الاختبار التكيفي المحوسب Computerized Adaptive Testing (CAT) "كتطبيق لتطوير أساليب التقويم والاختبارات الالكترونية"

ا.د.علي بن أحمد الصبيحي
أستاذ، مناهج البحث والإحصاء التطبيقي
كلية التربية بجامعة طيبة

د.شرف بنت حامد الأحمدي
أستاذ مساعد، القياس والتقويم
كلية التربية بجامعة طيبة

مجلة كلية التربية . جامعة دمنهور **المجلد الثالث العدد (١) لسنة ٢٠١١**

مستخلص

لقد تأثر تطور ونمو الاختبارات التربوية والنفسية وتطبيقاتها إيجاباً بثورة تقنية الحاسوب وصناعة البرمجيات، حيث ساعدت في إيجاد نظريات قياس متقدمة، وكان من ثمارها أيضا أن ظهرت أساليب جديدة للاختبار المعتمد على الحاسب احتلت مكان اختبارات الورقة والقلم، ومنها الاختبار التكيفي المحوسب، والذي يختلف في عدد أسئلته وصعوبتها باختلاف القدرة العلمية لِلْمُخْتَبِر، أي أنه اختبار مُفَصَّل **tailored** لكل فرد على حدة؛ بحيث لا يتعرض للمفردات السهلة جداً، ولا الصعبة جداً بالنسبة للفرد، بل تقدم له المفردات فقط التي تتناسب مع مستوى قدرته.

لذا كان الهدف الرئيس لهذه الورقة هو توضيح كيفية توظيف التقنية لتطوير أساليب التقويم في التعليم الجامعي ممثلة في الاختبار التكيفي المحوسب؛ من خلال: عرض الخلفية النظرية للاختبار التكيفي المحوسب، وتحديد بعض المعايير المهمة عند بنائه، ومن ثم عرض أهم خصائص تطبيقه مقارنة بالاختبارات المعتمدة على الحاسب سواء الخطية أو اختبارات الورقة والقلم، وبينت الورقة أهم التأثيرات الإيجابية لتطبيق الاختبار التكيفي المحوسب على المناهج التعليمية، وعرضت الورقة أيضاً بعض التوجهات الحديثة في هذا المجال، كما تم التطرق إلى العوامل التي أخرت اعتماد الاختبار التكيفي المحوسب في مؤسساتنا التعليمية حتى الآن، وأخيراً خرجت الورقة بأهم التوصيات المتعلقة بالموضوع؛ حيث أنه من المؤمل أن تقدم هذه الورقة فكرة عامة عن ميدان لتطبيق التقنية في تطوير أساليب التقويم والاختبارات الالكترونية، وتمهد لتبني برامج مستقبلية طموحة تهدف لتطوير منظومة التقويم في المرحلة الجامعية.

Abstract

Computerized Adaptive Testing : An assessment Method Development and E-Testing Application

The computer revolution has positively affected the development and application growth of educational and psychological testing. One of these outcomes is Computerized Adaptive Testing (CAT). CAT is an individually tailored test that designed in number and difficulty to match the person's ability. The main objective of this paper is to illustrate the role of the evolved semantic in CAT to develop academic achievement tests at the university level. To achieve this goal, the theoretical background of CAT is explored, and some important factors are considered in the construction of this type of test. Moreover, some distinctive characteristics of CAT are compared to other linear or paper and pencil tests. The paper presents positive effects in applying CATs on the curricula, and highlights its trends. It also discusses reasons of why CATs have not been implemented in our educational establishments. Lastly, the paper suggests several recommendations for implementation of CATs in universities.

الاختبار التكيفي المحوسب Computerized Adaptive Testing (CAT)
"كتطبيق لتطوير أساليب التقويم والاختبارات الالكترونية"

ا.د.علي بن أحمد الصبيحي د.شرف بنت حامد الأحمدي

مقدمة:

أدت التطورات المعاصرة في تقنيات الحاسوب وبرمجياته لتطوير الأساليب والممارسات التعليمية ومراجعتها بهدف الاستفادة من التقنية المتاحة لتحقيق أهداف التعليم وجعله أكثر كفاءة وفاعلية، ولقد تأثر تطور ونمو الاختبارات التربوية والنفسية وتطبيقاتها إيجاباً بالثورة في تقنية الحاسوب وصناعة البرمجيات، حيث ساعدت في إيجاد نظريات قياس متقدمة مثل نظرية الاستجابة للمفردة **Item Response Theory (IRT)**، وكان من ثمارها أيضا أن ظهرت أساليب جديدة للاختبار المعتمد على الحاسب احتلت مكان اختبارات الورقة والقلم **Paper-and-Pencil**، ومنها الاختبار التكيفي المحوسب**(Computerized Adaptive Testing (CAT**، والذي يختلف في عدد أسئلته وصعوبتها باختلاف القدرة العلمية للمُخْتَبِر، أي أنه اختبار مُفَصَّل **Tailored** لكل فرد على حدة؛ بحيث لا يتعرض للمفردات السهلة جداً، ولا الصعبة جداً بالنسبة للفرد، بل تقدم له المفردات فقط التي تتناسب مع مستوى قدرته. وكان لا بد من وجود نظرية قياس حديثة يتم الاستناد إليها في تقدير قدرة الفرد وانتقاء المفردة المناسبة له.

وكان من أبرز هذه النظريات الحديثة نظرية الاستجابة للمفردة، حيث تعد نظرية الاستجابة للمفردة **IRT** من أهم التطورات الحاسمة في القياس. حيث أنها تحسب احتمالية الإجابة الصحيحة لكل مفردة على حدة، من خلال نمذجة العلاقة بين المتغيرات غير الملاحظة (قدرة الأفراد) واحتمال الاستجابة الصحيحة للفرد على أي مفردة اختبارية، كما أنها تعد نظرية إحصائية عامة تهتم بدراسة المفردات الاختبارية، والأداء على الاختبار، وعلاقة كل منها بالقدرة المقاسة بذلك الاختبار**(Baker and Kim, 2004)**. لأنه في إطار نظرية الاستجابة للمفردة **IRT**،

خطأ القياس يختلف باختلاف مستوى القدرة وليس واحداً لجميع الأفراد، والنموذج لا يحتاج إلى اختبارات متوازية لتقييم الثبات، كما أنه يقيس مستوى دقة كل درجة من درجات القدرة على حدة عن طريق دالة المعلومات للمفردة **Information Function**، ومن خلال هذه الميزة يمكن تصميم اختبارات تعطي أفضل معلومات عند مستويات معينة من القدرة حسب الحاجة؛ حيث أن نظرية الاستجابة للمفردة **IRT** نظرية تهتم بسلامة ودقة المفردة أكثر من كونها نظرية تولي اهتمامها بالشكل العام للاختبارات وفقط ، ونتيجة للمميزات السابقة، قدمت نظرية الاستجابة للمفردة **IRT** حلولا مثالية لكثير من مشاكل الاختبارات، مثل: تعميم الاختبار، والتعرف على المفردات المتحيزة، ومعادلة درجات الاختبار، والاختبارات التكيفية المحوسبة **(Lawson, 2006; Thomas, 1997)**.

وقد حدث تحول في الاختبارات التكيفية المحوسبة سواء في الجانب النظري أو التطبيقي والتنفيذ الفعلي في البلدان المتقدمة منذ أوائل التسعينيات من القرن الماضي، وقد وصف علام (٢٠٠٥، ص ٣٢٤) هذا التحول بأنه جوهري ، وأضاف أن هناك عدداً من المؤسسات والهيئات العالمية اعتمدت أنظمة الاختبار التكيفي المحوسب **CAT System** وقامت بتحويل اختباراتها الخطية إلى تكيفية، مثل مؤسسة وزارة الدفاع الأمريكية التي تبنت إصدار بطارية الاستعداد المهني للقوات المسلحة الأمريكية**Armed Services Vocational Aptitude Battery(ASVAB)** وجعلتها بطارية تكيفية محوسبة. وكذلك مركز الاختبارات التربوية **Educational Test Services (ETS)** التي حولت عدداً من الاختبارات الشهيرة إلى اختبارات تكيفية محوسبة مثل: اختبار اللغة الإنجليزية كلغة أجنبية"التوفل" **Test of English as a Foreign Language(TOEFL)**، واختبارات القبول للدراسات العليا **Graduate Record Examination(GRE)**، واختبار الاستعداد المدرسي **(SAT) Scholastic Assessment tests** ، واختبارات الإجازة للتمريض **The Licensure Exam of National Council of State Board of Nursing(NCLEX)** ، وهذه الورقة تركز تحديدا على

موضوع الاختبارات التكيفية المحوسبة لتقدم فكرة عامة عن ميدان تطبيق التقنية والحاسوب في تطوير أساليب التقويم والاختبارات الالكترونية، ، ولعل في ذلك تمهيدا لتبني برامج مستقبلية طموحة تهدف لتطوير منظومة التقويم في المرحلة الجامعية.

مشكلة الورقة البحثية:

لقد كان من ثمار استخدام التقنيات الحديثة أن ظهرت أساليب جديدة للاختبار المعتمد على الحاسب احتلت مكان اختبارات الورقة والقلم التقليدية، وأدى ذلك لظهور مصطلحات عديدة لتعبر عن أشكال مختلفة من القياس وعلى رأسها مصطلح الاختبار المحوسب **Computerized Testing** حيث يشير هذا المصطلح إلى عدة جوانب يمكن أن يستخدم فيها الحاسب في الاختبارات مثل: تخزين البيانات؛ أو تطبيق الاختبارات؛ أو تفسير الدرجات؛ أو حتى اختيار أسلوب العلاج المناسب بعد إجراء التشخيص الملائم. إلا أن الورقة الحالية قد ركزت على عرض الاختبار التكيفي المحوسب (**CAT**) وهو أحد أهم التطبيقات الحديثة في مجال الاختبارات التحصيلية.حيث يقوم الاختبار التكيفي المحوسب بالتكيف مع قدرة الطالب لأن المختبرين يجيبون عن مجموعات مختلفة من المفردات، تطرح عليهم بناء على مستوياتهم المعرفية، حيث تؤثر إجابتهم عن مفردة ما على نوعية المفردات التالية التي يختارها الحاسب لهم، ففي حال الإجابة الصحيحة عن المفردة سيختار الحاسب مفردة أصعب قليلاً منها، ويطرحها على الُمْختَبِر، وبالعكس إذا كانت الإجابة خاطئة سيختار الحاسب مفردة أقل صعوبة من تلك التي أخطأ في الإجابة عنها أي يُفَصَّل الاختبار لكل فرد على حدة؛ بحيث لا يتعرض للمفردات السهلة جداً، ولا الصعبة جداً بالنسبة للفرد. بل تقدم له المفردات فقط التي تتناسب مع مستوى قدرته لأن الاختبار لا يهدف لتحدي قدرات الفرد ولكن يهدف لقياس مدى تمكنه من تحصيل المعلومات وفقا لمستوى قدرته، فلا يشترط الإجابة على نفس المفردات وبنفس الترتيب؛ إنما

يختلف عدد المفردات المطبقة وترتيبها من فرد لآخر؛ حسب قدرة كل منهم ويتم تطبيقه غالباً عن طريق الحاسب، ومن هنا جاءت تسميته بالاختبار التكيفي المحوسب، وتتحدد المشكلة التي ستعالجها الورقة الحالية في أنه: كيف يمكن توظيف التقنية لتطوير أساليب التقويم والاختبارات الالكترونية ممثلة في الاختبار التكيفي المحوسب؟

أهداف الورقة البحثية:

تهدف الدراسة للتعرف على كيفية توظيف التقنية لتطوير أساليب التقويم والاختبارات الالكترونية في التعليم الجامعي ممثلة في الاختبار التكيفي المحوسب، من خلال ما يلي :

- التعرف على الخلفية النظرية للاختبار التكيفي المحوسب
- تحديد المعايير التي يجب أن تؤخذ في الحسبان عند بناء الاختبار التكيفي المحوسب.
- توضيح الفروق بين الاختبارات المعتمدة على الحاسب سواء الخطية أو المتكيفة حاسوبياً واختبارات الورقة والقلم التقليدية من خلال تحديد بعض الايجابيات والسلبيات لتطبيق الاختبار التكيفي المحوسب.
- التعرف على بعض التوجهات الحديثة في مجال تطبيقات الاختبار التكيفي المحوسب.
- توضيح بعض التأثيرات الإيجابية لتطبيق الاختبار التكيفي المحوسب على المناهج التعليمية في مختلف التخصصات.
- التعرف على عدد من العوامل التي أدت إلى تأخر اعتماد الاختبار التكيفي المحوسب في مؤسساتنا التعليمية.
- الخروج ببعض التوصيات المتعلقة بالموضوع في ضوء ما تم عرضه في هذه الورقة.

