

الاتجاهات الحديثة فى بحوث مشكلات تقويم التحصيل الدراسى

بحث من إعداد
د. محمد أحمد محمد إبراهيم غنيم
الأستاذ المساعد بقسم علم النفس التربوى

مقدم إلى
اللجنة العلمية الدائمة لعلم النفس التربوى والصحة النفسية
(مستوى الأساتذة)

أبريل 2003

محتويات البحث

الصفحة	
1	مقدمة
2	مفهوم التقويم
5	مفهوم التحصيل الدراسي
6	النظرية الكلاسيكية للاختبارات
16-8	مشكلات تقويم التحصيل الدراسي
8	1 - مشكلات نظم الامتحانات
14	2 - مشكلات تفسير الدرجة الخام
14	(أ) القياس مرجعي المعيار
16	(ب) القياس مرجعي المحك
26-19	نظرية السمات الكامنة
37-26	التقييم الحقيقي
30	الصحائف الوثائقية
34	خرائط المفاهيم
41-37	تعليق عام
51-42	المراجع

مشكلات تقويم التحصيل الدراسي بين النظريتين الكلاسيكية والمعاصرة

في القياس النفسي

اعداد الاستاذ الدكتور / محمد أحمد ابراهيم غنيم*

مقدمة:

يفاس تقدم الأمم بقوة النظام التربوي بها والتي تكون مخرجاته على درجة عالية من الجودة، حيث يُعد أفراد مؤهلين على درجة عالية من الكفاءة ومبدعين قادرين على تطوير المجتمع، ولديهم مرونة عالية على تطوير أنفسهم ومواكبة التغيرات ومستجدات العصر، ويعتمد ذلك على جودة وسائل القياس والتقويم التي تساعد في اتخاذ قرارات موضوعية بناءً على أسس علمية، باعتبار التقويم جزءاً عضوياً من نسيج النظام التعليمي، فمن خلاله يتم الوقوف على مدى تحقيق أهداف النظام التعليمي، ويقدم تغذية راجعة مستمرة تسهم في تعديل وتطوير النظام وتزويد من كفاءته ومن ثم نوعية وجودة مخرجاته، وقد يرجع ضعف النظام التعليمي (المدخلات - الأنشطة والعمليات - المخرجات)، إلى ضعف في أدوات ووسائل التقويم المستخدمة، وعليه فإن إصلاح التقويم والتعرف على المشكلات المرتبطة بأدواته، يُعد مدخلاً جيداً لإصلاح التعليم والارتقاء بمستوى جودة مخرجاته، وذلك من خلال تقويم كل مكونات النظام والعناصر المرتبطة به ومنها عناصر الموقف التعليمي الأساسية والفرعية، وباعتبار التلميذ أحد العناصر الأساسية فإنه ينبغي أن يتضمن التقويم جميع جوانب شخصيته: (المعرفية - العقلية - الجسمية - الوجدانية .. الخ)، على اعتبار أنها وحدة واحدة، ولكن يلاحظ أن الجانب المعرفي - التحصيل الدراسي - يحتل مكان الصدارة بين هذه الجوانب، وعملية تقويم التحصيل الدراسي يرتبط بها العديد من المشكلات والصعوبات والتي تختلف باختلاف فلسفة القياس والهدف منه سواء مشكلات ترتبط بأدوات التقويم الاختبارية التحصيلية - أو مشكلات ترتبط بالمعايير التي يستند إليها في تفسير درجة التلميذ التي حصل عليها في اختبار ما- أو مشكلات ترتبط بطبيعة القياس "الكلاسيكي - الموضوعي" أو مشكلات ترتبط بالتقويم الحقيقي أو البديل. والقياس يعتبر الخطوة الأولى من عملية تقويم التحصيل الدراسي، وللقياس أدواته التي يتم بها تقدير أداء المتعلم في صورة كمية، لتبدأ عملية اتخاذ القرار عند موازنة الكمية - درجة المتعلم - بمعيار متفق عليه لتفسير وتحديد معنى لهذه الكمية، والتي يتحدد في ضوءها جوانب القوة والضعف - عملية تقييم - ثم تأتي المرحلة النهائية - العلاج - والتي يتم فيها تدعيم نواحي القوة وعلاج نواحي الضعف سواء أكان للمتعلم أو المعلم أو المحتوى أو طرق التدريس.

والدراسة الحالية تتناول مشكلات تقويم التحصيل متجسدة في مشكلات القياس وأدواته وخاصة الاختبارات التحصيلية، مما يتطلب عرضاً للنظرية الكلاسيكية للاختبارات، وتوضيح أهم مشكلات هذه النظرية والدراسات والبحوث التي تعرضت لها، مع إبراز لمشكلات نظم الامتحانات وتفسير الدرجة التي يحصل عليها الفرد في الاختبار التحصيلي، ومشكلة تحديد المستويات، ثم عرضاً للنظرية الحديثة في القياس باعتبارها توجهاً جاء للتغلب على مشكلات النظرية الكلاسيكية وأهمها عدم توفر شروط الموضوعية في القياس، ثم يتم عرض لأدوات التقويم الحقيقي مثل الصحائف الوثائقية أو سجلات الأداء Portfolios وخرائط المفاهيم Concept Maps باعتبارها أدوات تستخدم لعلاج القصور في النظرية الكلاسيكية للاختبار من ناحية بناء الاختبار وقصوره في تحقيق الأهداف التعليمية، ثم توضيح لأهم مشكلات التقويم الحقيقي، ويأتي ذلك بعد توضيح مفهوم التقويم وعلاقته بالمفاهيم الأخرى ذات الصلة، وتوضيح مفهوم التحصيل والاختبار التحصيلي.

* قدم البحث الى اللجنة العلمية الدائمة لعلم النفس التربوي والصحة النفسية لوظائف الاساتذة في 2003/4/8

مفهوم التقييم Evaluation :

لتوضيح مفهوم التقييم يتم تناوله في إطار علاقته بالمفاهيم الأخرى ذات الصلة وهي:
القياس - الاختبار - التقييم، لتوضيح دور كلٍ منها في عملية التقييم.
فالقياس Measurement هو عملية تعيين أرقام أو مستويات مختلفة للصفة المقاسة باختلاف الأفراد، (صلاح مراد، 2000 : 18).
والقياس كما عرفه Nannaly هو قواعد استخدام الأعداد بحيث تدل على الأشياء بطريقة تشير إلى كميات من صفة أو خاصية، في (فؤاد أبو حطب وآخرون، 1997 : 6-7). كما أن القياس هو العملية التي تحدد بواسطتها كمية ما يوجد في الشيء من الخاصية أو السمة التي نقيسها، (محمود غانم، 1997 : 19).

وتجدر الإشارة إلى أن القياس عملية لها أدواتها والتي تستخدم لمعرفة مقدار ما يوجد في الشيء من خاصية، والتي تعتبر أثراً دالاً على وجود هذه الخاصية، أي أن القياس عملية غير مباشرة بمعنى أنه يستدل على الصفة موضوع القياس من خلال عينة السلوك (الأداء) والتي تكون في صورة كمية. والاختبار أحد أدوات القياس عبارة عن مجموعة من الأسئلة المقننة والمراد الإجابة عليها مما يؤدي إلى عملية قياس تكون في صورة رقمية، (عبد الرحمن الطيرى، 1997 : 5).

والدرجة التي يتم التوصل إليها من القياس تسند إلى معيار لاتخاذ قرار تفسر من خلاله هذه الدرجة فيما يعرف بالتقييم Assessment .

ويستخدم التقييم ليعنى اختبار تحصيل الطلاب باستخدام أدوات معيارية مع التركيز على أهداف أداء الطالب، حتى تكون عملية الاختبار أكثر تحديداً، للوقوف على نوعية أو جودة التدريس ومدى ملائمة المنهج لقدرات الطلاب وتقدم حركة الإصلاح، (Stake, 1998, p2).

كما أن التقييم هو عملية الحصول على معلومات تستخدم في اتخاذ قرارات تربوية حول الطالب، وفي تقديم تغذية راجعة توضح مدى تقدم الطلاب ونواحي الضعف والقوة والحكم على الفاعلية Effectiveness التدريسية والكفاية المنهجية والدلالة على حكمتها أو شرعيتها العلمية، (Sanders et al., 1990).

وتجدر ملاحظة أن التقييم عملية تالية للقياس ومرتبطة عليه، فهو عملية تشخيص تتم في ضوء المعلومات التي يتم الحصول عليها من عملية القياس - فيما يخص التحصيل الدراسي - والتي يترتب عليها إصدار أحكام في ضوء معايير محددة مسبقاً ومتفق عليها.

والتقويم مصطلح جاء من الفعل "قوم" بمعنى عدل الشيء أو أصلح ما فيه من إعوجاج، وقد ورد في الأثر عن "عمر بن الخطاب" (*) رضى الله عنه وأرضاه، عندما كان يخطب المسلمين ذات يوم قال: .. إن أحسنت فأعينوني وإن أسأت فقوموني .. رد أحد المسلمين قائلاً والله لو رأينا فيك إعوجاجاً لقومناه بسيوفنا - وهذا يؤكد أن التقويم يعنى تعديل ما أعوج من الأشياء.

(*) ابن كثير: البداية والنهاية، الجزء الثانى.

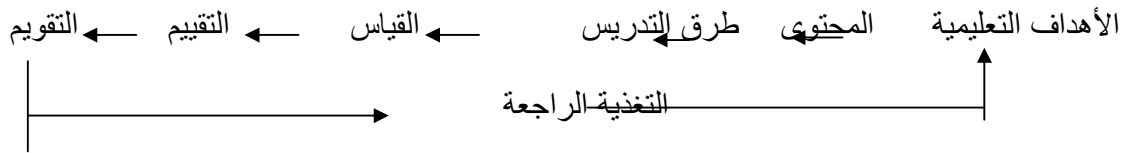
ولذلك فالتقويم هو الخطوة التى تأتى بعد تنفيذ أى برنامج تربوى للحكم على مدى تحقيق الأهداف المنشودة، وعليه فإن التقويم هو الحكم على مدى تحقيق الهدف من عدمه، (مصرى حنورة، 1998 : 10).

والتقويم هو العملية المنهجية التى تتضمن جمع المعلومات عن سمة معينة - بالقياس الكمى أو غيره - واستخدام المعلومات فى إصدار الحكم على هذه السمة فى ضوء أهداف محددة سلفاً لمعرفة مدى كفايتها، (محمود غانم، 1997 : 9).

كما أن التقويم هو عملية اتخاذ القرار التربوى على أساس من القياس والملاحظات بهدف التطوير، (السيد وهبى، 2001 : 223). وإذا كان القياس هو عملية تقدير كمى للأشياء أو خصائصها فى صورة درجات أو أرقام، فإن التقويم وفقاً للتعريف السابق يبدأ من هذه الدرجات ويفسرها.

ومن ثم فإن التقويم هو عملية إصدار حكم على الشئ أو الشخص فى ضوء درجة القياس وفى ضوء الأهداف المحددة وفى ضوء المعلومات الأخرى التى يتم الحصول عليها من مصادر مختلفة، (رجاء أبو علام، 2001 : 96).

يتضح من العرض السابق أن التقويم عملية اتخاذ قرار بشأن الحكم على قيمة موضوع ما فى ضوء نتيجة عملية القياس وفى ضوء معايير متفق عليها مثل: (مدى تحقيق الأهداف - مستوى الاتقان - المتطلبات المهنية .. الخ)، وذلك بهدف تقديم تغذية راجعة فورية تسهم فى تطوير العملية التعليمية بكامل عناصرها، ويقترح الباحث الشكل التالى الذى يوضح العلاقة بين التقويم والمحتوى الدراسى والأهداف المعرفية.



شكل (1) يوضح علاقة التقويم بالمحتوى والأهداف المعرفية

يتضح من الشكل السابق أنه كى تتم عملية تقويم التحصيل الدراسى، لا بد من وجود محتوى - خبرات تعليمية - تقدم للطلاب من خلال عمليتى التعليم والتعلم، يتم إعداده فى ضوء الأهداف التعليمية والتى تحول إلى أهداف سلوكية (إجرائية)، يتم قياسها من خلال أدوات القياس (الاختبارات - الملاحظة - قوائم التقدير .. الخ)، للوقوف على مدى تحقيقها - عملية تشخيص - من خلال إصدار أحكام على أسباب تحقق أو عدم تحقق الأهداف - تقييم - لتأتى مرحلة تقديم العلاج وإحداث عملية التطوير من خلال التغذية الراجعة.

وعلى الرغم من ذلك فإن عملية التقويم يمكن أن تتم دون أن تكون هناك عملية قياس - تقدير كمى - ويحدث ذلك عندما يكون الهدف إصدار أحكام قيمية، كما يحدث عندما يتم تقويم محتوى دراسى - تقويم دليل تقويم الطالب - تقويم ورقة اختبار مادة ما .. يكون الهدف التشخيص لمعرفة نواحى القوة لتدعيمها ونواحى الضعف لعلاجها فى ضوء معايير متفق عليها وليس بالضرورة أن تكون هناك تقديرات كمية.

وبناء على ذلك إذا أريد للتقويم أن يقوم بدور إيجابي في تحسين تعلم الطلاب فلا بد من أسس يرتكز عليها منها^(*):

- أن تكون أنظمة التقويم عادلة للطلاب.
 - أن يتعاون الجميع لدعم عملية التقويم.
 - أن يشارك المجتمع بدور يساهم في تطوير التقويم.
 - المراجعة الدورية لأدوات ووسائل التقويم.
- والتقويم عملية تساعد على قياس تحصيل التلاميذ لمحتوى دراسي وبيان مدى تقدمهم فيه، ويقدم للمعلم تغذية راجعة عن أدائه وفاعلية تدريسه، كما يزود القائمين على تطوير المنهج بتغذية راجعة عن مستويات المنهج المطبق في المدارس،

(Nicholas, 1999, 117)

والتقييم **Assessment** كأداة للإصلاح التربوي هو الذي يقدم صورة دقيقة لتحصيل المتعلم وانجازاته، يتناول العمليات كما يهتم بالنواتج النهائية - يهتم بمهارات التفكير العليا - يتضمن إبداعاً وحلاً للمشكلات، ورسماً تخطيطياً، ومناقشة وبرهنة، ومقابلات شخصية واندماجاً في مهام تعليمية نشطة، يوفر مصادر متعددة للتقويم تؤدي إلى التوصل إلى صورة أكثر دقة لتقويم المتعلمين، (جابر عبد الحميد، 2001).

وتجدر الإشارة إلى أن التقويم الأصيل هو الذي يعطى صورة حقيقية عن مدى تقدم الطلاب ويكشف عن مدى اكتسابهم للمهارات، ومستوى ما لديهم من معارف ومعلومات وقدرتهم على توظيف هذه المعارف في التغلب على مواقف الحياة وحل المشكلات التي تعترضهم.

(*) <http://www.fair Test Principles and Indicators.htm>.

مفهوم التحصيل الدراسي Academic Achievement :

يعرف التحصيل الدراسي بأنه درجة الاكتساب التي يحققها الفرد، أو مستوى النجاح الذي يحرزه أو يصل إليه في مادة دراسية أو مجال تعليمي، (صلاح علام، 2000 : 305).
والتحصيل يرتبط مباشرة بالأداء الدراسي للطلاب لتوضيح المدى الذي تحققت فيه الأهداف التعليمية لدى الطلاب، ويقاس باختبارات التحصيل وهي أدوات قياس لمدى تحصيل الفرد لما اكتسبه من معرفة أو مهارة معينة نتيجة التعليم أو التدريب، (عبد الرحمن الطيرى، 1997 : 280-281).

والاختبارات التحصيلية تقيس مدى استيعاب الطلاب لبعض المعارف والمفاهيم والمهارات المتعلقة بالمادة الدراسية، كما يدل التحصيل الدراسي على الوضع الراهن لأداء الفرد أو تعلمه أو ما اكتسبه بالفعل في برنامج تعليمي، (صلاح علام، 2000 : 305 - 306).

كما تهدف الاختبارات المدرسية - التحصيلية - إلى قياس مدى تحقيق الأهداف التعليمية، وهي إما اختبارات عادية تعد بواسطة المعلمين أو اختبارات عامة تعدها الوزارة في نهاية كل مرحلة دراسية، وبخاصة الثانوية العامة، (عبد الوارث الرازحي، 2001 : 110).
ولكي تكون نتائج التحصيل متنسقة وصادقة وغير متحيزة لا بد من أن يركز قياس التحصيل على بعض الافتراضات وهي:

- أن يقيس الاختبار نطاقاً سلوكياً يمكن تحديده بدقة، بمعنى توفر أهداف تربوية تصاغ في صورة سلوكية.
- أن يقيس الاختبار الأهداف المتعلقة بالمادة الدراسية.
- أن تتاح لجميع الطلاب فرص تعلم المجال الذي يقيس الاختبار، خاصة عند استخدام الاختبارات المقننة، (صلاح علام، 2000 : 309 - 310).

