مشكلات تقويم التحصيل الدراسي

في ضوء النظرية الكلاسيكية

الأستاذ بغداد ابراهيم محمد، جامعة تيارت

بإشراف الأستاذ الدكتور حبيب تليوين، جامعة أحمد بن بلة وهران 2

الملخص:

يختصر التقويم ألتحصيلي في المؤسسات التربوية على الاختبارات الكتابية التي تّعد من طرف الأستاذ وبالتالي تفتقد هذه الاختبارات للخصائص الموضوعية من حيث الشمولية والصدق والثبات. وأي اختبار صُمّم في المجال التربوي أو النفسي، ينتج عنه مشكلات عديدة: مشكلات نظم ألامتحانات، مشكلات تتعلق بتفسير درجات الأفراد، مشكلات نظرية تتجلى في أن درجات الأفراد تتباين بتباين مستويات صعوبة المفردات وكذلك تتباين معاملات الفقرات بتباين قدرات الإفراد المفحوصين. وللتغلب على هذه المشكلات لجأ تقويم التحصيل الدراسي إلى تبنى القياس الموضوعي والاختبارات المقننة، والاعتماد على التقويم المستمر والتقويم الذاتي وتقويم أنشطة الطلاب واعتماد سجلات الأداء وتبنى الاختبارات العملية أو الأدائية ما أمكن حسب طبيعة كل مادة وذلك لتكون عملية القياس شاملة ولا تقتصر على الاختبارات الكتابية، ولكن في الممارسات اليومية و الامتحانات الرسمية لا يمكن الاستغناء عنها.

كل ذلك دفع بالقياس النفسي والتربوي إلى تبنى الاتجاهات المعاصرة في القياس والاعتماد على النظرية الحديثة ونماذج الاستجابة للمفردة الاختبارية وبناء بنوك الأسئلة[1] التي توفر موضوعية القياس، حيث برهنت هذه النظرية على قدرتها على تحقيق الموضوعية وتوفير الشروط السيكومترية التي تجعل النتائج التي نحصل عليها أساسا لإصدار قرارات صائبة حول الأفراد،

والتغلب على مشكلات القياس الكلاسيكي ومنها تباين درجات الأفراد بتباين مستويات صعوبة المفردات وكذلك تباين معاملات الفقرات بتباين قدرات الأفراد المفحوصين.

الكلمات المفتاحية: الاختبار، التحصيل الدراسي، النظرية الكلاسيكية، نظرية الاستجابة للمفردة.

Abstract:

In educational assessment, the major tool to measure the knowledge is the paper-pencil test, which is administered by the teachers, and therefore it lacks objective characteristics in terms of comprehensiveness, validity and reliability. And after any test was designed in the educational or psychological testing, many problems raised: examination's system, explanation of the scoresof individuals and theoretical problems as an example:scores vary depending on the levels of difficulty of the items as well as these latter vary depending on the abilities of individual subjects.

To overcome these problems, educational assessment adopt an objective measurement and standardized tests, depending on continuous assessment and self-evaluation and evaluation of students' activities and the adoption of performance tests and also a practical tests or performing as much as possible depending on the nature of each courses' education. So as the assessment to be inclusive measurement process, and not limited to the written tests; but in daily practices and official exam cannot do without.

This situation actuate psychological and educational assessment to the adoption of contemporary trends in measurement and rely on modern item response theory models and construct item banks which provide objective measurement. This theory has demonstrated its ability to achieve the objectivity and provide psychometric properties that make the results that we get basis for the right decisions about individuals, and to overcome the classical measurement problems, including the scores of individual variation depending on levels of the item's difficulty, as well as the coefficients of these latter depend on individual abilities.

Key words: test, educational achievement, classical test theory, item response theory.

مقـــــدمة:أن الهدف من التقويم هو جمع مجموعة من المعلومات الصادقة والثابتة، وتحليل درجة وملاءمتها لمعايير خاصة بالأهداف المحددة في البداية بهدف اتخاذ القرار[2].

تهدف الاختبارات التحصيلية إلى قياس مدى تحقيق الأهداف ألتعليمية وهي إما اختبارات عادية تعد بواسطة الأساتذة أو اختبارات عامة تعدها الوزارة فى نهاية كل مرحلة دراسية، وبخاصة الثانوية العامة[3]. ولكي تكون نتائج التحصيل متسقة وصادقة وغير متحيزة لابد من أن يرتكز قياس التحصيل على بعض الافتراضات وهي:

1. أن يقيس الاختبار نطاقاً سلوكياً يمكن تحديده بدقة، بمعنى توفر أهداف تربوية تصاغ فى صورة سلوكية.

2. أن يقيس الاختبار الأهداف المتعلقة بالمادة الدراسية.

3. أن تتاح لجميع التلاميذ فرص تعلم المجال الذي يقيسه ألاختبار، خاصة عند استخدام الاختبارات المقننة[4].

يتضح مما سبق أن الاختبار التحصيلي هو أحد أدوات القياس والتي تقدم معلومات عن مدى اكتساب واستيعاب المتعلم للخبرات المعرفية بمستوياتها المختلفة والمهارات التى تم دراستها، وفى ضوء ما تسفر عنه نتائج الاختبارات تستمر عملية التقويم حتى يتحقق الهدف منها، وعليه فإن المشكلات المتعلقة بالاختبارات التحصيلية تنطبق على ألتقويم وعليه فإن الهدف من الدراسة الحالية تناول المشكلات المتعلقة باختبارات التحصيل الدراسي والتي تعرضت لها العديد من الدراسات السابقة.يمكن إيجاز مشكلات تقويم التحصيل الدراسي فيما يلي:

I. مشكلات نظم الامتحانات:

يقتصر التقويم في المؤسسات التعليمية على قياس التحصيل من خلال الاختبارات التحصيلية باعتبارها الأداة الأساسية وإن لم تكن الوحيدة، والتي يعتمد عليها في إصدار الأحكام واتخاذ القرارات التربوية بشأن مستويات التلاميذ والانتقال من صف إلى آخر ومن مرحلة دراسية

إلى أخرى. فقد أشارت نتائج دراسة طلعت الحاموليّ[5] والتي هدفت إلى تقييم القرارات الوزارية المتعلقة بسياسة التعليم في الفترة من 1985 م إلى 1990 م إلى أن سياسة الوزارة التقويمية تشير إلى أن نظام التقويم هو نظام امتحان التلاميذ مما يفقد عملية التقويم خصائصها المتمثلة في "الشمول . التكامل . الاستمرارية".وإذا كان ما يحدث في مدارسنا هو مجرد قياس التحصيل باستخدام الاختبارات، فإن عملية تحديد مواطن الضعف وعلاجها، وتحديد مواطن القوة لتدعيمها لم تكن ممارسة في الواقع التعليمي وذلك لعدم توفر تغذية راجعة حقيقية. وتؤكد نتائج بعض الدراسات أن الاختبارات التحصيلية هي أكثر أدوات التقويم شيوعاً لقياس التحصيل، كما قدمت هذه الدراسات إسهامات تفيد في تطوير الاختبارات التحصيلية لتكون أداة مناسبة في تقويم مخرجات التعلم.[6]