أهمية الورقة البحثية:

تستمد الدراسة أهميتها من الأهداف التي سعت لتحقيقها، حيث أن الهدف الرئيس للدراسة الحالية؛ هو توضيح كيفية توظيف التقنية في بناء وتطبيق الاختبار التكيفي المحوسب باعتباره تطبيق على لستخدام أساليب التقويم والاختبارات الالكترونية في التعليم.

حدود الورقة البحثية:

تتمثل حدود هذه الورقة بعرض لمفهوم الاختبار التكيفي المحوسب الذي يستخدم في تقدير بارامترات المفردة وقدرة الفرد على النماذج الرياضية ثنائية الاستجابة(٠،١) لنظرية الاستجابة للمفردة **IRT**، وتتحدد بالجوانب التي تم تناولها حول الموضوع.

المصطلحات المستخدمة في الورقة البحثية:

١. الاختبار التكيفي المحوسب**(Computerized Adaptive Testing (CAT** وهو أحد أهم التطبيقات الاختبارية الحديثة في مجال الاختبارات التحصيلية؛ حيث يتوافق الاختبار التكيفي المحوسب مع قدرة الطالب لأن المختبرين يجيبون عن مجموعات مختلفة من المفردات، تطرح عليهم بناء على مستوياتهم المعرفية ،فضلا عن أن صوابية إجابتهم عن مفردة ما تؤثر على نوعية المفردات التالية التي يختارها الحاسب الألي لهم، ففي حال الإجابة الصحيحة عن مفردة ما سيختار الحاسب مفردة أصعب قليلاً منها، ويطرحها على المتقدم إلى الاختبار، وبالعكس إذا كانت إجابة المختبر خاطئة سيختار الحاسب مفردة أقل صعوبة منها، أي أنه اختبار يُفَصَّل لكل فرد على حدة؛ بحيث لا يتعرض الفرد للمفردات السهلة جداً، ولا الصعبة جداً. بل تقدم له المفردات فقط التي تتناسب مع مستوى قدرته **(Lawson, 2006)**

٢. نظرية الاستجابة للمفردة **(Item Response Theory (IRT** إن نظرية الاستجابة للمفردة تُعد إطاراً نظرياً للقياس النفسي والتربوي لتحليل المفردات وبناء الاختبارات، ولها عدة افتراضات لا بد من استيفائها حتى يتسنى تطبيقها

والإفادة منها، وتشترط ملاءمة البيانات للنموذج المستخدم، وتفترض هذه النظرية أنه يمكن التنبؤ بأداء الأفراد في اختبار معين في ضوء خصائص مميزة لهذا الأداء تسمى سمات، وهذه السمات غير ملاحظة، ولكن تظهر من خلال أداء الأفراد، والذي يمكن ملاحظته وقياسه بصورة مباشرة من خلال مجموعة مفردات الاختبار. ولا تعتمد النظرية في تقدير درجة الأفراد على عدد الإجابات الصحيحة، إنما تعتمد على نمط الاستجابة بناء على دالة رياضية مستندة على نظرية الاحتمالات تسمى دالة الاستجابة للمفردة **Item Response Function** ، وتتميز هذه النظرية بخاصية عدم التباين بالنسبة لبارامترات المفردة والفرد(**AL-harbi, 2003; Hambleton et al., 1991, P 9**).

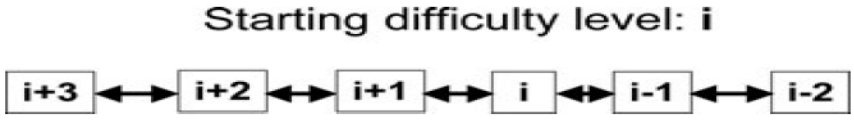
٣. البارامتر **Parameter** هو الخاصية العددية لمجتمع أو لنموذج (**Hambleton et al., 1991, P 12**) ، والتعريف الإجرائي للبارامتر في الدراسة الحالية يتحدد بالقيمة العددية لبارامترات المفردة وهي: (التمييز، والصعوبة، والتخمين)، وبارامتر قدرة الفرد.

٤. نماذج الاستجابة للمفردة **Item Response Models** هي مجموعة من النماذج الرياضية المرتبطة بنظرية الاستجابة للمفردة **IRT**، تهدف لإيجاد العلاقة بين أداء الأفراد على الاختبار وبين مجموعة القدرات التي تقف خلف هذا الأداء. وتعد هذه النماذج احتمالية، وتقوم على أربعة افتراضات عامة، هي: أحادية البعد، والاستقلال الموضعي، والمنحنى المميز للمفردة، وعامل السرعة، وتوجد افتراضات خاصة بكل نموذج على حدة، وهناك ثلاثة نماذج لوغارتمية شائعة الاستخدام في البيانات ذات التقسيم الثنائي الأصيل (٠ ، ١)، وتتمثل في النموذج: أحادي البارامتر، وثنائي البارامتر، وثلاثي البارامتر .(**Boeck and Wilson, 2004**)

الإطار العام للورقة البحثية

١ – الخلفية النظرية للاختبار التكيفي المحوسب

يركز الاختبار التكيفي المحوسب (CAT) على النمذجة الفردية للطالب بدلاً من تقديم نفس مجموعة المفردات لاختبار جميع الطلاب، فالاختبار التكيفي المحوسب يقدم آلية فعالة ليس فقط للحد من قلق المخْتَبِرين، إما لأن المفردات صعبة جداً أو سهلة جداً، ولكنه أيضا يساعد المربي على فهم قدرة الطالب، وبالتالي يمكنه من تقديم العون في الوقت المناسب، ويتكون الاختبار التكيفي المحوسب من مكونين، هما: اختبار محوسب، بالإضافة لمكونات متكيفة **Adaptive Components**، والمقصود بالتكيف هنا القدرة على تَفْصِيل مستوى صعوبة كل مفردة على أساس صحة إجابة المُخْتَبِر عن مفردة سابقة. ويوضح الشكل رقم (١) إستراتيجية الاختبار التكيفي في التعامل مع مستويات الصعوبة



حيث **i:** تمثل بداية مستوىالصعوبة

شكل رقم (١): إستراتيجية الاختبار التكيفي في التعامل مع مستويات الصعوبة

المصدر :(**Cheng and Basu ,2009**)

يتضح من الشكل رقم (١) مفهوم الاختبار التكيفي المحوسب باستخدام قائمة من البيانات المرتبطة على متصل من السمة المراد قياسها، بحيث يتكون هذا المتصل من أوعية على نقاط محددة من هذا المتصل يحوي كل وعاء مجموعة من المفردات متساوية الصعوبة، وعند ما يجيب المختبر على مفردة منها، فإن اختيار المفردة التالية يتم بطريقة متكيفة، إما أن تقع المفردات الأكثر صعوبة على يسار المتصل لأي من أوعية المفردات الإختبارية ، أو أن تقع أسهل المفردات على يمين المتصل ، وذلك بناء على مدى صحة استجابة المختبر على المفردة السابقة، وعملية اختيار المفردة التالية تتم بطريقة أكثر تعقيداً من اختيار مفردة من الوعاء المجاور، أما من أي وعاء سيتم اختيارها فإن التحكم في ذلك يلزمه نظرية قياس تستند على قدرة الفرد وبارامترات

المفردة(الصعوبة، والتمييز، والتخمين)، ولا يستند فقط على عدد الإجابات الصحيحة كما في النظرية التقليدية للاختبار **Classical Test Theory**.

ومن ثم فإن أغلب الاختبارات التكيفية المحوسبة تستند على نظرية الاستجابة للمفردة **IRT** وهي عبارة عن نظرية تحوي عدة نماذج رياضية تصف كيف يتفاعل المختبرين مع مفردات الاختبار، وتقوم بحساب احتمالية الاستجابة الصحيحة لكل مفردة على حدة، وبغض النظر عن مستوى الصعوبة المقدم للمختبر فإن قدرته في السمة المقاسة يمكن تقديرها بعد تطبيق عدد محدد من المفردات وذلك من خلال البحث في مدى تقاربها مع المنحنى المميز للمفردة **(Linden and Hambleton, 1997 , P4)**، حيث أنه في النظام التقليدي لتقدير درجات المختبرين يتم الحكم على مستويات الأداء على أساس النسبة المئوية للإجابات الصحيحة على مفردات الاختبار؛ لأن جميع المختبرين يأخذوا نفس المفردات، أما في نظرية الاستجابة للمفردة؛ فإن مستوى مهارة الطالب يتم تقييمه بشكل فردي حيث أن كل طالب يأخذ اختبار مختلف عن زملائه من حيث عدد المفردات ومستوى صعوبتها بحسب قدرته التي يتم تقديرها بعد تطبيق كل مفردة، ويتم هذا التقدير استناداً إلى المقاربة بين مستوى الأداء النهائي والمنحنى المميز للمفردة **(Cheng and Basu ,2009)**، الذي يصف العلاقة بين احتمال الاستجابة الصحيحة للمفردة **P(θ)** والمقياس المتدرج للقدرة، وفي هذه النظرية كل مفردة لها منحنى مميز يميزها عن المفردات الأخرى في الاختبار، والمنحنى المميز للمفردة هو الأساس في بناء مفاهيم نظرية الاستجابة للمفردة **IRT** مثل: القدرة، والصعوبة، والتمييز، والتخمين**(Lawson, 2006)**.

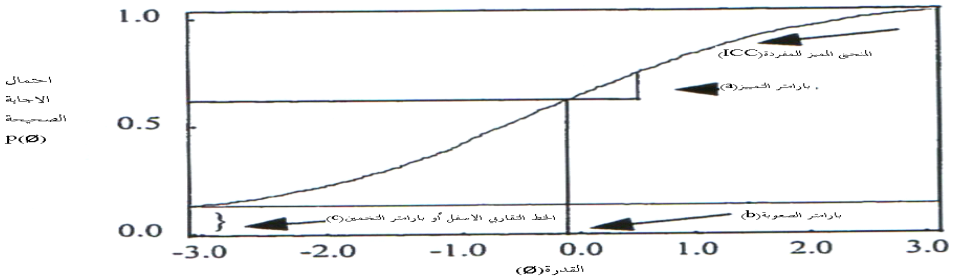
وعموماً، فإن أي منحنى من المنحنيات المميزة للمفردة يتبع مجموعة من المنحنيات لها نفس الشكل العام، ويختلف عدد البارامترات التي تحدد شكل هذه المنحنيات من نموذج لآخر سواء النموذج الأحادي أو الثنائي أو الثلاثي، وهو

النموذج الذي يأخذ فيه المنحنى المميز للمفردة الشكل العام والصيغة الرياضية له كالتالي: **(Hambleton and Swaminathan, 1985 , P37)**،

$$P_i(\theta) = b_i(\theta) + a_i(\theta)$$

حيث: $P_i(\theta)$ هو احتمال حدوث الاستجابة الصحيحة للفرد الذي تم اختياره عشوائياً الذي قدرته (θ) على المفردة (i).

