المجلة العربية للقياس والتقويم

دراسة أثر عدد فئات المفردات المشتركة ومستوى صعوبتها على تدريج مفردات بنك الأسئلة المصمم باستخدام نموذج التقدير الجزئي

إعـــــــــداد

د/ رضـــا ســـمير عـــوض عمـــر
مدرس بقسم علم النفس التربوي
كلية الدراسات العليا للتربية
جامعة القاهرة

ملخص البحث

هدف البحث الحالي إلى دراسة أثر عدد فئات المفردات المشتركة ومستوى صعوبتها على تدريج مفردات بنك الأسئلة المصمم باستخدام نموذج التقدير الجزئي، على عينة بلغ قوامها (1149) طالباً وطالبة من طلاب الصف الثالث الثانوي القسم العلمي في مدارس التعليم العام الحكومي في جمهورية مصر العربية، موزعين على (23) فصل دراسي تم اختيارهم بطريقة المعاينة العشوائية العنقودية من (6) مدارس، (3) للذكور، (3) للإناث تابعين لمديريات التربية والتعليم بالقاهرة والقليوبية.

واعتمد البحث على المنهج التجريبي لمعالجة المتغيرات واختبار الفروض وتم تطبيق اختبار تحصيلي محكي المرجع من إعداد الباحثة في الوحدتين الثانية والثالثة من كتاب الكيمياء للصف الثالث الثانوي. وتوصلت الباحثة إلى النتائج التالية:

وجود فروق بين تدريج مفردات بنك الأسئلة (التدريج المرجعي) وتدريج مفردات الاختبارين متكافئ 1، ومتكافئ 2 (التدريج المشترك) عند استخدام مفردات مشتركة ذات فئتين. لكن لا يوجد اختلاف بين تدريج مفردات بنك الأسئلة (التدريج المرجعي) وتدريج مفردات الاختبارين (التدريج المشترك) عند استخدام مفردات مشتركة ذات أربع فئات أو خمس فئات كذلك أوضحت نتائج الدراسة وجود فروق بين التدريج المشترك الناتج عند استخدام مفردات مشتركة ذات فئتين والتدريج المشترك الناتج عن استخدام مفردات مشتركة ذات أربع فئات. كذلك أوضحت النتائج عدم وجود أثر لمستوى صعوبة المفردات المشتركة على تدريج مفردات بنك الأسئلة، ولا يوجد فروق بين التدريجات المشتركة بينها وبين بعضها البعض والناتجة عن اختلاف مستوى صعوبة المفردات المشتركة (سهلة أو متوسطة أوصعبة المستوى).

الكلمات المفتاحية: تدريج مفردات – بنك الأسئلة – نموذج التقدير الجزئي

The effect of the Number of the common item categories and its level of Difficulty on the calibration of item bank designed According to partial credit model

Abstract

The purpose of the research was to find the effect of the categories number of common items and the difficulty parameters on calibration the Items of item Bank by using the partial credit model? The study sample consists of (1149) students in its final from in third grade secondary in the Practical department of the schools, and was represented by (23) Classrooms have been selected in a random cluster sampling from six schools represented two departments teaching ***(Cairo, Qalubia)***. The Current study used criterion referenced collective test Prepared by the researcher in the second and third units of the Book of chemistry for the third grade of secondary, where they from the core of item bank included in its find from two (33) different in the number of categories (two four and five categories) and multiple response (binary response and multiple response)., The study used experimental design to manipulate the variables, and test hypotheses.,The study found the following results: There is difference between the referential scale comparing with the resulting scales from equating of the two tests proportionate (I) and proportionate "2" in one common scale, if the common items had two categories, but no difference for four and five categories, There is not difference between the referential scale comparing with the resulting scales from equating of the two tests proportionate "l" and proportionate "2" in one common scale, if the difficulty parameters of common items differed. And there is not difference between the three common scales if the difficulty parameters of common items differed.

KeyWords: Item Scaling, Question Bank, Partial Credit Model.

مقدمـــة

تُعَّــد الاختبارات التربوية والنفسية من أهم الوسائل الرئيسية لعملية القياس والتقويم، والتي تزود المؤسسات التعليمية، والتربوية بالمعلومات اللازمة عن مدى تحقق الأهداف التعليمية، بالإضافة إلى تقديم معلومات يمكن من خلالها اتخاذ قرارات مصيرية للعديد ممن يخضعون لهذه الاختبارات سواء بالانتقاء أو التصنيف أو الأجازة أو التحويل إلى برامج علاجية أو تأهيلية، وترتبط دقة وسلامة هذه القرارات بمدى جودة وكفاءة هذه الاختبارات وإذا كانت هذه الاختبارات لا تتميز بالدقة والموضوعية والكفاءة المطلوبة فإن ذلك يؤدي إلى صدور قرارات غير صائبة، مما يترتب عليها إهدار العديد من القدرات وضياع حقوق العديد من الطلاب، الذين إن تيسرت لهم الوسائل والظروف الملائمة لأثبتوا قدراتهم وكفاءتهم في العديد من المجالات.

