

الفصل الثالث

الدراسات السابقة

المقدمة

أولاً: دراسات تناولت بناء وتقدير صدق وثبات الاختبارات محكية المرجع
ثانياً : دراسات تناولت تحليل مقدرات الاختبارات محكية المرجع .

الفصل الثالث

الدراسات السابقة

مقدمة :

نظراً لأهمية القياس محكي المرجع في الميدان العملي وما يقدمه من فوائد في تصنيف الأفراد إلى مستويات الإتقان وتشخيص مواطن القوة والضعف في التعليم ، اهتمت الدراسات السابقة ببناء الاختبارات محكية المرجع وطرق تقدير صدقها وثباتها وتحليل مفرداتها .

وبما أن أداة الدراسة الحالية هي اختبار محكي المرجع من إعداد الباحث ، وتحديد أفضل معامل لحساسية التعليم هو جزء من تحليل مفردات الاختبار محكي المرجع ، من هذا المنطلق جاء تركيز الباحث في مجال عرضه للدراسات السابقة ، التي أمكنه الحصول عليها، على النقاط المتعلقة بمجال بحثه ، وقسم الباحث الدراسات السابقة إلى الأقسام التالية :

أولاً: دراسات تناولت بناء وتقدير صدق وثبات الاختبارات محكية المرجع .

ثانياً : دراسات تناولت تحليل مفردات الاختبارات محكية المرجع .

أولاً : دراسات تناولت بناء وتقدير صدق وثبات الاختبارات محكية

المرجع .

دراسة **جانوبول (1978) Ganopole** استهدفت بناء وتقدير صدق اختبار محكي المرجع في الكفايات الأساسية للقراءة .

حددت الدراسة المهارات الأساسية للقراءة عن طريق مجموعة من الخبراء في المؤسسات التعليمية المتخصصة ، ثم أجرى تحليلاً سلوكياً لكل كفاية إلى مجموعة من الأهداف السلوكية ولبناء مفردات الاختبار تم إعداد جدول مواصفات الاختبار والذي اعتبر خطوة أساسية في هذا النوع من الاختبارات ، وبعد ذلك تم بناء مفردتين اختباريتين لقياس كل هدف في ضوء مواصفات الاختبار .

ثم التحقق من الصدق الوصفي **Descriptive Validity** لمفردات الاختبار ، والتحقق من الثبات تم بطريقتين .

الأولى اعتمدت على معامل الارتباط المستخدم في المقاييس مرجعية المعيار، وأما الطريقة الثانية فقد اعتمدت على معامل اتساق تكرار التضعيف إلى متقنين وغير متقنين ، وأشارت الدراسة إلى أن الطرق غير الارتباطية هي الأفضل لتقدير ثبات الاختبار محكي المرجع .

دراسة براتر (1978) Prater : هدفت هذه الدراسة إلى بناء وتقدير صدق اختبار محكي المرجع في المهارات الأساسية لكتابة Basic Writing Skills ، وقد مر بناء الاختبار بالمراحل التالية :

١- تحديد الكفايات الأساسية للكتابة Basic Writing Competencies حيث تم كتابة مجموعة من الكفايات وعرضت على مجموعة من المحكمين اشتملت على المتخصصين وأولياء الأمور والطلاب والمعلمين والهيئات المتخصصة بإعداد برامج تعليم الكتابة فتم التوصل إلى قائمة من الكفايات التالية :

١- فنيات الكتابة Mechanics ٢- الاستعمال Usag

٣- تنظيم الفقرات Paragraph Arganization ٤- التهجى Spelling

٥- تصحيح الجمل Sentence Correction ٦- الأشكال والأساليب Forms

٧- التعبير المكتوب Writing Sample

٢- إعداد مواصفات الاختبار: استخدم طريقة بابام Popham لإعداد مواصفات الاختبار وقد تضمنت الأمور التالية :

وصف عام للهدف السلوكي ، عينة المفردة ، خصائص المثير ، خصائص الاستجابة

٣- بناء المفردات : في ضوء المواصفات تم بناء صورتين أ وب للاختبار حيث اشتملت كل واحدة منها على عشر مفردات من نوع الاختيار من متعدد وذلك لكل كفاية من الكفايات الستة الأولى ، وأما مفردات الكفاية الأخيرة (التعبير المكتوب) فكانت من نوع المقال Essay .

٤- تقدير صدق وثبات الاختبار : قام الباحث بعرض قائمة المفردات الاختبارية بصورتها أ ، ب مع قائمة الأهداف السلوكية على أربعة من خبراء القياس وطلب منهم مضاهاة المفردات لكل من الصورتين مع الأهداف السلوكية ، وحسب معامل تجانس المفردات باستخدام الصيغة التي اقترحها روفينلي وهامبلتون .

أما بالنسبة للثبات فقد تم تطبيق الاختبار على عينة عشوائية اشتملت على ٣٠٠ طالب

من صفوف السابع والتاسع والثاني عشر ، وتم إعادة تطبيق الاختبار على نفس المجموعة بعد أسبوعين من التطبيق الأول ، وأوجد قيم معامل التكافؤ ومعامل الاستقرار ومعامل التكافؤ والاستقرار معاً .

وتم تصنيف الطلاب إلى متقنين وغير متقنين باستخدام مستويات تمكن مختلفة ٧٠% و ٨٠% و ٩٠% وحسب اتساق تصنيف الطلبة فوجد أن اتساق القرارات يتراوح بين ٧٨% و ١٠٠% وذلك بحسب مستوى التمكن .

أجرى والكر (1978) Walker دراسة هدفت إلى بناء وصدق اختبار محكي المرجع لمجموعة معارف مختارة من عناصر الأداء التدريسي لمعلمي التدريب المهني ، حيث اتبع الباحث في بناء الاختبار الخطوات التالية :

١- حدد ٣٠ نموذجاً مستخدماً في أحد مراكز التدريب المهني واختار ٨٣ هدفاً من بين ٣٨٤ هدفاً من الأهداف والمهارات المرجوة ، ثم صاغ هذه الأهداف صياغة سلوكية وشكل اختبارين متوازيين كل منهما يتكون من ٨٣ مفردة أي مفردة لكل هدف .

٢- ثم عرض الصورتين على المحكمين للتحقق من تكافؤ الصورتين ثم عرضهما على مجموعة أخرى من المحكمين لمطابقة المفردات مع الأهداف .

٣- أجرى دراسة تجريبية على ١٤ شخصاً ممن لديهم خبرة بمحتوى النماذج المحددة واستخدم الأساليب الإحصائية لتحليل المفردات لإيجاد قيمة معاملات الصعوبة والتمييز لكل مفردة ، وتعديل هذه المفردات في ضوء نتائج هذا التحليل .

٤- طبق الاختبار النهائي على عينة من ٣٠ معلماً ممن لديهم خبرة وتلقوا تدريباً وأتقنوا النماذج المحددة ، و ٣٠ معلماً لم يتلقوا هذا التدريب ، واختار درجة قطع ٨٠% وتراوحت قيمة الثبات بين ٠,٨٠ و ٠,٩٠ ولم تحدد الطرق التي تم حساب الثبات بها وأشارت الدراسة إلى أن درجة القطع ٨٠% مفضلة على باقي الدرجات دون ذكر السبب .

دراسة سمث (1980) Smith : هدفت هذه الدراسة إلى بناء اختبار محكي المرجع لكتاب القراءة وتمييز مفردات اللغة لطلاب الجامعة الجدد، واتبع في بناء الاختبار الخطوات التالية :

١- كتابة أهداف التحصيل الدراسي (لكتاب القراءة المقرر)

٢- تحليل هذه الأهداف إلى أربعة مجالات في تمييز مفردات اللغة وهي : الأحرف ،

المقاطع اللفظية ، الهجاء ، استخدام أجزاء الكلمة .و مجالين في المهارات الخاصة هما : استخدام أجزاء من الكتاب ، فهم مقاطع مختارة من الكتاب .

٣- تحديد ووصف كل اختبار فرعي .

٤- كتابة قائمة مفردات Pool of Items لكل اختبار فرعي .

٥- عمل دراسة تجريبية Pilot Study لتحديد صدق محتوى الاختبار وتحسين الاختبارات بناء على نتائج هذه الدراسة .

٦- تحديد ٨٠ ٪ كمستوى للإتقان في الاختبارات الفرعية بفترة ثقة ١٠ ٪ وبالتالي أصبح معيار المحك المقبول للأداء ٧٠ ٪ .

أما عينة الدراسة كانت ٤٠ طالباً في السنة الأولى بجامعة ولاية نيومكسيكو New Mexico State تم تقسيمها عشوائياً إلى مجموعتين تضم كل واحدة منها ٢٠ طالباً ، وجرى تسجيل درجات كل اختبار فرعي وفحص بشكل مستقل عن غيره من الاختبارات الأخرى ، ثم جرى مقارنة بين أداء المجموعتين على كل اختبار فرعي ، وباستخدام تحليل التباين وأسلوب التحليل الوصفي تم تحليل مفردات كل اختبار فرعي من حيث مستوى الدقة ومدى تمثيلها للمحتوى والكفاءة والفاعلية .