وهذه الدالة توصف من خلال بارامترين للمفردة هما صعوبة المفردة **Item Difficulty** (bi)، وتمييز المفردة **Item Discrimination(ai)** ، ويتحدد المنحنى المميز لمفردة ما تحديداً كاملاً إذا علمنا شكله العام، وقيم بارامترات المفردة التي يمثلها. وتختلف أشكال المنحنى المميز للمفردة حسب النموذج الذي تعتمد عليه البيانات من النماذج المستخدمة في النظرية، والرسم الذي يمثل المنحنى المميز للمفردة **Item Characteristic Curve(ICC)** في النموذج اللوغارتمي الثلاثي البارامتر، يمكن تمثيله بالشكل رقم (٢) **(Hambleton and Swaminathan, 1985 , P 39)**



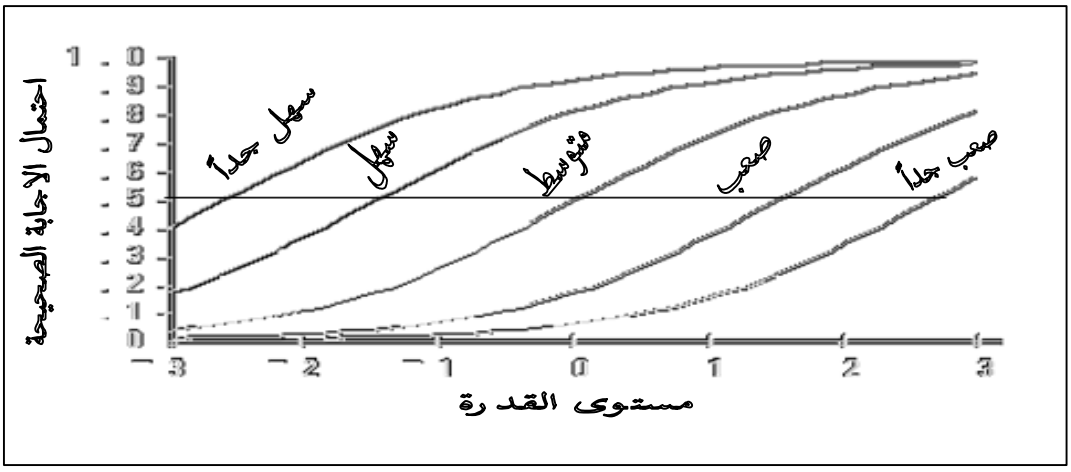
شكل رقم (٢): يوضح المنحنى المميز للمفردة (ICC) في النموذج اللوغارتمي الثلاثي البارامتر المصدر : (Linden and Hambleton, 1997,P4)

يتضح من الشكل رقم (٢) أن المنحنى المميز للمفردة ICC في النموذج اللوغارتمي ثلاثي البارامتر في نظرية الاستجابة للمفردة IRT، أن مستوى القدرة، وهو بارامتر قدرة الفرد (θ) يمثله المحور الأفقي، بينما احتمال الاستجابة الصحيحة في المحور الرأسي، وأن المنحنى يتزايد اطرادياً على طول متصل القدرة، كما أن قيم القدرة والصعوبة تقاس على نفس المتصل، ويمثل بارامتر

الصعوبة (b) بنقطة على ميزان القدرة، عندما يكون احتمال الاستجابة الصحيحة (٠,٥٠)، وكلما زادت صعوبة المفردة فإن المنحنى يتجه لليمين، والعكس صحيح.

وتختلف المنحنيات المميزة للمفردة في حالة النموذج الخطي في أجزائها المقطوعة من المحور الرأسي (bi)، والذي يمثل صعوبة المفردة، والميل (الانحدار) (ai) والذي يمثل تمييز المفردة، وهما يعكسان حقيقة أن مفردات الاختبار تختلف في جانبي الصعوبة وقوة التمييز بين المستويات المختلفة لقدرة الأفراد. وعلى قيمة خط التقارب السفلي للمنحنى المميز للمفردة (c)، ويشير خط التقارب السفلي إلى احتمال الاستجابة الصحيحة لذوي القدرة المنخفضة من الأفراد عن طريق التخمين، ويقابل الجزء المقطوع من محور الصادات في النماذج الخطية.

ويتم التعبير عن احتمال الاستجابة الصحيحة $P_i(\theta)$ بدالة في بارامتر قدرة الفرد (θ) وبارامترات المفردة (c, b, a)، و يقابل بارامتر القدرة (θ) القيمة التي يحصل عليها الفرد على السمة المقاسة، وبارامترات المفردة تقابل خصائص المفردة التي تؤثر على استجابة الفرد على المفردة، وعادةً ما يعبر عن مدى القدرة (θ) بوحدة الانحراف المعياري ، حيث يمتد القدرة من (−٣) إلى (+٣)، بالرغم من أنه يمكن زيادة أو تقليص هذا المدى، ويوضح الشكل رقم (٣) منحنيات مميزة للمفردة لها مستويات مختلفة من الصعوبة (سهل جدا–سهل–متوسط–صعب–صعب جدا) كالتالي:



شكل رقم (٣): يوضح منحنيات لها درجات صعوبة مختلفة

المصدر : **(Baker and Kim, 2004,p14)**

وفي ضوء ما سبق، يمكن القول أن تطبيق الاختبار التكيفي المحوسب في إطار نماذج نظرية الاستجابة على المفردة يعمل على مطابقة مستوى صعوبة مفردات الاختبار مع مستوى قدرة المخْتَبِر، وذلك يتطلب نطاقاً كبيراً من المفردات معلومة الخصائص الإحصائية لتزويد بنك المفردات داخل ذاكرة الحاسب بتقديرات لبارامترات كل مفردة(الصعوبة، والتمييز، والتخمين)، واحتمالية الإجابة الصحيحة لكل مفردة في الاختبار، وعند البدء في تطبيق الاختبار بعد برمجته على الحاسب يتم تكييفه من خلال ما تتيحه نظرية الاستجابة للمفردة من إجراءات حيث ذكر هاملتون وآخرون **Hambleton et al. (1991,P147)** أنها تقوم بحساب تقديرات قدرة الفرد على كل مفردة تقدم إليه في الاختبار، وبناءً على هذا التقدير تقوم بانتقاء المفردة المناسبة التالية إلى أن يتم إنهاء الاختبار بعد الوصول إلى تحديد الدرجة الرقمية التي تمثل قدرة الطالب المختبر.

٢ – معايير بناء الاختبار التكيفي المحوسب

الفكرة الأساسية للاختبار التكيفي المحوسب **CAT** هو أن يقوم الكمبيوتر باختيار المفردة التي تبدو ملائمة لمستجيب معين، ومن ثم تصحيح الاستجابة بطريقة تسمح بالمقارنة بين أداء المستجيب على مجموعة مختلفة من المفردات، ونتج عن ذلك أن أصبح التقييم أسرع وأكثر دقة، ويمكن تحديد الخطوات المنطقية النموذجية للاختبار التكيفي المحوسب، في الخطوات التالية: الخطوة (١) الاختبار يبدأ بتقدير أولي لدرجة المستجيب، وهذا يتطلب أن تكون هناك أسئلة مبدئية موحدة تقدم لجميع المستجيبين، أو يتطلب وجود معلومات قبلية عن المستجيبين، تلك الأسئلة أو المعلومات القبلية توفر معلومات عن متوسط أداء الأفراد، وعن المحتوى المناسب للمفردة الأولى التي ستقدم لهم، ثم تستخدم الدرجة الأولية في اختيار المفردة التالية، والتي تملك اكبر قدر من المعلومات، والتي يتم تطبيقها في الخطوة رقم (٢)، وهكذا تستخدم إجابتها في الخطوة رقم (٣) عند إعادة تقدير النتيجة، وفي الخطوة رقم (٤) يتم حساب فترات الثقة **Confidence Interval**للمستجيب المحدد وصولا لتقدير الدرجة، وفي الخطوة رقم (٥) يحدد الكمبيوتر ما إذا تم الوفاء بقاعدة إنهاء الاختبار أم لا، وإذا استوفيت القاعدة فإن الكمبيوتر يقدر ما إذا كانت فترة الثقة هي في حدود معينة، وحالما يتم استيفاء المعايير، يبدأ الكمبيوتر إما بتقييم مفهوم جديد، أو تنتهي البطارية. خلاف ذلك، يتم تكرار الخطوة رقم (٢) من أجل المفردة التالية بحيث تحوي أكبر قدر من المعلومات (**Bjorner et al., 2007**).

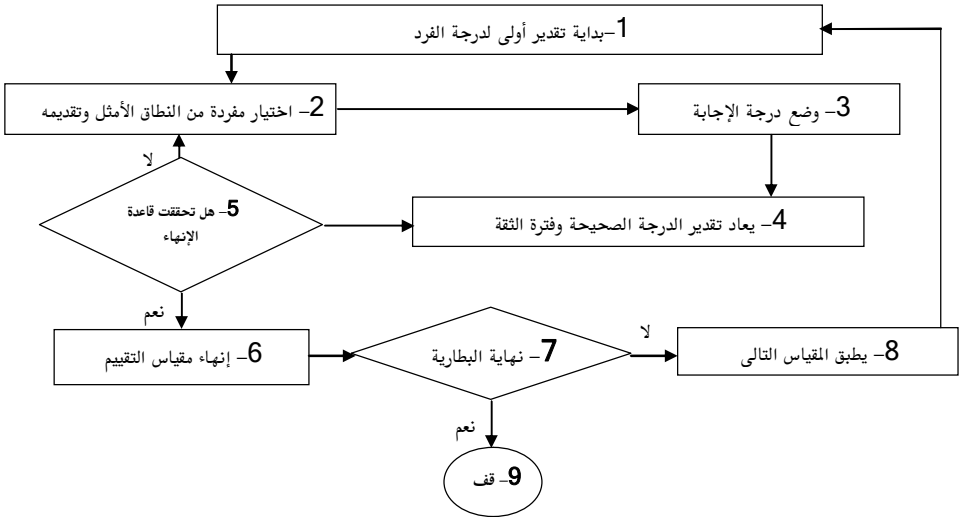
وحدد تشانق **Chang(2004)** ملخصاً لمتطلبات الاختبار التكيفي المحوسب الذي وصفه بالناجح فيما يلي:

١) تدريج قبلي لبنك المفردات باستخدام أحد نماذج نظرية الاستجابة للمفردة،
٢) تطبيق إجراءات معينة لاختيار المفردة الأولية،
٣) طريقة للتصحيح،
٤) طريقة لاختيار المفردة أثناء تطبيق الاختبار،

٥) قاعدة لإنهاء الاختبار،

٦) تقديم نظام محوسب يمكن الاعتماد عليه.

ويوضح الشكل رقم (٤) الخطوات المنطقية لتطبيق الاختبار التكيفي المحوسب كما جاءت في: **(Wainer and Mislevy , 2000, P106)**



الشكل رقم(٤): يوضح الخطوات المنطقية لتطبيق الاختبار التكيفي المحوسب.

المصدر : **(Wainer and Mislevy,2000,p106)**

مما سبق نجد أن الاختبار التكيفي المحوسب في حاجة لمجموعة من القواعد التي تساعد على اختيار انسب المفردات لوضعهما على تدريج مشترك من السهل للصعب،وبنك من المفردات يمكن الاختيار منه للاختبار بحيث يحتوي البنك على معلومات عن المحتوى الذي تقيسه كل مفردة وخصائص القياس التي يتمتع بها وفقاً لنموذج القياس المستخدم.

