نفسه والترتيب نفسه.

كما يعرف الاختبار التكيفي بأنه الاختبار المُفصّل أو المحبوك **Tailored testing** المتمثل في تقديم فقرات ذات صعوبة تتناسب مع قدرات المفحوصين على النحو **(Jain-quan , Dan-min xia and, Jing-jing, 2007)** الآتي:

١) بعد تقدير المهارة أو القدرة الأولى يقوم الحاسب باختيار فقرة جديدة تكون ذات قيمة في عملية تقدير المهارة أو القدرة الحقيقية، وهذا مرتبط بطريقة اختيار الفقرة.

٢) يتم تقديم الفقرة للمفحوص والإجابة عنها، ثم تصحيحها.

٣) يتم إعادة تقدير المهارة أو القدرة بناءً على جميع الفقرات التي تم تقديمها للمفحوص باستخدام إحدى طرق التقدير.

٤) تحديد فيما إذا استوجب التطبيق تقديم فقرة أخرى بناء على المحك المستخدم في قاعدة إنهاء الاختبار.

٥) إذا لم يتحقق محك إنهاء الاختبار يتم الرجوع للخطوة الأولى، وإذا تحقق يتم إنهاء الاختبار.

وينظر البعض إلى أن الاختبارات التكيفية هي: التي يُقدم فيها السؤال للطالب حسب إجابته على السؤال السابق فإن أصاب في الإجابة عنه قدم له سؤال أكثر صعوبة منه وإذا اخفق قدم له سؤالا أقل صعوبة.

❖ مسميات الاختبارات التكيفية :

يوجد العديد من المسميات للاختبارات التكيفية، تحمل جميعًا نفس المعنى وهى على النحو التالي:

- الاختبارات التكيفية.
- الاختبار الحاسوبي التكيفي.
- الاختبار الإلكتروني التكيفي.
- الاختبار المحوسب التكيفي.
- الاختبار الكمبيوتري التكيفي.
- الاختبار الإلكتروني التلائمي.
- الاختبار الإلكتروني التواؤمي.
- الاختبارات الإلكترونية التوافقية.

❖ أنواع الاختبارات :

نجد أن الوعاء الاختباري التعليمي المحوسب ينقسم إلى ثلاثة أشكال تتباين نوعًا أو كيفًا، فمن حيث الكيف: هناك الاختبار الشبكي الخارجي على مستوى المناطق أو الداخلي على مستوى المؤسسة، من حيث النوع هناك اختباران:

1- الاختبار التقليدي المحوسب.

2- الاختبار الأحدث المتكيف مع القدرات المعرفية والمهارية للمتقدمين للاختبار.

وبتوضيح أكثر يعنى **الاختبار الشبكي** تقديم الاختبارات من خلال شبكة الإنترنت **(Internet-Based Test, IBT)** من أحدث النقلات الكيفية في تقدم الاختبارات المحوسبة.

وبطبيعة الحال يمكن من خلال هذه القناة الاختبارية التحكم في تقديم الاختبار من مكان ما في وزارة التربية والتعليم في العاصمة مثلاً، أو من مكان ما في أرجاء المعمورة لينطلق الاختبار في مؤسسات تعليمية أخرى في المدينة أو في القطر أو في أطراف متباعدة من بلد الاختبار أو في أطراف العالم، ومن المكان نفسه يمكن إيقاف عرض الاختبار.

وتضمن الجهات المسئولة عن تقديم الاختبارات في هذه الصورة من الأوعية الاختبارية سرية عالية في حفظ معلومات الاختبار، وميكانيكية أسهل في إيصال الاختبارات للمواقع المطلوبة.

ويمكن أن يتم استخدام الإنترنت لتقديم الاختبارات الحاسوبية في صورتيها: **الاختبار الحاسوبي التقليدي المحوسب** وهو اختبار عادي ذو اتجاه واحد، **والاختبار التكيفي** مع قدرات المتقدمين للاختبار وهو اختبار يمكن أن يسير في اتجاهين، وهما طريقتان اختباريتان تختلفان نوعيًّا كما سيتم بيانه فيما يلي:

فالاختبار التقليدي المحوسب - أو ما اصطلح على تسميته بالاختبار المعتمد على الحاسب الآلي **(Computer-Based Test, CBT)** - يشبه الوعاء الورقي المعتاد من حيث مضمون الاختبار، وطريقة سير محتوى الاختبار في اتجاه واحد. تتمثل أبرز الفروق بينهما في الوعاء الجديد يحمل في طياته إمكانية أكبر للحفاظ على سرية الاختبار، وتتم عملية تقييم الإجابات آليًّا.

أمَّا أبرز أوجه التشابه بين الوعاءين التقليدي والاختبار الحاسوبي العادي أن الوعاءين يتسمان بمنهج ثابت في تقديم نفس المجموعة من الأسئلة لجميع أفراد المجموعة المتقدمة للاختبار.

أي أن كل المتقدمين لأخذ هذا الاختبار الحاسوبي يجيبون على الأسئلة بنفس الترتيب الذي تقدم فيه الأسئلة في النسخة الورقية. وعلى هذه الصورة من الممكن أن يكون لاختبار ما نسختان متماثلتان، واحدة ورقية وأخرى حاسوبية.

ومع التطور النوعي للاختبارات المحوسبة برز مؤخرا ظهور ما يسمى **بالاختبار الحاسوبي المُتكيف** مع المقدرة المهارية أو المعرفية للمتقدم للاختبار **CBT, (Computer-Adaptive Tests)** ففي هذا النوع من الاختبارات، إجابة أي من المتقدمين للاختبار عن سؤال ما من الأسئلة المقدمة ستؤثر على مستوى صعوبة الأسئلة التالية التي يختارها الحاسب لهم، وهذا يؤثر ضمنا على اتجاه سير الاختبار.

فقد يصعد اتجاه الاختبار للأعلى أو يرجع للخلف. ففي حالة إذا قدم الطالب إجابة صحيحة عن السؤال سيختار الحاسب سؤالاً أصعب قليلاً من الأول ويطرحه على المتقدم للاختبار.

وبالعكس إذا أخفق الطالب في الإجابة عن السؤال سيختار الحاسب سؤالاً أقل صعوبة، وهكذا يقوم الحاسب الآلي بتكييف الاختبار بناء على مستوى المتقدم للاختبار.

وتستمر عملية الصعود للأصعب والنزول للأسهل إلى أن يستقر أداء الطالب على مستوى معين، ربما دون الحاجة لعرض جميع الأسئلة في السمة المختبرة التي تتراوح عادة بين 20 و25 سؤالاً.