والمفردات التي ميزت بين مجموعتي المتقنين وغير المتقنين عند مستوى ٠,٠٥ كانت مقبولة تلقائياً واعتبرت دقيقة ومحددة المضمون وذات كفاءة وفاعلية ، ومن أهم التوصيات التي خلصت إليها هذه الدراسة : أن تحليل المفردات يجب أن يتم من خلال التباين والتحليل الوصفي ، وأن الصدق الوظيفي Functional Validity للاختبار يجب أن يحدد بعد استخدام الاختبار كأداة لما قبل وبعد التقييم .

دراسة عماد عبد المسيح يوسف (١٩٨٢) : هدفت لبناء وتقدير صدق وثبات مقياس محكي المرجع في مادة الطبيعة لطلاب الصف الأول الثانوي العام ، ومر بناء المقياس بالمراحل التالية :

١- تحديد محتوى المادة الدراسية المراد قياسها : حدد وحدة خواص المواد ثم حدد عناصر الموضوعات لمحتوى هذه الوحدة .

٢- تحديد الأهداف التعليمية لمحتوى الوحدة المراد قياسها : عين هدفين عامين ثم حللتهما إلى ١٥ هدفاً إجرائياً ثم عرضها على مجموعة من المحكمين ، وصنف هذه الأهداف حسب مستويات بلوم إلى تذكر ، فهم ، تطبيق .

٣- إعداد مواصفات الاختبار : اعتمد على طريقة بابام Popham في إعداد المواصفات

لكل هدف سلوكي وتكونت هذه المواصفات من خمسة عناصر هي : الوصف العام ، عينة المفردة ، خصائص المثير ، خصائص الاستجابة ، ملحق المواصفات .

٤- بناء مفردات الاختبار : تم صياغة مفردات الاختبار من نوع الاختيار من متعدد وبلغ عدد المفردات ٧٠ مفردة ، وذلك بعد حذف ١٢ مفردة منها استرشاداً بأراء المحكمين .

٥- تحديد مستوى الأداء : عن طريق الاستعانة بمجموعة من الخبراء المحكمين حدد الدرجة ٨٠ ٪ كمستوى أداء مقبول .

٦- تقدير صدق محتوى الاختبار وثباته : عرض الباحث قائمة المفردات والأهداف السلوكية على مجموعة من المحكمين للتحقق من صدق المحتوى واستخدام معادلة روفينيلي وهامبلتون (١٩٧٧) لحساب التطابق بين المفردة والهدف ، ولحساب الثبات طبق الاختبار على ٨٢٢ طالباً وطالبة ، ثم على ٣١٤ طالباً وطالبة من نفس الطلاب السابقين وحسب الثبات فوجد أنه يتراوح بين ٠,١٩ ، ٠,٥٨ .

واستخدم طريقة كوكس وفارجاس (١٩٦٦) لتحليل المفردات فوجد أن قيمها للمفردات تتراوح بين ٠,٠٤ و ٠,٥٨ .

دراسة محمد الدسوقي (١٩٨٥) : هدفت هذه الدراسة لبناء اختبار محكي المرجع في الهندسة الفراغية لطلاب الصف الثاني الثانوي ، لقياس درجة إتقان الطلاب لمادة الهندسة الفراغية المقررة عليهم ، وقد مر الاختبار بالخطوات التالية :

١- تحليل مجال ومحتوى المادة الدراسية المراد قياسها : حيث حدد الباحث عناصر الموضوعات المقررة وعرضها على مجموعة محكمين .

٢- تحديد الأهداف التعليمية لمجال المحتوى التعليمي المراد قياسه : فبلغ عدد الأهداف الإجرائية ٨٢ هدفاً (عرضوا على مجموعة محكمين) تقيس ثلاثة مستويات معرفية هي التذكر ، الفهم ، التطبيق .

٣- إعداد مواصفات الاختبار اعتمد على الطريقة التي اقترحها بابام Popham حيث اشتملت على : وصف الهدف السلوكي ، عينة المفردة ، خصائص المثير ، خصائص الاستجابة ، ملحق المواصفات .

٤- إعداد مفردات الاختبار : تم بناء مفردتي اختبارين متكافئين لقياس كل هدف سلوكي، وبلغ عدد المفردات ٨٢ مفردة في كل صورة منها ٢٤ مفردة اختيار من متعدد ، و ٤٣ مفردة إجابة قصيرة ، و ١٥ مفردة مقال .

٥- تحديد مستوى الأداء : عن طريق الاستعانة بمجموعة من الخبراء المحكمين حدد الدرجة ٨٠ ٪ كمستوى أداء جيد .

٦- تقدير صدق وثبات الاختبار: حقق صدق المحتوى بعرض قائمتي المفردات والأهداف السلوكية على مجموعة من المحكمين وطلب من كل منهم تحديد مدى مطابقة المفردة بالهدف ثم استخدم المعادلة التي اقترحها روفينيلي وهامبلتون فوجد أنها واحد صحيح لكل مفردة .

أما الثبات فقد تم حسابه من خلال تطبيق الصورة الأولى للاختبار على طلاب الصف الثاني الثانوي بمحافظة الدقهلية والغربية بلغت ٣٠٧ طالب ، وبعد مرور خمسة أيام طبق الصورة الثانية وقام بحساب الثبات باستخدام معادلة كيودر - رتشاردسون الصيغة (٢٠) فوجد أن قيمته ٠,٩٥ ، واستعمل أيضاً معادلة سواميناثان وهامبلتون لحساب ثبات قرارات التصنيف فوجد أنها ٠,٧٨ وبذلك أصبح الاختبار جاهزاً.

قدم سليمان محمد سليمان (١٩٨٥) دراسة بعنوان أثر بعض طرق تقدير الدرجات على ثبات الاختبارات مرجعية المحك ، وهدفت هذه الدراسة للإجابة على التساؤلين التاليين :

- ١- هل يختلف ثبات درجات الاختبارات محكية المرجع ذات الاختيار من متعدد باختلاف كل من تقدير درجات المفردات وطرق حساب درجات المفردات؟ .
- ٢- هل يوجد تأثير مشترك لكل من طرق تقدير الدرجات وطرق حساب الثبات على ثبات درجات الاختبارات محكية المرجع؟ .

وللإجابة على هذين التساؤلين أعد الباحث اختبار مرجعي المحك في مادة الجبر للصف الثالث الإعدادي في صورتين تشتمل كل منهما على مفردات الاختيار من متعدد . ولقد حدد الباحث أربع طرق لتقدير الدرجات هي : الطريقة التقليدية ، الدرجات الذاتية، وزن الثقة ، توزيع الثقة .

وكما حدد أيضاً ثلاثة طرق لحساب الثبات هي : طريقة ليفنجستون ، وطريقة هامبلتون وألجينا وطريقة هاريس .

وتم تطبيق الاختبار على ٥٩٠ طالباً من طلاب الصف الثالث الإعدادي ، وأهم النتائج التي توصلت إليها الدراسة هي :

تختلف معاملات ثبات درجات الاختبارات محكية المرجع باختلاف الطريقة المستخدمة

في تقدير الدرجات ، وعند استخدام الطريقة التقليدية في تقدير الدرجات فإنه يفضل استخدام طريقة ليفنجستون عن بقية الطرق الأخرى .

وأشارت النتائج أيضاً إلى أن الطريقة التقليدية لتقدير الدرجات أفضل الطرق الأربعة حيث تعطي أعلى ثبات لدرجات الاختبار عند استخدام الطرق الثلاثة لحساب الثبات .

دراسة نادية محمد عبد السلام (١٩٨٦) : وهي بعنوان بناء اختبار هدفي المرجع وجماعي المرجع " دراسة مقارنة بين النموذجين "

وكان الهدف من هذا البحث هو بناء اختبار تشخيصي هدفي المرجع لقياس إتقان الطالب لمجموعة من المهارات الأساسية التي ترتبط بمادة الإحصاء الوصفي ، حيث تكون الاختبار من ٢٢ هدفاً لقياس المهارات الأساسية اللازمة لتحصيل مادة الإحصاء الوصفي .

و يضم هذا الاختبار صورتين ، حيث أن الصورة تتكون من ٥١ مفردة ، ومن ثم قامت الباحثة ببناء صورة موازية له وتتكون من ٢٢ هدفاً ، وعدد مفرداته ٥٣ مفردة وكانت جميع المفردات من نوع الاختيار من متعدد .