ومعظم الاختبارات التكيفية المحوسبة تعتمد على نظرية الاستجابة للمفردة عند اختيار المفردات وتصحيح الاستجابات **(Bjorner, 2004؛ Bjorner et al., 2007)**،ومنه يمكن تحديد أهم المعايير عند بناء الاختبار التكيفي المحوسب واستخدامه، وهذه المعايير هي:

١- اختيار نموذج التحليل المناسب من نماذج نظرية الاستجابة للمفردة

تتكون نظرية الاستجابة للمفردة IRT من عدة نماذج (تسمى نماذج الاستجابة للمفردة Item Response Models) هدفها تحديد العلاقة بين أداء الفرد في الاختبار (الذي يمكن ملاحظته بصورة مباشرة) ومجموعة السمات أو القدرة التي تكمن وراء هذا الأداء، وتفسرها بطريقة أدق من أساليب القياس التقليدية آخذة بالحسبان أن النماذج الرياضية تخضع لنظرية الاحتمالات وليست حتمية (Hambleton and Swaminathan, 1985 , Pp10-11)

ويعتمد اختيار النموذج المناسب على طبيعة المفردات وعددها وحجم العينة المستخدمة في تقدير البارامترات، إلا أن طبيعة البيانات هي الموجه الأول لاختيار النموذج المناسب، وتنقسم البيانات إلى بيانات ثنائية الاستجابة (٠، ١) Dichotomous Data، وبيانات متعددة الاستجابة Polytomous Data، ومن أهم النماذج الرياضية ثنائية الاستجابة لنظرية الاستجابة للمفردة Wells, 2004؛ Bjorner et al., 2007؛ Fischer, 1997; Hambleton et al.,1991، P12) ما يلي:

أ) **النموذج اللوغارتمي أحادي البارامتر (نموذج راش)** The One-Parameter Logistic -Rasch Model وهو من النماذج البسيطة والشهيرة الاستخدام في نظرية الاستجابة للمفردة IRT، وهو أحد الحلول المتاحة عند تقدير بارامترات نظرية الاستجابة للمفردة IRT في حالة العينات الصغيرة لأنه يتطلب بيانات أقل لتقدير البارامترات بدقة، وهو حالة خاصة من النموذج اللوغارتمي الثلاثي والثنائي البارامتر بافتراض أن قيمة بارامتر التمييز (١) a =، وبارامتر التخمين يفترض أنه أصغر ما يمكن أي: (٠) c =، ويأخذ المنحنى المميز للمفردة في هذا النموذج الصيغة الرياضية التالية:

$$p_i(\theta) = \frac{\ell^{(\theta - b_i)}}{1 + \ell(\theta - b_i)}$$

$P_i(\theta)$= احتمالية أن يجيب الفرد الذي اختير عشوائياً بمستوى قدرة (θ) عن المفردة i استجابة صحيحة.

(θ)= مستوى قدرة الفرد.

bi= بارامتر صعوبة المفردة i.

حيث: i= 1, 2,...,n رقم المفردة، و n = عدد مفردات الاختبار.

e = دالة الثابت الرياضي تساوي تقريباً ٢,٧١٨

وبسبب القيود المفروضة على حجم العينة وطول الاختبار اللازم لتطبيق كل نموذج؛ ولأن النماذج التي تتضمن تقدير بارامترات أقل تحتاج لقدر أقل من البيانات من أجل دقة تقدير البارامتر؛ فإن النموذج اللوغارتمي أحادي البارامتر يعتبر من الاختيارات النموذجية عندما تكون البيانات المتاحة محدودة (Barenes and Wise, 1991).

ب) **النموذج اللوغارتمي ثنائي البارامتر** The Tow-Parameter Logistic Model وهو النموذج الذي يأخذ فيه المنحنى المميز للمفردة شكل التوزيع اللوغارتمي ذي البارامترين (التمييز، والصعوبة)، وهو حالة خاصة من النموذج اللوغارتمي ثلاثي البارامتر، ويشترط هذا النموذج أن تكون مفردات الاختبار متفاوتة في مستويات التمييز والصعوبة أيضاً؛ لكن لا يوجد بها مجال للتخمين، والصيغة الرياضية له كالتالي:

$$p_i(\theta) = \frac{e^{Da_i(\theta - b_i)}}{1 + e^{Da_i(\theta - b_i)}}$$

$P_i(\theta)$= احتمالية أن يجيب الفرد الذي اختير عشوائياً بمستوى قدرة (θ) عن المفردة i استجابة صحيحة.

(θ)= مستوى قدرة الفرد.

ai= بارامتر تمييز المفردة i.

bi= بارامتر صعوبة المفردة i.

D= عامل التدريج الذي يساعد في الوصول إلى أقرب لوغارتم ممكن لعملية التراكم الطبيعي.

حيث: i= 1, 2,...,n رقم المفردة، و n = عدد مفردات الاختبار.

e = دالة الثابت الرياضي تساوي تقريباً ٢,٧١٨

ج) **النموذج اللوغارتمي ثلاثي البارامتر The Three-Parameter Logistic Model**

وهو النموذج الذي يأخذ فيه المنحنى المميز للمفردة شكل التوزيع اللوغارتمي ذي البارامترات الثلاثة (الصعوبة، والتمييز، والتخمين) والصيغة الرياضية له كالتالي:

$$p_i(\theta) = c_i + (1 - c_i)\frac{e^{Da_i(\theta - b_i)}}{1 + e^{Da_i(\theta - b_i)}}$$

$P_i(\theta)$= احتمالية أن يجيب الفرد الذي اختير عشوائياً بمستوى قدرة (θ) عن المفردة i استجابة صحيحة.

(θ)= مستوى قدرة الفرد.

ai= بارامتر تمييز المفردة i.

bi= بارامتر صعوبة المفردة i.

Ci = بارامتر تخمين المفردة، ويمثل الخط التقاربي الأسفل lower Asymptote للمنحنى المميز للمفردة.

D= عامل التدريج الذي يساعد في الوصول إلى أقرب لوغارتم ممكن لعملية التراكم الطبيعي.

حيث: i= 1, 2,...,n رقم المفردة، و n = عدد مفردات الاختبار.

e = دالة الثابت الرياضي تساوي تقريباً ٢,٧١٨

وبالنسبة لحجم العينة المناسب عند اختيار أي نموذج من نماذج نظرية الاستجابة للمفردة يوجد اختلاف في النتائج والتوصيات الخاصة بكمية البيانات

المطلوبة لتطبيق نظرية الاستجابة للمفردة IRT، وفقاً للنموذج المستخدم في النظرية؛ (AL-harbi, 2003; Ramsay, 2000; Truskosky, 1999; Setiadi, 1997; Patsula, 1995)، كما يلي: النموذج اللوغارتمي الثلاثي البارامتر لا يجب أن يقل فيه عدد المفردات فيه عن (٥٠) مفردة، والأفراد من (١٠٠٠) إلى (١٠٠٠٠) فرد، أما النموذج اللوغارتمي الثنائي البارامتر فيتطلب على الأقل (١٠٠٠) فرد للحصول على أخطاء معيارية طفيفة، مع تقدير لبارامتر تمييز المفردة، وبالنسبة للنموذج اللوغارتمي الأحادي البارامتر فيتطلب كحد أدنى (٢٠) مفردة و(٢٠٠) فرد. كما أنه فيما يتصل بالعينات صغيرة الحجم، فقد أوصي باستعمال النموذج أحادي البارامتر (نموذج راش)، (Darabi, 1997; Setiadi, 1997; Barnes and Wise, 1991)

وتجدر الإشارة إلى أن معظم دراسات الاختبار التكيفي المحوسب تعتمد في تحليل بياناتها على استخدام نماذج نظرية الاستجابة للمفردة، كما أن الدراسات التي استخدمت نماذج الاستجابة للمفردة ثنائية الإجابة (٠،١)، قد ركزت على النموذج اللوغارتمي ثلاثي البارامتر(Bjorner et al., 2007; Bjorner, 2004;Hambleton et al., 1991 , Pp146-147). وتوجد دراسات تستخدم النموذج اللوغارتمي ثلاثي البارامتر المعدل، حيث تم تثبيت قيمة بارامتر التخمين عند قيمة ثابتة يحددها الباحث وذلك عند ما يكون حجم العينة لا يكفي لتقدير البارامترات، لأن افتراض وجود خط التقارب السفلي غير الصفري سيجعل التعامل مع بيانات اختيار من متعدد أكثر واقعية، وأن الخط التقاربي الأسفل يثبت عند قيمة أقل من احتمالية الاستجابة الصحيحة عن مفردة بطريقة عشوائية.خاصة أنه من الممكن التحكم في مشكلات التقدير الشائعة للخط التقاربي الأسفل (بارامتر التخمين) في النموذج اللوغارتمي ثلاثي البارامتر، وضبطها، مما قد ينتج عنه قدرة مستقرة، لأن تقدير بارامتر قدرة الفرد عند ذوي القدرة المنخفضة تتقلص دقته بوجود خط تقارب سفلي غير صفري (Barenes and Wise، ١٩٩١)، بالإضافة

للحصول على تقدير لبارامترات تمييز وصعوبة المفردة (Thomas, 1997). وقد جرى تحديد القيمة الثابتة للخط التقاربي الأسفل بمقلوب عدد بدائل الاستجابة،(Darabi, 1997; Setiadi, 1997; Barnes and Wise, 1991)، أو تحدد القيمة الثابتة للخط التقاربي الأسفل بمقلوب عدد بدائل الاستجابة مطروحا منه (٠,٠٥)، وقد ذكر كلا من بارنزووايز Barnes and Wise (1991)أن إجراء التعديل قد أثبت فعاليته في تقدير بارامترات المفردة في النماذج اللوغارتمية أحادية وثلاثية البارامتر عند ما تكون العينات صغيرة أي أقل من ٢٠٠ فرد.

٢ – بناء بنك المفردات Item bank المناسب

عرف تشانق Chang(2004) بنك المفردات بأنه أكثر من مجرد مجموعة من المفردات أو الأسئلة، وذكر باجورنر Bjorner(2004) وركاس Reckase(2007) بأن بناء بنك المفردات يتطلب أولا تعريف واضح للتكوين الفرضي الذي سيقيسه الاختبار، ومن ثم صياغة عدد كبير من المفردات الجيدة، واستخدام أساليب إحصائية دقيقة للمفردات(ويتضمن ذلك تقييم أحادية البعد، والتحقق من مدى ملاءمة نموذج نظرية الاستجابة للمفردة الذي تم اختياره مع المفردات، وتقييم التحيز)، وأخيرا تحديد مواصفات واضحة للاختبار التكيفي المحوسب النهائي. و ليس هناك قانون يحكم أو يحدد العدد الكافي من المفردات التي يجب إيداعها في بنك المفردات، خاصة أن الاختبار التكيفي المحوسب يقدم مفرداته بأعداد مختلفة من فرد لآخر، إلا أن بنك المفردات إذا احتوى على ١٠٠ مفردة فأكثر قد يعطي نتائج جيدة (Weiss and Kingsbury, 1984) ولكن عند النظر لاعتبارات السرية ودقة تقدير قدرة كل مختبر، والرغبة في أن تغطي المفردات محتوى السمة المراد قياسها فإن العدد الأكبر من المفردات قد يكون هو الأفضل.

٣–اختيار الأسلوب المناسب لاختيار المفردة

يتمثل المعيار الثالث لبناء الاختبار التكيفي المحوسب في اختيار الأسلوب المناسب لاختيار المفردات، وهناك طرق عديدة لاختيار المفردة الأولي التي تقدم للفرد، منها: تحديد مستواه القبلي بناء على معلومات سابقة عن المختبر، ومن الطرق الشائعة في هذا المجال تحديد المفردة التي تقدم أكبر قدر من المعلومات على متصل السمة التي يجري قياسها،ثم تأتي قدرة الاختبار التكيفي أن يختار المفردة اللاحقة بناء على مستوى قدرة الفرد التي تم تقديرها من خلال إجابته على المفردة الأولى المقدمة في الاختبار، ويتم ذلك باستخدام نظرية الاستجابة للمفردة والتي يمكن من خلالها التنبؤ باحتمالية استجابة الفرد على المفردة بناء على معلومات سابقة تتعلق بتقدير سابق للقدرة وبارامترات المفردة **(Bjorner et al., 2007)** ، وتوجد عدة خيارات لاختيار الأسلوب المناسب لاختيار المفردة ذكرها سلا وآخرون **Cella et al. (2007)**، ومنها: أسلوب المعلومات القصوى لدالة فيشر **(MFI) Maximum Fisher Information**، وأسلوب المعلومات الارجحي الأقصى المـوزون**Maximum Likelihood-Weighted Information (MLWI)**، وأسلوب المعلومات البعدي الأقصى الموزون**Maximum Posterior Weighted (MPWI) Information**،وأسلوب المعلومات الأقصى المتوقع **Maximum Expected Information (MEI)** ، أسلوب التباين البعدي الأدنى المتوقع **MinimumExpected Posterior Variance (MEPV)** ، وأسلوب المعلومات الأقصى لجداول فيشر **MaximumFisher Information Table (MFIT)**، وطبقا لما ذكره سلا وآخرون **(Cella et al. (2007** فإنه بالرغم من أن أسلوب المعلومات الأقصى لفيشر يعد أكثر الأساليب شيوعا إلا أن أسلوبي المعلومات الأقصى المرجح الموزون وأسلوب المعلومات البعدي الأقصى الموزون، أكثر كفاءة منه لأنهما يقدمان مزايا نظرية أفضل فهما يأخذان في الاعتبار جميع القيم المحتملة لتقديرات السمةبدلاً من أخذ تقدير النقطة الحالية فقط.