فيحصل الحاسوب بهذا الأسلوب على المعلومات الكافية التي تمكنه من الحكم على مستوى المتقدم للاختبار، وهذا يشير إلى أنه ليس بالضرورة أن يُسأل المتقدمون للاختبار في مجموعة اختباريه واحدة نفس الأسئلة، ذلك أنهم يجيبون عن أسئلة متباينة تطرح عليهم بناء على تباين قدراتهم.

وبالنسبة لتمييز مستوى صعوبة الأسئلة، فيتم عن طريق معادلات معامل التمييز التي يمكن أن يقوم بها الحاسوب آليًّا بسهولة، وذلك بمتابعة مدى أداء الطلاب المتفوقين والطلاب الضعفاء عن سؤال ما من بنود الاختبار، والمدى الذي أخطأت فيه المجموعتان من الطلاب في الإجابة عن نفس السؤال، وفق عمليات رياضية معروفة في كتب الإحصاء في مجال تمييز الصعوبة والسهولة.

وهذا يعني أن الاختبار المتكيف يتكون من أسئلة على شكل مجموعات متباينة، أي من (بحيرات) متعددة من الأسئلة التي تختلف في صعوبتها، وبهذا الأسلوب يكون الوعاء الحاسوبي عادة أقصر من حيث الوقت المستغرق للإجابة من الاختبارات الورقية العادية، لأن كثير من المتقدمين للاختبار يتحدد مستواهم من خلال مجموعة محددة من الأسئلة.

❖ الفرق بين الاختبارات CBT ، وIBT، وCAT :

* يقصد بـ **CBT = Computer Based Test** ويعنى الاختبارات المحوسبة (التي يتم تطبيقها بواسطة جهاز الحاسب) وغالبًا يبث الاختبار من خلال الشبكة الداخلية بنفس المكان أو من خلال جهاز محدد بالمعمل أو أسطوانة مدمجة. ولا يختلف الاختبار في صورته هذه عن الصورة الورقية، غير أن استخدام الحاسب يتيح استخدام بعض الأنواع المتحدثة من الأسئلة التي تعتمد على الوسائط المتعددة. كما يتيح هذا النوع تغيير ترتيب ظهور الأسئلة لكل ممتحن بشكل مختلف، ومن أمثلته اختبار كفاءة اللغة الإنجليزية **CBT TOEFL** والذي وقف استخدامه في كثير من دول العالم بعد ظهور **IBT**.

* يقصد بـ**IBT = Internet Based Test** ويعنى الاختبارات المحوسبة (التي يتم تطبيقها بواسطة جهاز الحاسب) والتي يبث الاختبار فيها من خلال شبكة الإنترنت الدولية. ولا يختلف الاختبار في صورته هذه عن الصورة الورقية، غير أن استخدام الحاسب يتيح استخدام بعض الأنواع المُحدثة من الأسئلة التي تعتمد على الوسائط المتعددة. كما يتيح هذا النوع تغيير ترتيب ظهور الأسئلة لكل ممتحن بشكل مختلف، ويواجه هذا النوع من الاختبارات عدة مشاكل تقنية خاصة فيما يخص سرعة نقل البيانات عبر شبكة الإنترنت والحفاظ على سرية محتوى الاختبار، ومن أمثلته اختبار كفاءة اللغة الإنجليزية **IBT TOEFL**.

* يقصد بـ **CAT = Computerized Adaptive Tests** ويعنى الاختبارات المحوسبة المتكيفة أي التي يختر فيها الحاسب الأسئلة وفق مستوى أداء الممتحن، فكلما أجاب الممتحن عن الأسئلة بشكل صحيح، قدم له الحاسب أسئلة أصعب، وإذا أخفق في إجابة الأسئلة، قدم له أسئلة أسهل حتى يستطيع الحاسب تقدير قدرة الممتحن بشكل دقيق، وعندها يوقف الحاسب الاختبار ويصدر النتيجة، ويلاحظ هنا أن الاختبار في صورته هذه يختلف تماما عن الصور الورقية، حيث يكون لكل ممتحن صورة اختباريه خاصة به تتفق مع مستواه وتختلف من حيث عدد الأسئلة وصعوبتها، ويعد هذا النوع من الاختبارات أحدث الأنواع؛ وما زال هناك العديد من التحديدات التي تواجه تطبيقه على نطاق واسع، إلاَّ انه يعد أدق أنواع الاختبارات حتى الآن في تقدير قدرة الفرد مستخدمًا نظرية الاستجابة للمفردة **IRT** وتطبيقاتها في هذا السياق، ومن أمثلته اختبار **GMAT**.

❖ الفرق بين الاختبارات الخطية والاختبارات التكيفية :

يكمن الفرق بينهما في عدة نقاط هي على النحو التالي:

1. يسير الطالب في الإجابة في **الاختبارات الخطية** بشكل تتابعي الواحد تلو الأخر أما في **الاختبارات التكيفية** يسير في الإجابة بحسب إجابته عن السؤال السابق.
2. في **الاختبار الخطى** يُطلب من جميع الطلاب الإجابة على كل أسئلة الاختبار بينما في **الاختبار التكيفي** يجيب الطالب فقط على عدد معين من الأسئلة ويختلف عن زميله المطبق عليه الاختبار ذاته.
3. في **الاختبار الخطى** زمن الاختبار موحد لجميع الطلاب ولكن يختلف في حالة **الاختبار التكيفي** على حسب قدرات كل طالب.

4. في **الاختبار الخطى** تقدر درجة الطالب بعدد الأسئلة الصحيحة التي أجاب عنها في حين تقدر في حالة **الاختبار التكيفي** بمستوى صعوبة الأسئلة التي أجاب عليها.
5. تقديم الأسئلة في **الاختبار الخطى** يظل مستمرًا طالما لم يستنفذ الوقت المخصص للإجابة بينما يتوقف تقديم الأسئلة في **الاختبار التكيفي** متى حدد الكمبيوتر المستوى التحصيلي للطالب.

❖ كيف تعمل الاختبارات التكيفية CAT :

تُختار الأسئلة لغرض تحقيق أقصى قدر من الدقة في الامتحان بناءًا على ما هو معروف عن المتعلم ومستواه في الأسئلة السابقة، وصعوبة الاختبار تكمن في تكيف نفسها مع مستوى المتعلمين وقدراتهم بكل دقة، فعلى سبيل المثال: إذا كان الممتحن يؤدى بشكل جيد بمستوى صعوبة متوسط، سيقدم له الاختبار التكيفي أسئلة في مستوى أكثر صعوبة، وإذا كان لم يؤدى في مستوى الصعوبة المتوسط بشكل جيد يقدم له الاختبار أسئلة أكثر بساطة، فمقارنة بالاختبارات الثابتة مثل الاختيار من متعدد والتي تقدم لجميع الطلاب على حد سواء، وبمجموعة ثابتة من البنود والأسئلة والفقرات تقدم لهم جميعًا في نفس الوقت وكأنهم طالب واحد، نجد أن الاختبارات الكمبيوترية التكيفية تتطلب عناصر وفقرات أقل للوصول إلى مستويات معينة ومطلوبة بالتساوي بين جميع الطلاب بدقة **(Weiss, D. J., & Kingsbury, G. G. 1984)** .