ولقد استرشدت الباحثة في بناء الاختبار بالخطوات التالية :

- ١- إعداد الأهداف .
- ٢- بناء مفردات الاختبار .
- ٣- تحليل المفردات .
- ٣- انتقاء المفردات / طول الاختبار / الدرجة الفاصلة .
- ٥- دراسة الثبات والصدق .

وقد طبق الاختبار على عينة استطلاعية بلغت ٢٢ طالباً وطالبة من طلاب الدبلوم الخاص و التمهيدي ماجستير في كلية البنات جامعة عين شمس .

دراسة صلاح الدين علام (١٩٨٧) وهي بعنوان : اختبار تشخيصي مرجعي المحك لقياس الكفايات الأساسية في الإحصاء التربوي والنفسي .

وتهدف هذه الدراسة إلى بناء اختبار تشخيصي مرجعي المحك لقياس تمكن طلاب علم النفس والتربية بالجامعات من الكفايات الأساسية للإحصاء التربوي والنفسي ، ويمتاز هذا الاختبار بأن له قيمة تشخيصية .

ونظراً لتعدد الكفايات فقد اشتمل الاختبار على عدد من الاختبارات الفرعية (سبعة اختبارات فرعية) لكل اختبار منها صورتين متوازيتين ، أي قام الباحث ببناء اختبارين متوازيين في كل منهما سبعة اختبارات فرعية ، وعدد مفردات كل من الاختبارات السبعة

مساوياً لعدد الأهداف المساعدة ، والعدد الكلي لمفردات الصورة الواحدة للاختبارات ١١٦ مفردة .

ولقد اتبع الباحث في بناء الاختبار الخطوات التالية :

١- مرحلة التحليل : وتتضمن التالي

✧ تحديد الكفايات الأساسية للإحصاء التربوي والنفسي .

✧ تحديد النطاق السلوكي للكفايات وتحليلها وتحديد الأهداف .

✧ صياغة الأهداف السلوكية وترتيبها في نسق هرمي والتحقق من صدق النطاق

السلوكي الذي يحتوي هذه الأهداف .

٢- مرحلة البناء : كتابة مفردات الاختبار .

٣- مرحلة التجريب : ١- التحقق من صدق محتوى الاختبار ٢- تقدير ثبات الاختبار .

٣- تجميع مفردات الاختبار .

ثم قام الباحث بتطبيق إحدى الصورتين على ٣٤ طالباً من طلاب الدراسات العليا بكلية التربية بجامعة الأزهر الذين سبق لهم دراسة الإحصاء التربوي والنفسي وانتهوا منه ، حيث أفترض أنهم اكتسبوا المعارف والمهارات المرتبطة بالكفايات الأساسية المحددة ، ثم طبق الصورة الأخرى بعد عدة أيام وحسب الثبات لكل اختبار ، حيث تراوحت قيمته بين ٠,٦٧ و ٠,٨٧ وبذلك أصبح الاختبار جاهزاً .

دراسة آمال أحمد محروس (١٩٨٨) وهي بعنوان : مقارنة عدة طرق لحساب

ثبات الاختبارات المرجعية المحك .

تمثلت مشكلة الدراسة في تحديد أفضل طرق حساب ثبات الاختبارات مرجعية المحك عند إجراء مقارنات بين أكثر من طريقتين ، وحددت الدراسة أربع طرق لحساب الثبات وهي :

لفنجستون (١٩٧٢) Livingston وهارس (١٩٧٢) Harris وهامبلتون ونوفيك (

١٩٧٣) Hambleton & Novick وصيكوفياك (١٩٧٦) Subkoviak .

وقامت الباحثة بإعداد ثلاث اختبارات تحصيلية مرجعية المحك في المواد التالية : علم

النفس ، التقويم والقياس ، علم النفس العام ، علم النفس التعليمي .

ولقد اتبعت الباحثة في بناء هذه الاختبارات الخطوات التالية :

١. تحديد المجال السلوكي الذي يقبسه الاختبار ويشتمل على :

- تحديد نطاق محتوى المادة الدراسية .
 - تحليل المحتوى الدراسي .
 - تحديد أهداف مجال المحتوى الدراسي وصياغتها إجرائياً .
٢. بناء مفردات الاختبار ويتضمن التالي :
- إعداد مواصفات الاختبار .
 - إعداد مفردات الاختبار .
 - تحديد مستوى الأداء (الدرجة الفاصلة) .
٣. حساب ثبات الاختبارات .

وعينة الدراسة اختوت على طالبات الصف الثاني بالكلية المتوسطة للبنات في المملكة العربية السعودية ، حيث كانت ٦٦ طالبة لاختبار مادة القياس النفسي و١٥٢ طالبة لاختبار مادة علم النفس العام و ١١ طالبة لاختبار علم النفس التعليمي .

وتوصلت الدراسة إلى ما هو آتي :

- ١- إذا تمت المقارنة بين الطرق الثلاثة التي تعتمد على صورة واحدة للاختبار في إيجاد معامل الثبات ، وهي (لفنجستون - هارس - صبكوفياك) نجد أن طريقة لفنجستون تعطي ثباتاً أعلى من ثبات الطريقتين المذكورتين .
- ٢- إذا تمت المقارنة بين طريقة هامبلتون التي تعتمد على صورتين للاختبار المتكافئتين والطرق الثلاثة الباقية ، نجد تفوق طريقة هامبلتون .

دراسة محمود محمد إبراهيم (١٩٩٠) وهي بعنوان : دراسة سيكومترية مقارنة - لطرق حساب معامل ثبات الاختبارات المرجعة إلى المحك .

سعت هذه الدراسة من أجل :

- التوصل إلى أفضل الطرق الإحصائية التي يمكن استخدامها في حساب معاملات الثبات للاختبارات المرجعة إلى المحك ، في ضوء طول الاختبار ودرجة القطع وحجم العينة .
- وضع خطوات واضحة لاستخدامها في بناء الاختبارات المرجعة إلى المحك .
- التوصل إلى توضيحات نظرية لكل من درجة القطع ، طول الاختبار ، صدق وثبات الاختبارات المرجعة إلى المحك .

وضمت عينة الدراسة ٨١٨ طالب وطالبة من طلاب الصف السابع بمرحلة التعليم الأساسي بمحافظة القاهرة ، أما أداة الدراسة فهي اختبار مرجع إلى المحك في باب المجموعات من كتاب الجبر ، وقد اعتمد في بنائه على المراحل التالية :

- ١- تحديد النطاق السلوكي .
 ٢- تحديد الأهداف التعليمية .
 ٣- إعداد مواصفات الاختبار .
 ٤- بناء مفردات الاختبار .

وحدد الباحث ثلاث طرق لحساب الثبات هي: طريقة ليفنجستون Livingsto Method .
 وطريقة هاينا Huynh Method . وطريقة صبكوفياك Method Subkoviak .

ومن أهم النتائج التي توصلت إليها هذه الدراسة أنه : عندما يكون الاختبار طويلاً وأحجام العينة متوسطة وصغيرة يفضل استخدام طريقة هاينا Huynh ، وأما إذا كان طول الاختبار متوسطاً فتستخدم طريقة ليفنجستون Livingston مع أحجام صغيرة ومتوسطة للعينة ، ومع العينات الكبيرة والطول المتوسط للاختبار تستخدم طريقة هاينا ، أما في حال الاختبارات القصيرة فإننا نستخدم طريقة صبكوفياك Subkoviak مع العينة الصغيرة (٣٠ طالباً) ومع العينة الكبيرة (٨١٨ طالباً) ، أما إذا كانت العينة متوسطة فيمكن استخدام طريقة هاينا .

وبالنسبة لدرجة القطع من ٣٠٪ إلى ٥٠٪ فيمكن استخدام طريقة ليفنجستون لجميع أحجام العينات ولجميع الأطوال ، وعند درجة القطع ٦٠٪ يفضل استخدام طريقة هاينا ، ومن درجة القطع ٧٠٪ إلى ٩٠٪ فيمكن استخدام طريقة ليفنجستون لأي حجم ولأي طول للاختبار .

دراسة صلاح الدين علام (١٩٩١) وهي بعنوان : اختبار تشخيصي مرجعي المحك في أساسيات القياس والتقويم التربوي والنفسي .

وتهدف هذه الدراسة لقياس تمكن طلاب كليات التربية والمعلمين الذين يتلقون تدريباً في مجال القياس والتقويم أثناء الخدمة ، من أساسيات القياس والتقويم التربوي والنفسي .
 حيث احتوى الاختبار على ٦٠ مفردة من نوع الاختيار من متعدد تقيس ٥٧ هدفاً ،
 واتباع الباحث في بناء الاختبار الخطوات التالية:

١. تحديد محتوى الاختبار والأهداف السلوكية المتعلقة به.
٢. بناء مفردات الاختبار والتحقق من صدقها .
٣. ثبات الاختبار .
٤. تحديد الحد الأدنى لمستوى الإنجاز في الاختبار.