أما أسلوب التباين البعدي الأدنى المتوقع فهو أحد الأساليب البيزية (Bayesian) وهو لا يعتمد على انتقاء المفردة بناء على دالة المعلومات عند القدرة المقدرة بل يستخدم التباين البعدي للقدرة المقدرة ويختار المفردة التي لها أقل تباين بعدي متوقع بالاستفادة من المعلومات القبلية في بنك الأسئلة والمتضمنة تقدير القدرة وبارامترات المفردة، ومنها يحسب التباين، وبالتالي فالمفردة التي تعطي أقل تباين هي التي يتم انتقاؤها وهكذا.

أما أسلوب المعلومات الأقصى لجداول فيشر فيعد تعديلا لأسلوب المعلومات الأقصى لفيشر لأنه أسرع، حيث يستخدم مصفوفة المعلومات التي ولدت على نطاق محدد من مدى القدرة بفواصل ثابتة، ثم في كل مرة يتم إعادة الإجراءات داخلها إلى أن تصل تقديرات القدرة للدقة المطلوبة ثم تستخدم هذه المعلومات في إعداد الجداول، ومن خلال هذه الجداول في كل مرة يطبق فيها الاختبار التكيفي المحوسب يتم البحث في قيم القدرة المحسوبة مسبقا لتحديد اكبر قدر من المعلومات التي تقدمها المفردة المناسبة لقدرة الفرد المختبر، لكل مفردة مهما كان عدد مفردات البنك كبيرا، وهذا الإجراء يساعد على الاقتصاد في الوقت اللازم لإجراء الحسابات. والذي يعالج أهم عيوب أساليب المعلومات القصوى والأساليب البيزية، حيث أنها عموما تستغرق وقتاً كبيراً في إجراء العمليات الإحصائية لتقدير قدرة الفرد عند البحث في جميع مفردات البنك عن أفضل مفردة.

٤ –اختيار الأسلوب المناسب لتقدير قدرة الفرد

يبدأ الاختبار التكيفي المحوسب بتقدير مبدئي لقدرة الفرد، ثم يجري الحاسب تعديلات متتالية على التقدير المبدئي للقدرة بناء على المفردات التالية التي تقدم للفرد في أثناء تطبيق الاختبار إلى أن يتم الحصول على أكبر قدر من الدقة في تقدير قدرة الفرد بأحد نماذج نظرية الاستجابة للمفردة التي تكون محددة مسبقا، لذا يعتبر اختيار الأسلوب المناسب لتقدير قدرة الفردمهماً لأن دقة كل

من اختيار المفردة التي يتم تقديمها للفرد، وإنهاء الاختبار، ومن ثم التقدير النهائي للقدرة المقاسة تتوقف عليه. وتتحدد طرق تقدير بارامترات النماذج أحادية البعد في نظرية الاستجابة للمفردة التي تعتمد على دالة الترجيح فيما يلي: أسلوب الأرجحية الهامشية القصوى **Marginal Maximum Likelihood Procedure (MML)**، وأسلوب الأرجحية القصوى المشتركة **Joint Maximum Likelihood Procedure (JML)**، وأسلوب الأرجحية القصوى المشروطة **Conditional Maximum Likelihood Procedure (CML)**، والأساليب البيزية **Bayesian Procedures** لتقدير المنحنيات المميزة للمفردة.

وتختلف هذه الأساليب في تعاملها مع مشكلة مستويات القدرة غير المعلومة، كالتالي: أسلوب دالة الترجيح الأقصى المشترك يعمل على نمذجة احتمالات أنماط الاستجابات باستخدام تقديرات مبدئية لمستويات قدرة الأفراد، بينما يعمل أسلوب دالة الترجيح الأقصى الهامشي على نمذجة احتمالات أنماط الاستجابات وكأنها ناجمة عن معاينات عشوائية من مجتمع توزيع مستويات قدرة أفراده معلومة.

أما أسلوب دالة الترجيح الأقصى المشروط، فإنه يعمل على نمذجة هذه الاحتمالات باستخدام احتمالات مختلف أنماط الاستجابات، التي تؤدي إلى الدرجات الكلية نفسها، ويعد أسلوب الترجيح الأقصى الهامشي من أكثر الأساليب استخداما في المجالات التربوية ، ويتميز بالعديد من المزايا التي من أهمها أن التقديرات الناتجة عنه تحقق محكات التحقق من جودة التقدير مثل؛ عدم التحيز، والاتساق، كما يمكنه التوصل للتقدير عندما يجيب فرد معين على جميع المفردات استجابات صحيحة أو استجابات خطأ. إلا أن أسلوب دالة الترجيح الأقصى المشترك لا يحقق محكات عدم التحيز، والاتساق.

بينما أسلوب الترجيح الأقصى الهامشي يحقق هذه المحكات إلا أن استخدامه يقتصر فقط على النموذج اللوغارتمي آحادي البارامتر (نموذج راش)،

بينما الأساليب البيزيةBayesian Procedures تتوقف على الاحتمالات القبلية المفترضة للبارامترات. (Cella et al., 2007 ، علام، ٢٠٠٥)،وتوجد عدة برامج خاصة بتقدير بارامترات النماذج IRT وتشمل برنامج بايلوج BILOG-MG(Zimowski et al., 2003) لتقدير بيانات الاستجابة الثنائية، أما بالنسبة لتقدير بيانات الاستجابة المتعددة فيستخدم برنامج مالتيلوجMULTILOG7(Thissen, 2003) وبرنامج بايلوجBILOG-MG يحتوي على ثلاثة خيارات لطرق تقدير بارامترات المفردة والفرد، وهي: الترجيح الأقصى Maximum Likelihood (ML)،والبيزي البعدي المتوقع Expected a Posteriori (EAP)، والبيزي البعدي الأقصى Bayes model(Maximum a Posteriori MAP). انظر (Zimowski et al., 2003) و(Baker and Kim, 2004).

وقد وجد كل من فوكس وجلاس Fox and Glas (2003) من خلال دراسة محاكاة أن الإجراءات البيزية تتفوق على الترجيح الأقصى من ناحية الدقة على الأقل في العينات الصغيرة، وأن تقديراتها متاحة لدرجات المفردة والفرد، إلا أنها تتطلب إجراءات إحصائية متقدمة، بينما أشار دَرابي Darabi (1997) أن معاملات الارتباط بين بارامترات المفردة المقدرة بالأساليب الثلاثة المتاحة في برنامج بايلوجBILOG-MG، وهي: الترجيح الأقصى Maximum Likelihood (ML)، والبيزيBayes البعدي المتوقع Expected a Posteriori (EAP)، والبيزيBayes البعدي الأقصى Maximum a Posteriori (MAP) كانت مساوية (١,٠٠٠)، ومعاملات الارتباط بين بارامترات قدرة الفرد المقدرة بنفس الطرق الثلاث السابقة مساوية (٠,٩٩٩٨)، مما يعني أن استخدام أي منهم سيعطي نفس مستوى الدقةتقريباً .

٥-تحديد الأسلوب المناسب لقاعدة إنهاء الاختبار

طبقا لما ذكره شوي وآخرونChoi et al. (2011) فإنه عند تطبيق الاختبار التكيفي المحوسب يتم عرض المفردة الأولى وبعد استجابة الفرد عليها، يتم تصحيحها، ومن ثم تقدير قدرة الفرد على المفردة وقبل انتقاء المفردة التالية المناسبة لقدرته يتم أولا الرجوع لقاعدة إنهاء الاختبار Stopping Ruleلمعرفة ما إذا

تحققت القاعدة، فإذا تحققت يتم إنهاء الاختبار، وإلا يتم انتقاء المفردة المناسبة التالية، ويمكن اعتبارأن الاختبار التكيفي قد أكتمل بعد أن يتم إعطاء عدد محدد سلفا من المفردات، وعند ما يتم الوصول لمستوى محدد سلفا من دقة القياس، أو عند انقضاء وقت محدد سلفا ولكن من أكثر الطرق التي تستخدم كمحكات لإنهاء الاختبار، هما:

أ. **محك الطول الثابت للاختبار The Fixed Test Length:** وهو أن ينتهي الاختبار بعد تطبيق عدد محدد سلفا من المفردات المناسبة لقدرة الفرد، وبناء عليه، سيتم إعطاء نفس العدد من المفردات لجميع الممتحنين، بغض النظر عن درجة دقة القياس التي استحصلت بعد تطبيق الاختبار، والميزة الأساسية لتطبيق هذا المحك هي بساطته، وإحدى نتائج تطبيق هذا المحك انه سيتم قياس الممتحنين بدرجات متفاوتة من الدقة، وبالتالي سيصبح خطأ القياس أكبر عند المستويات المتطرفة من القدرة المقاسة، لأن الممتحنين سيجيبوا على مفردات سهلة جدا أو صعبة جدا بالنسبة لهم،وبالتالي فإن استخدام هذا المحك قد يحد من كفاءة الاختبار التكيفي المحوسب وذلك عند ما يتم تطبيق مفردات لا داعي لها عند ما تكون مساهمتها قليلة في تقديم معلومات حول مستوى قدرة الفرد.

ب. **محك الطول المتغير للاختبارVariable Test Length:** ويهدف إلى تحقيق درجة معينة من الدقة لجميع الممتحنين وهذا يعني أنه سيتم إعطاء الممتحنين عدد متفاوت من المفردات، ويستخدم محك آخر لإنهاء الاختبار غير طول الاختبار، ويتم استخدام نوعين من قاعدة الإنهاء بالطول المتغير للاختبار، هما:

١. قاعدة الإنهاء بالخطأ المعياريThe Standard Errorstopping Rule وهو الأكثر شيوعا حيث يتم وقف الاختبار عند ما يصل الخطأ المعياري المصاحب لآخر قدرة مقدرة عند أقل من قيمة محددةسلفاً

٢. قاعدة الإنهاء بأدنى معلومات **The Minimum Information Stopping Rule** يتم وقف الاختبار عند ما يصل إلى أدنى معلومات عن المفردة المصاحبة لآخر قدرة مقدرة.

٣- خصائص الاختبار التكيفي المحوسب مقارنة بالاختبارات المعتمدة على الحاسب واختبارات الورقة والقلم التقليدية

توجد عدة خصائص للاختبار التكيفي المحوسب مقارنة بالاختبارات المعتمدة على الحاسب واختبارات الورقة والقلم التقليدية ساعدت على ازدهار صناعة الاختبارات الحاسوبية المُتَكّيفة في كثير من حقول المعرفة في العلوم الإنسانية، وغالبية المراجع تذكر الإيجابيات العامة للاختبارات الحاسوبية (وهي الصورة الموازية للنسخة الورقية للاختبارات)، والإيجابيات الخاصة بالاختبارات الحاسوبية المُتَكّيفة (وهي التي يختلف عدد أسئلتها وصعوبتها باختلاف القدرة العلمية للمُخْتَبِر)، فقد أوضح كل من **(Lovleret al.; 2011, Dominoand Domino; 2006, Kaplan, and Saccuzzo,2008)** هذه العوامل، بمايلي:

أولاً: الإيجابيات العامة للاختبارات الحاسوبية

١. **الكفاءة.** توفر الاختبارات الحاسوبية المُتَكّيفة على الجهةالمنفّذة لها من تكاليف الإعداد وإجراءات التطبيق حيثتخفّض من تكلفة طباعة الاختبارات وأجور البريد، وتظهر ميزة هذه الخاصية في حال تطبيق الاختبار على مجموعة كبيرة من المختبرين. كما توفر الاختبارات الحاسوبية المُتَكّيفة على المُخْتَبِر الوقت اللازم للحصول على نتيجة الاختبار، وتظهر هذه الخاصية في الحالات التي يكون فيها الوقت جوهري ومصيري في حياة أو مستقبل المُخْتَبِر.