وتكون خطوات وخوارزمية الاختبار الكمبيوتري التكيفي **(Thissen, D., & Mislevy, R.J. 2000)** على النحو التالي:

- يتم البحث في مجموعة من المصادر المتاحة لاختيار البند أو الفقرات المناسبة، استنادًا في ذلك على التقديرات الموجودة للمتعلم، وأساليبه وأنماطه في التعلم.

- يتم عرض الفقرة أو البند على المتعلم «الممتحن» ثم يجيب بشكل صحيح أو غير صحيح على العنصر أو السؤال المختار.
- يتم تحديد وتحديث وتخزين وتقييم قدرات المتعلم في النظام للاختبار التكيفي، على أساس كل الإجابات السابقة للمتعلم.
- يتم تكرار الخطوات من 1 – 3 حتى يتحقق المعيار بالنهاية.

فليس هناك معلومات معروفة عن المتعلم لدى نظام الاختبارات التكيفية قبل إجرائه واستجابته على البند أو السؤال الأول، لذلك تبدأ خوارزمية الاختبار التكيفي عن طريق اختيار عنصر من المستوى الذي يلاءم المتعلم في البداية،

ففي الولايات المتحدة الأمريكية، نجد أن اختبارات القبول في الإدارة العليا حاليًا تدار في المقام الأول بنظام الاختبار الحاسوبي التكيفي، فيوجد قائمة للاختبارات التكيفية ببرامج **CAT** النشطة في الرابطة الدولية للحوسبة، جنبًا إلى جنب مع قائمة أخرى من البرامج البحثية الحالية وببليوجرافيا شبه شاملة لجميع الأبحاث المنشورة عن الاختبارات التكيفية.

❖ فوائد ومميزات الاختبارات التكيفية :

تتمتع الاختبارات التكيفية بالعديد من المزايا والفوائد والتي تجعلها بالفعل أكثر فاعلية عن غيرها من الاختبارات ونذكر منها الآتي:

- أكثر مرونة من الاختبارات الخطية والاختبارات الإلكترونية التقليدية.
- تقلص عدد فقرات الاختبار الضرورية للوصول إلى مستوى معين من الدقة في القياس.
- تقلل الوقت المطلوب للاختبار: فيمكن أن توفر الاختبارات التكيفية الوقت المستغرق في الاختبار لجميع الممتحنين، في حين أن الاختبارات الثابتة يكون زمنها موحد للجميع.
- تقلل احتمالية معاناة المتعلم من التعب والملل وتراجع مستوى الدافعية.

- تقدم معلومات أكثر دقة عند نهاية الاختبار أفضل من أي نوع من الاختبارات الأخرى.
- توفر معلومات كافية عند مستويات متوسطة من الصعوبة للقدرات المعرفية والمهارية.
- تتطلب عدد فقرات أقل للحضور على المستوى المطلوب من الدقة والثبات وذلك مقارنة بالاختبارات الخطية.
- تقوم على فكرة تشخيص عملية التقويم.
- تتعامل مع كل متعلم وفقًا لقدراته ومستواه الاستيعابي.
- تقدم أسئلة في جميع المستويات «سهله، متوسطه، صعبه».
- تحقق الهدف من عملية التقويم بسهولة وسرعة ويسر.
- يشعر كل متعلم وكأن الاختبار أُعد خصيصًا لأجله.
- تراعى الفروقات الفردية بين المتعلمين في عملية التقويم.

❖ مزايا الاختبارات التكيفية مقارنة بالاختبارات الخطية :

بالطبع يوجد العديد من المميزات في الاختبارات التكيفية والتي تجعلها تختلف عن الاختبارات الخطية على النحو التالي:

1. نجد أن الاختبارات التكيفية تحتاج إلى زمن اقل في تطبيقها عن الاختبارات الخطية.
2. الاختبارات التكيفية أقل تكلفة في تطبيقها.
3. الاختبارات التكيفية تحدد المستوى التحصيلي بدرجة أكثر دقة من الاختبارات الخطية.
4. الاختبارات التكيفية تزيد من زيادة صدق الاختبار **Test Validity**، إذ يمكنها بسهولة متابعة الطريقة التي يستخدمها المتقدم للاختبار في أداء اختباره، في حين أن الوعاء الورقي يتيح لنا فقط الإطلاع على المنتج الأخير للأداء.

❖ عيوب ومعوقات الاختبارات التكيفية :

كونها أسلوب جديد في التقييم لا يشفع لها في عدم خلوّها من العيوب والمعوقات التي قد تقف أمام تطبيقها في المنظومة التعليمية والتي نذكر بعضًا منها على النحو التالي:

- عدم وجود مهارات تكنولوجية لدى المعلمين أو المتعلمين قادرة على التعامل مع مثل تلك الاختبارات فنحن ما زلنا نجد صعوبة في التعامل مع الاختبارات الخطية والإلكترونية التقليدية.
- ضعف مهارات المعلمين في إعداد الأسئلة على مستويات مختلفة وفقًا لقدرات المتعلمين وأساليبهم وسرعتهم في التعلم.
- عدم توافر البرمجيات التي تقوم عليها الاختبارات الإلكترونية التكيفية.
- عدم اقتناع القائمين على العملية التربوية بها في البداية، يعد من أهم المعوقات، فنحن لا نقبل التغيير بسهولة.
- عدم وجود إمكانيات تجهيزية ومالية وتقنية وبرمجية لدى المؤسسات التعليمية لتطبيق هذا الأسلوب في عملية التقييم.
- كونها تعتمد على الدقة في وضع الأسئلة، وبالتالي تحتاج إلى واضعي أسئلة على كفاءة عالية جدًّا، لكي تقيس مدى تحقيق الأهداف لدى المتعلمين بمستوياتها المختلفة.
- عدم القدرة لدى الكثيرين على تحديد أنماط وأساليب المتعلمين في التعلم، مما يجعل من الصعب التعرف على المستويات التي يتم وضع الأسئلة بها.
- كثرة وتعدد واختلاف أنماط المتعلمين، مما يجعل من الصعب وضع أسئلة اختباريه تناسب هذا الكم من الاختلافات بين المتعلمين.