يعاني النظام الحالي للامتحانات من قصور في تحقيق الأهداف التعليمية ويرتبط بمجموعة من المشكلات والمظاهر السلبية حددتها نتائج بعض الدراسات والبحوث ومنها:

1. تعتمد كثيرُ من الاختبارات على الأسئلة المقالية بدرجة كبيرة مما يجعل صعوبة في عملية التصحيح واتخاذ القرارات الصائبة، وأحيانا على الأسئلة الموضوعية مما يجعل الأهداف التعليمية تقتصر على المستويات الدنيا فقط.[7]

2. ارتفاع مستوى القلق والخوف المرتبط بالاختبارات ووجود هدر مالي كبير نتيجة إتاحة فرص أوسع للدروس الخصوصية.[6]

3. الاقتصار على الاختبار النهائي الذي يعقد في نهاية الفصل الدراسي وعدم توزيع الدرجات على مدار العام.

4. يؤدى النظام الحالي للامتحانات إلى انتشار الكتب الخارجية والدروس الخصوصية مما يؤدى إلى هدر الاقتصاد الوطني.[6]

5. كما توضح نتائج بعض الدراسات أن الاختبارات في معظم الدول العربية تفتقر إلى الصدق والشمول والموضوعية وتركز في معظمها على المستويات المعرفية الدنيا وإهمال المستويات المعرفية العليا.[2]

ومن أبرز المشكلات التي ترتبط بالاختبارات التحصيلية هي ما تتعلق بما تقيسه أسئلة الاختبارات ونوعها والتي توجز فيما يلي:

- أن أسئلة الاختبارات التحصيلية لا تعطى صورة حقيقية عما تقيسه من أهداف تعليمية، بمعنى عدم صدق الاختبارات في قياس الأهداف التعليمية كما تعكسها أهداف المحتويات الدراسية.
- لا تقيس الاختبارات الأهداف المهارية والوجدانية.
- لا تراعى كثيرُ من الاختبارات الوزن النسبي والأهمية النسبية لكلٍ من موضوعات المحتوى والأهداف التعليمية وتوزيع أسئلة الامتحانات على المحتوى.
- تأثر الدرجة الكلية للمتعلم في اختبار ما بمفرداته .
- عدم توافر وحدة قياس ثابتة .
- عدم إمكانية المقارنة بين تحصيل المتعلمين في حالة استخدام اختبارات تحصيلية مختلفة.

وحول هذه المشكلات أجريت العديد من الدراسات التي اهتمت بتحليل أسئلة الاختبارات التحصيلية لبعض المواد الدراسية بهدف تشخيص الوضع الراهن للامتحانات من أجل تطوير الامتحانات ومراعاة الأسس والمعايير إلى يجب أن تكون عليها أسئلة الاختبارات ومن بين هذه الدراسات: تلك الدراسات التي قام بها المركز القومي للامتحانات والتقويم التربوي بمصر منذ نشأته 1990م لتقديم حلول علمية لمشكلات التقويم بصفة عامة والامتحانات بصفة خاصة حيث تم إعداد العديد من المشاريع البحثية كدراسات تقويمية إما لأدوات تقويم الطالب أو تقويم الأوراق الامتحانية للوقوف على جوانب القوة وجوانب الضعف في سبيل تطوير أدوات التقويم، إيماناً بأن إصلاح التقويم يؤدى إلى إصلاح التعليم. وفيما يلي نتائج بعض هذه الدراسات:

دراسة فكرى ريان وآخرون(20) والتي سعت لتقويم نماذج أسئلة المقررات الدراسية لجميع المواد الدراسية فى المرحلة الثانوية، وذلك بهدف الارتقاء بها حتى تكون أداة فعالة لتطوير المنهج وتحقيق أهدافه وتطوير الأسئلة لاستخدامها فى بناء بنوك الأسئلة، وقد أسفرت النتائج عن أن نماذج الأسئلة في العام 1991/90 م تركز في معظم المواد الدراسية على المستويات الثلاثة الأولى للأهداف المعرفية "التذكر . الفهم . التطبيق" مع قدر محدود لقياس المستويات العليا للأهداف المعرفية، وتضمنت نماذج الأسئلة نوعى الأسئلة المقالية والموضوعية.

كما أن الغالبية العظمى من الأساتذة والتلاميذ يعتبرون نماذج أسئلة المقررات ـ دليل التقويم ـ أنها الرديف الحقيقي لاختبار نهاية العام والإطار المرجعي لها، وكونها تركز على المستويات الدنيا للأهداف المعرفية، فإن اختبارات نهاية العام سوف تركز على هذه المستويات مما لا يحقق أهداف العملية التعليمية بجميع مستوياتها، فتطوير نماذج الأسئلة يسهم في تطوير اختبارات نهاية العام.

وتجدر الإشارة إلى أن التذبذب في نسب تمثيل الأهداف المعرفية بمستوياتها في الاختبارات ربما يرجع لعدم الالتزام من قبل واضعي الاختبارات بإستراتيجية واضحة مبنية على أهداف محددة، تأخذ بعين الاعتبار مشكلات التقويم السابقة لتلافيها بهدف تطوير نظم الامتحانات في الوضع الراهن والمستقبلي، وهذه إشكالية الاختبارات التقليدية أو المعتادة التي هي أسيرة النظرية الكلاسيكية.

وتُعتبر دراسة وتقييم الورقة الامتحانية ونماذج الأسئلة ودليل التقويم، يؤدى إلى نتائج تعد مدخلاً جيداً لتشخيص واقع الاختبارات والتعرف على جوانب الضعف وما يتصل بها من مشكلات لاقتراح الحلول المناسبة لتطوير الاختبارات مستقبلاً مم يسهم في تطوير نظم التعليم وإصلاحها.

ومن الدراسات الأخرى دراسة أمينة كاظم وآخرون[8] والتي أوضحت أن أدوات التقويم المستخدمة في الدول العربية "دول الخليج ومصر" تتجسد في الاختبارات التحريرية المقالية

والموضوعية بأنماطها المختلفة وبنسب متفاوتة، وتعتمد أحياناً على الاختبارات الشفهية أثناء الفصل الدراسي من خلال التقويم المستمر، ونادراً ما تستخدم الاختبارات الأدائية أو العملية، كما أن البطاقة المدرسية غير مفعّلة، ويتم التركيز في تقويم التحصيل الدراسي حول الجانب المعرفي على مستوى التذكر وإن تناولت مستوى الفهم والتطبيق فبدرجة أقل، ونادراً ما تشمل الاختبارات على المستويات العليا للأهداف المعرفية.