يتضح مما سبق أن الاختبار التحصيلي هو أحد أدوات القياس والتي تقدم معلومات عن مدى اكتساب واستيعاب الفرد للخبرات المعرفية بمستوياتها المختلفة والمهارات التي تم دراستها، وفي ضوء ما تسفر عنه نتائج الاختبارات تتم عملية التقويم، وعليه فإن المشكلات المتعلقة بالاختبارات التحصيلية تنسحب على التقويم، وحتى يتحقق الهدف من الدراسة الحالية فإنه يتم تناول المشكلات المتعلقة باختبارات التحصيل الدراسي والتي تعرضت لها الدراسات السابقة.

النظرية الكلاسيكية للاختبارات Classical Tests Theory :

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:

- درجة الفرد التي يحصل عليها ليس من الضروري أن تمثل درجته الحقيقية، فهي قابلة للتغير حسب ظروف الموقف.
- درجة الفرد هي نتاج نوعين من الدرجات "الدرجة الحقيقية - الدرجة الخاطئة".
- الدرجة الحقيقية يفترض أن تكون ثابتة لأنها تمثل قدرة الفرد المقاسة.
- زيادة خطأ القياس يقلل من الدرجة الحقيقية والعكس صحيح.
- الدرجة الحقيقية يمكن معرفتها من خلال تكرار تطبيق الاختبار عدة مرات، وهي عبارة عن متوسط مرات التطبيق.

▪ لا يوجد اقتران بين الدرجات التي يحققها الأفراد وبين الدرجات الخاطئة.
(عبد الرحمن الطيرى، 1997 : 35-34)، (Rust and Golombx, 1995, 29-31)

- كما ارتكزت النظرية الكلاسيكية للاختبار في القياس النفسى على مجموعة من القواعد هي:
 - الخطأ المعياري للقياس ينطبق على جميع الدرجات في عينة محدودة.
 - الاختبارات الطويلة تكون أكثر ثباتاً من الاختبارات القصيرة.
 - تكون مقارنة درجات الاختبار أفضل عندما تتوفر عدة صورة متكافئة.
 - التقديرات غير المتحيزة Unbiased لخصائص المفردة تعتمد على مدى تمثيلها للعينة.
 - تكتسب درجات الاختبار مدلولها من خلال مقارنة وضعها بمعيار المجموعة الواحدة.
 - خصائص مقاييس المسافة Interval Scales تتحقق بالحصول على توزيع إعتدالى.
 - الصيغ المختلفة للمفردات تؤدي إلى أثر غير متوازن على الدرجة الكلية للاختبار.
 - التغيير في الدرجات لا يمكن أن يكون ذات معنى عندما تختلف درجات الاختبار الأولية (الدرجات الخام).
 - التحليل العاملي للمفردات الثنائية يؤدي إلى عوامل مصطنعة Artfacts وليست عوامل حقيقية.
 - ملامح Features مثيرات المفردة غير هامة إذا ما قورنت بالخصائص السيكمترية
- (Embretson and Reise, 2000, 15)

والخطأ في القياس الكلاسيكى 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:

- صعوبة إحداث تكافؤ بين عدة أنواع من الاختبارات.
- صعوبة تحديد مصادر خطأ القياس خاصة ما يرتبط منها بخصائص أداة القياس.
- اعتماد الخصائص السيكمترية للاختبارات على خصائص عينة التقنين، وعلى خصائص عينة المفردات، مما يؤدي إلى قياس غير مستقر، (شكرى سيد أحمد، 2001 : 24-27).

وقد أجريت محاولات لتقليل الخطأ التجريبي في القياس الكلاسيكى منها: الاختيار العشوائى 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 جزئين 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، (Van der Linden and Hambleton, 1997, 1-2).

وتجدر الإشارة إلى أن الأحكام التي تصدر بشأن مستويات الطلاب التحصيلية، قد لا تكون دقيقة بسبب مقدار خطأ القياس والذي يمكن الحد منه باعتماد القياس الموضوعى الذى يعتمد على النظرية الحديثة في القياس.

مشكلات تقويم التحصيل الدراسي في ضوء النظرية الكلاسيكية:

يمكن إيجاز مشكلات تقويم التحصيل الدراسي فيما يلي:

1 - 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 الحامولي، 1996) 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 الطلاب 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 تقييم مخرجات التعلم، (رجاء أبو علام، 2001)، (Tittle and Pape, 1995)، (Doran et al., 1998)، ويعانى النظام الحالي للامتحانات من قصور في تحقيق الأهداف التعليمية ويرتبط بمجموعة من المشكلات والمظاهر السلبية حددتها نتائج بعض الدراسات والبحوث ومنها:

- تتعتمد كثيراً من الاختبارات على الأسئلة المقالية بدرجة كبيرة وأحياناً على الأسئلة الموضوعية، (صديق عفيفي، 2001)، (محمد على نصر، 2001).
- ارتفاع مستوى القلق والخوف المرتبط بالاختبارات ووجود هدر مالي كبير نتيجة إتاحة فرص أوسع للدروس الخصوصية، (صديق عفيفي، 2001)، (رجاء علام، 2001).
- الاقتصار على الاختبار النهائي الذي يعقد في نهاية الفصل الدراسي وعدم توزيع الدرجات على مدار العام، (صديق عفيفي، 2001).
- يؤدي النظام الحالي للامتحانات إلى انتشار الكتب الخارجية والدروس الخصوصية مما يؤدي إلى هدر الاقتصاد الوطني، (حسين بشير، 2001).
- 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، (عبد الوارث الرازحي، 2001).

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

- أن أسئلة الاختبارات التحصيلية لا تعطي 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 كثيراً من الاختبارات الوزن النسبي والأهمية النسبية لكل من موضوعات المحتوى والأهداف التعليمية وتوزيع أسئلة الامتحانات على المحتوى.

وحول هذه المشكلات أجريت العديد من الدراسات التي اهتمت بتحليل أسئلة الاختبارات التحصيلية لبعض المواد الدراسية بهدف تشخيص الوضع الراهن للامتحانات من أجل تطوير الورقة

الامتحانية ومراعاة الأسس والمعايير الى يجب أن تكون عليها الورقة الامتحانية ومن بين هذه الدراسات: تلك الدراسات التي قام بها المركز القومي للامتحانات والتقويم التربوي منذ نشأته 1990م لتقديم حلول علمية لمشكلات التقويم بصفة عامة والامتحانات بصفة خاصة حيث تم إعداد العديد من المشاريع البحثية كدراسات تقييمية إما لأدلة تقويم الطالب أو تقويم الأوراق الامتحانية للوقوف على جوانب القوة وجوانب الضعف في سبيل تطوير أدوات التقويم، إيماناً بأن إصلاح التقويم يؤدي إلى إصلاح التعليم، وكان من أبرز الجهود للمركز عقده للمؤتمر العربي الأول (2001م) والذي عرض فيه عدد من البحوث وأوراق العمل حول مشكلات نظم الامتحانات الحالية في الوطن العربي والنظرة المستقبلية لتطويرها ودراسات حول الاتجاهات المعاصرة في التقويم "نظرية الاستجابة المفردة والاستفادة منها في بناء بنوك الأسئلة ومشكلات عدم الملائمة والتقويم الحقيقي" وفيما يلي نتائج بعض هذه الدراسات: دراسة (فكري ريان وآخرون، 1994) والتي سعت لتقويم نماذج أسئلة المقررات الدراسية لجميع المواد الدراسية في المرحلة الثانوية، وذلك بهدف الارتقاء بها حتى تكون أداة فعالة لتطوير المنهج وتحقيق أهدافه وتطوير الأسئلة لاستخدامها في 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 نوعي الأسئلة المقالية والموضوعية.

وتجدر الإشارة أن الغالبية العظمى من المعلمين والطلاب يعتبرون نماذج أسئلة المقررات - دليل التقويم - 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، فتطوير نماذج الأسئلة يسهم في تطوير اختبارات نهاية العام.

دراسة (حسن جامع وآخرون، 2001) والتي سعت لبيان مدى تطور الورقة الامتحانية في قياس المستويات العليا للقدرات العقلية والنسب المئوية للأهداف المعرفية في الفترة من 1992/91 وحتى 1999/98 م وقد شملت الدراسة جميع الأوراق الامتحانية لجميع المواد الدراسية للصف الثالث الثانوي، حيث دلت النتائج أن الأسئلة التي تقيس مستوى التذكر ومستوى الفهم كانت تحتل مكان الصدارة في معظم الأوراق الامتحانية وبنسب عالية، كما اختفت المستويات العليا للأهداف المعرفية: التحليل - التركيب - التقويم من بعض الأوراق الامتحانية لبعض السنوات، وفيما يلي مثال لمادة الفيزياء: تمثل الأسئلة التي تقيس مستوى الفهم والتذكر نسبة من 40 % إلى 50 % ، بينما الأسئلة التي تقيس المستويات العليا تراوحت نسبتها بين 10 % إلى 20 % تتفاوت من عام لآخر خلال الفترة 1992/91 إلى 1994/93 م. وفي عام 1995 م لا توجد أسئلة تقيس مستويات أعلى من مستوى التطبيق، واحتل مستوى التذكر والفهم نسبة 67 %، بينما جاءت الورقة الامتحانية في عام 1999 م لتقيس جميع مستويات الأهداف المعرفية وبنسب متفاوتة.

وتجدر الإشارة إلى أن التذبذب في نسب تمثيل الأهداف المعرفية بمستوياتها في الاختبارات ربما يرجع لعدم الالتزام من قبل واضعي الاختبارات باستراتيجية 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 أسرى النظرية الكلاسيكية، وفي هذا الصدد تشير نتائج دراسة (طلعت الحامولي، 1996) إلى أن القرارات الوزارية فيما يتعلق بإعداد الامتحانات تضع شروطاً يجب توافرها فيمن يضع الامتحانات تنتمي فقط إلى الجانب الشخصي أو الإداري، ولم تنطرق إلى المهارة والكفاءة التي ينبغي توافرها في واضع الامتحانات، ولم تحدد معايير الكفاءة التي يُختار على أساسها من يكلف بوضع الامتحانات.

وفى دراسة (عبد الوارث الرازحي، 2001) يعرض لنتائج بعض الدراسات حول مشكلات نظم الامتحانات فى الأقطار العربية، حيث توضح الدراسات أن الاختبارات فى معظم الدول العربية تقتصر إلى الصدق والشمول والموضوعية وتركز فى معظمها على التذكر والحفظ مع إهمال المستويات العليا.

وتجدر الإشارة إلى أن دراسة وتقييم الورقة الامتحانية ونماذج الأسئلة - دليل تقويم الطالب - يؤدى إلى نتائج تعد مدخلاً جيداً لتشخيص واقع الاختبارات والتعرف على جوانب الضعف وما يتصل به من مشكلات لاقتراح الحلول المناسبة لتطوير الاختبارات مستقبلاً - مم يسهم فى تطوير نظم التعليم وإصلاحها.

ومن الدراسات الأخرى دراسة (أمينة كاظم وآخرون، 1989) والتي أوضحت أن أدوات التقويم المستخدمة فى الدول العربية "دول الخليج ومصر" تتجسد فى الاختبارات التحريرية المقالية والموضوعية بأنماطها المختلفة وبنسب متفاوتة، وتعتمد أحياناً على الاختبارات الشفهية أثناء الفصل الدراسى من خلال التقويم المستمر، ونادراً ما تستخدم الاختبارات الأدائية أو العملية، كما أن البطاقة المدرسية غير مفعلة، ويتم التركيز فى تقويم التحصيل الدراسى على الجانب المعرفى وعلى مستوى التذكر وإن تناولت مستوى الفهم والتطبيق بدرجة أقل، ونادراً ما تشمل الاختبارات على المستويات العليا للأهداف المعرفية.

ويضيف (فؤاد أبو حطب، 1995) أن الأسئلة التى تقيس النواحي المعرفية هي نوعان هما:

- أسئلة انتاج الاستجابات مثل الاجابات القصيرة - ملء الفراغ - أسئلة المقال الطويلة.
- أسئلة انتقاء الاستجابة وهي أسئلة موضوعية "الاختيار من متعدد - المزاوجة - الاختيار من بديلين"، ومن أهم المشكلات المرتبطة بهذه الأنواع والتي تؤثر فى 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 البناء، 2001) من أن أسئلة امتحانات الكيمياء للثانوية العامة فى الأعوام 1998، 1999، 2000 م لا تعطى صورة حقيقية لما تقيسه من أهداف، حيث إن ما يظهر من صياغتها يختلف عما تقيسه بالفعل.

وقد أوضحت نتائج دراسة (عبد العزيز عبد الباسط، 1994) والتي اهتمت بتحليل أسئلة الاختبارات التحصيلية الفصلية لطلاب المرحلة المتوسطة بسلطنة عمان وذلك بهدف تحديد المستويات المعرفية التى تقيسها أسئلة عدد من المواد إلى أن أسئلة الاختبارات التحصيلية لمعظم المواد - عينة التحليل - تعطى اهتماماً كبيراً للمستويات المعرفية الدنيا وخاصة الأسئلة التى تقيس مستوى التذكر، فى حين أن الاختبارات تهمل إلى حد كبير الأسئلة التى تقيس المستويات المعرفية العليا.

وأوضحت دراسة (محمد خيرى محمود، 1994) أن أسئلة كتب العلوم بالمرحلة الإعدادية موضع التحليل والتي تم اختيارها عشوائياً، كانت تركز فى معظمها على قياس المستويات المعرفية الدنيا وبخاصة مستوى التذكر والفهم، وكان اهتماماً محدوداً بالمستويات المعرفية العليا.

وأسفرت نتائج دراسة (رضا درويش، 1995) حول تقويم "دليل تقويم التلميذ فى مادة العلوم للصف الثالث الإعدادى" إلى شمول الدليل على أسئلة لقياس الأهداف المعرفية بدرجة كبيرة، بينما شمل الدليل على أسئلة لقياس الجانب المهارى بدرجة ضئيلة جداً، ولم يتضمن الدليل أسئلة لقياس أهداف الجانب الوجدانى، 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.

كما أسفرت نتائج دراسة **(نادية على مسعود، 1999)** والتي سعت للكشف عن المستويات المعرفية التي تقيسها أسئلة امتحانات فروع اللغة العربية للصف الثالث الإعدادي لجميع المحافظات في الفترة من 1991 إلى 1994 م إلى ما يلي:

- بالنسبة لأسئلة التعبير كان الاهتمام الأكبر بمستوى التركيب، حيث بلغت نسبته 89 % عام 1992/91 م ونسبة 80 % عام 1993/92 م بينما بلغت نسبته 75 % عام 1994/93 م. وجاء مستوى التطبيق في المرتبة الثانية، بينما احتل مستوى الفهم المرتبة الثالثة.
- بالنسبة لأسئلة النصوص جاء مستوى الفهم في المرتبة الأولى ثم أسئلة التذكر - أسئلة التحليل.
- بالنسبة لأسئلة القراءة كانت أسئلة مستوى الفهم في المرتبة الأولى يليها أسئلة التذكر ثم أسئلة التركيب وكانت جميعها أسئلة مقالية.

وتوصلت دراسة **(حمدي البناء، 2001)** والتي اهتمت بتحليل أسئلة الكيمياء للثانوية العامة للفترة من 1998 حتى 2000 م إلى أن الأسئلة أولت اهتماماً بالأهداف المعرفية الدنيا، حيث بلغت نسبة أسئلة مستوى التذكر في الأعوام الثلاثة على الترتيب (89 % - 78 % - 83 %) بينما بلغت نسبة الأسئلة التي تقيس المستويات العليا (2 % - 9 % - 8 %) للعوام الثلاثة على الترتيب، كما بلغت نسب الأسئلة التي تقيس مهارة الرسم قيم ضئيلة للغاية.

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 وترتكز على الحفظ والتذكر مما يؤثر في جودة مخرجات التعلم.