- هنا قضية تواجهها الاختبارات التكيفية هي المعايير التي على أساسها يتم إعداد الفقرات والبنود والأسئلة، فيجب أن يتم تطبيق جميع بنود الاختبار على عينة كبيرة ثم تحليلها، ولتحقيق ذلك، يتم تقديم بنود متنوعة من حيث المستويات في السهولة والصعوبة والتي يعرف بالاختبار التجريبي، والذي يهدف إلى تحديد مستويات وأساليب وقدرات المتعلمين.
- نجد هناك قضية أخرى مهمة للغاية، وهى أنه يجب تنفيذ هذا النوع من الاختبارات على عينات كبيرة جدًّا في جميع المواد وذلك للوصول إلى معايير الجودة لذلك النوع من الاختبارات والوصول إلى وجود علامة تجارية، وذلك للوصول إلى بنود ومعايير مستقرة وثابتة، للوقوف على فاعلية الاختبارات التكيفية **(Wainer, H., & Mislevy, R. J. 2000)**
- فعلى الرغم من أن الاختبارات التكيفية تقوم على استخدام خوارزميات لمنع الإفراط في استخدام عدد قليل من البنود والأسئلة، إلاَّ أنه قد نجد أنه قد ينتشر ويشيع بين المستخدمين لتلك الاختبارات البنود المستخدمة في الاختبار تحديدا بين الأشخاص من نفس القدرة أو المستوى، مما يجعل ذلك مصدر قلق خطير جدًّا، لأن الأشخاص داخل المجتمع من ذوى القدرات المشتركة يتبادلون كل شيء، وبالتالي يكون الأفضل تقديم اختبارات عشوائية مما يجعلها أكثر أمانًا، ولكن قد تكون أقل كفاءة **(Thissen, D., & Mislevy, R. J. 2000).**

❖ مبررات اعتماد الاختبارات التكيفية في المدارس :

تعتبر تكنولوجيا الكمبيوتر التكيفي تطور حديث نسبيًّا، ويبدو استخدامه مهيأ للنمو بشكل كبير في الولايات المتحدة على مدى السنوات القادمة. فعلى سبيل المثال، مبادرتي التقييم الوطني الكبرى، واتحاد التقييم المتوازن الأذكى والشراكة من أجل تقييم الجاهزية للكلية ووظائف **(PARCC)** تنوي استخدام هذه التكنولوجيا.

وبشكل عام، سوف يتم اعتماد اختبار الحاسوب التكيفي من قبل الدول والمدارس بناء على المبررات التالية:

✓ استخدام تقييمات أكثر دقة وكفاءة على أن يستغرق وقتا أقل للإكمال، ويكون لدى المعلمين والطلاب المزيد من الوقت للتعليم والتعلم.

✓ الاختبارات التكيفية يُصمم فيها كل سؤال وفقا لمعرفة وقدرات كل متعلم، وإزالة الصراع عند الطلاب مع الأسئلة التي الصعبة جدا أو قضاء بعض الوقت مع الأسئلة التي هي سهلة جدا.

✓ الاختبارات التكيفية يمكن أن توفر معلومات أكثر دقة ومتاحة بسرعة عن احتياجات تعلم الطلاب، والتي يمكن استخدامها من قبل المعلمين ليتم تكيف التعليم معهم وتحسين الدعم الأكاديمي للطلاب.

✓ ليس كل المتقدمين للاختبار يمكنهم رؤية نفس البنود، مما يوفر مصداقية أكثر.

❖ مُؤيدي ومُعارضي الاختبارات التكيفية :

فمنذ ظهور اختبارات الكمبيوتر التكيفية دارت مناقشات حول استعمالها، ودرجة الموثوقية فيها، والفوائد من استخدامها، وأوجه القصور، فهي مجرد بداية للخروج. فنجد أن العديد من الدول تخطط حاليًا لاستخدام الاختبارات الحاسوبية الجديدة القائمة على التكيف عبر الإنترنت في السنوات المقبلة، وبالإضافة إلى الفوائد المحتملة المذكورة أعلاه يمكن التطرق إلى ذكر مبررات مؤيدي ومعارض هذا النوع الجديد من الاختبارات على النحو التالي:

❖ مبررات مؤيدين الاختبارات الحاسوبية التكيفية :

وفيما يلي عدد قليل من المبررات المنطقية التي قد يقدمها دعاة ومؤيدي استخدام اختبارات الحاسوب التكيفية:

- ✓ الاختبارات التكيفية يمكن أن تساعد على تحديد مستوى تعلم الطالب بصورة أكثر دقة من الامتحانات الثابتة، وخاصة للطلاب في النهايات الدنيا والعليا من الصف التعليمي.
- ✓ الاختبارات التكيفية تعطي المعلمين معلومات أكثر دقة حول الطلاب الذين هم بارعون للغاية أو بشكل استثنائي وما يقف وراء ذلك بكثير في إتقانهم للمعارف والمهارات المتوقعة.
- ✓ الاختبارات التكيفية قد تزيد مشاركة الطلاب في عملية الاختبار نفسها، وربما يؤدي ذلك إلى نتائج أكثر دقة بسبب الاختبارات القصيرة من حيث الزمن، والأقل من حيث التعب والجهد، والأفضل من حيث المعايير التي تحدد القدرات الفردية للطالب.
- ✓ تعتبر سجل محوسب من الأسئلة المفتوحة «بنك الأسئلة»، وسوف تصبح أكثر موثوقية، ويمكن أن تزيد من الكفاءة وخفض التكاليف على نطاق واسع بدلًا من الاختبار الموحد.

❖ مبررات معارضي الاختبارات الحاسوبية التكيفية :

وفيما يلي عدد قليل من المبررات التي يمكن أن تُقدم من قِبل من ينتقد استخدام الاختبارات التكيفية:

- ✓ تحتاج التكنولوجيا المتطورة اللازمة لتسجل أسئلة مفتوحة العضوية والفروع إلى وقت في إعدادها حيث أن اختبارات الكمبيوتر التكيفية ليست جاهزة بعد للاستخدام على نطاق واسع في المدارس. فبعض نظم الاختبارات التكيفية ربما لم يتم اختباره بما فيه الكفاية، والبعض الآخر قد يكون عرضه للأخطاء والمشاكل والتي يمكن أن تؤدي إلى نتائج غير دقيقة.
- ✓ استخدام الاختبارات المحوسبة عمومًا والتكيفية خصوصًا يحتاج إلى محو الأمية التكنولوجية الدنيا لدى الطلاب مما يمثل عيبًا فيها وكذلك الطلاب

الأقل قدرة على الوصول إلى التكنولوجيا الرقمية، مثل طلاب من الأسر ذات الدخل المنخفض والطلاب في المناطق الريفية فيصعب في تلك الأماكن الدخول إلى الإنترنت.