ويضيف فؤاد أبو حطب[9] أن الأسئلة التي تقيس النواحي المعرفية هي نوعان هما:

- أسئلة إنتاج الاستجابات مثل " الإجابات القصيرة . ملء الفراغ . أسئلة المقال الطويلة."
- أسئلة انتقاء الاستجابة وهى أسئلة موضوعية "الاختيار من متعدد . المزاوجة . الاختيار من بديلين".

ومن أهم المشكلات المرتبطة بهذه الأنواع والتي تؤثر في التحصيل هي مشكلة كتابة وصياغة الأسئلة، لأن عدم وضوح السؤال ودقته يؤدى إلى فهمه بطريقة تؤدى إلى إجابة خاطئة. وهذا ما أكدته نتائج دراسة حمدي البنا[10]، من أن أسئلة امتحانات الكيمياء للثانوية العامة في الأعوام 1998، 1999، 2000 م لا تعطى صورة حقيقية لما تقيسه من أهداف، حيث إن ما يظهر من صياغتها يختلف عما تقيسه بالفعل.

وأسفرت نتائج دراسة رضا درويش[11] حول تقويم "دليل تقويم التلميذ في مادة العلوم للصف الثالث الإعدادي" إلى شمول الدليل على أسئلة لقياس الأهداف المعرفية بدرجة كبيرة، بينما شمل الدليل على أسئلة لقياس الجانب المهاري بدرجة ضئيلة جداً، ولم يتضمن الدليل أسئلة لقياس أهداف الجانب الوجداني، وبالنسبة للأهداف المعرفية كان التركيز بدرجة كبيرة على أسئلة لقياس قدرة التلميذ على التذكر والفهم، بينما شمل الدليل على أسئلة لقياس قدرة التلميذ على التطبيق والتحليل والتركيب والتقويم بنسب ضئيلة للغاية، وتحقق المعيار الخاص بتنوع الأسئلة بين المقالية والموضوعية، كما اشتمل الدليل على أنواع متعددة من الأسئلة الموضوعية ولكن لم يتحقق المعيار الخاص بالتوازن حيث حازت الأسئلة الموضوعية على معظم الاهتمام.

يتضح مما سبق أن الغالبية العظمى من أسئلة الاختبارات أو أدلة التقويم أو الأسئلة المتضمنة في الكتب باعتبارها أحد المصادر الأساسية لأسئلة الاختبارات تركز على المستويات المعرفية الدنيا وتركز على الحفظ والتذكر مما يؤثر في جودة مخرجات التعلم.وهكذا نستنتج أنّ الاختبارات التي يعدّها الأساتذة، أو تلك التي موجودة في الدليل، تفتقد للموضوعية.

ويذكر علام [3] أن الموضوعية يقصد بها أن درجة الفرد فيالاختبار لا يجب أن تعتمد على عينة الأفراد ، ففي الاختباراتالنفسية المقننة تشتق الدرجات الخام من عينة التقنين ثم يوازنأداء الفرد الذي يطبق عليه الاختبار فيما بعد بمعايير مشتقة منهذه العينة ، فإذا تغيرت العينة فقدت هذه المعايير دلالتها ، أبيصبح الاختبار محكوماً بالعينة ، كما أن الفرد يحصل على الدرجةنفسها في أي اختبارين يقيسان نفس السمة ، إلا أن درجة الفردتختلف عادة باختلاف الاختبار الذي يطبق عليه ، أي أن درجته تصبحمحكومة بعينة المفردات التي يختبر بها .

وسعيا للقضاء على مشكلات تقويم التحصيل الدراسي وتطوير الورقة الامتحانية، قدمت بعض الدراسات وأوراق العمل تصورات مقترحة ومشاريع نماذج لتطوير الامتحانات كأدوات لتقويم التحصيل الدراسي ومنها:دراسات كل من أمينة كاظم وآخرون[7] وعبد الوارث الرازحي[2] وأوراق عمل رجاء أبو علام [5]، والتي أكدت على ضرورة تبنى القياس الموضوعي والتقويم المستمر إضافة إلى التقويم النهائي وأن يتضمن الاختبار النهائي أسئلة مقالية ذات إجابات محددة وأخرى موضوعية متنوعة، وتبنى مشروع بنوك الأسئلة والتركيز على المستويات العقلية العليا، مع الاعتماد على التقويم من خلال الأنشطة التي ترتبط بمواقف الحياة الواقعية.

ويضيف حسين بشير[12] ضرورة استخدام سجلات الأداء التي تضم عينات من عمل التلاميذ ومساعدة التلميذ على التقويم الذاتي على اعتبار أن التقويم عملية تعاونية يشترك فيها كل من لهم علاقة بهذه العملية.وهذا يدلّ على أنّ التقويم الذي هو شائع عند أغلب الأساتذة يتلخص في الاختبارات الكتابية والتي تفرضها الرزنامة السنوية للتقويم والتي تعتبر قطعية أي أنها تحقق ما ينتظر منها وهو الحصول على درجات المتعلمين وقائم على ترتيب التلاميذ من الأحسن إلى الأضعف دون النظر إلى الأهداف المرجو تحقيقها.

II. النظرية الكلاسيكية للاختبارات Classical Tests Theory:

من المناسب أن نبدأ هذا الفصل بالتساؤل التالي: ما معنى النظرية الكلاسيكية ؟ التعريف الاجرائي الوحيد جاء به Harold Gulliksen's (1950)عند حديثه عن نظرية الاختبارات العقلية. حيث لم يقدم مصطلح الكلاسيكي لوصف الموضوع لأنه لم يكن هناك نظرية أخرى للتمييز بينها، ومن جهة أخرى وبعد مرور ثمانية عشرة سنة، كان Fred Lord و Mel Novick بحاجة للتمييز بين النظريات الإحصائية لنتائج الاختبارات العقلية، وبالتالي ظهر مصطلح النظرية الكلاسيكية. تهدف هذه النظرية إلى تحديد العوامل المؤثرة على درجة الفرد والتي تسبب ما يسمى خطأ القياس، ولتحقيق ذلك استندت النظرية على مجموعة من الفروض منها[3]:

درجة الفرد التي يحصل عليها ليس من الضروري أن تمثل درجته الحقيقية، فهي قابلة للتغير حسب ظروف الموقف.

1. درجة الفرد هي نتاج نوعين من الدرجات "الدرجة الحقيقية . الدرجة الخاطئة".
2. الدرجة الحقيقية يفترض أن تكون ثابتة لأنها تمثل قدرة الفرد المقاسة.
3. زيادة خطأ القياس يقلل من الدرجة الحقيقية والعكس صحيح.
4. الدرجة الحقيقية يمكن معرفتها من خلال تكرار تطبيق الاختبار عدة مرات، وهى عبارة عن متوسط مرات التطبيق.
5. لا يوجد اقتران بين الدرجات التي يحققها الأفراد وبين الدرجات الخاطئة.
6. صعوبة إثراء الاختبارات المعدة باستخدام النظرية الكلاسيكية بمفردات جديدة .