٢. **الملاءمة.** يمكن للشخص الراغب في تطبيق أحد صور الاختبارات الحاسوبية المُتَكّيفة اختيار الوقت الملائم لظروفه دون الحاجة إلى الانتظار إلى تكّون عدد معّين من المُخْتَبِرين كما هو الحال في الاختبارات التقليدية.

فمثلاً، يمكن للطالب الراغب في الحصول على قبول للدراسة في إحدى الجامعات العالمية الحصول على اختبار اللغة الإنجليزية في صورته الإلكترونية (تحديداً في صورته الحاسوبية الُمتَكّيفة) في الوقت الملائم لظروفه قدر الإمكان.

٣. **الراحة.** توفر الاختبارات الحاسوبية الُمتَكّيفةللُمْختَبِر بيئة اختبار أكثر راحة عن الاختبارات التقليدية، حيث تمكّن الُمْختَبِر من أخذ الاختبار في غرفة ذات ضجيج أقل ودرجة حرارة وقوة إضاءة أكثر ملاءمة من قاعات الاختبارات الجماعية التقليدية التي عادة تراعي المتوسط العام للُمْختَبِرين.

٤. **التقنيين.** تُيّسر الاختبارات الحاسوبية الُمتَكّيفة تطبيقها على الُمْختَبِرين بطريقة تزيد من مستوى دقة التقنيين من خلال تقليل نسبة الأخطاء البشرية المصاحبة عادة عند تنفيذ الاختبارات المقننة بالطريقة التقليدية الورقية.

٥. **المساندة.** تساعد الاختبارات الحاسوبية المعلم على الاستفادة من إمكانية تضمين الاختبار بالوسائط المتعددة (الصوت والصورة المتحركة أو الثابتة) لتوضيح المعلومة وتجعلها أقرب للواقع بطريقة لا توفرها النسخة الورقية؛ لتقلل بذلك من أثر التصميم على درجة تقييم الُمْختَبِر.

٦. **تعدد النماذج.** توظيف للحاسب الآلي في تصميم الاختباراتيسهّل من عملية إنتاج عدة نماذج من الاختبار الواحد في الوقت الواحد، الأمر الذي يقلل من خطر تسرب أسئلة الاختبار بين الُمْختَبِرين.

ثانياً: الإيجابيات الخاصة للاختبارات الحاسوبية الُمتَكّيفة

وبالإضافة إلى ما ذكرناهسابقاً حول إيجابيات الاختبارات الحاسوبية عموماً، تتميز الاختبارات الحاسوبية الُمتَكّيفة بعدد من الميزات الخاصة، ومنها ما يلي:

١. **الجهد.** تختصر الاختبارات الحاسوبية الُمتَكّيفة على الُمْختَبِر جهداً كبيراً يصل إلى ٥٠% **(Kaplan, and Saccuzzo,2008; p330)** أو تصل إلى ٨٠%

(De Ayala,2008; p373) مقارنة للاختبارات التقليدية الحاسوبية منها أو الورقية دون التضحية بخاصية الثبات حيث قيامها على مبدأ اختيار عدد الأسئلة وصعوبتها بناءً على المستوى العلمي للطالب.

٢. **الدقة.** توفر الاختبارات الحاسوبية المُتَكَيِّفة قدراً عالياً من الدقة في القياس من خلال تقليل نسبة أخطاء القياس المصاحبة عادة لعملية تقدير المستوى العلمي للمُخْتَبِر، بمعنى أنها أكثر ثباتاً ومصداقية من نظيراتها الاختبارات التقليدية؛ حيث تشخّص قدرة كل مُخْتَبِر على حدة بناءً على خط سير إجاباته على أسئلة الاختبار.

٣. **الأمن.** تمتاز الاختبارات الحاسوبية المُتَكَيِّفة بمستوى الحماية الأمنية العالية جداً مقارنة بالاختبارات عموماً والاختبارات الحاسوبية التقليدية خصوصاً التي تقوم على مبدئها الأساسي الذي يحدد عدد الأسئلة ومستوى صعوبتها بناءً على مستوى للمُخْتَبِر العلمي، والأمر الذي يجعل سرقتها غير مجدية.

٤. **زيادة الدافع.** تنطلق الاختبارات الحاسوبية المُتَكَيِّفة عادة بطرح سؤال متوسط الصعوبة، ثم تتدرج في الصعوبة صعوداً (أو نزولاً) تبعاً لمستوى المُخْتَبِر العلمي، الأمر الذي يزيد من الدافعية لدى المُخْتَبِر، على عكس ما تتركه عادة الاختبارات التقليدية على نفسية المُخْتَبِر من توتّر (أو استرخاء) مبالغ فيه عندما تواجهه أسئلة صعبة (أو سهلة) متتالية.

٥. **اختلاف النمط المعرفي.** تراعي الاختبارات الحاسوبية المُتَكَيِّفة النمط المعرفي للمُخْتَبِرين من خلال اعتماد تحديد نوعية أسئلة بداية الاختبار وعدد أسئلته وصعوبته على المستوى المعرفي للمُخْتَبِرين كل على حدة، والتي تتأثر بالنمط المعرفي للشخص على خلاف الاختبارات التقليدية القائمة على وحدة النوع والحجم للاختبار.

وفي المقابل، ذكر كلا من (Triantafillou et al., 2008;Dominoand Domino, 2006; Linacre, 2000) عدد من السلبيات المرتبطة بالاختبارات الحاسوبية عموماً والاختبارات الحاسوبية المُتَكَيِّفة تحديداً، ومن هذه السلبيات ما يلي:

أولاً: السلبيات العامة للاختبارات الحاسوبية

١. **محدودية نوعية الأسئلة.** قد تعيق محدودية مكونات بعض أجهزة الحاسب الآلي القائمين على تصميم مفردات الاختبارات الحاسوبية في وضع مفردات تتطلب رسوم أو فنون تفصيلية، ووضع مفردات تتطلب قراءة مقاطع واسعة النطاق.

٢. **بنية تحتية إلكترونية.** تتطلب الاختبارات الحاسوبية عموماً بنية تحتية إلكترونية قادرة على استيعاب عدد المُخْتَبِرين، كما تتطلب قدراً معيناً من المهارات الحاسوبية للمُخْتَبِرين، وهذان الشرطان يضيقان من مدى الاستفادة من الاختبارات الحاسوبية عموماً، ويجعلانها متحيّزة ضد المجتمعات والمنظمات ذات الإمكانات الحاسوبية المحدودة.

٣. **الأخطاء التقنية.** تواجه الاختبارات الحاسوبية عموماً مشكلة تعرضها للأخطاء التقنية التي عادة تصاحب الحاسب الآلي مما له الأثر السلبي على نفسية المُخْتَبِرين، وعلى سمعة الجهة المنظمة للاختبار بين مجموعة المستفيدين من الاختبارات سواء كانوا مُخْتَبِرين أو جهات دراسية أو توظيفية.

٤. **القرصنة.** إمكانية تعرض الاختبارات الحاسوبية إلى جرائم إلكترونية (قرصنة أو الحرب الجرثومية)، يضع صدقها وأمنها في خطر دائم، وتزيد هذه الإمكانية في الدول والمنظمات المتقدمة تقنياً عن نظيراتها في البيئات الأقل تقدماً، وعلاج هذا العيب يزيد في تكلفة الإعداد، أي زيادة ظهور عيب أخر.

ثانياً: السلبيات الخاصة للاختبارات الحاسوبية المُتَكَيِّفة

١. **محدودية التطبيق.** تعتمد الاختبارات الحاسوبية المُتَكَيِّفة في تصميمها على نظرية الاستجابة على الفقرة التي لا يمكن تطبيقها على جميع المعلومات والمهارات التي يتعرض إليها عادة المُخْتَبِرون في مجال تعلّمهم.

٢. **صعوبة الإعداد.** يعاب على الاختبارات الحاسوبية المُتَكَيِّفةتحديداً تطلبها لجهد وخبرة ووقت كبير جداً لإعدادها بطريقة صحيحة متوافقة مع أهدافها ومبادئها النظرية (التي تقوم على مبدأ عرض السؤال المناسب لمستوى المُخْتَبِر العلمي) مقارنة بالاختبارات الحاسوبية والورقية التقليدية؛ فهي تتطلب عناية دقيقة في معايرة معالم الأسئلة.

٣. **الظهور المتكرر للسؤال.** انطلاقاً من الأساس الذي تبدأ به الاختبارات الحاسوبية المُتَكَيِّفة حيث تبدأ بالسؤال المتوسط الصعوبة، الأمر الذي يجعل الأسئلة الابتدائية أكثر انتشاراً وتوقعاً بين المُخْتَبِرين متشابهي المستوى العلمي خصوصاً وإذا سلّمنا بأن المُخْتَبِرين يتبادلون المعلومات بين بعضهم البعض.

٤. **عدم إمكانية مراجعة الإجابات.** يشتكي غالبية المُخْتَبِرين عن طريق الاختبارات الحاسوبية المُتَكَيِّفة من عدم تمكّنهم من مراجعة إجاباتهم عن الأسئلة السابقة قبل تصحيحها واعتماد الدرجة؛ فعدد أسئلة الاختبارات الحاسوبية المُتَكَيِّفة اللاحقة ومستوى صعوبتها يتحدداعتماداً على إجابة المُخْتَبِر للأسئلة السابقة؛ حيث تزداد صعوبة الأسئلة اللاحقة ويقل عددها كلما كانت الإجابة السابقة صحيحة، والعكس بالعكس.

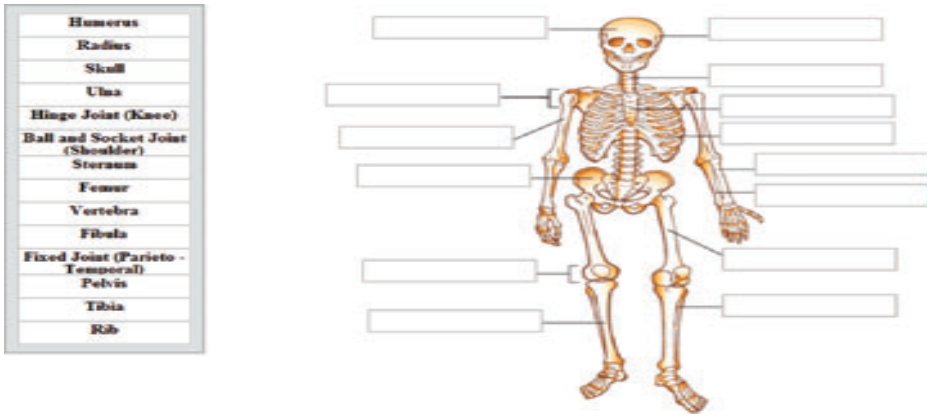
٥. **ضعف مستوى العدل.** يعتقد عدد من المُخْتَبِرين بأن اختلاف عدد ومستوى صعوبة أسئلة اختبار المادة العلمية الواحدة لمُخْتَبِرين المجموعة الواحدة يضع المساواة المطلوبة في الاختبارات عموماً والاختبارات المقننة خصوصاً على المحك ويعرضها للخطر.