- ✓ الانتقال من اختبارات الورقة والقلم لاختبارات الكمبيوتر التكيفية يحتاج في كثير من الحالات إلى تحديات لوجيستية كبيرة وأعباء مالية، لاسيما بالنسبة للدول التي تعاني ضائقة مالية ومشاكل اقتصادية وميزانيات تعليمية منخفضة، والمدارس العامة. لذلك قد يلجأ البعض إلى الشراء لبرمجيات هذا النوع من الاختبارات وبالطبع ستكون باهظة الثمن.
- ✓ بالنسبة للمدارس نجد بها عدد قليل من أجهزة الكمبيوتر وعدم الاتصال بشبكة الإنترنت الدولية أو عدم وجود شبكة داخلية بها، مما قد يكون من الصعب تخصيص الوقت وأجهزة الكمبيوتر اللازمة لجميع الطلاب لإكمال الاختبارات.
- ✓ تتطلب اختبارات الحاسوب التكيفية عادة الدعم الفني القوي لأنظمتها لمنع أي كسر أو خلل يمكن أن يعرقل إدارة الاختبار ويعرقل بشكل كبير خطة عمل المدرسة والعملية التعليمية.
- ✓ بالنسبة للمناطق والمدارس التي لا تزال تعتمد على العمليات الورقية والانتقال إلى شبكة الإنترنت، قد يكون اختبار الحاسوب التكيفي مرهق أو أن استخدامه يكون في حكم المستحيل لأن المدرسة قد لا تملك الموارد الكافية، والأجهزة اللازمة، وعلى الموظفين إجراء التدريبات التكنولوجية للتعامل مع مثل تلك التقنيات والاختبارات الجديدة.

❖ مكونات الاختبارات التكيفية :

هناك خمسة عناصر لتقنية الاختبارات التكيفية **CAT** عند بناءها، فهى لا تتضمن المسائل العملية مثل التطبيق المبدئي للاختبار (**Weiss, D. J., & Kingsbury, G. G. 1984**)، وهى على النحو التالي:

- **تجميع البنود وفقًا للمعايير :**

يجب أن يكون هناك مجموعة من المصادر والأدوات المتاحة لاختيار البنود والأسئلة من بينها، ولا تقتصر هذه الخطوة على التجميع فقط، بل يجب أن تكون هناك معايير يتم التجميع على أساسها، والتي ترتبط بأساليب تعلم المتعلمين، ومراعاة البُعد النفسي ومستوى قدرات المتعلمين، لكي يكون المتعلم والبند على نفس المقاس.

- **نقطة البداية أو مستوى الدخول :**

فيتم اختيار البنود والأسئلة بعد ذلك على أساس أداء الممتحن لكي يصل إلى نقطة معينة في الاختبار، ويتحدد فيها التقديرات الأولية لقدرة الممتحن ويعتبر ذلك ضروريًا في الاختبارات التكيفية، والوقوف أيضًا على معرفة الممتحن السابقة ووضعها في الاعتبار في الأسئلة والبنود القادمة، لذلك في الكثير من الأحيان يقدم البند أو السؤال الأول في مستوى الطالب المتوسط الصعوبة، وبناءًا عليه تتحدد نقطة الانطلاق والدخول للبنود التالية.

- **اختيار خوارزمية البند :**

فكما ذكرنا سابقًا، عن كون البند يجب أن يكون مع المتعلم على نفس المقاس، لذلك فإن نظام الاختبار التكيفي سيكون لديه تقدير لقدرة ومستوى الممتحن، وأنه قادر على تحديد العناصر والأسئلة الأنسب لهذا المستوى، ومن الناحية الفنية، يتم ذلك عن طريق اختيار البند وفقا للمستوى الذي تم تحديده في نقطة الانطلاق.

- **سجل إجراء الاختبار :**

فيقوم نظام الاختبارات التكيفية بتعديل وتحديث تقديره لمستوى قدرة الممتحن، إذا أجاب على هذا البند بشكل صحيح، فبالتالي يكون البند القادم بمستوى أعلى إلى حد ما، والعكس بالعكس.

- **نهاية الاختبار :**

يتم تصميم نظم الاختبارات التكيفية وخوارزمياتها لإدارة وتنفيذ الاختبارات وتحديث تقدير مستوى وقدرة الممتحن، ففي كثير من الأحيان ينتهي الاختبار عندما يقع الممتحن في خطأ يكون فيه القياس المعياري المُحدد بقيمة معينة أقل من القيمة المحددة من قبل المستخدم أو الممتحن، ويوجد هناك معايير أخرى مختلفة لغرض إنهاء الاختبار، تختلف وفقا لطبيعة لكل ممتحن، وطبيعة الاختبار نفسه.

❖ دراسات سابقة حول الاختبارات التكيفية :

- دراسة أحمد عودة وعمر عبيدات 2013م :

وهدفت هذه الدراسة إلى فحص فاعلية الاختبار التكيفي المحوسب في دقة تقدير القدرة العقلية باستخدام مصفوفات رافن باختلاف طرق تقدير القدرة (طريقة الأرجحية العظمى **MLE**، طريقة التقدير البعدي الأعظم **MAP**)، وقواعد إنهاء الاختبار (قاعدة الإنهاء: عدد محدد من الفقرات، قاعدة الإنهاء: أدنى خطأ معياري)، ولتحقيق أهداف الدراسة تم تكوين بنك أسئلة مكون من ١٠٥ من فقرة من تلك المصفوفات. وإجراء خمسة تطبيقات محوسبة، بواقع اختبارين لكل تطبيق، على عينات بلغ عددها ٦٣٨ طالبًا وطالبة. وأظهرت النتائج أن قاعدة إنهاء الاختبار بعدد محدد من الفقرات توفر تقديرات للقدرة أدق، ودالة معلومات أعلى من قاعدة أدنى خطأ معياري باختلاف طريقتي تقدير القدرة، كما توفر قاعدة أدنى خطأ معياري في عدد الفقرات المطبقة بنسبة تصل إلى 50٪ من قاعدة عدد محدد من الفقرات. كما أن الاختبار التكيفي يوفر تقديرات للقدرة أدق، ويوفر في عدد الفقرات المطبقة بنسبة تصل إلى ٧٠ ٪، وله دالة معلومات أعلى من الاختبار الخطي باختلاف طريقتي تقدير القدرة. وتعطي كل من طريقتي تقدير القدرة **(MLE, MAP)** تقديرات قدرة متساوية ومؤشرات دقة متساوية ولطريقة **MLE** دالة معلومات للاختبار أعلى من **MAP** ومن ابرز التوصيات استخدام الاختبارات التكيفية المحوسبة في الاختبارات العامة كاختبارات القبول واختبارات القدرات العقلية لفاعليتها من حيث الدقة، وفعاليتها الاقتصادية في زمن التطبيق، وتقليل الضغوطات على المفحوصين. واستخدام طريقة الأرجحية العظمى في تقدير القدرة بالاختبارات التكيفية لأن المؤشرات تعطي أفضلية لهذه الطريقة مقارنة بطريقة البعدي الأعظم. وإجراء دراسات مقارنة بين طريقتي التقدير **MLE, MAP** عند

اختلاف عدد الفقرات الأصلية في بنك الأسئلة، وعند اختلاف قيم أدنى خطأ معياري غير التي استخدمت في الدراسة.