كما ارتكزت النظرية الكلاسيكية للاختبار في القياس النفسي على مجموعة من القواعد هي:[13]

1. الخطأ المعياري للقياس ينطبق على جميع الدرجات في عينة محدودة.

2. الاختبارات الطويلة تكون أكثر ثباتاً من الاختبارات القصيرة.

3. تكون مقارنة درجات الاختبار أفضل عندما تتوفر عدة صورة متكافئة.

4. التقديرات غير المتحيزة Unbiased لخصائص المفردة تعتمد على مدى تمثيلها للعينة.

5. تكتسب درجات الاختبار مدلولها من خلال مقارنة وضعها بمعيار المجموعة الواحدة.

6. خصائص مقاييس المسافة Interval Scalesتتحقق بالحصول على توزيع إعتدالي.

7. الصيغ المختلطة للمفردات تؤدى إلى أثر غير متوازن على الدرجة الكلية للاختبار.

والخطأ في القياس الكلاسيكي يرجع لعوامل متعددة منها ما يرتبط بالمفحوص أو ظروف التطبيق أو بيئة المفحوص وما يحيط بها، ويفترض أن هذا الخطأ يتوزع اعتدالياً، وقد تطورت الأساليب الإحصائية للوصول إلى معادلات إحصائية ملائمة للحد من هذا الخطأ، وعلى الرغم من ذلك فإن هناك مشكلات أخرى ترتبط بهذه النظرية دلت عليها البحوث السابقة ومنها:

1. صعوبة إحداث تكافؤ بين عدة أنواع من الاختبارات.

2. صعوبة تحديد مصادر خطأ القياس خاصة ما يرتبط منها بخصائص أداة القياس.

3. اعتماد الخصائص السيكومترية للاختبارات على خصائص عينة التقنين، وعلى خصائص عينة المفردات، مما يؤدى إلى قياس غير مستقر.[14]

وقد أُجريت محاولات لتقليل الخطأ التجريبي في القياس الكلاسيكي منها: الاختيار العشوائي للأفراد وتحديد المعايير واستخدام الأساليب الإحصائية الملائمة ، وذلك لأن الدرجة التي يحصل عليها الفرد على اختبار ما لا تعبر عن مستوى قدرته الفعلية، بل تخضع لخطأ القياس، حيث إن درجة الفرد قد تزيد **أو تقل عن مستوى قدرته،** فالدرجة الملاحظة التي يحصل عليها الفرد تتضمن جزأين هما الدرجة الحقيقية ودرجة الخطأ والتي يعبر عنها بالمعادلة التالية: $X = T \pm E$.

حيث (E) تعبر عن الخطأ ، (T) تعبر عن الدرجة الحقيقية، (X) تعبر عن الدرجة التجريبية أو الملاحظة[15].

وتجدر الإشارة إلى أن الأحكام التي تصدر بشأن مستويات التلاميذ التحصيلية، قد لا تكون دقيقة بسبب مقدار خطأ القياس والذي يمكن الحد منه باعتماد القياس الموضوعي الذي يعتمد على النظرية الحديثة فى القياس. بينما يؤكد على أنه يجب الاعتماد على أكثر من معيار عند تفسير درجة التلميذ على الاختبار وعدم الاقتصار على معيار واحد فقط من المعايير التالية:[16]

1. المعيار السيكومتري "موقع الفرد بالنسبة لأقرانه".
2. المعيار الأديومتري "موقع الفرد بالنسبة لمحك محدد تحديداً دقيقاً".
3. موقع الفرد بالنسبة لنفسه من وقت لآخر.
4. موقع الفرد بالنسبة لما يطمح أن يكون عليه مستواه.

ويوصى كل من (Hambleton & Wedman,)[17] بضرورة التخلص من المشكلات المرتبطة بالاختبارات المعتادة التقليدية والتي تتمثل في عدم موضوعيتها وعدم قدرتها على تقديرات النطاق الواسع للمحتوى، كما أنها لا توفر تقويماً تشخيصياً جيداً، وذلك من خلال تبنى الاتجاهات الحديثة في القياس النفسي التي تمد الممارسين بأطر تسهم في تطوير عملية التقويم وتوفر موضوعية القياس وتركز على تقويم الكفاءات Competencies بدلاً من تقويم الأداء، والاستعانة بالتكنولوجيا الحديثة فى تطوير النماذج السيكومترية الحديثة التي تسهم في تطوير طرق بناء الاختبارات وتحليل بنياتها.ويضيف كل من (Hambleton & Slater)[18] أن الاختبارات في الوقت الحالي تمر بتغيرات أساسية، وإحدى هذه التغيرات هو الانتقال من الاعتماد على القياس الكلاسيكي إلى تبنى النماذج الحديثة للتغلب على عيوب الاختبارات التقليدية.

وتجدر الإشارة إلى أن علماء القياس النفسي لم يقفوا أمام هذه المشكلات مكتوفي الأيدي، بل كانت هناك جهودا كبيرة للتغلب على هذه المشكلات والتي تؤثر في الخصائص السيكومترية

للاختبارات التحصيلية، مما يؤثر في دقة وصحة النتائج، فقد تم التوصل إلى نماذج جديدة تجعل القياس موضوعيا بمعنى عدم تأثر درجة الفرد بخصائص عينة الأفراد ولا خصائص عينة الأسئلة المكونة لأداة القياس وهو ما يسمى بنماذج نظرية الاستجابة للمفردة.

III. الأنظمة المرجعية في القياس التربوي والنفسي

من المشكلات التي ترتبط بالنظرية الكلاسيكية هي مشكلة تفسير درجة الفرد التي يحصل عليها في اختبار ما والتي تتوقف على المعيار أو المحك الذي يكسب الدرجة معنى أو دلالة، وفى إطار هذه النظرية تم بناء اختبارات مرجعية المحك وأخرى مرجعية المعيار ولكل منها مشكلاتها التي ترتبط بالتحصيل الدراسي والتي يتم تناولها فيما يلي:

1. النظام مرجعي المعيار Norm-Referenced System:

يقوم القياس مرجعي المعيار على تقدير الفروق بين الأفراد والتمييز بينهم في الأداء، حيث تكتسب درجة الفرد على الاختبار معناها عندما تسند إلى معيار مستوى الجماعة التي ينتمي إليها الفرد [19]، ويحدد مستوى الجماعة بمتوسط درجاتها على الاختبار، ويتغير هذا المعيار بتغير الجماعة، حيث تعتمد فلسفة القياس على التمييز بين الأفراد بهدف الانتقاء والاختيار، وتعد الاختبارات في هذا النظام هي الأقدر على التمييز بين الأفراد[20]. وتجدر الإشارة إلى أن الاختبارات في ظل القياس مرجعي المعيار تعتمد على الفروق الفردية، حيث تتوزع درجات الأفراد في ظل النظرية الكلاسيكية توزيعاً اعتدالياً.والاختبارات مرجعة المعيار تكون فائدتها واضحة في الحالات التالية:[21]

•عندما يكون الموضوع الدراسي ليس بشكل متراكم أو سلسلة مترابطة يعتمد بعضها على البعض الآخر.