٤ – التوجهات الحديثة في مجال الاختبار التكيفي المحوسب

كان الاختبار التكيفي المحوسب يعتمد أساسا على النماذج ثنائية الاستجابة **Dichotomous Models** لنظرية الاستجابة للمفردة لكونه يعنى بقياس مفردات الاختيار من متعدد والصواب والخطأ فقط، والتي تفترض أن الفرد إما يعرف الإجابة أو لا يعرفها وتهمل مبدأ الأخذ بالمعرفة الجزئية **partial Knowledge** لدى الفرد عند الإجابة عن أسئلة مفردات لمقررات، مثل: الرياضيات، أو الأحياء، أو الكيمياء، أو التطبيقات اللغوية، أو المهارات اليدوية، أو اختبارات الذكاء حيث يمكن للمصحح أن يمنح درجات للطالب على خطوات الحل الصحيحة بالإضافة للاستجابة النهائية، ولكن أنظمة التقويم المعاصر تتحول تدريجيا إلى تقويم الأداء **Performance Assessment** الذي يتميز بالثراء وتنوع مهامه ومفرداته الاختبارية، لذا أصبح من أهم التوجهات الحديثة للاختبار التكيفي المحوسب استيعابه لتطبيقات النماذج متعددة الاستجابات **polytomous Models** لقياس التحصيل والسمات الشخصية والاتجاهات، مثل:نموذج الاستجابات **المتدرجة Graded**، ونموذج الدرجات الجزئية **Partial-credit** في تطوير خوارزميات للاختبارات التكيفية المحوسبة المستمدة من بنك أسئلة يتم تعييره استنادا إلى هذه النماذج(علام،٢٠٠٥، ص٢٢٥)، لأن الاختبارات في هذه المجالات تتألف من مفردات لفظية تتضمن حلولاً جزئية للمشكلة المتضمنة في المفردة، وكذلك اختبارات الشخصية، ومقاييس الاتجاهات، وبناء عليه فقد أجرى الباحثون عدداً من الدراسات النفسية والتربوية التي وظفت النماذج المتعددة الاستجابات في سياق الاختبار التكيفي المحوسب **(Thompson, 2007;Passoset al. , 2007).**

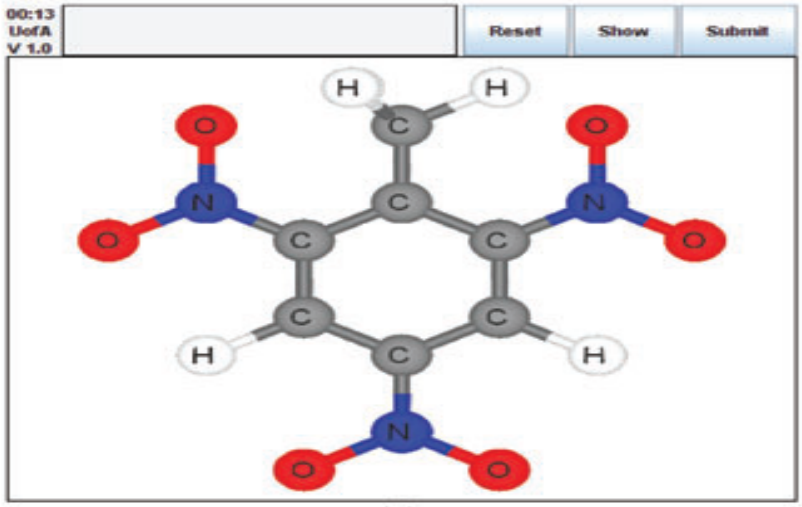
وقد أوضح بعض الباحثون بعض التطبيقات لهذه المنهجية، بما في ذلك إمكانية تطوير استبيانات تكيفية **محوسبة Computer Adaptive Questionnaires**، حيث تشتمل على فقرات مصممة لقياس التحصيل أو الاتجاهات أو الآراء، وتدريج هذه الفقرات، وانتقاء تلك التي تؤدي إلى تعظيم "المعلومات" المتاحة من الاستبيان (علام،٢٠٠٥، ص٢٢٥). كما أن استخدام التقنيات الحاسوبية مثل

الرسومات التفاعلية، والرسوم المتحركة، والألعاب التعليمية، والفيديو، والرسومات ثلاثية الأبعاد، والهاتف النقال في إعداد مفرداته سيساهم في إثرائه وجعله أكثر فعالية **(Zenisky and Sireci, 2002)**، وأيضا يساعد كلا من معدي المفردات والمربين في إعداد مفردات متعددة يمكن توليدهاتلقائياً وتتضمن تقدير قدرة الفرد بإعطاء علامات على المعلومات الجزئية من الإجابة الموجودة لدى الطالب، من منطلق أن القدرة على الاختبار ليست فقط معرفة الموضوع ولكن أيضا المهارات المعرفية المتضمنة فيه. كما أن استخدام الألعاب التعليمية في تصميم المفردات تمثل مصدر جذب لكي يكون الاختبار وسيلة تعليمية ويجعل الطلاب يشعرون كأننا نقدم لهم مكافأة، فيمكن أن تستخدم مفردات مبتكرة بالطرق السابقة لاختبار مجموعة متنوعة من العناصر، على سبيل المثال: الشكل رقم (٥) يوضح مفردات يمكن حلها بالسحب والإفلات عن طريق الماوس، مثلاً، لمقررات: (أ) مفردات لمقرر الأحياء، (ب)مفردة لمقرر الجغرافيا، (ج) مفردة لمقرر الكيمياء.



(أ) اسحب الاسم الصحيح في الفراغ المناسب

(ب) اسحب الاسم الصحيح في المكان المناسب



(ج) ادخل المعادلة الخاصة بالمركب

الشكل رقم (٥) مفردات من اختبار تكيفي محوسب يمكن حلها بالسحب والإفلات عن طريق الماوس لمقررات: (أ) الأحياء، (ب)الجغرافيا، (ج) الكيمياء

المصدر: **(Cheng and Basu ,2009)**

يتضح من المفردات في الشكل رقم (٥) أنها تسمح للطلاب بسحب النص أو الرسومات الموجودة في المفردة لمواقع الإجابة المناسبة على الشاشة، وهذا النوع من المفردات يختلف عن المفردات التقليدية من نوع الاختيار من متعدد

والصواب والخطأ وملء الفراغ، فمثلا في مفردات الاختبار من متعدد التقليدية يقوم معد الاختبار بوضع أربع بدائل يحدد احدها كإجابة صحيحة، ثم تكون هذه المفردة كقالب يغير فقط محتواه من مفردة لأخرى، أما الأمثلة السابقة فإنها تحتاج لمهارات إضافية في بناء المفردة وفي الإجابة عليها حسب متطلبات محتوى المفردة. كما أنها لا تهمل مبدأ الأخذ بالمعرفة الجزئية لدى الفرد وتساعد الطالب في الحصول على درجات على خطوات الحل الصحيحة بالإضافة للاستجابة النهائية، وهذا يوضح جانب من الإمكانات التي يمكن أن يوفرها دمج التطورات الحديثة في مجال تقنية المعلومات في صناعة وتطبيق الاختبار التكيفي المحوسب**(Cheng and Basu ,2009).**

ومما سبق نجد أنه بالاستفادة من إمكانيات الاختبارات التكيفية المحوسبة أصبح من الممكن قياس قدرات كان من غير الممكن قياسها بالاختبارات التقليدية وفي إطار النماذج التقليدية أو الحديثة أحادية البعد، حيث تساعد الاختبارات التكيفية المحوسبة المعلم على الاستفادة من إمكانية تضمين القياساتوفقاً للنماذج المتعددة في نظرية الاستجابة للمفردة، بالإضافة لتضمين الاختبار بالوسائط المتعددة (الصوت والصورة المتحركة أو الثابتة) لتوضيح المعلومة وتجعلها أقرب للواقع بطريقة لا توفرها النسخة الحاسوبية الخطية أو الورقية؛ لتقلل بذلك من أثر التصميم على درجة تقييم الُمْختَبِر.

٥- أهم التأثيرات الإيجابية لتطبيق الاختبار التكيفي المحوسب على المناهج التعليمية.

ذكر ديفيدسون وآخرون (٢٠٠٥) أن إحدى فوائد الاختبار التكيفي المحوسب تتلخص في أنه يتضمنتأثيراً إيجابياً في المناهج التعليمية، لأنه يزود الفردفورياً بالتغذية الراجعة الشاملة فأسئلة الاختبارات الفردية التي تتصل بمهارة محددة أو بمحتوى محدد من المعرفة يمكن أن تزود المتقدمين إلى الاختبار بالتغذية الراجعة المتصلة ببعض مناطق القوة والضعف لديهم، كما أن قاعدة

بيانات الاختبار تزود المعلمين بمعلومات أكثر حول قدرة طلابهم المعرفية، وذلك بهدف التكيف مع الاحتياجات الخاصة لدى الطلاب، كما يمكن أن يقارن مطوروا المناهج التعليمية هذا الاختبار ويحلّلوه من أجل إيجاد مناهج مناسبة تهدف إلى تعزيز عملية التعلم لدى الطالب،وأكد ديفيدسون وزملائه (٢٠٠٥) بأنه يجب ألا نستخف بقدرة تأثير الاختبار التكيفي المحوسب المتوقعة في عملية تطوير المناهج التعليمية، فمن المحتمل في المستقبل القريب أن يعطى الطلاب الاختبار النهائي في اليوم الأول من الفصل الدراسي، حيث ستحدد نتائج الطلاب في هذا الاختبار تفاصيل المنهج التعليمي الذي سيتم تدريسه، وذلكوفقاً لاحتياجات الطلاب المعرفية من أجل تحقيق الأهداف التعليمية المحددة المتوقع الحصول عليها في نهاية الفصل بالإضافة إلى تحديد المواد التعليمية التي يحتاجون إليها خلال هذا الفصل الدراسي.

وعلاوة على ما سبق أضاف علام (٢٠٠٥، ص٣٢٥) أنه مع توافر الوسائط الإلكترونية الجديدة، وانخفاض كلفتها، وزيادة مستوى الخبرة الفنية في مجال تصميم وبناء نظم الاختبارات التكيفية المحوسبة، سوف يؤدي ذلك إلى دمج تقنيات المعلومات مع تطبيقات العمليات الاختبارية، مما يسهم في انتشار هذه الاختبارات وزيادة مرونتها، وفعاليتها في القياس والتقويم في المؤسسات التربوية وغيرها من المؤسسات المجتمعية، وسوف ينعكس ذلك أيضا انعكاسا إيجابيا على تطوير وتحسين عملية التعليم.

٦-أهم العوامل التي أدت إلى تأخر اعتماد الاختبارات التكيفية المحوسبة في المؤسسات التعليمية في العالم العربي

إن التقييم باستخدام الاختبار التكيفي المحوسب في الاختبارات التربوية يحتاج إلى بيئة يمكن السيطرة عليها عند التطبيق، كما أن محتوى المفردة يجب أن يحاط بالسرية لتجنب الغش، وهذه الاعتبارات تستلزم توفر مواقع اختبار خاصة وبنك كبير للمفردات،وإجراءات صارمة للسيطرة على تقديم المفردات

(Bjorner,2004) وقد حدد علام(٢٠٠٥، ص٣٢٢)؛ أهم العوامل التي أدت إلى عدم اعتماد الاختبار التكيفي المحوسب على نطاق واسع في المؤسسات التعليمية في العالم العربي، بالنقاط التالية:

١. يعتمد تصميم وبناء الاختبارات التكيفية المحوسبةاعتماداً أساسياً على نظرية الاستجابة للمفردة؛ لذلك فإن جميع المشكلات المتضمنة في استخدام هذه النظرية تعد أيضا مشكلات لهذه الاختبارات والتي من ابسطها كونها تتطلب خبرة فنية عاليةوجهداً كبيراً، وتمويلاً كافياً .

٢. تتطلب إجراءات تطبيق الاختبارات التكيفية المحوسبة إنشاء بنك أسئلة يشتمل على عدد كبير من المفردات يبلغ عدة مئات تختزن في الحاسوب لتيسير انتقاء المفردات المناسبة لقدرة الأفراد من ذوي المستويات المختلفة، والتوصل إلى تمييزات دقيقة بينهم في القدرة المقاسة. وبناء هذه المفردات وتقدير بارامتراتها باستخدام أحد نماذج الاستجابة للمفردة يتطلب –كما ذكر سابقا–خبرة فنية عاليةوجهداً كبيراً، وتمويلاً كافياً . كما يتطلب توافر حواسيب، وتقنيات متقدمة، وبرمجيات حديثة لتطبيق الاختبارات، وتصحيحها، وتقديم نتائجها الفورية للأفراد أثناء الجلسات الاختبارية وتحليل هذه النتائج باستخدام الحاسوب. وهذه الخبرات والتقنيات ربما لا تكون متوافرة في كثير من الدول العربية ، مما يحد من انتشار استخدام الاختبارات التكيفية المحوسبة في الوقت الحاضر على الأقل.

٣. تحتاج المشكلات المتعلقة بتأثير سياق عرض المفردات، وترتيبها، وتوازن محتواها إلى مزيد من الدراسات والبحوث التطويرية، حيث إن هذه المشكلات، أشارت إليها كثير من الدراسات المتعلقة بالاختبارات التكيفية المحوسبة، وأوضحت ما يترتب عليها من تأثيرات سلبية متعددة في تقدير قدرة الأفراد المختبرين خاصة على طرفي متصل القدرة..