- دراسة أمين محمد صبري نور الدين 2002م :

والتي جاءت تحت عنوان «فاعلية استخدام الاختبار الموائم باستخدام الحاسب في تقدير قدرة الأفراد وتحديد الخصائص السيكومترية للمقياس».

لقد كان للقياس النفسي تأثيره الهام والحيوي في المجتمعات المتحضرة في المجالات التربوية والتقويمية والعلاجية. ويهدف القياس النفسي -فيما يهدف إليه- إلى تطوير الأدوات والمقاييس وتحسين عملية القياس وصولاً إلى دقة أفضل للتقويم.

ولقد تعرضت مفاهيم القياس النفسي في الآونة الأخيرة لطفرة هائلة. وساعد ظهور نظريات القياس المتقدمة مثل "نظرية الاستجابة للمفردة **Item Response Theory IRT)** في الخمسينيات من القرن الماضي على تطوير أساليب القياس النفسي من كافة جوانبه. حيث قدمت النظرية حلولاً مناسبة لمعالجة وتحليل المفردات، ومعادلة الدرجات، ووصل الاختبارات بعضها البعض، وإنشاء بنوك الأسئلة. مما دعا إلى التوسع في تطبيقات عديدة للقياس النفسي.

كما أسهم تطور صناعة الحاسبات والبرمجيات في الفترة الأخيرة في إمكانية تطبيق الاختبارات على الحاسب بصورة تواؤمية غير خطية، وتصحيحها، وتحليلها بالاستناد إلى مفاهيم نظرية الاستجابة للمفردة. وقد أدى ذلك إلى ظهور مفهوم «الاختبار التواؤمي المحوسب» الذي يعد أهم وأحدث تطبيقات نظرية الاستجابة للمفردة في مجال القياس النفسي الحديث.

ففي الاختبار التواؤمي المحوسب يتفاوت عدد وترتيب مفردات الاختبار التي يتلقاها كل مفحوص عن الآخر؛ وفقا لتقدير قدرة كل منهم. ونظرًا لأن الحاسب يعيد تقدير قدرة الفرد بعد كل مفردة، وينتقي له المفردات الأكثر ملائمة؛ فإنه من

الممكن الوصول إلى مستويات عالية من الدقة لتقدير القدرة باستخدام عدد محدود من مفردات الاختبار.

وقد أجريت مئات البحوث حول الاختبارات التواؤمية المحوسبة من مختلف جوانبها. والنتيجة العامة لتلك الدراسات هي أن للاختبار التواؤمي المحوسب معاملات ثبات وصدق مكافئة لمعاملات ثبات وصدق الاختبارات التقليدية، أو أعلى منها حتى مع خفض طول الاختبارات إلى حوالي 50٪، بالرغم أنها تقيس نفس المتغيرات. وتراوح انخفاض زمن الاختبار من 25٪ إلى 75٪ من زمن الاختبار الأصلي دون فقدان دقة القياس. كما تم الوصول إلى تقديرات ثبات بلغت 0.80 عن طريق تطبيق 9 مفردات فقط.

وهناك عدد من المؤسسات والهيئات في الولايات المتحدة الأمريكية والدول الغربية اعتمدت الاختبار التواؤمي المحوسب في أنظمتها مثل مؤسسة الدفاع الأمريكية التي تبنت إصدار بطارية الاستعداد المهني للقوات المسلحة. وكذلك مؤسسة الاختبارات التربوية التي حولت عددًا من الاختبارات الشهيرة إلى اختبارات تواؤمية محوسبة مثل «اختبار اللغة الإنجليزية كلغة أجنبية» **(TOEFL)** واختبارات **(GRE)** وغيرها من الاختبارات. كما تبنت شركة مايكروسوفت الشهيرة في نظم الحاسبات والبرمجيات الاختبارات التواؤمية المحوسبة في اختباراتها لنظم التشغيل وبناء البرامج والشبكات.

وبالرغم من هذا التطور الهائل الذي طرأ على القياس النفسي في الغرب؛ ما زال القياس في بيئتنا العربية يخطو خطوات متواضعة في ذلك المجال. لذا بات من الأهمية بمكان أن نلحق بركب حركة القياس المعاصرة، وأن نسير بصورة موازية معه إن لم نسبقه. فتطوير حركة القياس النفسي في البيئة العربية لا تكون فقط باستيراد تكنولوجيا جديدة للقياس. أو مجرد الانقياد لصيحة معاصرة في العلم؛ وإنما يكون بالاستعداد لتقبلها داخل السياق الثقافي الملتحمة معه، والاستيعاب

الكامل لأهدافها، والإتقان التام لمكوناتها والنظرية التي تعمل وراءها. من ثم تتلخص مشكلة هذه الدراسة في الأسئلة التالية:

1- ما مدى فاعلية الاختبار التواؤمي المحوسب باستخدام قاعدة إنهاء الاختبار «بأدنى خطأ معياري» مقارنة بالاختبار الخطي المحوسب في دقة تقدير القدرة كما تتمثل في مؤشرات الدقة المختلفة؟

2- ما مدى فاعلية الاختبار التواؤمي المحوسب باستخدام قاعدة إنهاء الاختبار «بعدد محدد من المفردات» مقارنة بالاختبار الخطي المحوسب في دقة تقدير القدرة كما تتمثل في مؤشرات الدقة المختلفة؟

3- ما مدى فاعلية الاختبار التواؤمي المحوسب باستخدام أسلوب تقدير الأرجحية القصوى مقارنة بالاختبار التواؤمي المحوسب بأسلوب تقدير البعدي الأقصى كما تتمثل في مؤشرات الدقة المختلفة؟

عينة الدراسة: تألفت عينة الدراسة لتقدير مَعْلَمات مفردات الاختبار من أكثر من (5000) طالب وطالبة من الفرقة الثانية والثالثة والرابعة وطلاب الدراسات العليا بكلية التربية بجامعة عين شمس. بينما تجاوزت عينة دراسة تطبيق الاختبارات التواؤمية المحوسبة أكثر من (500) طالب وطالبة آخرين من مختلف التخصصات.