وللتحقق من ثبات الاختبار طبق على ٤٣ طالباً من طلاب السنة الرابعة بكلية التربية جامعة الأزهر الذين انتهوا من دراسة المقرر وكذلك ٣٦ طالباً من طلاب الدبلوم العامة الذين

درسوا المقرر نفسه ، وكانت قيمة معامل الثبات ٠,٧٤ ، وبذلك اعتبر الاختبار صالحاً .

دراسة جرافيك (Griffec (1995) : وهي بعنوان إجراء اختبار صفي لمعلمين يكرهون إجراء الاختبار (بناء وتقييم اختبار محكي المرجع)
تقدم هذه الدراسة مقارنة للاختبارات محكية المرجع مع الاختبارات معيارية المرجع ، وكيفية مراجعة وتقويم الاختبارات محكية المرجع .

حيث كانت عينة الدراسة عبارة عن صف دراسي حقيقي لطلاب يدرسون اللغة الإنكليزية كلغة غريبة (في اليابان) ، وذلك باستخدام اختبار من إعداد معلم له خبرة لسنوات طويلة في تعليم اللغة الإنكليزية كلغة ثانية .

ولقد أشارت الدراسة إلى أن الاختبارات معيارية المرجع هيمنت على منهجية وضع الاختبار حتى منتصف السبعينيات ، حيث كانت الاختبارات محكية المرجع أقل انتشاراً .

وأشارت نتائج هذه الدراسة إلى ما يلي :

• الاختبارات معيارية المرجع تقيس كفاءة (تقدم) عامة في اللغة ، في حين أن الاختبارات محكية المرجع تقيس أهداف معينة (محددة) .

• وكذلك الاختبارات معيارية المرجع ذات فائدة ضئيلة للمعلمين في تشخيص نقاط القوة والضعف في تحصيل الطلاب أو تقييم البرامج ، أما الاختبارات محكية المرجع التي يمكن تصميمها وتقييمها باستخدام تحليل المفردة تقدم فائدة أكثر في هذه المجالات .

دراسة واي (Wie (1995) وهي بعنوان : بناء اختبار لحنى لطلاب الصف الرابع الملغنين بمنهجية خاصة في تعليم الموسيقى .

هدفت هذه الدراسة إلى بناء اختبار تحصيل لحنى لتقييم فاعلية تعلم طلاب الصف الرابع الذين يدرسون مدخل توليدي في تعليم الموسيقى .

تم تصميم الاختبار لإعطاء صورة كاملة عن إنجاز الطلاب بواسطة :

١- استخدام أمثلة موسيقية مستقاة من ذخيرة لحنية خاصة بالطلاب .

٢- استخدام أمثلة موسيقية شاملة أكثر من أن تكون مبتورة .

٣- استخدام أنشطة موسيقية نموذجية خاصة بالأنشطة الموسيقية اليومية للطلاب في

الصف الدراسي .

وقد تكون الاختبار من فرعين : الأول من اختبار وصفي والثاني يشمل على اختبار أداء واختبار إيداعي ، وتم حساب صدق المحتوى الذي يضم أهداف عملية الاختبار والأمثلة الموسيقية ومهام الاختبار من خلال خبراء مختصين .

طبق الاختبار الأساسي على ٨٠ طالباً في الصف الثالث والرابع والخامس ، وتمت مراجعة بنود مجموعة (بطارية) الاختبارات للتركيز على أهداف الصف الرابع حيث تم حذف المفردات التي تميزها أقل من ٠,٢٥ .

وطبق الاختبار النهائي على ٨٦ طالباً من الصف الرابع و٦١ طالباً من الصف الخامس حيث كانت الارتباطات المتبادلة بين بعض الاختبارات الفرعية متوسطة (٠,٨٨ - ٠,٦٥) ، وأشارت معاملات ثبات التجزئة النصفية التي تراوحت بين ٠,٨١ إلى ٠,٩٣ بأن الاختبار يتمتع بدرجة مقبولة من الاتساق الداخلي ، وقد وجد أن الارتباطات بين علامات الطلاب وتصنيف المعلمين للصف الرابع كانت $r = ٠,٨٨$ في حين كان هذا الارتباط لطلاب الصف الخامس $r = ٠,٧٥$ وهذا الفرق لصالح طلاب الصف الرابع ، دليل على أن هذا الاختبار المحكي المرجع يتوافق مع أهداف التعليم الخاصة بالصف الرابع .

وقدم محمد محمد فتح الله (١٩٩٥) دراسة بعنوان : بناء اختبار محكي المرجع في العلوم لتلاميذ نهاية مرحلة التعليم الأساسي .

وهدفت هذه الدراسة لبناء اختبار محكي المرجع في مادة العلوم الأساسية لتلاميذ نهاية التعليم الأساسي (الصف الثالث الإعدادي) .

حيث احتوت عينة الدراسة على ٢٠٠ طالب من طلاب الصف الثاني الإعدادي والصف الثالث الإعدادي الذين تم تقسيمهم إلى مجموعتين ، المجموعة الأولى وهم طلاب الصف الثالث الإعدادي الذين تلقوا التعليم لهذه المادة ، والمجموعة الثانية هي مجموعة طلاب الصف الثاني الإعدادي الذين لم يتلقوا التعليم لهذه المادة .

وقد قام الباحث ببناء اختبار محكي المرجع لمقرر العلوم يتكون من ٦٠ مفردة من نوع الاختيار من متعدد وهو مؤلف من صورتين .

واتبع الباحث في بناء الاختبار الخطوات التالية :

١- تحديد النطاق السلوكي للمحتوى التعليمي. ٢- تحليل النطاق السلوكي إلى مكوناته السلوكية.

٣- إعداد مواصفات الاختبار . ٤- تحديد الخصائص السيكومترية للاختبار .

وبناء الاختبار تكون الدراسة حققت هدفها .

دراسة إسماعيل حسن فهيم (١٩٩٦) وهي بعنوان : أثر استخدام بعض طرق التغذية المرتدة على ثبات درجات اختبار هدفي المرجع .

تهدف هذه الدراسة على وجه التحديد إلى التعرف على أثر ثلاث طرق للتغذية المرتدة الاختيارية على ثبات درجات اختبار هدفي المرجع ، وهذه الطرق هي :

١- تغذية مرتدة كلية . ٢- تغذية مرتدة جزئية . ٣- تغذية مرتدة مؤجلة .

وأما أداة البحث فهي اختبار تحصيلي من إعداد الباحث ، يتكون من صورتين متكافئتين تشمل كل منهما على ٣٢ مفردة من نوع اختيار من متعدد في وحدة الدوال المتثلثة للصف الأول الثانوي ، و اتبع في بناء هذا الاختبار الخطوات التالية :

- ١- تحديد محتوى الموضوعات المتضمنة في الوحدة الدراسية .
- ٢- تحديد المهارات الرئيسية المرجوة ومحتواها .
- ٣- تحليل المهارات الرئيسية إلى مكوناتها .
- ٤- صياغة الأهداف السلوكية في ضوء نواتج تحليل المهارات الرئيسية .
- ٥- تقييم الأهداف السلوكية .
- ٦- إعداد مواصفات الاختبار
- ٧- بناء مفردات الاختبار .
- ٨- تقييم مفردات الاختبار
- ٩- التحقق من صدق محتوى الاختبار وتقدير ثبات درجاته .

أما عينة الدراسة فقد ضمت ٢٨٥ طالباً من الصف الأول الثانوي في مدرسة بلبس الثانوية بمحافظة الشرقية ، قسمت إلى أربع مجموعات .

وكان من أهم النتائج التي توصلت إليها هذه الدراسة : أن التغذية المرتدة الكلية تؤثر في ثبات درجات الاختبار أكثر من التغذية المرتدة الجزئية ، في حين أنه لا توجد فروق بين التغذية المرتدة الكلية والتغذية المرتدة المؤجلة على ثبات درجات الاختبار .

دراسة اعتدال غازي عبد الله (٢٠٠٠) وهي بعنوان : القياس محكي المرجع لأحد مقررات علم النفس التعليمي " دراسة ميدانية "

والهدف من هذه الدراسة بناء صورة اختباريه محكية المرجع متعادلة القياس في أحد مقررات علم النفس التعليمي ، تقيس مدى تحقق الأهداف المعرفية بهذا المقرر وتسعى لتحديد موقع الأداء للأفراد على محكات الإثقان المحددة وتحديد النسب المئوية لتحقيق الأهداف عند كل مستوى معرفي (تذكر ، فهم ، تطبيق) .

وعينة الدراسة تحتوي ١٩٤ طالبة من طالبات الفرقة الرابعة في الشعب التربوية للأقسام العلمية بكلية البنات جامعة عين شمس ، وأما أداة الدراسة فقد تمثلت في الصور الأولية لبنك الأسئلة والتي تقيس الأهداف الإجرائية لمادة علم النفس التعليمي المقررة .