•في الحالات التي لا يلزم الطلاب الوصول لمستوى محدد من الكفاءة.

•عندما تكون بحاجة لاختبار عدد من الأفراد المتميزين.

•عندما يكون الهدف التنبؤ بدرجة النجاح للطلاب، لأنها تعتمد على الفروق الفردية.

•عندما يكون هناك حاجة لقياس عمليات عقلية عليا .

وعلى الرغم من كل هذه الفوائد وما يتميز به المعيار السيكومتري من الخصائص السيكومترية من موضوعية وصدق وثبات وقدرة على التمييز إلا أن عيوب الاختبار أو مشكلاتها في ضوء القياس مرجعي المعيار يحددها Glaser 1963فيما يلي:

■ يقتصر على التمييز بين الأفراد ولا يتجاوز ذلك إلى معرفة ماذا تحقق من أهداف العملية التعليمية، مما يؤدى إلى حذف عناصر ووحدات من الاختبار إذا كانت غير مميزة بين الأفراد بغض النظر عن قيمتها مما يؤثر في بناء الاختبار.

■ الاقتصار على اتخاذ الجماعة كمحك لتفسير درجة الفرد على الاختبار يؤدى إلى عدم التيقن من موقع الفرد وجماعته بالنسبة لأهداف التعلم ومدى تحقيقها.

■ سطوة المنحنى ألاعتدالي المعياري باعتباره النموذج الأمثل لتوزيع الدرجات، وهو الذي يتحقق عندما يكون مستوى الأسئلة في مستوى التلاميذ، وعندما تكون الأسئلة صعبة أو سهلة تحصل على منحنى ملتو التواءً سالباً أو موجباً، مما يدعو إلى التركيز على مستويات التلاميذ بدلاً من الأهداف التعليمية [22]

■ ومن الاختبارات التي تستخدم معيار الجماعة كنظام مرجعي لتفسير النتائج هي الاختبارات التحصيلية المقننة والتي تعتمد على موازنة أداء المتعلم بأداء أقرانه وفقا للمجموع الكلى للدرجات والتي تخضع لخصائص عينة التقنين، كما تخضع لخصائص عينة الأسئلة التي يشتمل عليها الاختبار، مما يؤثر في صدق الموازنة وبالتالي صعوبة تعميم النتائج، ففي ظل القياس مرجعي المعيار يختلف تدريج صعوبة الأسئلة باختلاف قدرات الأفراد، كما أن قياس قدرات الأفراد يتباين بتباين صعوبة الأسئلة.[23]

2. النظام مرجعي المحكReferenced System-Criterion:

يهدف القياس مرجعي المحك إلى تقويم أداء الفرد وتحديد مدى تقدمه بالنسبة لمجموعة من المهارات أو الأهداف ومعرفة جودة الأداء في ضوء محكات محددة مسبقا[24] .

والاختبار محكي المرجع كما حدده Glaser هو الاختبار الذي صمم عن قصد لإعطاء قياسات تفسر مباشرة بالنسبة إلى معايير أداء محددة [25] بمعنى أنه يهدف إلى تقدير الفرد بصرف النظر عن مستوى أداء أقرانه، حيث تفسر درجة الفرد على الاختبار في ضوء محك، يحدد حسب الأهداف الموضوعية للقياس.[19]

والاختبار مرجعي المحك يسهم في تقويم كل تلميذ بناءً على مستواه وقدراته، حيث يقوم أداء التلميذ على ما يحتاجه هو بالنسبة لقدرته الفردية، وليس بالنسبة لمقارنته بمتوسط أداء مجموعته، أي أن التفسيرات مرجعية المحك تساعد على شرح ماذا يستطيع عمله الفرد بدون أداء الآخرين[24]. والاختبارات المرجعة إلى المحك تكون مناسبة في الحالات:

•عندما تكون الموضوعات الدراسية تراكمية ومرتبطة مع بعضها البعض.

•عندما يكون الهدف الوصول إلى مستوى تمكن محدد، واكتساب مهارات لمنح درجة علمية .

•عندما يكون الهدف تحديد نواحي القوة والضعف لدى الفرد (أي عملية التشخيص)[26] .

ويستخدم قياس محك المرجع في البرامج التعليمية القائمة على الأهداف أو الكفاءات للتحقق من مدى إتقان التلميذ للمخرجات المعبرة عن الأهداف بالنسبة لمجال سلوكي محدد جيداً، ويتكون هذا المجال من مجموعة من المهارات أو الميول أو الاستعدادات التي يؤديها المختبرون في موقف الاختبار وذلك في ضوء مستوى الإتقان كما ينعكس في درجة فاصلة[27].

كما تستطيع الاختبارات مرجعية المحك أن تحقق ما يلي[28]:

•قياس بعض المهارات الدقيقة.

•قياس بعض الأهداف التربوية المتخصصة التي ترتبط بمجموعات معينة من الطلاب، والتي تفسر في ضوء مستويات محددة.

•قياس أقصى أداء يمكن أن يحققه الفرد.

•الاستفادة منها في حل بعض المشكلات التي يصعب حلها باستخدام الاختبارات مرجعية المعيار.

وعلى الرغم من كل هذه المميزات إلا أن هذه الاختبارات أقل انتشاراً لصعوبة تقدير الثبات والصدق لها ولندرة الدراسات حول هذا النوع للتغلب على هذه المشكلات[24] .

وفى دراستي نادية عبد السلام[29] توضح أن القياس المرجع إلى المحك تعرقله مشكلات علمية من أهمها مشكلة تحديد المحك، فاختلاف المحك يؤدى إلى اختلاف القرار الذي يتخذ بشأن الحكم على المتعلم عند تقدير مستوى الإتقان، والذي يختلف باختلاف الدرجة الفاصلة.

ويتفق أنور الشرقاوي[27] مع ما توصلت إليه نتائج دراستي نادية عبد السلام[28] و[30]، إلا أنه يضيف مشكلة أخرى وهى أن تفسير الدرجات في ضوء القياس محكي المرجع يحتاج إلى معلومات إضافية، بمعنى رد درجة الفرد على الاختبار إلى المستوى المقابل لها والذي يتضمن معلومات عن المحك.وفى دراسة أمينة كاظم [31] توضح أن من أهم مشكلات التقويم الأكاديمي:

•عدم موضوعية تقدير المستوى الأكاديمي نتيجة استخدام الاختبارات المختلفة وتباين مستوياتها.

•عدم موضوعية تقدير السلوك الأكاديمي للتلميذ نتيجة اختلاف مستوى الجماعة التي ينتمي إليها.