أدوات الدراسة: استخدم الباحث (137) مفردة من اختبار رافن للمصفوفات المتدرجة بصوره المختلفة الملونة، والعادية، والمتقدمة، في اشتقاق المَعْلَمات المختلفة للمفردات، ثم إعدادهم للتطبيق بصورة تواؤمية محوسبة بناء على المَعْلَمات التي تم حسابها من قبل. وقد تم تقسيم المفردات بصورة مبدئية إلى اختبارين يطبق كل منهما بصورة منفصلة.

إجراءات الدراسة: قام الباحث بتطبيق الاختبار الأول للمصفوفات على مجموعة من (2902) طالب وطالبة في الفصل الدراسي الأول من العام الجامعي

2001/2000. ثم طبق الاختبار الثاني على (2147) طالب وطالبة في الفصل الدراسي الثاني من نفس العام الجامعي. وكان الطلاب من الفرق الثانية والثالثة والرابعة والدبلوم العام بكلية التربية. وقد تم التطبيق في معامل علم النفس بالإجراءات المقررة.

ثم قام الباحث بإجراءات تحليل البيانات من خلال التأكد من أحادية البعد للاختبارين الأول والثاني. ثم اختار نموذج الاستجابة للمفردة الأكثر مطابقة لبيانات الاختبار وهو النموذج ثلاثي المَعْلَم بعد حذف سبع مفردات غير ملائمة من الاختبار الثاني. وبعد ذلك قام الباحث بوصل الاختبارين معًا ليصيرا اختبارًا واحدًا من خلال مجموعة من المفردات المشتركة بينهما. وبناء على ذلك؛ تم اشتقاق مَعْلَمات نهائية ذات قيم مناسبة لمفردات الاختبار البالغ عددها (131) مفردة.

وبعد إنشاء مستودع المفردات للاختبار التواؤمي المحوسب من المفردات السابقة؛ أجرى الباحث تطبيقًا للاختبارات التواؤمية المحوسبة من خلال ثلاثة دراسات مختلفة. قارنت الدراسة الأولى بين فعالية الاختبار التواؤمي المحوسب الذي يتوقف بعد درجة محددة من الدقة مقارنة بالاختبار الخطي المحوسب. والدراسة الثانية قارنت بين فعالية الاختبار التواؤمي المحوسب الذي يتوقف بعد عدد محدد من المفردات مقارنة بالاختبار الخطي المحوسب. أمَّا الدراسة الثالثة فقد قارنت بين اختبارين تواءمين محوسبين يستخدم أحدهما تقدير القدرة بالأرجحية القصوى والثاني بالبعدي الأقصى. وكل منهما ينتهي بعد تطبيق عدد محدد من المفردات. وقد طبق على كل طالب اختباران. وكان التطبيق داخل معمل الحاسب بقسم علم النفس بكلية التربية.

نتائج الدراسة: أسفرت النتائج أن الاختبار التواؤمي المحوسب أكثر فعالية من الاختبار الخطي المحوسب. فالاختبار التواؤمي يعطي دقة أعلى بدلالة من الخطي لأنه ينتقي المفردات المناسبة لقدرة الفرد فقط. كما أنه أقصر بصورة دالة حيث أنه تمكن من الوصول إلى دقة أعلى من خلال تطبيق (25) مفردة في المتوسط بينما

كان عدد مفردات الخطي (60) مفردة. بل إنه بلغ دقة أعلى مع تطبيق (13) مفردة فقط، ووصل إلى معامل ثبات تجاوز (0.80) من خلال تطبيق (5) مفردات فقط. كما أمكن للاختبار التواؤمي المحوسب من خلال تطبيق (50٪) من عدد المفردات في الاختبار الخطي الوصول إلى معامل ثبات (0.94)

كما أسفرت النتائج أيضا أن أسلوب تقدير القدرة البعدي الأقصى أكثر دقة من أسلوب الأرجحية القصوى لتقدير القدرة حيث أنه يقدم معلومات أكثر وله من الخطأ المعياري أقل بدلالة من الأرجحية القصوى. غير أن الأرجحية القصوى أعطى متوسطا لتقدير القدرة أعلى بدلالة من البعدي الأقصى.

وأثبتت الدراسة أن الاختبار التواؤمي المحوسب يمتاز في أنه:

- يوفر في عدد مفردات الاختبار بنسبة تزيد عن (50٪) في المتوسط عن الاختبار الخطي.
- يزيد من دقة القياس بنسبة يصل متوسطها إلى (20٪) عن الاختبار الخطي.
- يعطي مقدارًا من المعلومات أعلى بنسبة تتجاوز (32٪) في المتوسط عن الاختبار الخطي.
- يعطي تقديرًا للقدرة لا يختلف متوسطه عن متوسط تقدير القدرة للاختبار الخطي.

مراجع الفصل الخامس

❖ مراجع عربية :

1- **أحمد سليمان عودة وعمر سليمان عبيدات (2013م).** فاعلية الاختبار التكيفي المحوسب في تقدير القدرة العقلية باستخدام مصفوفات رافن، مجلة العلوم التربوية، المجلد40، العدد 2.

2- **أمين محمد صبري نور الدين (2002م).** فاعلية استخدام الاختبار الموائم باستخدام الحاسب في تقدير قدرة الأفراد وتحديد الخصائص السيكومترية للمقياس. رسالة دكتوراه، كلية التربية، جامعة عين شمس.

❖ مراجع أجنبية :

1- **Alkhader, 0-, Clark, D- and 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.

2- **Chen, S-, Hou, L- and Dodd, B- 1998.** A Comparison of maximum likelihood estimation expected a posterior estimation in CAT using the partial credit model. Educational and Psychological Measurement, 58, 569-595.

3- **Cisar, D-, Radosav, D-, Markoski, B-, Pinter, R and Cisar, P- 2010.** Computer adaptive testing of Student knowledge- Acta Polytechnic Hungarica, 7(4).

4- **Clarke, D- Mackinnon, F. Mckenzie, F- and Herrman, H. 2000.** Dimensions of psychopathology in the medically ill: a latent trait analysis. Psychosomatic; 41(5), 420.

5- **Dodd, H, Koch, W- and De Ayala, R 1989.** Operational characteristics of adaptive testing procedures using the graded response model, Applied Psychological Measurement, 13, 129-143.

6- **Embretson, S- and Reiase, S- 2000.** Item Response Theory for Psychologists- Mahwah, New Jersey: Lawrence Erlbaum Associates, Publishers.

7- **Fliege, H Becker, 1, Walter, 0-, Bjorner, J-, Klapp, B- and Rose, M- 2005.** Development of a computer-adaptive test for depression (D-CAT)- Quality of Life Research, 14, 2277-2291.