وقد اتبعت الباحثة في بناء الصور الأولية لبنك الأسئلة الخطوات التالية :

١- تحديد المجال السلوكي الذي يقيسه بنك الأسئلة : ويتضمن

• تحديد محتوى المقرر الدراسي المراد قياسه وتحليله إلى عناصره الأساسية .

٢- تحديد الأهداف السلوكية لعناصر المقرر الدراسي .

• بناء مفردات بنك الأسئلة .

• إعداد جدول للمواصفات .

• صياغة المفردات التي تقيس الأهداف السلوكية .

• صياغة التعليمات .

٣- الصدق الوصفي لأداة القياس (صدق المحتوى) .

واستخدمت الباحثة نموذج رايش (نموذج أحادي البارمتر) في تدريج هذه الأداة ،

وتصنيف بلوم للأهداف المعرفية المستخدمة ، ومن أهم النتائج التي توصلت إليها :

تشخيص نقاط القوة والضعف في تحقيق أهداف القياس ، حيث حددت الأهداف التي

تحققت بدرجة عالية ومتوسطة ومنخفضة ، وكانت النسب المئوية لاستجابات الأفراد عند

المستويات المعرفية للأهداف متقاربة إلى حد ما .

وتبين أن أعلى نسبة لتحقيق الأهداف هي عند مستوى التطبيق ، وأرجعت السبب إلى أن

إعطاء المقرر يعتمد على الحوار والمناقشة وإجراء التطبيقات اللازمة .

ثانياً : دراسات تناولت تحليل مفردات الاختبارات محكية المرجع

دراسة هالاديننا (1976) Haladyna وهي بعنوان : النوعية لمفردات الاختبار

محك المرجع .

كانت أهداف هذه الدراسة :

أولاً : تحديد إذا ما كان تحليل المفردة التجريبي Empirical Item Analysis مناسباً

للاختبار مرجعي المجال أم لا ؟.

ثانياً : في حال حدوث ذلك فما هي الإجراءات المطلوبة والتي تكون أكثر فاعلية في

تحديد نوع المفردة ؟.

ولقد اختارت الدراسة أربع طرق لاختيار المفردة وهي :

١- إحصائية راش Rasch Statistics .

٢- معاملات الحساسية للتعليم Instructional Sensitivity Indexes .

٣- الإحصاءات التقليدية Traditional Statistics . ٤- معاملات بايزن Bayesian Indexes

ودلت نتائج الدراسة إلى أن إجراء التحليل التجريبي مرغوب به كثيراً ومطلوب ، وعندما طبقت إجراءات التحليل التجريبي على بيانات الاختبار فإن النتائج أشارت إلى أن التقنيات الأربعة السابقة تعطي معلومات متماثلة تقريباً ، ويظهر أساس هذه النتائج أن أية واحدة من هذه التقنيات الأربعة تكون كافية عند إجراء تحليل المفردة .

دراسة ثراش (1977) Thrash وهي بعنوان : نموذج للقياس محكي المرجع ومقارنة أساليب تحليل المفردة .

ولقد هدفت هذه الدراسة إلى :

١- تقديم إطار نظري للقياس محكي المرجع ، وشرح طريقتين لتحليل المفردات هما معامل كوكس وفارجاس ، ومعامل رودابوش .

٢- مقارنة ثلاث طرق لتحليل مفردات الاختبار مرجعي المحك باستخدام بيانات حقيقية ، وهذه الطرق هي :

-معامل كوكس وفارجاس (C-V) .Cox&Vargas -معامل رودابوش (R) Roudabush
-طريقة برينان وستوليرو (B-S) Brennan & Stolurow .

وحيث تم استخدام اختبار محكي المرجع في مجالين مختلفين هما الرياضيات والقراءة لطلاب الصف المتوسط والأعلى ، وبلغ عدد المفردات التي حسب بها معامل (C-V) و (R) ١٢٨ مفردة و ٦٤ مفردة لحساب معامل (B-S) .

ونتائج هذه الدراسة أظهرت أن طريقة كوكس وفارجاس هي أفضل الطرق الثلاثة لتحليل مفردات القياس مرجعي المحك، لأنها تعطي تقديراً معتدلاً بغض النظر عن الافتراضات، وهذه التقديرات تكون أكثر ثباتاً عند استخدام أحجام كبيرة من البيانات ، وأن طريقة (R) تعطي تقديرات أكثر دقة وثباتاً عندما تكون الافتراضات متوائمة من أن تكون غير متوائمة.

دراسة سميث (1978) Smith وهي بعنوان : التأثير لطرق اختيار المفردة المتعددة على دقة التصنيف واتساق التصنيف في الأدوات محكية المرجع .

والهدف من هذه الدراسة الإجابة على الأسئلة التالية :

- كيف نقارن طرق اختيار المفردة في الاختبار محكي المرجع؟ وكيف نقارن مدى فاعلية هذه الطرق في رفض المفردات الضعيفة الفاعلية؟.
- كيف تختلف الاختبارات المكونة من مفردات هي الأفضل وفقاً لطريقة اختيار المفردة فيما يتعلق بتصنيف الأفراد وفق حالات الإلتقان الحقيقي؟.
- إلى أي مدى تتسق الاختبارات المكونة من مفردات هي الأفضل في تصنيف الأفراد؟
وحددت الدراسة الطرق التالية لاختبار المفردة:

- * معامل الارتباط الثنائي التقليدي RPB .
- * معامل كوكس - فارجاس C-V .
- * معامل برينان Br .
- * معامل فاي Phi .
- * معامل الاختيار العشوائي RDM .

ومن ثم أخذ استجابة ١٠٠ فرد على ٧٠ مفردة ، وولدت منها ثلاث مجموعات مختلفة الاستجابات في فاعليتها للتعليم ، ثم رتب المفردات من الأفضل إلى الأسوأ وذلك حسب القيم الناتجة عن تطبيق الطرق السابقة عليها ، وذلك لمعرفة المفردات الضعيفة المرفوضة وفق كل طريقة من هذه الطرق في ظل مستويات مختلفة من الفاعلية للتعليم ، أما تقييم دقة التصنيف واتساقه تمت من خلال مقارنة الإلتقان الحقيقي للفرد مع الإلتقان الملاحظ على الاختبار .

وتوصلت الدراسة إلى أن طريقة الاختيار العشوائي RDM قدمت تصنيفات دقيقة إلا أنها غير متسقة ، والأداء الضعيف يرجع إلى طريقة برينان ومعامل فاي Phi وذلك بسبب عدم اختيار هذه الطرق لمفردات ذات تمييز متوسط وعال ، في حين أن طريقتي كوكس - فارجاس والارتباط الثنائي كانت الأفضل من باقي الطرق الأخرى لأنها حققتا تصنيف دقيق ومتسق ، وعلى وجه التحديد كانت طريقة الارتباط الثنائي هي الأفضل إذا كان هناك تباين في درجات الاختبار وطريقة كوكس - فارجاس هي الأفضل إذا لم يكن هناك تباين في درجات الاختبار .

دراسة أينسلي (1984) Ainsley وهي بعنوان : دراسة مقارنة للعلاقة بين معاملات التمييز المتنوعة التي ينصح المدرسون باستخدامها في بناء الاختبارات محكية المرجع.

ولقد أشارت الدراسة إلى وجوب إجراء نوعين من التحليل عند بناء مفردات الاختبار محكي المرجع :

النوع الأول : وهو التحليل المنطقي حيث يستخدم لتحديد صدق المحتوى في الاختبار محكي المرجع .

النوع الثاني : وهو التحليل التجريبي (الإمبيرقي) ويركز على معرفة المفردات التي تميز بين المتقنين وغير المتقنين في الاختبار ، حيث يربط معامل تمييز المفردة في الاختبار مرجعي المحك درجة المفردة بقرار الإتيان وعدم الإتيان .

واختارت الدراسة ثلاث طرق لتمييز مفردات اختبارات معيارية المرجع ، وينصح أيضاً باستخدامها في الاختبارات محكية المرجع ، وهذه الطرق هي :

١. الارتباط الثنائي النقطي : (Crehan 1974) Point-Biserial Correlation
(r(pbi)) .

٢. الارتباط الثنائي للعينات الموحدة : Combined - Samples Point - Biserial
Correlation (COMPBI) (Haladyna 1974)

٣. معامل فاي (Swezey & Pearlstein 1975) Phi Coefficient

وهناك أربع طرق تستخدم في التحليل الإمبيرقي للاختبارات محكية المرجع وهي :

١- معامل الاختلاف القبلي والبعدي لكوكس وفارجاس Pretest - Posttest
Difference (Dis(PPD))

٢- معامل مجموعة الدارسين وغير الدارسين لكوسكوف وكلاين ١٩٧٦
Uninstructed - Instructed Groups (Dis(UGD))

٣- معامل التحصيل الفردي لرودايوش (Individual Gain (Dis(IG))

٤- معامل التحصيل الصافي لكوسكوف وكلاين ١٩٧٤ (Net Gain (Dis(NG))

وبحثت هذه الدراسة في العلاقات بين الطرق الأربعة السابقة لمعامل الحساسية في الاختبار محكي المرجع ، والطرق الثلاث السابقة لمعامل تمييز الاختبارات معيارية المرجع وكيف تعمل هذه الطرق في اختيار مفردات الاختبار .