وفى دراسة مصطفى كامل[26]تم إعداد اختبار محكي المرجع وقد استخدم محك الإتقان 80 % من الدرجة ليمثل الحد الأدنى الذي ينبغي على المتعلم الوصول إليه لكي يعتبر متمكناً، وقد أسفرت النتائج عن فعالية إستراتيجية التعلم حتى التمكن في الحكم على مستوى تمكن الطلاب من المهارات التدريسية لمقرر التقويم التربوي.

وحددت دراسة أحلام ألباز ورضا حجازي[32] أربعة مستويات لإتقان الطلاب تمثل محكات لها درجات قطع محددة في ضوء تقديرات محددة هي (مستوى التمكن . مستوى الإجادة . مستوى الاجتياز . المستوى الضعيف)، وأسفرت نتائج الدراسة عن اختلاف نسب الطلاب في المستويات الأربع فكانت أقل نسبة للطلاب هي التي حققت مستوى التمكن وحصلت على 85 % فأكثر من المحتوى المحدد للاختبار، وكانت أعلى نسبة هم الطلاب الذين حققوا المستوى الضعيف وهم الذين استطاعوا تحصيل 50 % من محتوى الوحدة موضع الاختبار.

وتجدر ملاحظة أن تحديد المحك يعتمد إلى حد كبير على ذاتية صاحب القرار، ومن ثم فإن القرارات التى تتخذ بشأن الطالب المتقن وغير المتقن تكون مبنية على خطأ في القياس والذي يعتمد على دقة أو عدم دقة تحديد المحك بصورة موضوعية، ويظل السؤال قائماً كيف يمكن الحد من أخطاء القياس؟فالمحك يشير إلى محتوى أو سلوك ترجع إليه درجات الاختبار كما حدده Glaser 1963، وقد يقصد بالمحك: مجال السلوك ـ درجة معيار ثابت ـ درجات مطلقة ـ درجة الإتقان، حيث إن درجة التلاميذ على القياس مرجعي المحك تزود بمعلومات واضحة بخصوص ما يستطيع وما لا يستطيع أن يفعله الفرد، وهو بذلك يختلف باختلاف المحك[28].وتضيف نادية عبد السلام «أن الدرجة الكلية على الاختبارات مرجعية المحك ربما لا تكفى لتمييز أداء التلاميذ، فربما يحصل تلميذان على نفس الدرجة الكلية، لكن نمط استجاباتهم والمهارات المتعلمة ربما تختلف بدرجة كبيرة، فالتلميذ الأول (س) قد يجيب على بعض المفردات التي ترتبط بمهارات غير تلك التي يجيب عليها التلميذ الثاني (ص) ويحصلان على نفس الدرجة، ولحل هذه المشكلة يقسم الاختبار إلى مقاييس فرعية يحدد لكل مقياس فرعى محك خاص به»[33].

والاختبار مرجعي المحك يستخدم للتأكد من مستوى الفرد بالنسبة إلى مجال سلوكي محدد جيداً، وقد يكون المجال السلوكي مجموعة من المهارات أو الاستعدادات التي يؤديها المختبر فى موقف ألاختبار وبالتالي يمكن أن تسمى اختبارات مجالية المرجع، أو تسمى اختبارات هدفية المرجع عندما يتكون الاختبار من مفردات تقابل أهداف محددة تكون مصاغة في صورة إجرائية.[24]

والمستوى يشير إلى الحد الأدنى من الكفاءات المطلوب تحقيقها لغرض معين، وهذا الحد الأدنى يمثل أقل الكفاءات أو المهارات الواجب توافرها لدى التلميذ للانتقال إلى المستوى الدراسي الأعلى.[34]أي أن المستويات هي تحديد الكفاءات والمهارات والأداءات المطلوب توفرها للحكم على اجتياز مستوى دراسي معين، ومن أنواعها **مستوياتالمحتوى** والتي تهتم بتحديد المعارف والمهارات التي يجب أن يمتلكها التلاميذ في نهاية تعلم محتوى دراسي[35]. **ومستوىالأداء** نوع آخر من المستويات يصف الكفاءات المطلوب تحقيقها بصورة إجرائية، بمعنى أنها تضع مستويات المحتوى في صورة أداءات مطلوب تحقيقها للحكم على اجتياز التلميذ لمرحلة دراسية أو صف دراسي.

ومن المشكلات المتعلقة بالاختبارات التى يتم بناؤها فى إطار النظرية الكلاسيكية هى:[3]

- . الخصائص السيكومترية للاختبارات تعتمد على خصائص عينة الأفراد التي يجرى عليها الاختبار، وعلى خصائص عينة المفردات التي يشتمل عليها الاختبار، مما يؤدى إلى نظام قياس نفسي وتربوي غير مستقر، بمعنى تغير معاملات الصدق والثبات وصعوبة المفردات بتغيير قدرات الأفراد.
- . يفترض أن الاختبار يقيس متغيراً أحادى البعد وربما لا يكون ذلك صحيحاً في كثير من الأحيان.
- . يفترض أن الدرجة التي يحصل عيها فرد على مفردات الاختبار يمكن جمعها، كما لو كانت تمثل ميزاناً خطياً والحقيقة أن هذا الميزان منحنياً، فالفرق الثابت بين درجتين من درجات الاختبار يختلف معناه بالنسبة لدرجة القدرة المقاسة بحسب موقع الدرجات على متصل القدرة. فالفرق بين الدرجتين 16−13 ≠12 − 9 بالنسبة للقدرة المقاسة.
- . يفترض أن درجات الاختبار التي تمثل القدرة المقاسة دالة خطية مطردة، بمعنى كلما زادت درجة الفرد دل ذلك على زيادة مقدار القدرة لديه ولكن يلاحظ أحيانا أن بعض الأفراد مرتفعي القدرة يحصلون على درجات منخفضة أو العكس.
- . تغير تكوين الاختبار بمرور الزمن حسب الإضافة أو الحذف أو ظروف الموقف الاختباري.

3. نظام القياس الموضوعي:

يُعدّ هذا النظام من التطورات المعاصرة في القياس النفسي والتربوي، وارتبط هذا النظام بمدخل جديد يطلق عليه "مدخل السمات الكامنة". وهو يشمل نظريات ونماذج سيكومترية مستحدثة، وجاء كرد فعل للنقد الذي يوجه إلى فلسفة القياس التربوي بعامة والنظام مرجعي المعيار بخاصة.[36] وكما أشرنا سابقا، يتلخص النقد الموجه في:

- الاختبارات تعتمد على موازنة أداء المتعلم بأداء أقرانه وفقا للمجموع الكلي لدرجاتهم،
- تخضع الدرجات لخصائص عينة الأفراد، كما تخضع لخصائص عينة الأسئلة التي يشتمل عليها الاختبار،
- هذا الاختلاف يؤثر في صدق إجراء ألموازنات، وبذلك يصعب تعميم النتائج أو الاستفادة العملية منها،[35]
- تدريج صعوبة فقرات الاختبار يتباين بتباين قدرات الأفراد، وقياس قدراتهم أو سماتهم يتباين بتباين صعوبة الأسئلة،
- هذا الاختلاف يؤثر في درجات الثبات،