وسعيًا 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، 1989م)**، **(محمد غنيمه، 1996)**، **(عبد الوارث الرازحي، 2001)** وأوراق عمل منها **(محمد صبور، 1995)**، **(رجاء أبو علام، 2001)**، **(محمد على نصر، 2001)**، **(صديق عفيفي، 2001)** 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، 2001)** ضرورة استخدام سجلات الأداء التي تضم عينات من عمل الطلاب ومساعدة التلميذ على التقويم الذاتي على اعتبار أن التقويم عملية تعاونية يشترك فيها كل من لهم علاقة بهذه العملية.

بينما يؤكد **(رشدي فام، 1995: 194)** على أنه يجب الاعتماد على أكثر من معيار عند تفسير درجة الطالب على الاختبار وعدم الاقتصار على معيار واحد فقط من المعايير التالية:

- المعيار السيكومتری "موقع الفرد بالنسبة لأقرانه".

- المعيار الأديومتري "موقع الفرد بالنسبة لمحك محدد تحديداً دقيقاً".
- موقع الفرد بالنسبة لنفسه من وقت لآخر.
- موقع الفرد بالنسبة لما يطمح أن يكون عليه مستواه.

ويوصى كل من **Hambleton & Wedman, 1997** بضرورة التخلص من المشكلات المرتبطة بالاختبارات المعتادة - التقليدية - والتي تتمثل في عدم موضوعيتها وعدم قدرتها على تقديرات النطاق الواسع للمحتوى، كما أنها لا توفر تقييماً تشخيصياً جيداً، وذلك من خلال تبني الاتجاهات الحديثة في القياس النفسى التي تمد الممارسين بأطر تسهم في تطوير عملية التقويم وتوفر موضوعية القياس وتركز على تقويم الكفايات **Competencies** بدلاً من تقويم الأداء، والاستعانة بالتكنولوجيا الحديثة في تطوير النماذج السيكومترية الحديثة التي تسهم في تطوير طرق بناء الاختبارات وتحليل بنياتها.

ويضيف كل من **Hambleton & Slater, 1997** أن الاختبارات في الوقت الحالى تمر بتغيرات أساسية، وإحدى هذه التغيرات هو الانتقال من الاعتماد على القياس الكلاسيكى إلى تبني النماذج الحديثة للتغلب على عيوب الاختبارات التقليدية ومنها تباين درجات الأفراد بتباين مستويات صعوبة المفردات.

يتضح مما سبق أنه لتقويم التحصيل الدراسى يجب تبني القياس الموضوعى والاختبارات المقننة، والاعتماد على التقويم المستمر والتقويم الذاتى وتقويم أنشطة الطلاب واعتماد سجلات الأداء، وتبنى الاختبارات العملية أو الأدائية ما أمكن حسب طبيعة كل مقرر، وكل ذلك يتجه بمن يقوم بالتقويم إلى تبني الاتجاهات المعاصرة فى القياس والاعتماد على النظرية الحديثة ونماذج الاستجابة للمفردة الاختبارية وبناء بنوك الأسئلة التى توفر موضوعية القياس والاعتماد على القياس مرجعى المحك وهو ما يتم عرضه فى الجزء التالى.

2 - مشكلة تفسير الدرجة التى يحصل عليها الطالب فى الاختبار التحصيلى:

من المشكلات التى ترتبط بالنظرية الكلاسيكية هى مشكلة تفسير درجة الفرد التى يحصل عليها فى اختبار ما والتى تتوقف على المعيار أو المحك الذى يكسب الدرجة معنى أو دلالة، وفى إطار هذه النظرية تم بناء اختبارات مرجعه المحك وأخرى مرجعه إلى معيار لكل منها مشكلاتها التى ترتبط بالتحصيل الدراسى والتى يتم تناولها فيما يلى:

القياس مرجعى المعيار **Norm-Referenced Measurement** :

يقوم القياس جماعى المرجع على تقدير الفروق بين الأفراد والتمييز بينهم فى الأداء، حيث تكتسب درجة الفرد على الاختبار معناها عندما تسند إلى معيار مستوى الجماعة التى ينتمى إليها الفرد، 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 هى الأقدر على التمييز بين الأفراد، (أمينة كاظم وآخرون، 1989: 19) (أمينة كاظم، 1995: 250)، (نادية عبد السلام، 1996أ).

وتجدر ملاحظة أن الاختبارات فى ظل القياس مرجعى المعيار تعتمد على الفروق الفردية، حيث تتوزع درجات الأفراد فى ظل النظرية الكلاسيكية توزيعاً اعتدالياً.

والاختبارات مرجعة المعيار تكون فائدتها واضحة في الحالات التالية:

- عندما يكون الموضوع الدراسي ليس بشكل مترام 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، (عبد الرحمن الطيرى، 1997: 330-331).

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

- أ - يقتصر على التمييز بين الأفراد ولا يتجاوز ذلك إلى معرفة ماذا تحقق من أهداف العملية التعليمية، مما يؤدي إلى حذف عناصر ووحدات من الاختبار إذا كانت غير مميزة بين الأفراد بغض النظر عن قيمتها مما يؤثر في بناء الاختبار.
 - ب - الاقتصار على اتخاذ الجماعة كمحك لتفسير درجة الفرد على الاختبار يؤدي إلى عدم التيقن من موقع الفرد وجماعته بالنسبة لأهداف التعلم ومدى تحقيقها.
 - ج - سطوة المنحنى الاعتدالي 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، (رشدى فام، 1995 ب :
- (209-208).

ومن الاختبارات التي تستخدم معيار الجماعة كنظام مرجعي لتفسير النتائج هي الاختبارات التحصيلية المقننة والتي تعتمد على موازنة أداء المتعلم بأداء أقرانه وفقاً للمجموع الكلي 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 تدرج صعوبة الأسئلة باختلاف قدرات الأفراد، كما أن قياس قدرات الأفراد يتباين بتباين صعوبة الأسئلة، (صلاح علام، 1995: 15-19).

وتجدر الإشارة إلى أن علماء القياس النفسي لم يقفوا أمام هذه المشكلات مكتوفى الأيدي، بل كانت هناك جهوداً متعاظمة 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 موضوعي بمعنى عدم تأثر درجة الفرد بخصائص عينة الأفراد ولا خصائص عينة الأسئلة المكونة لأداة القياس وهو ما يتم عرضه في الجزء الخاص بنظرية الاستجابة المفردة.

القياس مرجعي المحك Criterion-Referenced Measurement :

يهدف القياس مرجعي المحك إلى تقويم أداء الفرد وتحديد مدى تقدمه بالنسبة لمجموعة من المهارات أو الأهداف ومعرفة جودة الأداء في ضوء محكات محددة مسبقاً، (Carey, 1994, 261-262) والاختبار محكى المرجع كما حدده Glaser هو الاختبار الذى صمم عن قصد لإعطاء قياسات تفسر مباشرة بالنسبة إلى معايير أداء محددة، فى (نادية عبد السلام، 1996، ب، هـ). بمعنى أنه يهدف إلى تقدير الفرد بصرف النظر عن مستوى أداء أقرانه، حيث تفسر درجة الفرد على الاختبار فى ضوء محك، يحدد حسب الأهداف الموضوعية للقياس، (أمينة كاظم، 1995 : 251).

والاختبار مرجعي المحك يسهم فى تقويم كل تلميذ بناءً على مستواه وقدراته، حيث يقوم أداء التلميذ على ما يحتاجه هو بالنسبة لقدراته الفردية، وليس بالنسبة لمقارنته بمتوسط أداء مجموعته، أى أن التفسيرات مرجعه المحك تساعد على شرح ماذا يستطيع عمله الفرد بدون أداء الآخرين، (نادية عبد السلام، 1996 هـ). والاختبارات المرجعة إلى المحك تكون مناسبة فى الحالات:

- عندما تكون الموضوعات الدراسية تراكمية ومرتبطة مع بعضها البعض.
- عندما يكون الهدف الوصول إلى مستوى تمكن محدد، واكتساب مهارات لمنح إجازة (درجة علمية).
- عندما يكون الهدف تحديد نواحي القوة والضعف لدى الفرد (تشخيص)، (عبد الرحمن الطيرى، 1997 : 331).

ويستخدم لقياس محك المرجع فى البرامج التعليمية القائمة على الأهداف للتحقق من مدى إتقان الطالب للمخرجات المعبرة عن الأهداف بالنسبة لمجال سلوكى محدد جيداً، ويتكون هذا المجال من مجموعة من المهارات أو الميول أو الاستعدادات التى يؤديها المختبرون فى موقف الاختبار وذلك فى ضوء مستوى الإتقان كما ينعكس فى درجة فاصلة، (مصطفى كامل، 1999).

كما تستطيع الاختبارات مرجعية المحك أن تحقق ما يلى:

- قياس بعض المهارات الدقيقة.
- قياس بعض الأهداف التربوية المتخصصة التى ترتبط بمجموعات معينة من الطلاب، والتى تقس فى ضوء مستويات محددة.
- قياس أقصى أداء يمكن أن يحققه الفرد.
- الاستفادة منها فى حل بعض المشكلات التى يصعب حلها باستخدام الاختبارات المرجعة إلى معيار (أنور الشرقاوى، 1996 : 38).

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، (نادية عبد السلام، 1996 ب : 84).

وفى دراستى (نادية عبد السلام، 1992 أ، ب) توضح أن القياس المرجع إلى المحك تعرقله مشكلات علمية من أهمها مشكلة تحديد المحك، فاختلاف المحك يؤدى إلى اختلاف القرار الذى يتخذ بشأن الحكم على المتعلم عند تقدير مستوى الاتقان، والذى يختلف باختلاف الدرجة الفاصلة، ويتفق (أنور الشرقاوى، 1996) مع ما توصلت إليه نتائج دراستى (نادية عبد السلام، 1992 أ، ب)، إلا أنه يضيف مشكلة أخرى وهى أن تفسير الدرجات فى ضوء القياس محكى المرجع يحتاج إلى معلومات إضافية، بمعنى رد درجة الفرد على الاختبار إلى المستوى المقابل لها والذى يتضمن معلومات عن المحك.

- وفى دراسة (أمينة كاظم، 1996 ب) توضح أن من أهم مشكلات التقويم الأكاديمي:
 - عدم موضوعية تقدير المستوى الأكاديمي نتيجة استخدام الاختبارات المختلفة وتباين مستوياتها.
 - عدم موضوعية تقدير السلوك الأكاديمي للطالب نتيجة اختلاف مستوى الجماعة التي ينتمى إليها.
- وتوصلت نتائج دراسة (عصام الدسوقي، 1998) إلى فعالية نموذج (أنجوف) فى تحديد المستوى لاختبار محكى المرجع فى وحدة من مقرر الجبر للصف الأول الثانوى، حيث كانت درجة القطع (المستوى) 74 %، بينما بلغت درجة القطع لنفس الاختبار باستخدام نموذج (هوفستى) 68 %، حيث إن نموذج (أنجوف) يعتمد على أحكام المتخصصين فى مجال الاختبار لتحديد درجة القطع بينما يعتمد نموذج (هوفستى) على أحكام المتخصصين بالإضافة إلى الأداء الفعلى للمفحوصين أى أن عينة الدراسة تلعب دوراً فى تحديد درجة القطع للاختبار والتي تتوقف على مستوى العينة.

وفى دراسة (مصطفى كامل، 1999) تم إعداد اختبار محكى المرجع وقد استخدم محك الإلتقان 80 % من الدرجة ليمثل الحد الأدنى الذى ينبغى على المتعلم الوصول إليه لكى يعتبر متمكناً، وقد أسفرت النتائج عن فعالية استراتيجية التعلم حتى التمكن فى الحكم على مستوى تمكن الطلاب من المهارات التدريسية لمقرر التقويم التربوى.

وحددت دراسة (أحلام الباز ورضا حجازى، 2002) أربعة مستويات لإتقان الطلاب تمثل محكات لها درجات قطع محددة فى ضوء تقديرات محددة هى (مستوى التمكن - مستوى الإجابة - مستوى الاجتياز - المستوى الضعيف)، وأسفرت نتائج الدراسة عن اختلاف نسب الطلاب فى المستويات الأربع فكانت أقل نسبة للطلاب هى التى 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 فى القياس 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 - درجة معيار ثابت - درجات مطلقة - درجة الإلتقان، حيث إن درجة الطالب على القياس مرجعى المحك تزود بمعلومات واضحة بخصوص ما يستطيع وما لا يستطيع أن يفعله الفرد، وهو بذلك يختلف باختلاف المحك، فى (نادية عبد السلام، 1992 ب).

وتضيف (نادية عبد السلام، 1996 د) أن الدرجة الكلية على الاختبارات مرجعية المحك ربما لا تكفى لتمييز أداء الطالب، فربما يحصل طالبان على نفس الدرجة الكلية، لكن نمط استجاباتهم والمهارات المتعلمة ربما تختلف بدرجة كبيرة، فالطالب الأول (س) قد يجيب على بعض المفردات التى ترتبط بمهارات غير تلك التى يجيب عليها الطالب الثانى (ص) ويحصلان على نفس الدرجة، ولحل هذه المشكلة يقسم الاختبار إلى مقاييس فرعية يحدد لكل مقياس فرعى محك خاص به.

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 في موقف الاختبار، 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، (نادية عبد السلام، 1996 ج).

والمستوى يشير إلى الحد الأدنى من الكفاءات المطلوب تحقيقها لغرض معين، وهذا الحد الأدنى يمثل أقل الكفاءات أو المهارات الواجب توافرها لدى الفرد للانتقال إلى المستوى الدراسي الأعلى، (Ysseldyke et al., 1994). أى أن المستويات هي تحديد الكفاءات والمهارات والأداءات المطلوب توافرها للحكم على اجتياز مستوى دراسي معين، ومن أنواعها مستويات المحتوى والتي تهتم بتحديد المعارف والمهارات التي يجب أن يمتلكها الطلاب في نهاية تعلم محتوى دراسي، (Hespe, 2002, p.2). ومستوى الأداء نوع آخر من المستويات يصف الكفايات المطلوب تحقيقها بصورة إجرائية، بمعنى أنها تضع مستويات المحتوى في صورة أداءات مطلوب تحقيقها للحكم على اجتياز الطالب لمرحلة دراسية أو صف دراسي، (Ysseldyke et al., 1994).

نظرية السمات الكامنة Latent Trait Theory :

ويطلق عليها نظرية الاستجابة المفردة Item Response Theory (IRT) أو نظرية الاختبار الحديثة في القياس النفسي. وقد جاءت هذه النظرية لمواجهة جوانب الضعف أو التغلب على المشكلات في النظرية الكلاسيكية في القياس والتي كان من بينها:

- أداء المفحوصين على الاختبار يختلف باختلاف بنود الاختبار لاختلاف مستويات الصعوبة للمفردات مما ينعكس على الأداء من اختبار لآخر.
- تفترض النظرية الكلاسيكية تساوى تباين خطأ القياس لجميع الأفراد الذين يطبق عليهم الاختبار، مما ينتج عنه أن متوسط خطأ القياس لجميع الأفراد صفراً، لأن تباين خطأ القياس الإيجابي يساوى تباين خطأ القياس السلبي، (عبد الرحمن الظريوي، 1997 : 58-59) Embretson and (Reise, 2000, 45).

ومن مشكلات القياس السلوكي ما عرضته (أمينة كاظم، 1996 أ : 293-295) هو:

- اتخاذ الدرجة الكلية أساسياً للقياس أو دالة لتقدير مستوى أداء الأفراد على المتغير موضوع القياس، وهذا الأمر لا يتحقق إلا في ظل القياس الموضوعي حيث لا تختلف الدلالة الكمية لأداء الفرد باختلاف اختبار المستخدم.
- لا تحدد درجات الاختبار مواضع القياس على متصل المتغير بصورة خطية.

ونظرية الاستجابة المفردة تتضمن عدة نماذج تسمح بالقياس الموضوعي، حيث يتم التركيز على المفردة الاختبارية مما يسمح بإضافة أو حذف أو تعديل مفردات دون أن يتأثر الاختبار ككل، مما يساعد في التغلب على مشكلات القياس النظرية والتطبيقية التي عجزت عنها النظرية الكلاسيكية. ومن المتطلبات التي يلزم توافرها لكي يكون القياس النفسي أكثر موضوعية والتي حددها Wright and Stone هي:

- توفر مفردات صادقة يمكنها تعريف المتغير موضوع القياس إجرائياً.
- صدق التدرج لهذه المفردات، بحيث يمكنها تمثيل هذا المتغير في صورة خطية.

- أنماط استجابات الأفراد يمكنها تحديد مواضعهم على متصل المتغير.
- توافق بين درجات الأفراد وخصائص المفردات بحيث تؤدي إلى تقديرات لمستويات الأفراد لا تعتمد على اختبار معين.
- قياسات خطية يمكن استخدامها لدراسة النمو أو للمقارنة بين الأفراد في (أمينة كاظم، 1988: 39)، (أمينة كاظم، 1996 ب : 307) .