8- **Huo, Y- 2009.** Variable-length Computerized Adaptive Testing: adaptation of the a-sn'atified strategy in item selection with content balancing. Dissertation of doctor of philosophy in psychology in the graduate college of the University of Illinois at Urbana-Champaign.

9- **Jacobusse, G- and Buuren, S. 2007.** Computerized adaptive testing for measuring development of young children- Statistics in Medicine, 26.

10- **Jain-quail, T-, Dan-min, M Xia, Z. and Jing-jing, G- 2007.** An Introduction to the computerized adaptive testing. US-china Education Review, 4(1): 72-81.

11- **Kingsbuiy, G. and Zara, A. 1989.** Procedures for selecting item for computerized adaptive testing. Applied Psychological Measurement, 2.

12- **Linden, W. and Pushley. P. 2003.** Item Selection and ability estimation in adaptive testing. In Linden, W. and Glass, C. (eds). Computerized Adaptive T esting Theory and Practice. Kluwer Academic Publishers.

13- **Magis, D. and Raiche, G. 2011.** CatR: An R package for computerized adaptive testing. Applied Psychological Measurement, 33(7): 576-577.

14- **McKinley, R and Reckase, M. 1981.** A Comparison of a Bayesian and a maximum likelihood tailored testing procedure. Office of Naval Research. Arlington, VA-Personnel and Training Research Programs Office.

15- **Milhnan, J- and Arter, J- 1984.** Issues in item banking. Journal of Educational Measurement. 21 , 3 15-330.

16- **Murphy, K. and Davidshofer, C. O. 1994.** Psychological testing: Principles and applications. 3". New Jersey: Prentice-Hall.

17- **Roex, A and Degryse, J. 2004**- A Computerized adaptive knowledge test as an assessment tool in general practice: a pilot study, Medical Teacher, 26 (2).

18- **Rosso, M- and Reckase, M 1981.** A comparison of a maximum likelihood and a Bayesian ability estimation procedure for tailored testing. Paper presented at the annual meeting of the national council on measurement in education, Missouri Univ, Columbia.

19- **Stone, E. and Davey, T. 2011.** Computer-Adaptive Testing for Students with Disabilities: A Review of the Literature. ETS, Princeton, New Jersey, Retrieved at online: http:// WWW.ets. org/ research/ contact. html.

20- **Vispoel. W- 1988.** An adaptive test of musical memory: An application of item response theory to the assessment of musical ability. Dissertation Abstracts, DAI-A 49101, P 79.

21- **Vispoel, W. 1993.** Computerized adaptive and fixed-item testing versions of the ITED vocabulary subtest. Educational and Psychological Measurement, 53, 779-788.

22- **Vispoel, W., Wang, T. and 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.

23- **Wang, S. and Wang, T- 2001.** Precision of weighted likelihood estimates for a polytomous model in computerized adaptive testing, Applied Psvchological Measurement, 25(4): 317-331.

24- **Wang, S. and Wang, T. 2002.** Relative precision of ability estimation in polytomous CA T: A Comparison under the generalized partial credit model and graded response model. ED 477926, shudon wang, 19500 Bulverde Road, San Antonio, TX 78259-3701.

25- **Wang, T. 1995.** The precision of ability estimation methods in computerized adaptive testing (item response). Dissertation Abstracts, DAI-A 56/ 06, P2212.

26- **Wang, T. and Vispoel, W. 1998.** Properties of ability estimation methods and computerized adaptive testing, Journal of Educational Measurement, 35 (2): 109- 135.

27- **Ward, W. 1984.** Using microcomputers to administer measurement. Issues and Practices, 3, 16-20.

28- **Weiss, D. 2004.** Computerized adaptive testing for effective and efficient measurement in counseling and education- Measurement and Evaluation in Counseling and Development, 37, 70-84.

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

30- **Thissen, D., & Mislevy, R.J. (2000).** Testing Algorithms. In Wainer, H. (Ed.) Computerized Adaptive Testing: A Primer. Mahwah, NJ: Lawrence Erlbaum Associates.

31- **Green, B.F. (2000).** System design and operation. In Wainer, H. (Ed.) Computerized Adaptive Testing: A Primer. Mahwah, NJ: Lawrence Erlbaum Associates.

32- **Wainer, H., & Mislevy, R.J. (2000).** Item response theory, calibration, and estimation. In Wainer, H. (Ed.) Computerized Adaptive Testing: A Primer. Mahwah, NJ: Lawrence Erlbaum Associates.

33- **Lin, C.-J. & Spray, J.A. (2000).** Effects of item-selection criteria on classification testing with the sequential probability ratio test. (Research Report 2000-8). Iowa City, IA: ACT, Inc.

34- **Wald, A. (1947).** Sequential analysis. New York: Wiley.

35- **Reckase, M. D. (1983).** A procedure for decision making using tailored testing. In D. J. Weiss (Ed.), New horizons in testing: Latent trait theory and computerized adaptive testing (pp. 237-254). New York: Academic Press.

36- **Weitzman, R. A. (1982).** Sequential testing for selection. Applied Psychological Measurement, 6, 337-351.

37- **Kingsbury, G.G., & Weiss, D.J. (1983).** A comparison of IRT-based adaptive mastery testing and a sequential mastery testing procedure. In D. J. Weiss (Ed.), New horizons in testing: Latent trait theory and computerized adaptive testing (pp. 237-254). New York: Academic Press.

38- **Eggen, T. J. H. M, & Straetmans, G. J. J. M. (2000).** Computerized adaptive testing for classifying examinees into three categories. Educational and Psychological Measurement, 60, 713-734.

39- **Spray, J. A., & Reckase, M. D. (1994).** The selection of test items for decision making with a computerized adaptive test. Paper presented at the Annual Meeting of the National Council for Measurement in Education (New Orleans, LA, April 5-7, 1994).

40- **Sympson, B.J., & Hetter, R.D. (1985).** Controlling item-exposure rates in computerized adaptive testing. Paper presented at the annual conference of the Military Testing Association, San Diego.

41- **For example: van der Linden, W. J., & Veldkamp, B. P. (2004).** Constraining item exposure in computerized adaptive testing with shadow tests. Journal of Educational and Behavioral Statistics, 29, 273-291.

❖ مواقع إلكترونية :

- https://www.facebook.com/Agiteq/posts/1392602751054942
- http://www.eulc.edu.eg/eulc_v5/Libraries/Thesis/BrowseThesisPages.aspx?fn=ThesisPicBody&BibID=9628676&TotalNoOfRecord=315&PageNo=7&PageDirection=Next
- http://www.al-jazirah.com.sa/digimag/02042006/maaa29.htm

- http://edglossary.org/computer-adaptive-test/
- https://en.wikipedia.org/wiki/Computerized_adaptive_testing#cite_note- ThissenMislevy-2