أما عينة الدراسة فتكونت من ٦٦ طالباً في الصف السابع حيث تراوحت أعمارهم بين ١١ و ١٥ سنة ، طبق عليهم اختبار محكي المرجع مكون من ٢٥ مفردة اختيار من متعدد في مادة العلوم ، وتم تقسيم عينة الدراسة بشكل عشوائي إلى ثلاث مجموعات :

١- مجموعة القياس القبلي والبعدي Postinstruction - Preinstruction Group

٢- مجموعة غير المتعلمين Uninstructed Group

٣- مجموعة المتعلمين Instructed Group

و طبق اختبار على المجموعتين الأولى والثالثة وبعد المعالجة تم تطبيق اختبار ثان .
وأشارت نتائج الدراسة من خلال الاختبارين إلى أن علاقات الارتباط الداخلي لكل من
معامل التمييز القبلي والبعدي (Dis(PPD) ، والتحصيل الفردي (Dis(IG) ، والتحصيل
الصافي (Dis(NG) ، والارتباط الثنائي للعينات الموحدة COMPBI جميعها تميز المفردات
بشكل متماثل ، بينما طريقة مجموعة الدارسين وغير الدارسين (Dis((UiGD تختلف عن
الطرق السابقة في قياس التمييز للمفردة .

وكانت المعاملات في الاختبار محكي المرجع أكبر بشكل دال من معامل ، الارتباط
الثنائي $r(pbi)$ وذلك من أجل بيانات الاختبار الثاني .

وأشارت النتائج أنه في جميع الحالات عندما يكون التعليم فعالاً بشكل معتدل فالمعاملات
تختار مفردات أكثر مما لو كان التعليم فعالاً بشكل كبير ، وفي الاختبارات محكية المرجع
معاملات الارتباط الثنائي للعينات الموحدة تميل لاختيار مفردات أكثر نسبياً من معاملات
الارتباط الثنائي وهذا الميل يكون أكبر مع التعليم الفعال .

دراسة هانا وبنيت (Hanna & Bennett (1984 وهي بعنوان : توسيع معامل
الحساسية للتعليم .

أظهرت هذه الدراسة دور وفائدة قياس الحساسية للتعليم وأكدت عليها ، وقدمت عرضاً
للأسباب التي تجعل قياس الحساسية للتعليم يطبق على جميع الاختبارات التحصيلية ، وأن لا
تبقى حصراً في اختبارات التمكن محكية المرجع.

وقدمت هذه الدراسة أيضاً تفسيراً بالأمتثلة لنظرية اعتماد السياق Context
Dependence في محتوى الاختبار ، وأشارت الدراسة إلى أن منطق هذه النظرية يتكامل
مع الأساس المنطقي لمعامل الحساسية ، وأن العلاقات العملية المتبادلة بين صدق المحتوى
Content Validity والثبات Reliability ومعامل الحساسية للتعليم Instructional
Sensitivity ونوقشت ووضحت باستخدام عينات مختلفة من المفردات الاختيارية
والبيانات.

دراسة سلفا (Silva (1985 وهي بعنوان : مقارنة الطرق التقليدية وطرق
الاستجابة للمفردة لمشكلة اختيار المفردة في القياس محكي المرجع .

قامت هذه الدراسة بمقارنة الأساليب التقليدية لتحليل المفردة مع أساليب تحليل المفردة في نظرية الاستجابة للمفردة .

وتم استخدام ثلاث طرق تقليدية لتحليل المفردة هي : ١- معامل كوكس وفارجاس (Cox & Vargas V) . ٢- الارتباط الثنائي النقطي على البيانات القبلية والبعديّة المشتركة . ٣- معامل فاي Phi - Coefficient .

وطريقتين مشتقتين من نظرية الاستجابة للمفردة وهما : ١- نموذج أحادي البارامتر One Parameter Model . ٢- نموذج ثلاثي البارامتر Three Parameter Model . أما عينة الدراسة فقد تكونت من ٩٤٥ طالباً من طلاب الصف الأول والثاني و١٧٩٦ طالباً من طلاب الصف الرابع حتى السادس .

حيث استخدم اختبارات القراءة محكية المرجع ، وقدمت قائمة الإرشادات للقراءة من خلال أنظمة قراءة الأهداف التعليمية ، واستخدم معامل كايا لمعرفة تصنيف الاتساقات .

وأشارت نتائج هذه الدراسة إلى أن نماذج الاستجابة للمفردة تعطي تصنيفاً متسقاً أكثر من الطرق التقليدية الثلاثة ، وبالإضافة إلى ذلك فإن نموذج ثلاثي البارامتر يظهر تصنيفاً متسقاً أعلى من نموذج أحادي البارامتر .

دراسة هاتسون وآخرون (1986) Hanson وهي بعنوان : الاختلاف في معامل الحساسية للتعليم بين صيغ المفردات وبين مفردات الاختيار التحصيلي .

قدمت هذه الدراسة الإطار العملي والإجراءات الإمبريقية (التجريبية) لبناء الاختبارات التحصيلية الحساسية للمؤثرات التعليمية في البرامج المدرسية ، وتم استخدام مكونات نموذج التباين لتقويم الحساسية للتعليم Instructional Sensitivity .

وأشارت نتائج هذه الدراسة من خلال مقارنة البيانات المستمدة لبدائل صيغ المفردات للاختبار محكي المرجع إلى وجود اختلاف في الحساسية للتعليم بين نفس المفردات المقامة بصيغ Format مختلفة وبين مفردات مختلفة مقدمة ضمن الصيغة المحرزة .

دراسة هارس وصبكوفياك (1986) Harris & Subkoviak وهي بعنوان : تحليل المفردة : إحصائية مختصرة لاختبارات الإلتقان .

هدفت هذه الدراسة إلى معرفة العلاقة بين ثلاث طرق إحصائية لاختبار مفردات اختبار الإلتقان :

- الطريقة الأولى : طريقة الاختبار القبلي - البعدي Pretest - Posttest Methods لكوكس وفارجاس (Cox & Vargas 1966) وتعتبر هذه الطريقة بسيطة حسابياً .
 - الطريقة الثانية : هي طريقة السمات الكامنة Latent Trait Method وهي لفان دير ليندن (Van der linden 1981) وهي معقدة حسابياً ، ولكن لها العديد من الميزات النظرية .
 - الطريقة الثالثة : تم اقتراحها في هذه الدراسة وهي توازي طريقة السمات الكامنة في العديد من الأوجه ولكنها أبسط حسابياً فهي تشابه طريقة كوكس وفارجاس ، وهذه الطريقة هي معامل التوافق Agreement Statistic لهارس وصبكوفياك (Harris & Subkoviak 1986)
- جعلت هذه الدراسة معامل السمات الكامنة معياراً للمقارنة ، وسعت لتحديد إذا ما كان اختيار مفردات الاختبار بنظرية السمات الكامنة وبمعامل التوافق متشابهاً بحيث يمكن استخدام معامل التوافق مكان معامل السمات الكامنة .
- ولمقارنة الطرق الثلاثة استخدم برنامج كمبيوتر لتوليد استجابات المفردات لاختبارات إتقان مكونة من ٣٠ ، ٥٠ ، ١٠٠ مفردة ، طبقت كاختبار قبلي - بعدي على عينات ٣٠ ، ٦٠ ، ١٢٠ شخصاً ، وتم تحديد درجتي قطع ٧٥ % و ٨٥ % .
- وأشارت نتائج الدراسة إلى أن العلاقة بين معامل السمات الكامنة ومعامل التوافق كانت قوية ، حيث بلغت نسبة الارتباط ٠,٩١ أي أن الأمثلة التي تختار بطريقة السمات الكامنة هي نفسها تقريباً تختار بطريقة معامل التوافق وبذلك يمكن إحلال معامل التوافق مكان معامل السمات الكامنة .
- بينما الارتباط بين معامل السمات الكامنة ومعامل كوكس - فارجاس كان سالباً ، وهذا يدل على أن المفردات التي تكون صعبة في معامل السمات الكامنة تميل لأن تختار أولاً بمعامل كوكس - فارجاس والمفردات التي تكون سهلة بمعامل السمات الكامنة يتم اختيارها في النهاية .
- دراسة شانون وكلفر (Shannon & Cliver 1987) وهي بعنوان : تطبيق نظرية الاستجابة للمفردة في المقارنة بين أربعة معاملات تمييز تقليدية للاختبارات محكية المرجع .