ومن المشكلات المتعلقة بالاختبارات التي يتم بناؤها في إطار النظرية الكلاسيكية هي:

- 1 - الخصائص السيكمترية للاختبارات تعتمد على خصائص عينة الأفراد التي يجرى عليها الاختبار، وعلى خصائص عينة المفردات التي يشتمل عليها الاختبار، مما يؤدي إلى نظام قياس نفسي وتربوي غير مستقر، بمعنى تغير معاملات الصدق والثبات وصعوبة المفردات بتغيير قدرات الأفراد.
- 2 - يفترض أن الاختبار يقيس متغيراً أحادي البعد وربما لا يكون ذلك صحيحاً في كثير من الأحيان.
- 3 - يفترض أن الدرجة التي يحصل عليها 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.
- 4 - 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 أحياناً أن بعض الأفراد مرتفعي القدرة يحصلون على درجات منخفضة أو العكس.
- 5 - تغير تكوين الاختبار بمرور الزمن حسب الإضافة أو الحذف أو ظروف الموقف الاختباري، (صلاح علام، 2000، 680-681) .

وفي إطار نظرية الاستجابة المفردة فإن المتغيرات الكامنة هي التي تحدد السلوك وهي غير ملاحظة ويتم الاستدلال عليها من ملاحظة أداء الفرد أو الاستجابة على المفردة، ومن ثم فإن خصائص المفردات والأفراد يستدل عليها من السلوك، وبالتالي فإن نظرية القياس في علم النفس تؤدي إلى تحسين السلوك، (Embretson and Reise, 2000, 40-41) وتفترض نظرية الاستجابة للمفردة وجود سمة أو قدرة هي التي تحدد استجابة الفرد للمفردات، وهذه السمة يشترك فيها جميع الأفراد، (Molenaar and Hoijtink, 1996) .

- ومن القواعد التي ارتكزت عليها نظرية الاستجابة للمفردة (IRT) ما يلي:
- 1 - الخطأ المعياري للقياس يختلف باختلاف الدرجات أو أنماط الاستجابة ولكنه يعمم للعينات الكبيرة.
 - 2 - الاختبار الأقصر من الممكن أن يكون أكثر ثباتاً من الاختبار الأطول.
 - 3 - تكون مقارنة درجات الاختبار في صورته المتعددة أفضل عندما يختلف مستويات الصعوبة للاختبار بين الأشخاص.
 - 4 - يمكن الحصول على تقديرات غير متحيزة لخصائص المفردة من عينات غير ممثلة.
 - 5 - يكون لدرجات الاختبار مدلول إذا قورنت ببعضها عن المفردات.
 - 6 - خصائص مقاييس المدى يمكن تحقيقها بتطبيق نماذج قياس سابقة.
 - 7 - الصيغ المختلطة للمفردات يمكن أن تقدم درجات أفضل للاختبار.

8 - التغيير فى الدرجات يمكن مقارنته بشكل مفهوم عندما تختلف مستويات الدرجات الأولية (الخام).

9 - التحليل العاملى لبيانات المفردة ينتج عنه تحليل عاملى كامل المعلومات.

10 - ملامح مثيرات المفردة يمكن ربطها مباشرة بالخصائص السيكومترية.

(Embretson and Reise, 2000, p. 15)

ونظرية الاستجابة المفردة توفر مؤشرات إحصائية ثابتة للاختبار ومفرداته لا تختلف باختلاف عينة الأفراد، مما يوفر قياساً أكثر ثباتاً مع أخطاء أقل للقياس، (Nitko, 2001) وتتضمن نظرية الاستجابة للمفردة ثلاثة نماذج أساسية تختلف حسب عدد البارامترات وطبيعتها هى نموذج راش Rasch Model أحادى البارامتر - نموذج لورد Lord ثنائى البارامتر - ونموذج Birnbaum ثلاثى البارامتر، وجميعها من النوع الثنائى الاستجابة، بمعنى أنها تعتمد على بيانات المفردات التى يجاب عليها "صفر/واحد" أى المفردات من نوع "الصواب/الخطأ" - "نعم/لا" - أو الاختيار من متعدد، (Van-der Linden and) (Embretson and Reise, 2000, 65-70) (Hambleton, 1997, 29-30).

ويعد نموذج راش أحادى البارامتر من أكثر النماذج شيوعاً فى تصميم وبناء الاختبارات والمقاييس التربوية والنفسية، ويهتم بتحديد موقع المفردة الاختبارية على ميزان صعوبة جميع المفردات التى تشكل الاختبار، كما يهتم بتدريج مستويات قدرة الفرد باختبار معين على نفس مستويات القدرة المقاسة، كما يفترض النموذج أن الفرد لا يلجأ إلى التخمين العشوائى فى إجابته على مفردات الاختبار، (صلاح علام، 2000 : 693) (Hashway, 1998, pp.87-90).

ونموذج Rasch يوفر متطلبات القياس الموضوعى عندما تتحقق فروضه وهى:

- أحادية البعد: حيث يعرف المتغير "السمة أو القدرة" بواسطة مجموعة من بنود الاختبار ذات صعوبة أحادية البعد، أى أن بنود الاختبار لا تختلف فيما بينها إلا من حيث مستوى الصعوبة فقط، كما يكون الأفراد ذوى قدرة أحادية البعد، تحدد وحدها مستوى أدائهم على الاختبار.
- استقلالية القياس: ويعنى ذلك استقلال صعوبة البند عن كل من صعوبة بنود الاختبار الأخرى وقدرات الأفراد الذين يجيبون عليها، وأيضاً استقلال تقدير قدرات الأفراد عن قدرات أى مجموعة أخرى من الأفراد الذين يؤدون الاختبار أو على صعوبات البنود التى يؤدونها.
- توازى المنحنيات المميزة للبنود: بمعنى أن جميع بنود الاختبار يكون لها قوة تمييزية متساوية. (أمينة كاظم، 1996 : 310)

وقد تعرضت دراسات كل من : (أمينة كاظم، 1988، 1996 ب)، (صلاح علام، 1989 : 451-496) (صلاح علام، 1995) (أمينة كاظم وآخرون، 1989)، (سليمان الخضرى، وأمينة كاظم، 1997، 1998، 2000) إلى أهمية بنوك الأسئلة المكونة من مجموعات من الأسئلة تم تدريجها بواسطة نموذج "راش" بحيث تشكل مقياساً واحداً له صفراً واحد مشترك وتدرج واحد مشترك، ولهذا التدرج وحدة قياس واحدة ثابتة على جميع مستويات المتغير، كما أوضحت تلك الدراسات كيف يؤدى استخدام نموذج "راس" فى تدرج البنود أو الأسئلة إلى أن تتميز هذه البنوك بأحادية القياس بمعنى تدرج بنود أسئلة البنك تبعاً لصعوبتها بحيث تعرف فيما بينها متغيراً واحداً، كما تتميز بنوك الأسئلة باستقلالية القياس بمعنى استقلال قدرة الأفراد وصعوبة البنود عن جميع البنود، كما يستقل تقدير صعوبة البنود عن مجموعة الأفراد والذين يجرون الاختبار بمعنى توفر موضوعية القياس، الأمر الذى يؤدى إلى الثقة

العلمية فى نتائج القياس المبنية على اختبارات تم اشتقاقها من بنوك الأسئلة التى تم تدريج بنودها باستخدام نموذج راش.

كما أسفرت نتائج دراسة (أمينة كاظم، 1996 ج) عن فاعلية نموذج "راش" فى تحويل التدريج المستقل لاختبارين إلى تدريج واحد مشترك، من خلال عملية معادلة لتدريج البنود المكونة لاختبارين (سهل - صعب) فى ضوء بيانات لعينتين مختلفتين من حيث المستوى (مرتفعة القدرة - منخفضة القدرة)، أى أن نموذج راش بما يوفره من أحادية واستقلالية القياس يسهم فى دقة التعادل الرأسى Vertical Equating لتدريج اختبارين.

وتجدر الإشارة إلى أن الدراسات السابقة والتي اعتمدت على نموذج "راش" ومدى فاعليته فى بناء بنوك الأسئلة، قد ارتكزت جميعها على مفردات أو بنود من نوع الاختيار من متعدد أو الصواب والخطأ بمعنى أن المفردات ثنائية الاستجابة (صفر أو واحد) وليبان مدى فاعلية نموذج "راش" لأسئلة من نمط الاستجابة المتعددة (اختبارات المقال التحصيلية)، هدفت دراسة (محمد منصور الشافعى، 2001) إلى التحقق من صدق استبعاد الأفراد غير الملائمين وفقاً لحالات الغش (فردى - جماعى)، وحالات عدم الجدية، ومدى تأثير ذلك على مؤشرات الملائمة للمفردات وتقديرات الصعوبة لاختبار يتضمن أسئلة من نوع المقال فى الرياضيات، حيث اعتمد على مؤشرى الملائمة (التقاربى Infit - التباعدى Outfit) المتضمن فى برنامج مايكروسكيل Microscale، وقد أسفرت الدراسة عن عدم قدرة برنامج التحليل مايكروسكيل فى التعرف على أغلب حالات عدم الملائمة لحالات (غش فردى - غش جماعى - عدم جدية) وذلك عند تحليل استجابات 610 طالب، كما أسفرت الدراسة عن وجود تأثير سلبي لاستجابات حالات عدم الملائمة على مؤشرى الملائمة الاحصائية لمفردات اختبار المقال مما يعطى صورة غير حقيقية عن هذه المفردات، كما تأثرت تقديرات الصعوبة الخاصة بمفردات الاختبار لعدم حذف حالات عدم الملائمة. يتضح مما سبق أن الاعتماد على بنوك الأسئلة من نوع المقال والتي درجت باستخدام نموذج "راش" والتي اعتمدت فى تحليل استجابات الأفراد لأسئلة هذا النوع من الاختبارات باستخدام مؤشرى "التقارب - التباعده" يعتبر عملية مضللة لعدم توفر شروط الموضوعية وبالتالي الاعتماد على نتائجها فى تقويم التحصيل الدراسى يؤدى إلى اتخاذ قرارات غير دقيقة، وبالتالي الحاجة ملحة لمزيد من البحوث والدراسات للتغلب على مشكلات عدم الملائمة.

وهدف دراسة (Davison and Change, 1992) على بحث مدى ملائمة النموذج أحادى البعد "راش" والنماذج الثنائية ومتعددة الأبعاد لنظرية الاستجابة المفردة فى تدرج مفردات اختبارات القراءة والرياضيات من بطارية Woodcock-Johnson النفس - تربوية على عينة من 1000 فرد. وقد أسفرت الدراسة عن ملائمة النموذج الثنائى متعدد الأبعاد فى تدريج مفردات الاختبارات التى تتضمن قدرات أحادية وثنائية مع خطأ معيارى محدود، كما لا ترتبط درجات النموذج أحادى البعد خطياً مع تقديرات القدرة متعددة الأبعاد للاختبارات الفردية التى تتضمن أكثر من قدرة كما تعطى خطأ معيارياً كبيراً. وهنا تظهر محدودية نموذج "راش" عندما تكون الاختبارات تتضمن أكثر من بعد يقيس أكثر من قدرة، بينما تظهر فاعلية النماذج المتعددة رغم محدودية انتشارها لصعوبة التعامل معها كونها تعتمد على تعقيدات رياضية تحتاج إلى تدريج.

دراسة (Bay-Luz, 1998) تهدف إلى بحث الفروق فى أداء الطلاب على مفردات اختبار اختيار من متعدد (انتقاء الاستجابة) والاستجابة المبنية (إنتاج الاستجابة) على المفردات

المرتبطة بالتقييم القومى لمستويات التحصيل المتقدم فى التربية (إنتاج الاستجابة)، حيث يتم تقدير أداء الطالب فى إطار نظرية الاستجابة المفردة IRT وقد حددت سبع تجمعات للمفردات وحددت نقاط قطع Cut point لكل منها لتحديد مستويات التحصيل وذلك بأخذ متوسط تقديرات المشاركين لتحديد النسب المئوية المتضمنة والتي تستخدم لتحديد المنحنيات المميزة للاختبارات، حيث أسفرت النتائج عن أن أداء الطلاب على مفردات الاختيار من متعدد يكون أفضل فى ظل نظرية (IRT) والمرتبطة بنقاط قطع محددة من أدائهم على اختبارات إنتاج الاستجابة، كما وجد أن توقعات الأداء لتحليل البيانات للمفردات للاختبار (MC) تكون أقل ارتباطاً بالأداء الواقعى أو الحقيقى، بينما توقعات الأداء للمفردات المرتبطة بإنتاج الاستجابة تكون أكثر مناسبة للأداء الحقيقى.

ويحدد (صلاح علام، 2000 : 683 - 864) عدداً من المشكلات المرتبطة بالنظرية

الحديثة فى القياس وهى:

- المشكلات المتعلقة بتقدير بارامترات النماذج وتقييم حسن مطابقة البيانات لهذه النماذج وإمكانية استخدامها عندما تكون المتغيرات التى تقيسها الاختبارات متعددة الأبعاد وغير ذلك من القضايا السيكومترية والتطبيقية.
- تحتاج النظرية إلى قدر كبير من الرياضيات العالية الإحصاء متعدد المتغيرات.
- البحوث والدراسات حول نظرية القياس وجهت إلى منظرى القياس، وليس للذين سيقومون بتطبيقه فى بناء الاختبارات وتحليل وتفسير نتائجها.
- عدم توفر برامج للحاسب الألى فعالة ومناسبة لتقدير بارامترات نماذج السمات الكامنة التى تتضمنها النظرية الحديثة (IRT).

وقد توصلت دراسة (أمينة كاظم، 1996 أ) إلى عدد من المشكلات التى ترتبط بنتائج الدراسات والبحوث السابقة حول تحذير الباحثين عند استخدامهم لنموذج "راش"، واعتمادهم عليه كوسيلة موضوعية للقياس، حيث يكمن التحذير فى اختيار عينة التدرج عند بناء اختبار ما، وكذلك فى اختبار مجموعة البنود التى يقدر بها قدرات الأفراد، صعوبة تطويع النموذج للتطبيق وبناء الاختبارات وتفسيرها بالنسبة للمدرس العادى الذى اعتاد على بناء الاختبارات الموضوعية الشائعة.

وتعرض دراسات كل من (صلاح علام، 1989)، (أمينة كاظم وآخرون، 1989)،

(شكرى سيد أحمد، 2001) بعض المشكلات المرتبطة بنظرية الاستجابة المفردة (IRT) والنماذج المنبثقة عنها وهى:

- المقاومة ورفض التجديد والإبقاء على ما هو شائع وسبق قبوله.
- الكلفة المالية العالية (كلفة اقتناء الحاسبات الآلية - كلفة تصميم أو شراء البرمجيات الحديثة - كلفة تدريب الخبراء وغير المتخصصين على إعداد وبناء بنوك الأسئلة وتطبيقها والتحقق من شروطها).
- الحاجة إلى خبراء ومتخصصين على مستوى عالى من الخبرة فى مجال نظرية (IRT).

ويضيف **Lawrence, 1998** أن نظرية الاستجابة المفردة (IRT) وبنوك الأسئلة

ليست الدواء الشافى لكل مشكلات القياس الكلاسيكى، بل توجد بعض الصعوبات أو المشكلات التى ترتبط بنظرية الاستجابة المفردة وبنوك الأسئلة منها: أنها تحتاج إلى مستوى عال من الرياضيات والتقنية المتقدمة، وقدر من المهارات على التعامل مع التقنية الحديثة، ومهارة فائقة وجهد كبير فى إعداد المفردات وصياغتها مما يتطلب ممارسين على درجة عالية من الكفاءة، ولديهم خبرة بالأسس النظرية والتطبيقية المرتبطة بالنماذج الحديثة. ويحدد

(Ercikan, 2003) بعض المشكلات المرتبطة بنظرية (IRT) منها ما يرتبط بالجوانب النظرية والمفاهيم وأخرى ترتبط بالإجراءات وثالثة ترتبط بنمط الاستجابات القصيرة وتحليل وتفسير البيانات المرتبطة بهذه الاستجابات حتى باستخدام الحاسب الآلى.