ونتيجة لهذه الأسباب، يصعب إجراء العمليات الحسابية اللازمة للموازنة بين الأفراد وتقدير التحسن أو التغير في سلوكاتهم أو قدراتهم. فتم التوصل إلى نماذج جديدة تجعل القياس موضوعيا، ومعنى الموضوعية أن "درجة الفرد في الاختبار لا يجب أن يُعتمد تفسيرها على عينة الأفراد المختبرين التي يوازن على أساسها، كما أن الفرد يجب أن يحصل على نفس الدرجة تقريبا في اختبارين يقيسان نفس السمة أو القدرة" [35] والسؤل الذي نطرحه، كيف نجعل القياس التربوي والنفسي موضوعيا؟

ينبغي أولا أن يكون التقويم شاملا أي لا يقتصر على الاختبارات الكتابية فقط خصوصا في المراحل الابتدائية والاعتماد على التقويم الواقعي أي المرتبط بمختلف الأنشطة المرتبطة بواقع التلميذ. أما في المراحل المتقدمة ينبغي بناء الاختبارات التحصيلية بطريقة حديثة لتفادي

المشكلات السابقة، فعلماء القياس التربوي والنفسي اتجهوا نحو مدخل جديد يعتمد على النظرية المعاصرة للاختبارات :Modern Test Theory، والتي تعرف بنظرية السمات الكامنة Latent Trait Theory، ويفضل الكثير تسميتها بنظرية الاستجابة للمفردة Item Response Theory، وذلك لاهتمامها بربط بين استجابة الفرد لمفردة اختباريه وبين خصائص هذه المفردة. تقوم هذه النظرية على فرضية أساسية مؤداها أن القيمة الاحتمالية لاستجابة فرد لمفردة اختباريه تكون دالة لكل من السمة أو القدرة أو الخاصية Trait التي يفترض أن الاختبار يقيسها لدى الفرد، وخصائص المفردة التي يحاول الإجابة عليها [35]. ومعنى الدالة مصطلح رياضياتي وهي نتيجة قدرة الفرد (أي احتمال إجابة صحيحة على الفقرة) ومعامل صعوبة الفقرة وثوابت عددية. (Fan, 1998)

$$P_i(\theta) = \frac{e^{D(\theta - b_i)}}{1 + e^{D(\theta - b_i)}}$$

P_i : احتمالأنيجيبالفردالذيقدرتهعلىالفقرةiإجابةصحيحة.

b_i معامل صعوبة الفقرة،

e الأساس اللوغارتمي ويساوي 2.718

D معامل القياس أو التدريج وهو مقدار ثابت ويساوي 1.7

خاتـــــمة:

يتضح مما سبق أن التقويم ألتحصيلي في المؤسسات التربوية ينحصر على الاختبارات الكتابية والتي تّعد من طرف الأستاذ، وبالتالي تفتقد هذه الاختبارات للخصائص الموضوعية من حيث الشمولية والصدق والثبات. والسبب في ذلك يعود إلى فلسفة القياس التي هي سائدة، حيث يُعتمد في تفسير درجة الأفراد في المقارنة بينهم وترتيبهم حسب الدرجات، وبالتالي لا يمكن

الجزم حول مستوى فرد معين إلا من خلال المجموعة التي ينتمي إليها، وأقصى ما يمكن معرفته هو ترتيبه بين الآخرين. وهناك نظام آخر، يتمثل في تحديد الأهداف التي تحققت والتي لم تتحقق، وبالتالي لا نهتم بترتيب الأفراد، وهذا النظام مشكلاته تتمثل في قدرات الأفراد ومستوى صعوبة الأسئلة حيث يؤثر كل منها في الآخر.

كمحاولة للتغلب على هذه المشكلات لجأ المختصون إلى التقويم الشامل للتحصيل الدراسي فيجب الاعتماد على التقويم المستمر والتقويم الذاتي وتقويم أنشطة الطلاب واعتماد سجلات الأداء وتبنى الاختبارات العملية أو الأدائية ما أمكن حسب طبيعة كل مقرر. ورغم ذلك لا يمكن الاستغناء عن الاختبارات الكتابية في الاختبارات والامتحانات الرسمية ،مما دفع بالمختصين إلى تبنى القياس الموضوعي القائم على تبنى الاتجاهات المعاصرة في القياس والاعتماد على النظرية الحديثة ونماذج الاستجابة للمفردة الإختبارية وبناء بنوك الأسئلة التي توفر موضوعية القياس والاعتماد على هذه النظرية برهن على قدرتها على تحقيق الموضوعية وتوفير الشروط السيكومترية التي تجعل النتائج التي نحصل عليها أساسا لإصدار قرارات صائبة حول الأفراد، والتغلب على مشكلات القياس الكلاسيكي.

الهوامش:

بنوك الأسئلة عبارة عن مجموعة ملفات تضمّ الأنواع الملائمة من مفردات الاختبارات، والتي يتم تصنيفها وترميزها تبعا للمواصفات الدراسية والمستوى التعليمي والأهداف التعليمية التي يتم قياسها، والعديد من الخصائص المميزة للمفردات مثل صعوبة المفردات والقدرة على التمييز. (Rudner,1998 :1)

2 Rogiers, X. La pédagogie de l'intégration en bref, 2007.

3 عبد الوارث عبده الرازحي : تطوير نموذج معياري لتقويم كفاءة نظام إعداد الاختبارات العامة. المؤتمر العربي الأول للامتحانات والتقويم التربوي : رؤية مستقبلية، المركز القومي للامتحانات والتقويم التربوي، القاهرة: 22-24 ديسمبر 2001.

4صلاح الدين علام: القياس والتقويم التربوي والنفسي . أساسياته وتطبيقاته وتوجهاته المعاصرة. (2000)القاهرة: دار الفكر العربي.

5 طلعت الحاموني: تقويم طلاب التعليم الأساسي والثانوي "دراسة تقويمية". مجلة علم النفس، السنة العاشرة،1996 العدد (39).

6رجاء أبو علام: النظريات الحديثة في القياس والتقويم وتطوير نظام الامتحانات "ورقة عمل" المؤتمر العربي الأول للامتحانات والتقويم التربوي: رؤية مستقبلية، المركز القومي للامتحانات والتقويم التربوي، القاهرة 22-24 ديسمبر 2001.

محمد على نصر: رؤية مستقبلية لتفعيل الامتحانات والتقويم التربوي بالتعليم العام في إطار مفهوم الجودة الشاملة. (ورقة عمل)، المؤتمر العربي الأول للامتحانات والتقويم التربوي: رؤية مستقبلية، المركز القومي للامتحانات والتقويم التربوي، القاهرة، 22-24 ديسمبر 2001.