هدفت هذه الدراسة لدراسة الارتباط بين نظرية الاستجابة للمفردة والمشتقة من دوال معلومات المفردة (IIFS) [معلومات عن الفرد وهي قدرة الفرد ومعلومات عن المفردة هي : معامل الصعوبة - معامل التمييز - معامل التخمين (حسب النموذج)] أو بين أربعة معاملات تمييز تقليدية وهي :

١. معامل فاي Phi - Coefficient .

٢. معامل برينان Br .

٣. معامل Phi / Phi Max .

٤. معامل التوافق .

ورببت هذه المعاملات ، وتم استخدام معامل ارتباطات الرتب لسبيرمان ، وأظهرت نتائج هذه الدراسة أن الارتباطات بين نظرية الاستجابة ومعامل Phi كانت عالية ، وارتباطات نظرية الاستجابة مع معامل التوافق ومعامل Br كانت جيدة ، باستثناء العامل Phi / Phi Max كان منخفضاً .

دراسة لين (1988) Lin وهي بعنوان : مقارنة ثلاث طرق لاختيار المفردة في الاختبارات محكية المرجع.

هدفت هذه الدراسة لمقارنة ثلاثة طرق لاختيار المفردة ، وذلك لمعرفة أفضل مفردات مميزة للاختبار ، وتعطي نتائج التصنيف إلى متقنين وغير متقنين .

والطرق التي تمت المقارنة بينها هي :

١. طريقة التوافق Agreement Approach .

٢. طريقة معامل فاي Phi - Coefficient .

٣. طريقة الاختيار العشوائي Random Selection .

وأداة الدراسة كانت اختبار محكي المرجع مؤلف من ٥٠ مفردة في مادة العلوم الطبيعية، طبق على عينة مؤلفة من ١٨٣٦ طالباً ، وتم استخدام عينات جزئية من هذه العينة (٧٥ ، ١٥٠ ، ٣٠٠) وكذلك تم استخدام اختبارات فرعية عدد مفرداتها (١٥ ، ٢٥ ، ٣٥) .

وأشارت نتائج هذه الدراسة إلى أن طريقة معامل التوافق هي الأفضل في اختيار مفردات الاختبار محكي المرجع وذلك على مستوى الصف الدراسي ، بينما كانت طريقة معامل Phi هي الأفضل للاستخدام على مستويات المقاطعة أو الولاية .

وفي حين أن طريقة الاختيار العشوائي لمفردات الاختبار لم يكن لها خصائص مماثلة في اختيار المفردات ، وأعطت أدنى ثبات عند مقارنتها بمعامل التوافق ومعامل فاي Phi .

دراسة عبد الله القاضي (١٩٩٣) وهي بعنوان : دراسة مقارنة لبعض طرق تحليل بنود الاختبارات مرجعية المحك وفاعليتها في الاختبار .

واختار الباحث ثلاث طرق لتحليل مفردات الاختبار هي :

١. طريقة معامل التمييز القبلي - بعدي Pretest - Posttest Statistic (كوكس وفارجاس)

٢. طريقة معامل التمييز لمجموعة الدارسين وغير الدارسين - Uninstructed Instructed Statistic (ليفين ومارتن) .

٣. طريقة معامل التوافق المرجعي Agreement Statistic (مارس وصبكوفياك).

وهدفت هذه الدراسة إلى :

أ- معرفة العلاقة بين معامل التوافق المرجعي وكل من طريقة معامل التمييز القبلي - بعدي ومعامل التمييز لمجموعة الدارسين وغير الدارسين .

ب- معرفة العلاقة بين معامل التمييز القبلي - بعدي ومعامل التمييز لمجموعة الدارسين وغير الدارسين .

ج- معرفة مدى التقارب بين الطرق الثلاثة من حيث اختيارها لبنود الاختبار النهائي .

ولكي تتحقق هذه الأهداف بنى الباحث اختباراً مكوناً من ٢٠ سؤالاً موضوعياً عن مفهوم الثبات وطرق حسابه ، وطبقه على مجموعة من الطلاب المسجلين في أربع شعب في مادة التقويم التربوي في جامعة الملك سعود ، حيث قام بتقسيم هؤلاء الطلاب إلى فئتين رئيسيتين بناءً على مدى التقارب في القدرة العقلية ، وقسمت أيضاً كل فئة إلى مجموعتين متجانستين في القدرة العقلية .

أي تم تطبيق الاختبار على أربع مجموعات ، ثم قدم الخبرة التعليمية (محاضرة عن الثبات) وبعد أربعة أيام أعيد تطبيق الاختبار السابق بعد تحديد الدرجة الفاصلة ٦٠ ٪ .

ومن أهم النتائج التي توصلت إليها الدراسة هي :

• عدم وجود علاقة بين معامل التوافق المرجعي ومعامل التمييز القبلي - بعدي حيث كان الارتباط بينهما ضعيفاً جداً ، والعلاقة بين معامل التوافق المرجعي مع معامل

التمييز لمجموعة الدارسين وغير الدارسين كانت تتراوح بين ٠,٥٠ و ٠,٧٥ وذلك عندما تكون المجموعات غير متكافئة ، وهذا يعكس مدى حساسية معامل التمييز للمجموعة الدراسية المتكافئة وغير المتكافئة .

- وجود علاقة موجبة دالة إحصائياً بين معامل التمييز القبلي - بعدي ومعامل التمييز لمجموعة الدارسين وغير الدارسين .
- فيما يتعلق بدرجة الارتباط في الاختبار وجد أنها ٠,٥١ بين معامل التمييز القبلي - بعدي ومعامل التوافق المرجعي ، ونسبة التطابق بين معامل التمييز لمجموعة الدارسين وغير الدارسين ومعامل التوافق المرجعي هي ٠,٥٢ بينما كانت هذه النسبة بين معامل التمييز القبلي - بعدي ومعامل التمييز لمجموعة الدارسين وغير الدارسين ٠,٧٨ .

دراسة ديرير (1993) Dirir

إن كثير من طرق القياس المستخدمة حالياً في إجراء الاختبار تتطلب من المفردات أن تصنع في مجال اختبار يقيس فضاء أحادي البعد ، ولكن افتراض أحادية البعد يصعب تلبية في الممارسة .

وطالما أن مجالات المفردة متعددة الأبعاد ومن أسباب تعدد أبعاد الاختبار وجود أبعاد ثانوية متجاوزة للمقدرة السائدة التي يفترض أن الاختبار يقيسها ، وأن نتائج إهمال البعد الواحد ربما يكون له تأثير على طرق اختيار المفردة المستخدمة لبناء الاختبارات وسيكون له تأثيرات غير معروفة ومتفاوتة على صدق وثبات الاختبارات .
لذلك هدفت هذه الدراسة إلى :

- ١- مراجعة البحوث الخاصة بأبعاد (بعديّة) Dimensionality الاختبار .
- ٢- فحص تأثير بعد الاختبار على تقييم القدرة ودقة القرار للاختبارات محكية المرجع .
- ٣- فحص تأثيرات تفاعل طرق اختبار المفردة مع بعد الاختبار وفئات المحتوى على تقييم القدرة ودقة القرار باختبار محكي المرجع .

والبحث التجريبي تكون من جزئين :

- في الجزء (أ) تم استخدام ثلاث مجموعات من المفردات ببيانات مختلفة الأبعاد لاختبارين مختلفين واستخدمت أربع طرق لاختيار المفردة لبناء اختبارات من كل مجموعة مفردات ، وتمت مقارنة تقييم القدرة ودقة القرار في كل اختبار من الاختبارات الاثنتي عشر .

• وفي الجزء (ب) تم استخدام بيانات حقيقية كبنك للأسئلة ، واستخدمت أربع طرق لاختيار مفردات لبناء اختبارات قصيرة من بنك الأسئلة ، ثم جرت مقارنة دقة القياس وصحة القرار للاختبارات الناتجة .

ولقد تبين أن قوة الأبعاد الثانوية تؤثر على دقة تقييم المقدره وصحة القرار لاختبارات الإلتقان وأن الطرق الأمثل Optimal Item Selection Methods في اختيار المفردة تؤدي على نحو أفضل من طرق أخرى للمفردة وخاصة عندما تكون بيانات الاختبار غير أحادية البعد ، وأن الاختلافات في دقة القياس وصحة القرار داخل بيانات ودرجات مختلفة من أبعاد متعددة وضمن طرق اختيار المفردة المختلفة كانت بارزة من الناحيتين الإحصائية والعملية .

وأشارت الدراسة إلى أن وجود أبعاد ثانوية في الاختبار قد يؤدي إلى سوء تصنيف الممتحن ومن ثم يقيد الانتفاع من الاختبار .