التقييم الحقيقى (الأصيل) Authentic Assessment :

يعتبر التقييم الحقيقى أحد الاتجاهات المعاصرة فى مجال تقويم التحصيل كأحد مخرجات عملية التعلم، ويأخذ التقييم الحقيقى مسميات متعددة منها التقييم الواقعى - التقييم الأصيل والتقييم البديل Alternative Assessment . ويهدف هذا النوع من التقييم إلى قياس إمكانات عقلية عليا، ويركز على عمليات تعلم مهمة يمكن تنميتها فى إطار العمل المدرسى وخارجه، ومتابعة تطورها، (صلاح علام، 2000 : 746). ويعرف التقييم الحقيقى بأنه تقييم أداء المتعلم من خلال مواقف الحياة الواقعية والتي ترتبط بمشروعات تكشف عن مهارات حل المشكلات وتحليل المعلومات وبنائها فى سياق جديد، (Meyar, 1992).

ويساعد التقييم الحقيقى فى تهيئة المتعلمين للحياة، حيث يطلب من المتعلم إنجاز مهام لها معنى وترتبط بحياته الواقعية، وأيضا حل مشكلات واقعية (Henson and Eller, 1999).

لذلك يجب التركيز على تنمية مهارات وكفايات لدى الطلاب بالإضافة إلى المهارات الأساسية، حتى تكون لديهم القدرة على مواجهة مواقف الحياة الواقعية، حتى يتمكنوا من استعراض تلك المهارة أثناء أدائهم على وسائل التقييم البديلة بدلا من الاختبارات التحصيلية المقننة(*) .

وإذا كانت المعلومات التى يتعامل معها الطالب تنقسم إلى ثلاثة أنواع هى:

- معلومات أو معارف تقريرية وتشمل الحقائق والمفاهيم والمبادئ وهى المعلومات التى تدور حول ماهية الشئ، واختبارات التحصيل التقليدية لها قدرة استدلالية عالية فى قياس نواتجها.
- معلومات أو معرفة اجرائية وهى التى تدور حول معرفة كيفية أداء الأشياء، وتشمل الإجراءات المعرفية التى يقودها البحث العلمى مثل : "التحكم فى بعض المتغيرات"، وتعتبر تقديرات الأداء وسيلة لها قدرة استدلالية جيدة لقياس هذا النوع من المعارف.
- معلومات أو معارف استراتيجية وبمعرفة متى ولماذا وكيف نعمل الأشياء؟ حيث تكامل بين المعرفة الإجرائية والمعرفة التقريرية، مما يساعدهم على صنع مواقف نوعية.

(Ruiz-Primo and Shavelson, 1996b)

فإن الحاجة تتطلب أنواعاً من التقويم تتناسب وأنواع هذه المعلومات، بحاجة إلى تقييم الأداء وبحاجة إلى تقييم يرتبط بمواقف الحياة الواقعية، حتى نستطيع الاجابة على الأسئلة متى، ولماذا، وكيف نعمل الأشياء؟ والتقييم الحقيقى هو ذلك النوع الذى يتناسب مع المعارف الإجرائية والاستراتيجية، لذلك ما هى معايير هذا النوع من التقييم؟ وما هى خصائصه؟ وما هى أهدافه؟ وفيما يلى توضيح لذلك.

(*) <http://www.fair Test Principles and Indicators.htm>.

معايير التقييم الحقيقي:

- تتحدد المعايير فيما يلي:
- التركيز على الوثائق الفردية والتي توضح النمو المستمر للطلاب، وذلك أفضل من مقارنة الطالب بغيره من الطلاب.
- التأكيد على جوانب القوة لدى الطلاب، كم يعرفون؟ وما الذى يعرفونه؟ أكثر من التأكيد على جوانب الضعف "كم لا يعرفون"؟ وما الذى لا يعرفونه؟

خصائص التقييم الحقيقي:

- يظهر نموذج التقييم الحقيقي فى حجرة الدراسة بعض الخصائص العامة وهى:
- يطلب من الطالب انجاز عمل خلاق Create أو انتاج شىء ما.
- تشجيع الطلاب على التأمل الذاتى Self-Reflection .
- قياس النواتج ذات الدلالة.
- التأكيد على مهارات حل المشكلات ومستويات التفكير العليا.
- استخدام مهام توضح أنشطة التدريس التى لها قيمة.
- يستشهد بالتطبيقات التى ترتبط بالعالم الواقعى.
- يتطلب تدريساً جيداً وأدوات لتقييم المعلمين.
- يزود الطلاب بفرص التقييم الذاتى.
- يتيح فرصاً مناسبة للفرد ومجموعة العمل.
- يشجع الطلاب على التواصل مع أنشطة التعلم المرتبطة بما وراء الأهداف.
- يحدد بوضوح محك الأداء على المهمة.

(Barton and Collins, 1993) (Aly, 2001)

- ويذكر Wiggins 1998، أن التقييم الحقيقى يتميز بـ:
- الأصالة: حيث يرتبط هذا النوع من التقويم بمهام تعلم الطلاب الأعمال التى تواجه الكبار فى مجال عملهم.
- التغذية الراجعة الفورية لكل من الطلاب والمعلمين، لمراجعة أدائهم بالنسبة للأعمال التى يقومون بها وأعمال مشابهة لها، فى (رجاء أبو علام، 2001 : 110).

كما أن التقييم الحقيقى يتم قبل وبعد وأثناء عملية التعلم وهو محكى المرجع، يجعل التمكن من محكات الأداء هدفاً له، ويساعد المعلم على اتخاذ قرارات تربوية سليمة (Brosnan, 1998).

وفى دراسة (Sanders and Horn, 1995) تم تحديد العديد من خصائص التقييم الحقيقى منها:

- مقدرته على جعل التقييم وفقاً لحاجات كل فرد وقدراته "تفريد" التقييم.
- يؤدى إلى استخدام ممارسات تجعل التدريس جيداً ويؤدى إلى تحسين عملية التعلم.
- تساعد الطلاب على عرض مهارات التفكير العليا مثل: التفكير الناقد - التحليل - التركيب - الاستدلال وحل المشكلات.
- يساعد على تقييم جودة مهارات التفكير العليا.
- تفسير الاستراتيجيات المستخدمة فى حل المشكلات المعقدة.

أهداف التقييم الحقيقي:

تعددت أهداف التقييم الحقيقي كما حددها (Pett, 1990) في (كمال زيتون وعادل البناء، 2001) ومنها:

- تنمية قدرة المتعلم على الاستجابة لمهام التعلم أو مشكلات الواقع الحياتية.
- اختبار مهارات التفكير العليا.
- تقويم المشاريع الجماعية بصورة مباشرة وحقيقية.
- تجميع عدد من عينات عمل الطلاب خلال فترة زمنية طويلة.
- إتاحة الفرصة للمتعلمين لتقييم أعمالهم بأنفسهم.

يتضح مما سبق أن التقييم الحقيقي يساعد في الوقوف على تقدم تحصيل الطلاب من خلال أدائهم على مهام ترتبط بمشكلات الحياة الواقعية تسهم في استعراض الطلاب لمهارات التفكير العليا، على أن تتاح الفرصة لأن تقدم مهام لكل طالب في ضوء مستوى قدرته، كما تتيح الفرصة للطلاب أن يشتركوا في تقييم أعمالهم، مما يجعلها أكثر فاعلية وكفاءة من الاختبارات التحصيلية المقننة.

وبالتالي التقييم الجيد هو الذي يؤدي إلى تطوير التعليم من خلال استخدام طرائق تدريس جيدة، وعليه فإن كل تقييم يقيس عينة ممثلة للمحتوى بطريقة ملائمة وكانت أدواته على درجة عالية من الجودة الفنية، نستطيع أن نستنتج أن الطالب الذي يؤدي أداءً جيداً على أداة التقييم تلك يمكن أن يؤدي أداءً جيداً في باقي المحتوى الذي لم يتم قياسه، وإذا كان الهدف من التدريس هو من أجل الاختبار، فإننا لا نحصل على تلك النتيجة وتفقد فاعلية وفائدة التقييم كأداة لقياس المجال الواسع، (Bond, 1994).

وتعتمد فاعلية التقييم الحقيقي على خصائص المهام والتي تؤدي إلى الكشف عن العمليات العقلية العليا لدى الطلاب، فالتقييم الحقيقي يعتمد على الاختبارات القائمة على الأعمال التي يقوم بها الأفراد، وهو التقييم الفعلي للأداء والتي تساعد الطلاب على التجديد والابتكار، لذلك من أهم الخصائص التي يجب أن تتوفر في مهام التقييم الحقيقي ما يلي:

- الواقعية: أن ترتبط بمواقف الحياة الفعلية.
- الحكم والتجديد: قيام الطلاب باستخدام المهارات التي تسهم بفاعلية في حل المشكلات.
- الممارسة العملية: سعى الطالب للاكتشاف والعمل ضمن إطار المقرر.
- المحتوى: تكون ضمن محتوى المقرر على أن ترتبط بالحياة الواقعية.
- التكامل: إظهار الطالب لأكثر من مهارة والتكامل بينها لحل المشكلات المعقدة.
- يتيح فرص التدريب والممارسة والحصول على التغذية الراجعة، (رجاء أبو علام، 2001 : 111-112).

وتستخدم اختبارات التقييم الحقيقي في بريطانيا لانتقاء الطلاب للقبول بالجامعات، كما يستخدمها أصحاب العمل للوقوف على المهارات التي يجب أن تتوفر في المتقدم وتناسب طبيعة العمل، كذلك يجب أن تتضمن اختبارات التقييم الحقيقي الاحتياجات الواقعية المرتبطة بأى مجال، (Rust and Golombok, 1995: 102).

لذلك فتقدير الأداء من خلال عينات عمل الطلاب عمل ملائم وممتع يساعد المعلمين في تحديد المهارات التي يستخدمونها في الأداء، وتعتبر إحدى استراتيجيات التقييم التي تساعد على تحديد تقديرات الطلاب ومن ثم التطوير، كما تساعد الطلاب على التواصل مع الآخرين من خلال المشاركة في تنفيذ المشروع وقد تستخدم الملاحظة كوسيلة إضافية لتقييم هذا الأداء. (Roeber, 1996)

ويذكر (حسين بشير، 2001 : 81) أن التقييم الأصيل هو الذى يعتمد على عمل حقيقى وتقويم الأداء الأمر الذى يتطلب أن يكون التعليم قائما على الأداء، مع ضرورة أن تحدد معايير ومستويات الأداء والتي ينبغي على التلميذ التمكن منها بعد دراسته في نهاية كل مرحلة، وأن يتم التعرف على ما يتحقق من أهداف تربوية من خلال قياس نواتج التعلم.

وإذا كانت الاختبارات التحصيلية التحريرية تؤدي إلى عملية تقييم محدودة بالنسبة للمعرفة البنائية التقريرية، وإذا كان ما نحتاجه هو صور عقلية لدى المتعلم توضح كيف ينظم المفاهيم الأساسية، فإننا بحاجة إلى الأدوات البديلة التي تحقق البنية المعرفية بشكل مباشر ومنها خرائط المفاهيم، (Richard and Ruiz-Primo, 1998).

والتقييم الحقيقى يشتمل على مجموعة من المقاييس المتنوعة التي تناسب المواقف المتباينة يتناول الباحث منها الصحائف الوثائقية أو سجلات الأداء Portfolios وخرائط المفاهيم Concept Maps والتي يتم تناولها فيما يلى:

سجلات الأداء أو الصحائف الوثائقية Portfolios :

وتشمل الصحيفة الوثائقية تسجيل جميع الأنشطة والأعمال المدرسية للطلاب طوال العام، أو تشمل على عينة ممثلة لأعمال الطالب توضح إنجازاته لكي يطلع عليها ولي الأمر ويتعرف على نوعية هذه الانجازات، كما يمكن أن تشمل هذه الصحيفة على مجموعة منتقاه من الأعمال لغرض معين، مثل عينة من كتابات الطالب أو مهاراته في حل المشكلات أو بعض أعماله لعرضها على أقرانه أو أسرته، والتقييم لهذه الصحيفة يهدف إلى التحقق من جودة ونوعية تحصيل الطالب واكتسابه المعارف والمهارات والقدرات العقلية من خلال فحص أعمال الطالب في ضوء محكات متفق عليها ومرتبطة بواقع المجتمع مع التركيز على شكل ومحتوى الوثيقة. (صلاح علام، 2000 : 746-747).

ومن أنواع المواد التي يمكن أن يشملها سجل الأداء يكون في صورة عمل فنى كرسوم بيانية أو خرائط أو لوحات - تسجيلات صوتية - شرائط فيديو - عينات لكتابة حوار مسرحى أو كتابة مشاريع - الأعمال الجماعية التعاونية - تسجيل المقابلات - استخدام قوائم التقدير بواسطة المعلم أو الطالب نفسه أو ولي الأمر - تسجيل نتائج الاختبارات التحصيلية وغيرها من الأنشطة، (كمال زيتون، عادل البناء، 2001) (Aly, 2001). كما أنها مجموعة من الأعمال للطلاب تستخدم كوثائق توضح مدى التقدم في الأهداف التعليمية والأغراض الرئيسية وتحسين التواصل بين الطلاب وأولياء الأمور، وتسهم في تحسين التدريس عندما تكون عينات العمل للطلاب مختارة بعناية،

(Carey, 1994, 378).

والصحائف الوثائقية يمكن تحسينها وتطويرها من خلال:

- التخطيط الجيد لاستراتيجية التقييم والتي يتم من خلالها تقدير مدى التقدم.
- تطوير صيغ تقييم الصحائف الوثائقية.
- اختيار الحافظات أو الحاويات التي توضع فيها الصحائف الوثائقية.
- وضع المعلومات وعينات العمل داخل الحاويات.
- مقارنة وضع الطالب قبل وبعد الأداء - تقييم الأعمال - لتحديد مقدار التحسن.
- كتابة ملخص يصف وضع الطالب والنمو الحادث له، (Carey, 1994, 378).

تقييم الصحائف الوثائقية:

- الخطوات التالية تستخدم لتقدير محتوى الصحيفة من حيث: التحصيل - الانتاج - الأداء أو السلوك باستخدام مقاييس التقييم Rating Scales
- إعداد جدول مختصر يضم النقاط الهامة التي يحددها الطلاب لكل المكونات.
 - استخدام معايير توضح النقاط أو النسب لكل مكون فرعى من قبل المهني (المقيم).
 - تجميع النسب على النقاط الهامة لكل مكون فرعى.
 - تحديد المكونات الفرعية الأقل أهمية والهامة في ضوء المحكات الرسمية.
 - تقييم البرامج التدريسية والتعليمية في ضوء النتائج على الصحائف الوثائقية.
- (Carey, 1994, 493)

وعند التقييم تستخدم التمرين العملي التالي كمرشد لعملية تقييم الصحائف الوثائقية:

- ما هو الغرض الرئيسي من استخدام الصحائف الوثائقية؟
- ما نوع المعلومات التي تلخص صيغ الصحائف الوثائقية؟
- ما نوع النواتج التي تستطيع تسجيلها في الوثائق والتي تكشف عن المهارات الأولية بالنسبة للأهداف التعليمية؟
- ما نوع النواتج التي تستخدم كوثائق توضح تقدم الطلاب؟
- ما هي أفضل استراتيجية تستخدم لمعرفة جوانب القوة والضعف في مهارات الطلاب؟
- استخدام مقاييس التقييم لمقارنة صيغ تقييم الصحائف واستخدام التغذية الراجعة

(Carey, 1994, 494)

والصحائف الوثائقية وسيلة لتقييم أداء الطالب تستخدم للتغلب على عيوب الاختبارات المقننة وعيوب الاختبارات محكية المرجع خاصة عند تقييم تحصيل الطالب في الدراسات الاجتماعية، ولكن لتحقيق الأهداف يحتاج الأمر إلى تقييم متكرر لأنشطة نوعية من عمل الطلاب مما يحتاج إلى خبرات تسهم في تقييم بنية تنظيم المناهج والتركيز على الموضوعات المنفصلة، (Ediger, 2000).

والصحائف الوثائقية باعتبارها أحد أنواع التقييمات البديلة، فإن من بين خصائصها وأهدافها إتاحة الفرصة للمتعلمين لتقييم أعمالهم بأنفسهم، بمعنى التأكيد على دور التقويم الذاتي وبذلك يعتبر التقييم الذاتي أحد أنواع التقييم البديل.

التقويم الذاتي Self Evaluation وهو تلك العملية التي تمكن الفرد بنفسه من إجراء تقويم شامل لمختلف عناصر الأداء التعليمي تقويمياً يمكنه التأكد من رصد واستحداث

التطويرات والتجديدات الشاملة لتقابل في النهاية متطلبات معايير الاعتماد، (إسماعيل دياب، وعادل البناء، 2001).