8 أمينة محمد كاظم وصلاح مراد وإسحاق بطرس:تطور نظم الامتحانات بمراحل التعليم العام وبناء بنوك الأسئلة فى بعض الدول العربية. حلقة دراسية إقليمية حول تطوير الامتحانات وبناء بنوك الأسئلة فى الدول العربية، برنامج التجديد التربوي من أجل التنمية في الدول العربية (إبيداس) القاهرة 25-29 نوفمبر 1989.

9فؤاد أبو حطب : أنواع الأسئلة. في الأسس التربوية لإعداد المعلم الجامعي. ط2، جامعة عين شمس، 1995.

10 حمدي عبد العظيم ألبنا: دراسة تحليلية للمستويات المعرفية في امتحانات الكيمياء للشهادة الثانوية العامة، مجلة التربية العلمية، الجمعية المصرية للتربية العلمية، المجلد الرابع، العدد الأول، يناير 2001.

11رضا درويش،تقويم "دليل تقويم التلميذ في مادة العلوم العامة للصف الثالث الإعدادي" في ضوء أهداف تدريس العلوم. مجلة كلية التربية ببنها، الجزء الثاني، يناير 1995 .

12حسين بشير محمود : تطوير نظم التقويم التربوي والامتحانات للتعليم قبل الجامعي "ورقة عمل" المؤتمر العربي الأول للامتحانات والتقويم التربوي: رؤية مستقبلية، المركز القومي للامتحانات والتقويم التربوي، القاهرة 22-24 ديسمبر 2001.

13 Embretson, S.E. and Reise, S.P Item response theory for psychologists. New Jersey: Lawrence Eribaum Associates, Inc., Publishers. 2000.

14 شكري سيد أحمد: العلاقة بين ثلاثة اتجاهات معاصرة في مجال القياس التربوي النفسي ومشكلات تعميمها والإفادة منها. "ورقة عمل" المؤتمر العربي الأول للامتحانات والتقويم التربوي: رؤية مستقبلية، المركز القومي للامتحانات والتقويم التربوي، القاهرة 22-24 ديسمبر 2001 .

15 Van Der Linden, W.I. and Hambleton, R. K. (1997): Item response theory: brief history, common models, and extensions, in (Van der Linden and Hambleton) (Eds) Handbook of Modern Item Response Theory, New York, Springer, pp.(1-27). 1997

16 رشدي فام: تعدد زوايا الرؤية عند تفسير درجات الاختبار وعند بنائه . ضرورة تربوية. في الأسس التربوية لإعداد المعلم الجامعي، ط2، جامعة عين شمس،1995ب .

17 Hambleton, R.K. and Wedman, I. (1997): Introduction of the special issue: Advances in assessment practices, European Journal of Psychological assessment, 13(1),

18 Hambleton, R.K. and Slater, S.C. (1997): Item response theory models and testing practice: Current international status and future directions, European Journal of Psychological assessment, 13(1),

19صلاح الدين محمود علام، الاختبارات التشخيصية "مرجعية المحك" في المجالات التربوية والنفسية والتدريبية ، دار الفكر العربي .2007

20أمينة محمد كاظم: اتجاهات معاصرة في بناء بنوك الأسئلة، في الأسس التربوية لإعداد المعلم الجامعي، ط2، جامعة عين شمس،1995 .

21عبد الرحمن الطريري : القياس النفسي والتربوي: نظريته، أسسه، تطبيقاته. الرياض: مكتبة الرشد.1997

22رشدي فام منصور: التقويم وبناء الاختبارات. في الأسس التربوية لإعداد المعلم الجامعي، ط2، جامعة عين شمس، 1995أ.

23صلاح علام،الاختبارات التشخيصية "مرجعية المحك" في المجالات التربوية والنفسية والتدريبية، القاهرة، دار الفكر العربي. 1995.

24 Carey, L.M. Measuring and evaluating school learning, (2Ed), London: Allyn and Bacon. (1994).

25نادية محمد عبد السلام :بناء اختبار هدفي المرجع وجماعي المرجع "دراسة مقارنة بين النموذجين". في أنور الشرقاوي وآخرون، اتجاهات معاصرة في القياس والتقويم النفسي والتربوي، القاهرة: الأنجلو المصرية. 1996 جـ .

26عبد الرحمن الطريري : القياس النفسي والتربوي: نظريته، أسسه، تطبيقاته. الرياض: مكتبة الرشد.1997.

27مصطفى محمد كامل : استخدام إستراتيجية التعلم حتى التمكن في تدريس مقرر التقويم التربوي "لإكساب الطلاب المعلمين مهارات بناء الاختبارات التحصيلية". مجلة علم النفس، السنة الثالثة عشر، العدد (51)،1999.

28أنور محمد الشرقاوي : الاختبارات المرجعة إلى محك: وسائل جديدة في القياس النفسي والتربوي. في أنور الشرقاوي وآخرون، اتجاهات معاصرة في القياس والتقويم النفسي والتربوي، القاهرة: الأنجلو المصرية،1996.

29نادية محمد عبد السلام : مشكلات عند بناء الاختبارات محكية المرجع "تحليل وتقويم"، مجلة علم النفس، السنة السادسة، العدد الثالث والعشرون،1992 أ.

30نادية محمد عبد السلام: ثلاثة محكات رئيسية لتحديد مستوى الإتقان في الاختبارات محكية المرجع "دراسة إمبريقية". المجلة المصرية للدراسات النفسية، العدد الثالث، ديسمبر. 1992 ب .

31أمينة محمد كاظم: استخدام نموذج "راش" في بناء اختبار تحصيلي في علم النفس وتحقيق التفسير الموضوعي للنتائج، في أنور الشرقاوي وآخرون، اتجاهات معاصرة في القياس والتقويم النفسي والتربوي، القاهرة: الأنجلو المصرية، 1996 ب .

32 أحلام ألباز ورضا السيد حجازي: تقويم تحصيل تلاميذ مرحلة التعليم الأساسي في العلوم باستخدام نظرية السمات الكامنة، الجمعية المصرية للتربية العلمية، المؤتمر العلمي السادس للتربية العلمية وثقافة المجتمع، الإسماعيلية (أبو سلطان)، 28-31 يوليو، 2002

33نادية محمد عبد السلام : دراسة في حساب ثبات وصدق اختبار هدفي المرجع. أنور الشرقاوي وآخرون، اتجاهات معاصرة فى القياس والتقويم النفسي والتربوي، القاهرة: الأنجلو المصرية، 1996 د.

34 Ysseldy K.E.J.; Thutlow, M. and Shviner, J. (1994): Students with disabilities and educational standards: Recommendations for Policy and practice published by the National Center on Educational Outcomes, from http://www.education.umm.edu/ NCEO/online. pubs/policy2.htnl

35 Hespe, D.C. 2002: Standard and assessment for student achievement, , p.2 75. New Jersy Department of Education http://www.srste.nj. Us/njded/adopted/standards.

36صلاح الدين محمود علام، الاختبارات التشخيصية "مرجعية المحك" في المجالات التربوية والنفسية والتدريبية، دار الفكر العربي .2007