تعليق : هذه الدراسة تبين أهمية تحديد المجال الذي يقيسه الاختبار تحديداً جيداً وأن يكون هذا المجال أحادي البعد أي لا يضم أكثر من بعد بالإضافة إلى ذلك فإن الأبعاد الثانوية (مثل الدافعية وسرعة الأداء والقدرة على القراءة) يكون لها تأثير على الأداء وبالتالي على تصنيف الأفراد ، وكذلك الطرق في اختيار المفردة لها تأثير على تصنيف الأفراد ولكن هذه الدراسة لم تحدد الطرق التي استخدمتها في اختيار مفردات الاختبارات .

دراسة تايلر ولي (1995) Taylor & Lee وهي بعنوان : مراجعة لطرق ومشاكل قياس ثبات الاختبارات ومفردات محكية المرجع .

هدفت هذه الدراسة إلى مراجعة الكتابات حول المشاكل الفريدة لقياس ثبات الاختبارات محكية المرجع وثبات مفردات الاختبار ، ولمعرفة أكثر الطرق شيوعاً وعملية .

وقد أظهرت الدراسة أنه في نهاية الثمانينيات ظهر إجماع ضمن تلك الكتابات يقول بتوافر مقاييس معينة دقيقة لثبات الاختبارات محكية المرجع ولكنها غير عملية من الوجهة الحسابية .

وأشارت نتائج هذه الدراسة إلى أن التقدير لمعامل التوافق والإحصاء اللازمة التي تصحح معامل التوافق المتوقع بسبب الصدفة هو الأكثر استخداماً بين تلك المقاييس في ثبات الاختبار وثبات المفردة .

دراسة سعاد أحمد حساتين (٢٠٠٠) وهي بعنوان : مدى فاعلية نوع المفردة وبعض طرق اختيارها على ثبات الاختبار محكي المرجع .
وهدفت هذه الدراسة إلى الكشف عن مدى اختلاف معامل ثبات الاختبار محكي المرجع باختلاف :

- ١- طريقة اختيار المفردة .
- ٢- نوع المفردة .
- ٣- تغير ترتيب المفردة .
- ٤- طرق حساب معامل الثبات .

وقد حددت الباحثة لاختيار المفردة الطرق التالية :

- ١- قوة تمييز المفردة Item Discrimination Power .
- ٢- معامل صعوبة المفردة Item Difficulty .
- ٣- معامل الفروق القبلية - البعدية Pre - Post Test Differences Index .
- ٤- معامل برينان Brennan .

أما أداة الدراسة فهي عبارة عن ثلاثة اختبارات محكية المرجع في مادة علم النفس لطلاب الصف الثاني الثانوي ، ومر بناء الاختبارات بالمراحل التالية :

- ١- تحديد المجال السلوكي ويتضمن : أ) تحديد محتوى المادة الدراسية .
ب) تحديد أهداف محتوى النطاق إجرائياً .
 - ٢- بناء مفردات الاختبار يتضمن : أ) إعداد مواصفات الاختبار.ب) إعداد مفردات الاختبار .
 - ٣- تحديد مستوى الأداء المطلوب (الدرجة القاطعة) .
 - ٤- تقدير صدق وثبات مفردات الاختبار .
- وحددت الباحثة صدق الاختبار اعتماداً على صدق المحتوى من خلال المحكمين ، وطبق الاختبار على ٢٨٠ طالباً وطالبة في الثانوي العام حيث قامت بالتطبيق قبل التعليم وبعد الانتهاء منه .

أما أهم النتائج التي أسفرت عنها هذه الدراسة فهي :

- عند حساب الثبات بطريقة هارس فإن المفردات المختارة على أساس معامل كوكس وفارجاس Cox & Vargas أعطت أعلى معامل ثبات يليها المفردات المختارة على أساس معامل برينان ، ثم المفردات المختارة على أساس معامل التمييز ، ثم المفردات المختارة على أساس معامل الصعوبة .

- عند حساب الثبات بطريقة هاميلتون فإن المفردات المختارة على أساس معامل الصعوبة أعطت أعلى معامل ثبات يليها المفردات المختارة على أساس معامل برينان ، ثم معامل كوكس وفارجاس .
- عند حساب الثبات بطريقة ليفنجستون ، لا تختلف قيم معاملات الثبات الناتجة عبر طرق اختيار المفردات الأربعة .

• تعليق عام على الدراسات السابقة :

- أولاً : يلاحظ الباحث من خلال العرض السابق للدراسات التي تناولت بناء وصدق ثبات الاختبارات محكية المرجع ما يلي :
- أجمعت معظم الدراسات إن لم تكن جميعها على أن تحديد النطاق السلوكي هو الخطوة الأولى والأساسية في بناء الاختبار محكي المرجع .
- استخدمت أغلب الدراسات طريقة مواصفات الاختبار التي اقترحها بابام لتحديد النطاق ، وبناء مفردات الاختبار محكي المرجع مثل دراسة جانوبل ، براتر ، عماد عبد المسيح ، محمد الدسوقي .
- أكدت الدراسات السابقة على أن الصدق الوصفي (صدق المجال) هو أفضل الأنواع بالنسبة للاختبار محكي المرجع ، واستخدمت في ذلك آراء المحكمين والخبراء في المجال السلوكي في صدق مجال الاختبار وصدق المحتوى ، واستخدمت معادلة روفينيلي لمطابقة الهدف بالمفردات كما في دراسة براتر ، والكر ، عماد عبد المسيح ، محمد الدسوقي .
- أما بالنسبة للثبات فبعض الدراسات استخدمت الطرق المستخدمة في ثبات الاختبارات المعيارية كاستخدام معامل الاستقرار ومعامل التكافؤ كما في دراسة براتر ، وبعضها استخدم معادلة كيوبر رتشارسون إلا أن معظم الدراسات أوصت بعدم استخدام هذه الطرق ، وضرورة استخدام الطرق الخاصة بالاختبارات محكية المرجع .
- أغلب هذه الدراسات استخدمت درجة القطع ٨٠ ٪ لمستوى الإتيقان .
- وقد أفاد الباحث من هذا الجزء في خطوات بناء الاختبار محكي المرجع الذي قام بإعداده ، حيث قام بتحديد النطاق السلوكي واتباع في بناء مفردات الاختبار مواصفات الاختبار لبابام ، لأنها تعطي تفاصيل واضحة لبناء المفردات المرتبطة بالمجال ، وبالنسبة للثبات أفاد الباحث في اختيار الطرق التي تناسب الاختبار محكي المرجع وكذلك بالنسبة للصدق .

- وبالنسبة لدرجة القطع فقد اختلف الباحث مع الدراسات السابقة في تحديده درجة القطع ، حيث اختار الدرجة ٤٠ ٪ والتي تعتبرها وزارة التربية والتعليم في جمهورية مصر بأنها درجة النجاح في مادة الرياضيات للصف الأول الإعدادي .
 - ثانياً : ويلاحظ الباحث من عرض الجزء الثاني للدراسات السابقة والتي تناولت تحليل مفردات الاختبارات محكية المرجع ما يلي :
 - أكدت دراسة هالدنيا أن التحليل التجريبي للمفردات ضروري في الاختبارات محكية المرجع وأن طريقة راش ومعاملات الحساسية والمعاملات التقليدية وطريقة بايزن متماثلة في اختيار المفردة .
 - أشارت دراسة تراس أن أفضل طريقة لتحليل مفردات الاختبار محكي المرجع هي طريقة كوكس وفارجاس (C-V) . في حين أشارت دراسة سميث إلى أن طريقة كوكس وفارجاس وطريقة الارتباط الثنائي هما الأفضل ، وإذا كان هناك تباين في الدرجات فإن الارتباط الثنائي هي الأفضل بينما إذا لم يكن هناك تباين فإن طريقة كوكس وفارجاس تكون هي الأفضل .
 - دراسة اينسلي أشارت إلى أن تماثل الطرق التي درستها لاختيار مفردة الاختبار محكي المرجع ، وهي طريقة كوكس وفارجاس وطريقة كوسكوف وكلاين ١٩٧٤ وطريقة رودابوش والارتباط الثنائي النقطي .
 - بينما أثبتت دراسة سلفا أن نظرية الاستجابة للمفردة هي الأفضل في اختيار مفردات الاختبار محكي المرجع . في حين أن شانون أشار إلى وجود ارتباط بين معامل Phi (ليابام) والسمات الكامنة . أما لين فقد دل أن معامل التوافق هو أفضل من Phi والاختبار العشوائي للمفردات عندما يكون حجم العينة صغيراً ، ولكن عندما يكون كبيراً فإن معامل Phi يكون أفضل .
- يلاحظ الباحث مما سبق أنه يوجد تعارض بين نتائج هذه الدراسات ، وربما يكون سبب هذا التعارض العينة التي يطبق عليها أو حجم هذه العينة واختلاف مستوى الهدف السلوكي .