والتقييم الذاتى ينمى لدى الطلاب الثقة بالنفس - وتقدير الذات والتي بدورها تشجعهم على إدراك الفجوات فى معارفهم وتفكيرهم والعمل بسرعة لتقليل تلك الفجوات، (Stateman, 1993).

وفى دراسة (El-Koumy, 2001) لبحث أثر التقييم الذاتى على التحصيل المعرفى والتفكير الأكاديمى على عينة من 94 طالباً يدرسون مقرر طرق تدريس اللغة الانجليزية صنفوا فى مجموعتين "تجريبية - ضابطة" تستخدم المجموعة التجريبية التقييم الذاتى من خلال استراتيجية تبنتها الدراسة لمدة ثلاثة أشهر بواقع محاضرة واحدة كل أسبوع، وأسفرت الدراسة عن عدم وجود فروق ذات دلالة إحصائية بين متوسطى أداء مجموعتى الدراسة فى كلٍ من التحصيل المعرفى والتفكير الأكاديمى، بمعنى عدم دلالة أثر التقييم الذاتى على التحصيل الدراسى، وعليه قدم الباحث مجموعة من الفنيات تساعد على تفعيل استراتيجية التقييم الذاتى من خلال : مساعدة الطلاب على تصميم فنيات للتقييم الذاتى تساعدهم فى التحصيل وتحقيق أهداف المقرر - إمداد الطلاب بالمساعدة أثناء تقييم أنفسهم تحصيلياً وخفض مستوى المساندة تدريجياً بعد تنمية ثقتهم بأنفسهم - تقديم تغذية راجعة للطلاب ومساعدتهم على تعديل استراتيجياتهم بأنفسهم فى تقييم التدريس - العمل على تكوين اتجاهات إيجابية لدى الطلاب نحو التقييم الذاتى لتطوير مستوى تحصيلهم.

وقد جاءت نتائج تلك الدراسة متناقضة مع نتائج العديد من الدراسات التى توصلت إلى وجود علاقة ارتباطية موجبة ذات دلالة إحصائية بين التقييم الذاتى ودرجات الاختبارات الحقيقية، ومن ثم فإن التقييم الذاتى يساعد الطالب ليكون مشاركاً فعالاً فى بناء المعرفة وتطوير التفكير، (O'Neil, 1992).

مشكلات تقييم التحصيل باستخدام الصحائف الوثائقية:

على الرغم من المميزات التى تم عرضها فى الجزء السابق إلا أنه هناك بعض المشكلات المرتبطة بالصحائف الوثائقية كأحد أنواع التقييم الحقيقى منها:

1 - عدم اتفاق المسئولين حول ما يجب تقييمه فى الصحائف الوثائقية، فالبعض اعتمد على تقييم الأهداف والمستويات القومية - توجيهات خبراء المواد الدراسية .. الخ بدلا من الاقتصار على وضع قائمة بالمهارات المرجوة.

2 - دور المعلم والذى يتطلب تنمية مهنية للقيام بهذه الأدوار الجديدة وتصميم ما يجب أن يتناوله الطالب فى الصحيفة - إرشاد الطالب - تحديد الجدول الزمنى - المشاركة فى بناء محكات التقييم .. الخ.

3 - تباين تقييم صيغ الصحائف الوثائقية المتنوعة فى ضوء أهدافها، فعندما يكون الهدف من التقييم قياس نوعية التحصيل فإن ذلك يؤثر على تحسين عملية التعلم. فتطوير أداء البرنامج لتحقيق هدف ربما يقلل من فاعلية أدائه فى تحقيق هدف آخر مما ينعكس على بنية الصحيفة.

4 - صدق وثبات درجات الصحائف، حيث يشير الصدق المرتبط بالمحك - درجات الاختبارات المعتادة فى نفس المجال - أن قيم معاملات الارتباط كانت منخفضة، كما

اعتمد فى تقدير الثبات على تقديرات المحكمين وكانت القيم أيضا منخفضة. (صلاح
علام، 2000، 749-750).

ومن أبرز المشكلات اختيار المعلم لعينات عمل كوئائق للصحائف لا ترتبط مباشرة
بمحتوى المقرر وأهداف التدريس بمعنى عدم التركيز على المواد والأعمال المنتقاه - عدم
التركيز على المقرر طول الوقت، (Carey, 1994, 392-393).

ويضيف (Sewell et al., 2000) مجموعة من عيوب الصحائف الوثائقية تحد
من انتشارها وهى:

- أن الصحائف الوثائقية أقل ثباتاً وعدلاً من أنواع التقويم التى تعتمد على التقدير الكمي "درجات
الاختبار".
- تستهلك وقتاً وجهداً كبيراً من قبل المقدرين فى تقييم محتوياتها.
- المعايير المستخدمة فى تقدير محتويات الصحائف صعبة التطبيق وتحتاج لتدريب عليها.
- قد تحتوى الصحائف على مجموعة متنوعة من الأعمال أو الأنشطة التى لا توضح التقدم أو
النمو لدى الفرد، حينما تكون المعايير والأهداف غير واضحة ومحددة تماماً.
- صعوبة التعامل مع البيانات الكيفية وصعوبة تحليلها.

خرائط المفاهيم Concepts Maps :

وهي إحدى الأدوات البديلة للتقييم الحقيقي وتتكون - خريطة المفهوم - من نقاط التقاء Nodes تمثل المفهوم، وخطوط موصوفة Labeled بين النقاط تمثل العلاقات، ونقاط الالتقاء على الخطوط تمثل العلاقة الطبيعية بين مفهومين، كما يحدد الخط الموصوف الذي يربط بين زوج من نقاط الالتقاء المقترح - الفكرة - والتي تمثل الوحدة الأساسية لخريطة المفهوم، (Ruiz-Primo and Shavelson, 1996a).

ويذكر Horton et al. 1993 أن خرائط المفاهيم بمثابة تمثيل للمعنى، كما أنها تمثل هياكل تصويرية خاصة بمجال معين من المعرفة. في (كمال زيتون وعادل البناء، 2001 : 212). ولخريطة المفهوم مكونات تخضع للتقييم هي : المهمة - الاستجابة - نظام الدرجات. كل مكون يتضمن ثلاثة أبعاد فرعية، لكل بعد مطالب وهي:

المكون الأول: المهمة وتنقسم إلى تباينات ثلاثة هي:

- أ - متطلبات المهمة حيث يستطيع الطالب تحديد هيكل الخريطة: ملئ مخطط - بناء خريطة بالرسم - تنظيم الكروت - تقييم العلاقة بين زوج المفهوم - كتابة مقال - الاستجابة أثناء المقابلة.
- ب - يستطيع الطالب أو لا يستطيع عمل: بناء خريطة هرمية - تجهيز المفاهيم التي تستخدم في المهام ، إعداد روابط المفاهيم المستخدمة في المهمة - تخصيص أكثر من استخدام للرابطة الواحدة بين نقاط الالتقاء - تحديد المسافات في الخريطة - تبرير الاستجابات - رسم خريطة كلية - تحديد المفاهيم التي تدور حول البنية المقدمة.
- ج - بنية محتوى المهمة "نقط التقاطع" : تحديد متطلبات المهمة والالتزام بمجال بنية الموضوع.

المكون الثاني: الاستجابة ولها صور ثلاث هي:

- أ - استجابة الطالب سواء ب: الورقة والقلم - شفهيًا - باستخدام الكمبيوتر.
- ب - الشكل : يكون ملائم ودقيق ويرتبط بالمهمة.
- ج - معد الخريطة: قد يكون الطالب - المعلم أو الباحث.

المكون الثالث: نظام الدرجات وتنقسم إلى ثلاثة أنواع هي:

- أ - درجات مكونات الخريطة، حيث يتم التركيز على المقترحات أو الابتكارات - المستويات الهرمية - الأمثلة - عدد نقاط الالتقاء - عدد الارتباطات.
- ب - استخدام معايير للخريطة: تقدير أداء الطالب من خلال مقارنة درجاته عن طريق: واحد أو أكثر من الخبراء في المجال - واحد أو أكثر من المعلمين - واحد أو أكثر من طلاب القمة.
- ج - الجمع بين مكونات الخريطة ومعايير الخريطة حيث يشكلان درجات الطالب على خريطة المفهوم، (Ruiz-Primo and Shavelson, 1996b).

وتوجد العديد من الدراسات تناولت استخدام التقييم الحقيقي باستخدام خرائط المفاهيم فى تقويم التحصيل الدراسى منها:

دراسة Henson and Eller, 1999 – Ruiz-Primo and Schavelson, 1996a، كما أن عدداً من الدراسات استخدمت التقييم الحقيقى باعتباره إستراتيجية فى التدريس لها تأثير فعال على التحصيل الدراسى خاصة فى مادة العلوم وفروعها منها دراسات كل من: (محمد عبد الرؤوف وأسامة عبد العظيم، 1994)، (سحر عبد الكريم، 2000)، (نجيب محفوظ، 2001)، (إبراهيم شعير، 2002)، (أحلام الباز ورضا حجازى، 2002)، (Henson and Eller, 1999; Wolf and Others, 1997).

مشكلات التقييم الحقيقى باستخدام خرائط المفاهيم:

- تعد المشكلة الأكثر بروزا هى ما يتصل بتقديم الدليل الامبيريقى حول ثبات وصدق خرائط المفاهيم على الرغم من وجود معاملات ارتباط مرتفعة بين درجات تقدير خرائط المفاهيم وأشكال الاختبارات التقليدية مثل اختبارات الاختبار من متعدد. فى (كمال زيتون - عادل البناء، 2001 : 214).
- عدم توفر معلمين مدربين لتقدير الدرجات على خرائط المفاهيم، بمعنى عدم توفر كفايات مناسبة لدى المعلمين تسهم فى تقييم صادق للخرائط.
- لا توجد حاجة تفرض بنية هرمية محددة إذا كان المحتوى بناؤه غير هرمى.

(Ruiz-Primo and Shavelson, 1996b)

ويرى الباحث أن خرائط المفاهيم تسهم فى تقييم أداء الطالب وتكشف عن المهارات التى ترتبط بالعمليات العقلية العليا، وتساعد فى تقدير المعرفة الاجرائية والمعرفة التقريرية، لكن تظل فاعليتها محدودة فى بعض المواد الدراسية التى يمكن بناء هياكل هرمية لها نظراً لتشعب بعض المفاهيم لياخذ النظام الشجرى وخاصة فى مجال العلوم: "فيزياء - كيمياء - أحياء" ويظل الأمر صعب التنفيذ فى كثير من مواد الدراسات الانسانية والعلوم الشرعية، كما أنه يتطلب خبراء فى مجال خرائط المفاهيم ومعلمين مدربين أو تم إعدادهم من خلال برامج إعداد المعلمين لتكون لديهم كفايات تسهم فى تقييم صادق لخرائط المفاهيم.

عيوب أو مشكلات التقييم الحقيقى بصفة عامة:

- إذا كان بالإمكان إجراء فحص تفصيلى لاستيعاب الطالب المهام أو الأنشطة، فإن ذلك يعد أمراً محدوداً بسبب طول الزمن الذى يستغرق فى عملية التقييم وعدد الأنشطة المحدود فى الاختبار الواحد لطول الزمن اللازم لتنفيذ النشاط الواحد. (Sanders and Horn, 1995, 8)
- والأمر لا يقف عند توفر عدد كاف من الأنشطة لتقييم أداء الطالب، لكن المواقف الأدائية لها عمليات خاصة تتغير وتتنوع على نفس النشاط من موقف لآخر مما يؤدي إلى تغير درجة الطالب. (Schavelson et al., 1992, 26).
- الأداء على الأنشطة يعتمد بدرجة كبيرة على نوع النشاط.
- صعوبة تعميم النتائج لأنها تعتمد على عدد محدود من الأنشطة.
- صعوبة إجراء تجارب استطلاعية لارتفاع التكاليف وطول الوقت المستغرق وبالتالي صعوبة تعميم النواتج، حيث قدم George تقريراً يذكر فيه أن تكلفة استخدام أساليب تقييم الأداء فى جميع المواد الدراسية بأمريكا يتكلف من 2.5-3 بليون دولار سنوياً مما يؤدي إلى صعوبة إجراء عملية التقنين نظراً لارتفاع التكاليف.
- صعوبة إعداد أنشطة يتم تطبيقها بصورة جماعية.

- يذكر Worthon 1963 أن مؤيدى التقييم البديل لا يملكون أدلة دقيقة في أن الطلاب يستخدمون مهارات التفكير العليا عندما يؤدون عملاً يدوياً على أى نشاط، على الرغم من أنها اختبارات مفتوحة تقدم مساحة أوسع للاستجابة مقارنة بغيرها من الاختبارات الأخرى (Sanders and Horn, 1995).
- ويذكر Gregoire, 1997 أن التطور في القياس النفسى الحديث، أدى إلى تغيرات ذات دلالة في وسائل وأدوات التقويم التى تعتمد على الكفايات والتي توفر معلومات مفيدة وجيدة للممارسين فى فهم صعوبات التعلم ولكنها ترتبط ببعض المشكلات المنهجية التى تحد من فاعليتها.

تعليق عام:

- لقد جاءت النظرية الحديثة فى القياس للتغلب على مشكلات النظرية الكلاسيكية، على أن الاختبارات التحصيلية التى تصمم فى إطار النظرية الحديثة تكون على درجة عالية من الموضوعية، يمكن تعميم نتائجها وتتيح الفرصة للمقارنات بين مجموعات متباينة .. الخ، وعلى الرغم من ذلك فالحاجة ماسة إلى دراسة ميدانية للمقارنة بين الدرجات الخام والتي تعبر عن التحصيل الدراسى كما يقاس بالاختبارات التحصيلية التى تم إعدادها فى إطار النظرية الكلاسيكية، وبين قدرات الطلاب والتي تعبر عن مستوى التحصيل الدراسى كما يقاس بالاختبارات التحصيلية والتي تم تدرجها باستخدام نموذج "راش" أحد النماذج فى النظرية الحديثة، لتسهم فى اتخاذ قرارات مبنية على أسس علمية.
- تفترض النماذج الحديثة فى القياس (أحادية - ثنائية - ثلاثية) البارامتر إلى أن أداء الفرد على المفردات لا يختار ما يعود إلى قدرة واحدة، وهذا يتطلب كفاءة عالية عند إعداد الاختبارات وصياغة المفردات بحيث أن الاستجابة على كل مفردة لا تعكس إلا سمة كامنة واحدة لدى الفرد، وقد تكمن الصعوبة عند صياغة المفردات التى يتطلب الاستجابة عليها قدرة أو سمة واحدة لدى الفرد، لوجود قدرات متعددة تعمل معاً وفى أن واحد، لذلك يجب الحذر والحيطه والدقة عند بناء مفردات الاختبارات.
- صعوبة تحديد الأفراد غير الملائمين والمفردات غير الملائمة عند إعداد بنوك أسئلة، وذلك بسبب حالات الغش أو التخمين أو عدم الجدية والتي ينتج عنها إجابة على مفردات تزيد صعوبتها عن مستوى قدرة الفرد وعدم الإجابة على مفردات تقل فى صعوبتها عن مستوى قدرة الفرد مما يؤثر على مؤشرات الملائمة للمفردات.

وللتغلب على هذه المشكلات يجب الاستفادة من التطور الهائل والمستمر فى تكنولوجيا المعلومات والتقنية الحديثة، فالتغيرات المستمرة هذه يفرض على المؤسسات التربوية أن تعيد النظر فى برامجها والاستفادة منها، لأن إصلاح التعليم المعتمد على التقنية الحديثة أصبح ضرورة ملحة من أجل تحسين جودة التعليم ومخرجاته، واستخدام وسائل تقويم تعتمد على الحاسب الآلى تيسر إجراء الاحصاءات المتقدمة وتوفير برامج تسهم فى حل مشكلات نماذج النظرية الحديثة فى القياس.

ويذكر (Ediger, 2000) أن من أبرز طرق تقييم تحصيل الطالب هي:

- الاختبارات المقننة والتي تعد بواسطة الخبراء والمهنيين والتي تستخدم مع جميع الطلاب وفى كل الأوقات وذلك بجانب الاختبارات المعتادة التى يقوم بإعدادها المعلم، وتتميز الاختبارات المقننة بموضوعية وصدق مفرداتها وملاءمتها لقياس تحصيل الطالب.