٤. عدم توافر البنى التحتية التي يستند إليها في تنفيذ برامج الاختبارات التكيفية المحوسبة على نطاق واسع **Large-Scale Computerized Adaptive Testing** . فهذه البنى تحتاج إلى تكلفة باهظة غير أن بعض مراكز الاختبارات في الولايات المتحدة الأمريكية وأستراليا استطاعت تأسيس هذه البنى التحتية، وقامت بتصميم صيغ جديدة من الاختبارات العالمية التنافسية التي تطبق على نطاق واسع.

مما سبق يمكن القول بأن الاختبار التكيفي المحوسب يتطلب بنية تحتية إلكترونية قادرة على استيعاب عدد الُمْختبَرين، كما تتطلب قدراً معيناً من المهارات الحاسوبية للُمْختَبرين. وهذان الشرطان يضيقان من مدى الاستفادة من تطبيقات الاختبار التكيفي المحوسب، ويجعلها متحّيزة ضد الدول والمنظمات ذات الإمكانات الحاسوبية المحدودة.

التوصيات والمقترحات

وختاماً ؛ حتى يمكن تفعيل تطبيقات الاختبار التكيفي المحوسب لتطوير أساليب التقويم والاختبارات الالكترونية في التعليم الجامعي، ويوصي الباحثان بما يلي:

١. ضرورة تطوير أساليب التقويم والاختبارات الالكترونية في التعليم الجامعي؛ حيث أنه من الأهمية بمكان اعتماد الاختبارات الالكترونية ممثلة في الاختبارات التكيفية المحوسبة، وتقترح الدراسة الحالية أن يتم البدءأولاً بتحويل اختبارات اللغة الانجليزية ومهارات الحاسب التي تقدم لطلاب السنة التحضيرية واختبارات المواد العامة التي تصنف كمتطلبات للجامعة، مثل: مواد الثقافة الإسلامية، واللغة العربية العامة، والسيرة النبوية إلى اختبارات تكيفية محوسبة ثم التوسع في ذلكوصولاً لاختبارات المواد الجامعية المختلفة، وعلى مستوى وزارة التعليم العالي أن يتم تحويل اختبار القدرات والاختبار التحصيلي الذي يقدمه المركز الوطني للقياس والتقويم في التعليم

العالي لخريجي الثانوية العامة إلى اختبار تكيفي محوسب، وذلك للإفادة من مزاياه المتمثلة في أن الأسئلة يتم تحديدها على أساس مستوى قدرة الطالب الفردية، كما أن فوائده تشمل خفض التكاليف الكبيرة في تطبيق الاختبار، والحد من وقت تطبيقه، ويتم تقدير مستوى الطالب بوقت أسرع بكثير، فهو اختبار فوري التسجيل، وجاهز تحت الطلب، كما أنه يوفر إمكانية استخدام الوسائط المتعددة في تصميم مفرداته ويجعلها قريبة من الواقع.

٢. أن تعمل الجامعات على إنشاء مراكز متخصصة في القياس والتقويم، تتوفر بها البنية التحتية والبرامج الخاصة بتطوير وتحليل الاختبارات وتضم كفاءات متخصصة للإشراف على تطوير منظومة التقويم في كل جامعة، لتدريب أعضاء هيئة التدريس على استخدام الأساليب المتقدمة في التقويم، وتقديم الدعم الفني، وتشجع على تبني البرامج المتقدمة في التقويم مثل الاختبارات تكيفية محوسبة.

٣. إجراء دراسات للمقارنة بين نتائج الطلاب على الاختبارات الورقية والاختبارات التكيفية المحوسبة تطبق محلياً في جامعاتنا على مواد، مثل: اللغة الانجليزية في السنة التحضيرية.

المراجع

أولاً : المراجع العربية

١. ديفيدسون وآخرون (٢٠٠٥).**التقييم في العالم العربي**، ترجمة: خالد بن عبد العزيز الدامغ، منشورة على شبكة الانترنت:
http://faculty.ksu.edu.sa/Khalid/Research_word_A/Forms/AllItems.aspx.Retrieved: 21/3/2011

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

ثانياً المراجع الأجنبية

3. **Al-harbi, K. A. (2003). An Empirical Investigation of Comparability and Invariance of Classical Test Theory and Item Response Theory Person and Item Parameter Estimates. Unpublished doctoral dissertation, Ohio: Ohio University.**
4. **Baker, F. B.; and Kim, S. (2004). Item Response Theory: Parameter Simulation Techniques. New York: Marcel Dekker, Inc.**
5. **Barnes, L. B.; and Wise, S. L. (1991). The Utility of a Modified One Parameter IRT Models with Small Samples. Applied Measurement in Education, 4,2, 143-157.**
6. **Bjorner, J.B. (2004). Developing tailored instruments: Item banking and computerized adaptive assessment. Paper presented at the conference "Advances in Health Outcomes Measurement: Exploring the Current State and the Future of Item Response Theory, Item Banks, and Computer-Adaptive Testing," Bethesda MD. From: http://www.psych.umn.edu/psylabs/catcentral. Retrieved: 25/3/2011**
7. **Bjorner, J. B., Chang, C. H., Thissen, D., & Reeve, B. B. (2007).Developing tailored instruments: item banking and computerized adaptive assessment. Quality of Life Research, 16(Suppl 1), 95–108.**
8. **Boeck, P. and Wilson, M. (2004). Descriptive and Explanatory Item Response Models. In Paul De Boeck, Mark Wilson. (Eds.), Explanatory Item Response Models: A Generalized Linear and Nonlinear Approach, 343-373, New York: Springer-Verlag.**
9. **Cella, D.; Gershon, R.; Lai, JS.; Choi, S.(2007). The future of outcomes measurement: item banking, tailored short-forms, and computerized adaptive assessment. Quality of Life Research, 16(Suppl 1), 133–144.**
10. **Chang, C.H. (2004). Developing tailored instruments: Item banking and computerized adaptive assessment. Paper presented at the conference "Advances in Health Outcomes Measurement: Exploring the Current State and the Future of Item Response Theory, Item Banks, and Computer-Adaptive Testing," Bethesda MD. From: http://www.psych.umn.edu/psylabs/catcentral.Retrieved:25/3/2011.**

11. Cheng, I. and Basu, A. (2009). Interactive Graphics for Computer Adaptive Testing. Computer Graphics Forum, 28,8, 2033–2045.
12. Choi, Seung W.; Grady, Matthew W.; and Dodd, Barbara G..(2011) .ANew Stopping Rule for Computerized Adaptive Testing. Educational and Psychological Measurement, 71, 1, 37-53.
13. Darabi, M. (1997). Paraeter Estimation in A Modified One-Parameter Item Response Theory Model Using Monte-Carlo Methods. Unpublished doctoral dissertation, Ohio: Ohio University.
14. Domino, George, and Domino, Marla L. (2006). Psychological testing: an introduction. Cambridge University Press.
15. de Ayala, Rafael J. (2008).The Theory and Practice of Item Response Theory. Guilford Press.
16. Fischer, G. H. (1997). Unidimensional Linear Logistic Rasch Models. In Wim J. van der Linden and Ronald K. Hambleton (Ed.), Handbook of Modern Item Response Theory, 225-245, New York: Springer.
17. Fox, J. P.; and Glas, C. A.. W. (2003). Bayesian Modeling of Measurement Error in Predictor Variables Using Item Response Theory. Pychometrika, 68, 169-191.
18. Hambleton, R. K.; and Swaminathan, H. (1985). Item Response Theory: Principles and Applications. The Netherlands: Kluwer Nijhoff Publishing.
19. Hambleton, R. K.; Swaminathan, H.; and Rogers, H. J. (1991). Fundamentals of Item Response Theory. California, Newbury Park: Sage Publications, Inc.
20. Kaplan, Robert M. and Saccuzzo, Dennis P. (2008).Psychological Testing: Principles, Applications, and Issues, Seventh Edition.USA: Cengage Learning
21. Lawson, D. M.(2006). Applying the Item Response Theory to Classroom Examinations. Journal of Manipulative and Physiological Therapeutics, 29,5, 393-397
22. Linacre, J.M. (2000). Computer-adaptive testing: a methodology whose time has come. MESA memorandum no. 69. Published in SunheeChae, Unson Kang, EunhwaJeon and J. M. Linacre. Development of computerised middle school achievement test, Seoul, South Korea: Komesa
23. Linden V., Hambleton R. K. (1997). Handbook of Modern Item Response Theory. New York: Springer
24. Lovler, Robert L., Miller, Leslie A., and McIntire, Sandra A. (2011). Foundations of Psychological Testing: A Practical Approach ,Third Edition. SAGE Publications, Inc.
25. Patsula, l. (1995). A Comparison of Item Parameter Estimates and ICCs Produced with TESTGRAF and BILOG Under Different Test Lengths and Sample Sizes. Unpublished thesis, Ottawa: Ottawa University.
26. Passos, V. L., Berger, M. P. F., & Tan, F. E. (2007). Test design optimization in CAT early stage with the nominal response model. Applied Psychological Measurement, 31, 213-232.
27. Ramsay, J.O. (2000). TESTGRAF: A Program for the Graphical Analysis of Multiple Choice Test Data. [Technical Manual]. Montreal: McGill University. From: http://ego.psych.mcgill.ca/pub/ramsay.Retrieved:12/10/2010.

28. Reckase, M. D. (2007). The design of p-optimal item bank for computerized adaptive tests. In D. J. Weiss (Ed.). Proceedings of the 2007 GMAC Conference on Computerized Adaptive Testing. From: www.psych.umn.edu/psylabs/CATCentral.Retrieved: 25/3/2011
29. Setiadi, H. (1997). Small Sample IRT Item Estimates. Unpublished doctoral dissertation, Amherst: Massachusetts University.
30. Thissen, D. (2003). MULTILOG 7: Multiple categorical item analysis and test scoring using item response theory. [computer program]. Chicago, IL: Scientific Software.
31. Thomas, R. L. (1997).Amonte Carlo Investigation of Parameter Estimation Efficacy Using Modified Fixed "C" Three Parameter log(3pl) Item Response Theory Models with Small Sample Sizes. Unpublished doctoral dissertation, Michigan: Wayne State University.
32. Thompson, Nathan A.(2007). A comparison of two methods of polytomous computerized classification testing for multiple cutscores. Unpublished doctoral dissertation, MINNESOTA: UNIVERSITY OF MINNESOTA.From: http://proquest.umi.com/pqdlink?ver=1&exp=03-29-2016&fmt=7&did=1335354 151&RQT=309&attempt=1&cfc=1.Retrieved:25/3/2011.
33. Triantafillou, E., Georgiadou, E., & Economides, A. A. (2008). CAT-MD: Computerized adaptive testing on mobile devices. International Journal of Web-Based Learning and Teaching Technologies, 3, 1, 13-20.
34. Truskosky, D. M. (1999). An Empirical Examination of Classical Test Theory and Item Response Theory Parameters: Implications for Research and Practice in Small- and Large-Sample Assessments. Unpublished doctoral dissertation, Illinois: Southern Illinois University at Carbondale .
35. Wainer, H., &Mislevy, R.J. (2000). Computerized Adaptive Testing: A Primer. Mahwah, New Jersey : Lawrence Erlbaum Associates.
36. Wells, C. S.(2004). An Investigation of Model Misfit in IRT and a New Approach Based on Simultaneous Parametric and Nonparametric IRF Estimation. Unpublished doctoral dissertation, Madison: Wisconsin University.
37. Weiss, D. J., & Kingsbury, G. G. (1984). Application of computerized adaptive testing to educational problems. Journal of Educational Measurement, 21, 361-375.
38. Zenisky A. L., Sireci S. G. (2002). Technological Innovations in Large-Scale Testing. Applied Measurement in Education, 15, 4, 337–362.
39. Zimowski, M. F.; Muraki, E.; Mislevy, R. J.; and Bock, R. D. (2003). BILOG-MG3: Multiple-Group IRT Analysis and Test Maintenance for Binary Item. [computer program]. Chicago, IL: Scientific Software. From: http://www.assess.com.Retrieved:6/7/2010.