- الاختبارات مرجعية المحك (CRT) والتي تعتبر أكثر مناسبة وأهمية من الاختبارات المقننة، لأنها ترتبط بأهداف المقرر وبالتالي فهي أكثر موضوعية وأكثر جودة في تقييم التحصيل بعد الانتهاء من كل وحدة دراسية.
- الصحائف الوثائقية وهي المدخل الثالث الذى يستخدم فى تقييم أداء الطلاب، وتستخدم لعلاج القصور فى الاختبارات المقننة ومرجعية المحك، لأنها تعتمد على أنشطة وعينات لعمل الطلاب ترتبط بالحياة الواقعية. ويضيف Ediger أنه لى يكون التقييم أكثر موضوعية يجب أن يعتمد على أدوات ووسائل متعددة مثل : قوائم التقدير - مقاييس التقدير - التقييم الذاتى - التقييم من قبل المعلم .. الخ.

وبعد:

يلاحظ أن تقويم تحصيل الطلاب فى المؤسسات التعليمية اليوم يقتصر على مجرد قياس التحصيل، والذى يقتصر على قياس أداء الطلاب على الاختبارات سواء المقننة أو المدرسية أو مرجعية المحك، ولكن أكثر هذه الاختبارات شيوعاً هي الاختبارات المدرسية التى يقوم باعدادها المعلم والتي يتخذ فى ضوءها قرارات تربوية بشأن نقل الطلاب إلى الصفوف الأعلى أو من مرحلة لأخرى.

وبالتالى تعتبر الاختبارات على درجة كبيرة من الأهمية، مما يجعل من الضرورى التركيز فى بناء وتطوير الاختبارات الموضوعية واتباع الأساليب العلمية الحديثة فى عملية الاعداد بالاستعانة بالنماذج الحديثة التى تتضمنها نظرية الاستجابة المفردة لتحقيق أكبر قدر من الموضوعية، على اعتبار أن الاختبارات تساعد فى الكشف عن مدى تحقيق الأهداف التعليمية ومدى التقدم الأكاديمى ومعرفة نواحي القوة والضعف فى كل من: الأهداف - المحتوى - طرائق التدريس، والاستفادة من نتائج الاختبارات لإحداث تغذية راجعة لتدعيم جوانب القوة وعلاج أوجه القصور.

وعلى الرغم من المشكلات المرتبطة بالاختبارات التحصيلية كأدوات للتقويم، إلا أنها تظل الأكثر استخداماً وتداولاً ومألوفية لدى كل من المعلمين والطلاب، ومن ثم لا يمكن الاستغناء عنها، بل يمكن أن نرتقى بها لتكون أكثر موضوعية، وذلك عندما تعتمد النماذج الاحتمالية لنظرية الاستجابة المفردة ومنها نموذج "راش" لبناء بنوك الأسئلة والتي يمكن سحب نماذج اختبارات منها تصلح لعينات من الطلاب متباينة فى مستوى القدرات، كما أنها تسمح بالمراجعة والتطوير والإضافة والحذف، مع الأخذ بعين الاعتبار القصور فى بعض البرامج التى تستخدم فى حذف المفردات غير الملائمة أو الأفراد غير الملائمين، مما يؤثر على تقدير صعوبة المفردات نتيجة الغش أو التخمين أو عدم الجدية فى الإجابة، مع السعى لتطوير هذه البرامج اعتماداً على التقنية الحديثة وتكنولوجيا المعلومات واستخدام الحاسب الآلى، على أن تتضمن نماذج الاختبارات مستويات الأهداف المعرفية وخاصة التى تقيس العمليات العقلية العليا. ولكن لا يجب الاعتماد على الاختبارات بشكل فردى ونهاى لقصورها فى التقدير الدقيق والشامل لجميع الخبرات المعرفية والمهارية فى كل المستويات والتي يمر بها الطلاب.

وعلى الرغم من أن التقويم الحقيقى يحتاج إلى معلمين مدربين وخبراء لديهم دراية وعلم بكيفية التعامل مع وسائل التقويم الحقيقى وإصدار الأحكام بشأن أعمال الطلاب وإنجازاتهم من مهام وأنشطة، الأمر الذى يتطلب إعادة النظر فى برامج إعداد المعلمين لإكسابهم كفايات تتناسب مع النماذج الجديدة لتقويم التحصيل الدراسى، كما يتطلب إعداد برامج تدريبية للمعلمين

الحاليين للارتقاء بمستوى أدائهم كمقيمين في ضوء توفر أدوات التقييم البديلة، إلا أن المؤسسات التعليمية يجب أن تتبنى التقييم الحقيقي بجانب الاختبارات التحصيلية للتغلب على أوجه قصورها في تقييم الأهداف المهارية والمعرفية التي تستخدم في قياس العمليات العقلية العليا. وعليه أى شكل من أشكال التقييم يكون أكثر دقة في تحقيق الأهداف التعليمية؟ ما هو النموذج في التقييم الأكثر كفاءة بالنسبة لمدى تحقيق الأهداف التعليمية وبالنسبة للتكلفة وإمكانية التعميم؟

في ضوء ما سبق يجب ألا يقتصر تقييم التحصيل الدراسي على الاختبارات التحصيلية فقط أو اعتماد الأدوات البديلة، بل يجب أن تتنوع أدوات التقييم وأن تكون على درجة عالية من الموضوعية والثبات والصدق حتى يمكن الاطمئنان إلى نتائجها عند اتخاذ القرارات التربوية، وبالتالي يمكن الاعتماد على الاختبارات التحريرية والشفهية المقننة والموضوعية والتي تستل من بنوك الأسئلة التي تم إعدادها في ضوء النماذج الاحتمالية لنظرية الاستجابة المفردة. واستخدام أدوات التقييم الحقيقي "الصحائف الوثائقية - خرائط المفاهيم" والتي تستخدم في القياس أو الكشف عن مدى تحقيق الأهداف التعليمية التي يصعب على الاختبارات التحصيلية الكشف عنها.

- كما يجب ألا يقتصر تقييم التحصيل على الاختبارات النهائية التي تعقد في نهاية تدريس المقرر، بل يجب التكامل بين التقييم البنائي أو المستمر - والتقييم النهائي على أن تكون هناك استراتيجية واضحة ومتفق عليها على المستوى القومي، وكذلك تدريب الطلاب على التقييم الذاتي أى تقييم أنفسهم بأنفسهم مما يساعد على تنمية الثقة بالنفس من خلال الارتقاء بمستوى التحصيل.
- تقييم أداء الطلاب لجميع جوانب العملية التعليمية كما تتمثل في الأهداف المعرفية والمهارية والوجدانية واستخدام أدوات متنوعة للتقييم مثل: الاختبارات - قوائم الملاحظة - سجلات الأداء مع تنوع في أساليب التقييم (التحريري - الشفهي - العملي) بما يتناسب وطبيعة كل مقرر دراسي مع التأكيد على الموضوعية عند إعداد الأسئلة واثناء تقدير الدرجات وتبني النظريات الحديثة في القياس.
- بالنسبة للاختبارات التحريرية يجب أن تتضمن أسئلة متنوعة موضوعية ومقالية حتى تتيح الفرصة أمام الطلاب للتعبير عن أفكارهم وآرائهم بحرية وإظهار ما لديهم من قدرات عليا على أن تكون الأسئلة المقالية ذات إجابات محددة من أجل الحفاظ على موضوعية تقدير الدرجات.
- يجب أن تراعى الأهمية النسبية والوزن النسبي لكل موضوعات المحتوى عن إعداد اختبار تحصيلي وإعداد جداول مواصفات تتضمن عدد الأهداف المعرفية في كل موضوع وفق للمستويات المختلفة للأهداف التعليمية مع مراعاة الوزن النسبي لأهداف كل موضوع.
- يجب الارتقاء بمستويات الأداء من خلال وضع محكات عالية للتقييم تضمن الوصول إلى مستوى إتقان مرضى بناءً عليه يتم اتخاذ القرار بشأن إتقان أو عدم إتقان الطالب لوحدة أو مقرر دراسي.
- يجب قبل استخدام بنوك الأسئلة وأدوات التقييم الحقيقي عقد دورات تدريبية للمعلمين وذلك لتعريفهم بهذه الأدوات وكيفية التعامل معها بهدف إكسابهم الكفايات والمهارات اللازمة للتعامل ببسر وسهولة مع هذه الأدوات.

مما سبق يمكن وضع تصور مقترح لتقويم التحصيل فى مقرر دراسى على

النحو التالى:

- 1 - تفعيل التقويم البنائى (المستمر) خلال الفصل الدراسى على أن يخصص له 20 % من درجة المقرر.
- 2 - تكليف الطلاب بالقيام ببعض الأنشطة والمهام ذات الصلة بالمقرر الدراسى قد تكون فى صورة : مقالات - قصص قصيرة - أوراق عمل - بحوث وتقارير - تصميم تجارب - مشاريع .. الخ، ويخصص لها 20 % من درجة المقرر.
- 3 - العناية بالجزء العملى أو التطبيقى وتفعيل تقويم الأداء - الجانب المهارى - حسب طبيعة المقرر ويخصص له 10 % من درجة المقرر، وفى المقررات النظرية توزيع هذه النسبة بالتساوى على البندين 1، 2 .
- 4 - الاختبار النهائى ويخصص له 50 % من درجة المقرر ويجب أن يتضمن جزءاً تحريرياً وآخر شفهيًا، على أن يتضمن الجزء التحريرى أسئلة من نوع المقال وأخرى موضوعية يراعى فيها شروط ومواصفات الاختبار الجيد.

مراجع البحث

1. إبراهيم محمد شعير (2002): فعالية استخدام خرائط المفاهيم البارزة المدعومة بالمواد التعليمية للمسية على تحصيل التلاميذ المكفوفين واتجاهاتهم نحو مادة العلوم، الجمعية المصرية للتربية العلمية، المؤتمر العلمي السادس للتربية العلمية وثقافة المجتمع، الاسماعيلية (أبو سلطان) ، 28-21 يوليو، ص ص 259-339 .
2. أحلام الباز ورضا السيد حجازى (2002): تقويم تحصيل تلاميذ مرحلة التعليم الأساسي في العلوم باستخدام نظرية السمات الكامنة، الجمعية المصرية للتربية العلمية، المؤتمر العلمي السادس للتربية العلمية وثقافة المجتمع، الاسماعيلية (أبو سلطان)، 28-31 يوليو، ص ص 611-648 .
3. إسماعيل محمد دياب، عادل السعيد البنا (2001): تقويم جودة الأداء الجامعي، بناء نموذج رياضى وتطبيقه على بعض كليات جامعات الاسكندرية والمنصورة وأسيوط. الاسكندرية: المكتبة المصرية.
4. أمينة محمد كاظم (1988): دراسة نظرية نقدية حول القياس الموضوعى للسلوك "نموذج راش". الكويت، مؤسسة الكويت للتقدم العلمي، ص ص 1-161 .
5. أمينة محمد كاظم (1995): اتجاهات معاصرة في بناء بنوك الأسئلة، فى الأسس التربوية لإعداد المعلم الجامعي، ط2، جامعة عين شمس، ص ص 247-267 .
6. أمينة محمد كاظم (1996 أ): دراسة نظرية نقدية حول القياس الموضوعى للسلوك - نموذج راش". فى أنور الشرقاوى وآخرون، اتجاهات معاصرة فى القياس والتقويم النفسى والتربوى، القاهرة: الأنجلو المصرية، ص ص 281 - 430 .
7. أمينة محمد كاظم (1996 ب): استخدام نموذج "راش" فى بناء اختبار تحصيلى فى علم النفس وتحقيق التفسير الموضوعى للنتائج، فى أنور الشرقاوى وآخرون، اتجاهات معاصرة فى القياس والتقويم النفسى والتربوى، القاهرة: الأنجلو المصرية، ص ص 431-546 .
8. أمينة محمد كاظم (1996 ج): مستوى العينة وتدرج بنك الأسئلة باستخدام نموذج "راش" دراسة تجريبية فى أنور الشرقاوى وآخرون، اتجاهات معاصرة فى القياس والتقويم النفسى والتربوى، القاهرة: الأنجلو المصرية، 547-583 .
9. أمينة كاظم، رجاء أبو علام، نادية عبد السلام وعبد الفتاح القرشى (1992): تحليل نتائج امتحان الثانوية العامة فى جمهورية مصر العربية فى الفترة من 1986-1990، قسم البحوث، المركز القومى للامتحانات والتقويم التربوى بالقاهرة.
10. أمينة محمد كاظم وصلاح مراد وإسحاق بطرس (1989): تطور نظم الامتحانات بمراحل التعليم العام وبناء بنوك الأسئلة فى بعض الدول العربية. حقة دراسية إقليمية حول تطوير الامتحانات وبناء بنوك الأسئلة فى الدول العربية، برنامج التجديد التربوى من أجل التنمية فى الدول العربية (إبيداس) القاهرة 25-29 نوفمبر، ص ص 1-43.
11. أنور محمد الشرقاوى (1996): الاختبارات المرجعة إلى محك: وسائل جديدة فى القياس النفسى والتربوى. فى أنور الشرقاوى وآخرون، اتجاهات معاصرة فى القياس والتقويم النفسى والتربوى، القاهرة: الأنجلو المصرية، ص ص 19-39.
12. السيد إسماعيل وهبى (2001): بحث الفعل: تقويم جودة التعليم والتعلم، المؤتمر العربى الأول، الامتحانات والتقويم التربوى: رؤية مستقبلية، المركز القومى للامتحانات والتقويم التربوى، القاهرة 22-24 ديسمبر، ص ص 219-244 .

13. جابر عبد الحميد جابر (2001): التقييم كأداة للإصلاح التربوي (ورقة للمناقشة) ورقة عمل المؤتمر العربي الأول لامتحانات والتقويم التربوي: رؤية مستقبلية، المركز القومي لامتحانات والتقويم التربوي، القاهرة 22-24 ديسمبر، ص ص 1-4.
14. حسن حسيني جامع وآخرون (2001): تقويم الأوراق الامتحانية للثانوية العامة في الفترة من 1992-1999م - قسم البحوث، المركز القومي لامتحانات والتقويم التربوي بالقاهرة.
15. حسين بشير محمود (2001): تطوير نظم التقويم التربوي والامتحانات للتعليم قبل الجامعي "ورقة عمل" المؤتمر العربي الأول لامتحانات والتقويم التربوي: رؤية مستقبلية، المركز القومي لامتحانات والتقويم التربوي، القاهرة 22-24 ديسمبر، 79-85.
16. حمدى عبد العظيم البنا (2001): دراسة تحليلية للمستويات المعرفية فى امتحانات الكيمياء للشهادة الثانوية العامة، مجلة التربية العلمية، الجمعية المصرية للتربية العلمية، المجلد الرابع، العدد الأول، يناير ص ص 9-35.
17. رجاء محمود أبو علام (2001): النظريات الحديثة فى القياس والتقويم وتطوير نظام الامتحانات "ورقة عمل" المؤتمر العربي الأول لامتحانات والتقويم التربوي: رؤية مستقبلية، المركز القومي لامتحانات والتقويم التربوي، القاهرة 22-24 ديسمبر، ص ص 93-119.
18. رشدى فام منصور (1995أ): التقويم وبناء الاختبارات. فى الأسس التربوية لإعداد المعلم الجامعي، ط2، جامعة عين شمس، ص ص 193-205.
19. رشدى فام منصور (1995 ب): تعدد زوايا الرؤية عند تفسير درجات الاختبار وعند بنائه - ضرورة تربوية. فى الأسس التربوية لإعداد المعلم الجامعي، ط2، جامعة عين شمس، ص ص 207-214.
20. رضا عبد اقادير درويش (1995): تقويم "دليل تقويم التلميذ فى مادة العلوم العامة للصف الثالث الإعدادى" فى ضوء أهداف تدريس العلوم. مجلة كلية التربية ببها، الجزء الثانى، يناير ص ص 227-255.
21. سحر محمد عبد الكريم (2000): فعالية التدريس باستخدام اساليب التقويم الواقعي فى تحصيل الطالبة المعلمة للمفاهيم المتضمنة فى مقرر طرق تدريس العلوم وأدائها أثناء التربية العملية، مجلة التربية العلمية، الجمعية المصرية للتربية العلمية، المجلد الثالث، العدد الثانى، يونية، ص ص 43-79.
22. سليمان الخضرى وأمينة كاظم (1997): المشروع العربى للتقويم المقارن لمستويات التحصيل الدراسى فى التعليم العام، "المرحلة الأولى" المنظمة العربية للتربية والثقافة والعلوم بالتعاون مع المركز القومى لامتحانات والتقويم التربوي بالقاهرة.
23. سليمان الخضرى وأمينة كاظم (1998): المشروع العربى للتقويم المقارن لمستويات التحصيل الدراسى فى التعليم العام، "المرحلة الثانية" المنظمة العربية للتربية والثقافة والعلوم بالتعاون مع المركز القومى لامتحانات والتقويم التربوي بالقاهرة.
24. سليمان الخضرى وأمينة كاظم (2000): المشروع العربى للتقويم المقارن لمستويات التحصيل الدراسى فى التعليم العام، "المرحلة الثالثة" المنظمة العربية للتربية والثقافة والعلوم بالتعاون مع المركز القومى لامتحانات والتقويم التربوي بالقاهرة.
25. شكرى سيد أحمد (2001): العلاقة بين ثلاثة اتجاهات معاصرة فى مجال القياس التربوي النفسى ومشكلات تعميمها والإفادة منها. "ورقة عمل" المؤتمر العربي الأول لامتحانات والتقويم التربوي: رؤية مستقبلية، المركز القومي لامتحانات والتقويم التربوي، القاهرة 22-24 ديسمبر، ص ص 25-61.

26. **صديق عفيفي (2001):** اتجاهات تطوير نظم تقويم الامتحانات (ورقة عمل)، المؤتمر العربي الأول للامتحانات والتقويم التربوي: رؤية مستقبلية. المركز القومي للامتحانات والتقويم التربوي، القاهرة، 22-24 ديسمبر، ص ص 63-70 .
27. **صلاح أحمد مراد (2000):** الأساليب الإحصائية في العلوم النفسية والتربوية والاجتماعية، القاهرة: الأنجلو المصرية.
28. **صلاح الدين علام (1989):** بعض الاتجاهات المعاصرة في مجال إعداد وبناء نظم بنوك الأسئلة، برنامج التجديد التربوي من أجل التنمية في الدول العربية (إبيداس)، حلقة دراسية إقليمية حول تطور نظم الامتحانات وبناء بنوك الأسئلة في الدول العربية، القاهرة، 25-29 نوفمبر، ص ص 1-31 .
29. **صلاح الدين علام (1995):** الاختبارات التشخيصية "مرجعية المحك" في المجالات التربوية والنفسية والتدريبية، القاهرة، دار الفكر العربي.
30. **صلاح الدين علام (2000):** القياس والتقويم التربوي والنفسى - أساسياته وتطبيقاته وتوجهاته المعاصرة. القاهرة: دار الفكر العربي.
31. **طلعت كمال الحامولى (1996):** تقويم طلاب التعليم الأساسى والثانوى "دراسة تقويمية". مجلة علم النفس، السنة العاشرة، العدد (39) 84-108.
32. **عبد الرحمن سليمان الطيرى (1997):** القياس النفسى والتربوى: نظريته، أسسه، تطبيقاته. الرياض: مكتبة الرشد.
33. **عبد الوارث عبده الرازحى (2001):** تطوير نموذج معيارى لتقويم كفاءة نظام إعداد الاختبارات العامة. المؤتمر العربي الأول للامتحانات والتقويم التربوي: رؤية مستقبلية، المركز القومي للامتحانات والتقويم التربوي، القاهرة: 22-24 ديسمبر، ص ص 107-120 .
34. **عصام الدسوقي إسماعيل (1998):** مدى فاعلية نموذج "أنجوف" فى تحديد المستوى لاختبار محكى المرجع. مجلة كلية التربية بالمنصورة، العدد (36)، يناير ص ص 43-76 .
35. **فؤاد أبو حطب (1995):** أنواع الأسئلة. فى الأسس التربوية لإعداد المعلم الجامعى. ط2، جامعة عين شمس، ص ص 235-246 .
36. **فؤاد أبو حطب وسيد عثمان وأمال صادق (1997):** التقويم النفسى. ط4، القاهرة: الأنجلو المصرية.
37. **فكرى حسن ريان وآخرون (1994):** تقويم نماذج أسئلة المواد الدراسية بالتعليم الثانوى لعام 1991/1990 قسم البحوث، المركز القومي للامتحانات والتقويم التربوي بالقاهرة.
38. **كمال عبد الحميد زيتون وعادل السعيد البنا (2001):** سجلات الأداء وخرائط المفاهيم: أدوات بديلة فى التقويم الحقيقى من منظور الفكر البنائى. المؤتمر العربي الأول للامتحانات والتقويم التربوي: رؤية مستقبلية، المركز القومي للامتحانات والتقويم التربوي، القاهرة عام 22-24 ديسمبر، ص ص 187-218 .
39. **محمد خيرى محمود (1994):** دراسة تقويمية تحليلية للأسئلة المتضمنة فى كتب العلوم للمرحلة الاعدادية. المجلة المصرية للتقويم التربوي، المجلد الثانى، العدد الأول، ص ص 55-105 .
40. **محمد صادق صبور (1995أ):** إصلاح التقويم هو المدخل لإصلاح التعليم. فى الأسس التربوية لإعداد المعلم الجامعى، ط2، جامعة عين شمس، ص ص 183-191.
41. **محمد صادق صبور (1995ب):** خصائص الاختبار الجيد. فى الأسس التربوية لإعداد المعلم الجامعى، ط2، جامعة عين شمس ص ص 215-234 .

42. **محمد على نصر (2001):** رؤية مستقبلية لتفعيل الامتحانات والتقويم التربوى بالتعليم العام فى إطار مفهوم الجودة الشاملة. (ورقة عمل)، **المؤتمر العربى الأول للامتحانات والتقويم التربوى: رؤية مستقبلية**، المركز القومى للامتحانات والتقويم التربوى، القاهرة، 22-24 ديسمبر، ص ص 5-24 .
43. **محمد متولى غنيمه (1996):** تقويم نظم وبرامج إعداد المعلم فى مصر فى ضوء بعض التجارب العالمية المعاصرة، **قسم البحوث**، المركز القومى للامتحانات والتقويم التربوى بالقاهرة.
44. **محمد منصور الشافعى (2001):** أثر استجابات حالات الغش وعدم الجدية على مؤشرات الملائمة الإحصائية وتقديرات الصعوبة لمفردات بنوك الأسئلة من نوع المقال، **المؤتمر العربى الأول للامتحانات والتقويم التربوى: رؤية مستقبلية**، المركز القومى للامتحانات والتقويم التربوى، القاهرة 22-24 ديسمبر، ص ص 147-186 .
45. **محمود محمد غانم (1997):** القياس والتقويم، السعودية (حائل): دار الأندلس للنشر والتوزيع.
46. **مصرى عبد الحميد حنورة (1998):** أهمية المعالجات الإحصائية فى العلوم التربوية، **المجلة التربوية: الكويت**، إصدار خاص (5)، إبريل ص ص 5-31 .
47. **مصطفى محمد كامل (1999):** استخدام استراتيجية التعلم حتى التمكن فى تدريس مقرر التقويم التربوى "لاكساب الطلاب المعلمين مهارات بناء الاختبارات التحصيلية". **مجلة علم النفس**، السنة الثالثة عشر، العدد (51)، ص ص 6-31 .
48. **نادية على سعود (1999):** دراسة تتبعية تقويمية لأسئلة امتحانات إتمام الشهادة الإعدادية فى جمهورية مصر العربية من عام 1991 إلى 1994 فى مادة اللغة العربية. **المجلة المصرية للتقويم التربوى**، المجلد السادس، العدد الأول، ص ص 63-104 .
49. **نادية محمد عبد السلام (1992أ):** مشكلات عند بناء الاختبارات محكية المرجع "تحليل وتقويم"، **مجلة علم النفس**، السنة السادسة، العدد الثالث والعشرون، ص ص 30-40 .
50. **نادية محمد عبد السلام (1992 ب):** ثلاثة محكات رئيسية لتحديد مستوى الاتقان فى الاختبارات محكية المرجع "دراسة إمبريقية". **المجلة المصرية للدراسات النفسية**، العدد الثالث، ديسمبر، ص ص 88-105 .
51. **نادية محمد عبد السلام (1996 أ):** حول معايير التقويم: نظرة جديدة. **فى أنور الشرقاوى وآخرون**، اتجاهات معاصرة فى القياس والتقويم النفسى والتربوى، القاهرة: الأنجلو المصرية، ص ص 41-62 .
52. **نادية محمد عبد السلام (1996 ب):** ثبات وصدق القياس محكى المرجع. **فى أنور الشرقاوى وآخرون**، اتجاهات معاصرة فى القياس والتقويم النفسى والتربوى، القاهرة: الأنجلو المصرية، ص ص 63-85 .
53. **نادية محمد عبد السلام (1996 ج):** بناء اختبار هدفى المرجع وجماعى المرجع "دراسة مقارنة بين النموذجين". **فى أنور الشرقاوى وآخرون**، اتجاهات معاصرة فى القياس والتقويم النفسى والتربوى، القاهرة: الأنجلو المصرية، ص ص 86-120 .
54. **نادية محمد عبد السلام (1996 د):** دراسة فى حساب ثبات وصدق اختبار هدفى المرجع. **فى أنور الشرقاوى وآخرون**، اتجاهات معاصرة فى القياس والتقويم النفسى والتربوى، القاهرة: الأنجلو المصرية، ص ص 121-152 .
55. **نادية محمد عبد السلام (1996 هـ):** القياس محكى وتطوير التعليم ودراسة تجريبية. **فى أنور الشرقاوى وآخرون**، اتجاهات معاصرة فى القياس والتقويم النفسى والتربوى، القاهرة: الأنجلو المصرية، ص ص 153-195 .

56. نجيب محفوظ بالفقية (2001): أثر استخدام خرائط المفاهيم في تدريس مفاهيم في الكيمياء العضوية على تحصيل طلاب المرحلة الثانوية واتجاهاتهم نحو مادة الكيمياء بدولة الإمارات العربية المتحدة. مجلة التربية العلمية تصدر عن الجمعية المصرية للتربية العلمية، المجلد السابع، العدد الأول، ص ص 157-182 .

57. **Aly, M.M. (2001):** Authentic versus traditional assessments in the “EFL” classroom: what and why? *First Arab Conference, National Center Examination and Education Evaluation*, Egypt, 22-24 Dec., pp. (327-334).
58. **Barton, J. and Collins, A. (1993):** Portfolios in teacher’s education, *Journal of Teacher Education*, 44(3), 200-210.
59. **Bay-Luz (1998):** Comparing student performance on different item formats relative to achievement levels cut-points. Paper presented at the Annual Meeting of the National Council on Measurement in Education, *Eric No.* (ED 421532).
60. **Bond, L.A. (1994):** Critical Issue: Rethinking assessment and its role in supporting educational reform, *North Central Regional Educational Laboratory. All Rights Reserved.*
61. **Brosnan, P.A. (1998):** Interaction between assessment and instruction in science: a teacher’s decision-making process”. *D.A.T.*, 58(10), 3877a.
62. **Carey, L.M. (1994):** *Measuring and evaluating school learning*, (2Ed), London: Allyn and Bacon.
63. **Davison, M. L. and Chang, Y.W. (1992):** Measurement accuracy, an application of multidimensional item response theory to the Woodcock-Johnson Psycho-Educational Battery-Revised Achievement Scales. *Paper Presented at the Annual Meeting of the American Educational Research Association* (San Francisco, C.A., April 20-24).
64. **Doran, R.; Chan, F. and Tamir, P. (1998):** “Science educator’s guide to assessment”. *Eric, No* (ED 423126).
65. **Ediger, M. (2000 A):** Assessing student achievement in the social studies. *Eric No.* (ED 449180).
66. **Ediger, M. (2000 B):** Assessment of student achievement and the curriculum, *Eric No.* (ED 447202).
67. **EL-Koumy, A.S.A. (2001):** Effects of student self-assessment on knowledge achievement and academic thinking. *First Arab. Conference, National Center Examination and Education Evaluation*, Egypt, 22-24 Dec., pp. 313-326.
68. **Embretson, S.E. and Reise, S.P. (2000):** *Item response theory for psychologists*. New Jersey: Lawrence Erlbaum Associates, Inc., Publishers.
69. **Ercikan, K. (2003):** Advanced measurement topics: Item response theory, <http://www.ecps.educ.ubc.ca>.
70. **Gregoire, J. (1997):** Diagnostic assessment of learning disabilities: From assessment of performance to assessment of competence, *European Journal of Psychological assessment*, 13(1), 1-4.
71. **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), 1-4.

72. Hambleton, R.K. and Wedman, I. (1997): Introduction of the special issue: Advances in assessment practices, *European Journal of Psychological assessment*, 13(1), 1-4.
73. Hashway, R.M. (1998): *Assessment and evaluation of developmental learning*: Qualitative individual assessment and evaluation models. London West port, Connecticut's" Praiger Publishers.
74. Henson, K.T. and Eller, P.F. (1999): *Educational psychology for effective teaching*, U.S.A. Wedsworth Publishing Company, Division of International Thomson Publishing, Inc.
75. Hesse, D.C. (200): Standard and assessment for student achievement, New Jersey Department of Education <http://www.srste.nj.us/njded/adopted/standards>.
76. Jolly, A. (1999): "The effectiveness of learning with concept mapping on the science problem solving on sixth-grade children" *D.A.I.*, 59(9), 3356-3357.
77. Lawrence, R. (1998): Item banking-learninghouse on assessment and evaluation, Washington DC. *Eric (AE) No.* 423310.
78. Linn, R.L.; Baker, E.L. and Dumbar, S.B. (1991): Complex, performance-based assessment: expectations and validation criteria. *Educational Researcher*, 20(8), 15-21.
79. Meyar, J.A. (1992): What's different between "authentic" and "performance" assessment? *Educational Leadership*, 49(6), 39-40.
80. Molenaar, I. and Hoijtink, H. (1996): Person-fit and the Rasch model, with an application to knowledge of logical quantors. *Applied Measurement in Education*, 9(1), 27-45.
81. Nicholls, G. (1999): *Learning to teach, A handbook for primary and secondary school teacher's*, London, Bell and Bain Ltd.
82. Nitko, A.J. (2001): *Educational assessment of student*, (3Ed), New Jersey, Merrill Pentice Hall.
83. O'Neil, J. (1992): Putting performance assessment to the test. *Educational Leadership*, 9, 14-19.
84. Richard, J.S. and Ruiz-Primo, M.A. (1998): On the assessment of science achievement conceptual underpinnings for the design of performance assessment: *Report of year "2" Activities, Technical Report in Center for the Study of Evaluation National Center for Research on Evaluation, Standards and Student Testing, Graduate School of Educational of California, Los Angeles*.
85. Ruiz-Primo, M.A. and Shavelson, R.J. (1996A): Problems and issues in the use of concept maps in science assessment. *Journal of Research in Science Teaching*, 33(6), 569-600.
86. Ruiz-Primo, M.A. and Shavelson, R.J. (1996b): Rhetoric and reality in science performance assessments: An update. *Journal of Research in Science Teaching*, 33(6), 1045-1063.
87. Rust, J. Golombok, S. (1995): *Modern psycometrics, the science of psychological assessment*. London; Routledge.
88. Sander, J.R.; Hills, J.R.; Nitko, A.J.; Mewin, J.C.; Trice, C.; Dianda, M. and Schneider; J. (1990): Standards for teacher competence in educational assessment of students, *American Federation of Teachers*,

National Council on Measurement in Education, National Education Association, U.S.A.

89. **Sanders, W.L. and Horn, S.P. (1995):** Educational assessment reassessed: the usefulness of standardized and alternative measures of student achievement as indicators for the assessment of educational outcomes, ***Education Policy Analysis Archives***, 3(6), 1-24.
90. **Sewell, M.; Marczak, M. and Horn, M. (2000):** The use of portfolio assessment in evaluation, [http://ag.arizona.edu/fcr/fs/cyfar/portfos"3htm](http://ag.arizona.edu/fcr/fs/cyfar/portfos)
91. **Shavelson, R.; Baxter, G. and Pine, J. (1992):** Performance assessment: political rhetoric and measurement reality. ***Educational Researcher***, 21(4), 22-27.
92. **Stake, R.E. (1998):** Some comments on assessment in U.S. education. ***Education Policy Analysis Archives***, 6(14), 1-12.
93. **Statman, D. (1993):** "Self-assessment, self-esteem and self acceptance". ***Journal of Moral Education***, 22, (55-62).
94. **Tittle, C.K. and Pape, S. (1995):** "Selected references for assessment and evaluation of mathematics and science teaching reforms in classrooms a bibliography". ***Eric, No.*** (ED 395797).
95. **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).
96. **Wolf, K. and Others (1997):** "Portfolio in teacher evaluation". ***Paper Presented at Annual Meeting of The American Educational Research Association***, (Chicago, 24-28 March).
97. **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.html>.