ولقد كان خبراء القياس النفسي والتربوي يعتمدون في تصميم وبناء الاختبارات النفسية والتربوية وتحليل البيانات المستمدة منها على المدخل التقليدي في القياس والمسمى بنظرية القياس الكلاسيكية. إلا أن هذه النظرية لم تستطع التغلب على كثير من المشكلات السيكومترية المعاصرة، وقد أدى ذلك إلى محدودية فاعلية هذه النظرية وقلتها في كثير من التطبيقات العملية، مثل إنشاء بنوك الأسئلة، ومعادلة درجات الاختبارات، وتصميم أنظمة الاختبارات المحوسبة، والكشف عن تحيز المفردات وسلوكها المتمايز، وقياس النمو، وتشخيص صعوبات التعلم، وقد وجه إلى هذه النظرية الكثير من الانتقادات منها إن أساليب تحليل المفردات تتأثر تأثراً ملحوظًا بخصائص عينة الأفراد الممتحنين وصعوبة المفردات، مما يؤدي إلى تغيير الإحصاءات بتغير هذه العينة، كما أن درجات الأفراد تعتمد على عينة. المفردات التي يختبرون بها، أي أن الأساليب الكلاسيكية في تحليل المفردات تكون محكومة بعينة الأفراد والمفردات، ولذلك فإن الإحصاءات الناتجة لا يمكن تعميمها عبر العينات المختلفة. ***(محمد الشافعي 2008، Hambelton, swaminthan & Rogers 1991)***.

ونتيجة لأوجه القصور التي أظهرتها نظرية القياس الكلاسيكية إتجه علماء القياس والتقويم جاهدين إلى تطوير مدخل آخر يتحقق من خلاله الموضوعية في القياس ويمكن من

خلاله تلافي أوجه القصور السابقة، وقد أسفرت هذه الجهود عن ظهور بعض الاتجاهات الحديثة في القياس والتي حققت تقدماً ملموساً في تحقيق الموضوعية في القياس ***(Embretson & Reise, 2000)***.

ومن بين هذه الاتجاهات ظهور نظرية السمات الكامنة، أو ما يُسمى بنظرية الاستجابة للمفردة Item Response Theory (IRT)، والتي تهدف إلى تحديد العلاقة بين أداء الفرد في اختبار ما، وبين السمة أو القدرة الكامنة وراء هذا الأداء. كما تهدف إلى تحديد مقدار هذه السمة، وذلك للتنبؤ بسلوك الفرد في مواقف اختبارية متشابهة، وهذه السمة أو القدرة غير ملحوظة إنما يستدل عليها من خلال الأداء على المفردات، ولهذا سميت بنظريات السمات الكامنة. كما أن نظرية الاستجابة للمفردة تحقق متطلبات القياس الموضوعي المتمثل في تحرر تدريج أدوات القياس من خصائص الأفراد، وتحرر تقديرات قدرة الأفراد من خصائص المفردات التي استخدمت في هذا القياس.

مشكلة البحث:

أوضحت البحوث السابقة التي اهتمت بإنشاء بنوك الأسئلة باستخدام نماذج الاستجابة للمفردة، وخاصة التي أجريت في مصر والأردن والسعودية أنها اعتمدت على نماذج الاستجابة ثنائية التدريج في تصميم وإنشاء تلك البنوك، وفي أغلب هذه البحوث تم الاعتماد على نموذج راش الذي يتميز بسهولته، وقلة العمليات الإحصائية المستخدمة فيه، وهذا ما جعل الفائدة محدودة لهذه البنوك، بسبب إهمال قياس كثير من الأهداف السلوكية في الأسئلة المستخدمة فيها مثل الأهداف المهارية كإجراء البراهين الرياضية، أو حل المسائل الرياضية أو التعبير عن أفكار كتابية أو شفهية، أو تطبيقية كتجميع الأجهزة، وغيرها والتي يتطلب قياسها استخدام مفردات تتكون من أكثر من خطوة للوصول إلى الحل النهائي، حتى يتم التمكن من قياس المعلومات الجزئية التي اكتسبها المتعلم في مادة دراسية معينة، ومن هنا ظهرت الحاجة إلى

إنشاء بنوك أسئلة تقيس أهداف تعليمية متعددة لمحتوى دراسي معين، لذلك كان التركيز في هذا البحث على إنشاء بنك أسئلة باستخدام مفردات ثنائية الاستجابة ومفردات متعددة الاستجابة لفتح المجال أمام الباحثين والمهتمين ببنوك الأسئلة من استخدام هذه الدراسة في تطوير بنوك الأسئلة بالاعتماد على نماذج متعددة التدريج. ولتقدير معالم الصعوبة الفئوية لكل خطوة من خطوات الاستجابة للمفردات متعددة التدريج فإن ذلك يتطلب:

1- استخدام أحد نماذج الاستجابة متعددة التدريج لإجراء عملية التقدير، لذلك تم استخدام نموذج التقدير الجزئي لاعتماده على مُعلم واحد وهو معلم الصعوبة.
2- استخدام أحد تصاميم جمع البيانات، وهو تصميم المفردات المشتركة لمعادلة مفردات الاختبارين المتكافئين لجعلهما على تدريج واحد مشترك بصفر واحد مشترك، من خلال طريقة معادلة رايت وستون. والسبب أيضاً في اختيار هذا التصميم أن غالبية الدراسات والبحوث السابقة أوضحت أنه أدق من تصميم الأفراد المشتركين في تدريج بنوك الأسئلة.

لذلك تتحدد مشكلة البحث بالتساؤل الرئيسي التالي:

- ما أثر عدد فئات المفردات المشتركة ومستوى صعوبتها على تدريج مفردات بنك الأسئلة المصمم باستخدام نموذج التقدير الجزئي؟

ويتفرع من هذا التساؤل الرئيسي التساؤلات الفرعية التالية:

1- هل يختلف تدريج مفردات بنك الأسئلة في التدريج المرجعي عنه في التدريجات الناتجة عن معادلة مفردات اختباري البنك متكافئ 1، ومتكافئ 2 في تدريج مشترك وذلك باختلاف عدد فئات المفردات المشتركة (فئتين – أربع فئات – خمس فئات)؟
2- هل يختلف تدريج مفردات بنك الأسئلة في التدريج المرجعي عنه في التدريجات الناتجة عن معادلة مفردات اختباري البنك متكافئ1، ومتكافئ 2 في تدريج مشترك وذلك باختلاف مستوى صعوبة المفردات المشتركة (سهلة – متوسطة – صعبة) المستوى؟

أهداف البحث:

- هدف البحث الحالي إلى معرفة أثر بعض خصائص المفردات المشتركة على تدريج مفردات بنك الأسئلة وبالتحديد:
- معرفة أثر عدد فئات المفردات المشتركة على تدريج مفردات بنك الأسئلة المصمم باستخدام نموذج التقدير الجزئي.
- معرفة أثر صعوبة المفردات المشتركة على تدريج مفردات بنك الأسئلة المصمم باستخدام نموذج التقدير الجزئي

أهميـة البحـث:

أولاً: الأهمية النظرية:

تتضح الأهمية النظرية للدراسة في إمداد القائمين على إعداد وإنشاء بنوك الأسئلة بالمعلومات الضرورية عن كيفية إعدادها وتصميمها وفقًا لنظريات القياس الحديثة ومنها نموذج التقدير الجزئي، بحيث تقيس الأهداف التعليمية التي تتطلب مفردات متعددة الاستجابة، وحتى لا تفقد المعلومات الجزئية التي اكتسبها المتعلم.

ثانيًا: الأهمية التطبيقية:

1- تزويد القائمين على إعداد بنوك الأسئلة بمعلومات عن عدد فئات المفردات المشتركة التي تعطي أدق تدريج لبنك الأسئلة.
2- تزويد القائمين على إعداد بنوك الأسئلة بمعلومات عن مستوى صعوبة المفردات المشتركة الذي يعطي أدق تدريج لبنك الأسئلة.
3- توجيه القائمين على إعداد بنوك الأسئلة باستخدام برنامج PARSCALEV (4.1) في عملية تقدير معالم المفردات والأفراد وذلك عند استخدام نماذج الاستجابة متعددة التدريج.

مصطلحات البحث:

- **المفردات المشتركة** Common Item:

هي المفردات التي تشكل اختبار الربط بين الاختبارين المراد معادلة درجاتهما.

- **فئات الاستجابة** Response Categories:

هي عبارة عن الدرجات التي تفصل بين خطوات الاستجابة.

- **التدريج المرجعي** Referential Scaling:

هو التدريج الناتج من خلال تقدير معالم الصعوبة لجميع مفردات بنك الأسئلة بوحدة قياس مشتركة وصفر مشترك.

- **التدريج المشترك** Common Scaling:

هو التدريج الناتج من خلال معادلة صعوبة مفردات الاختبارين المتكافئين في مستوى الصعوبة باستخدام مفردات مشتركة.

- **المعادلة** Equating:

هي عملية الحصول على درجات متكافئة لاختبارين متكافئين في الصعوبة واللذان يقيسان نفس المتغير.

- **بنك الأسئلة** Item Bank:

هو مجموعة من المفردات ثنائية الاستجابة ومتعددة الاستجابة الممثلة لمتغير التحصيل في محتوى الوحدتين الثانية والثالثة من كتاب الكيمياء للصف الثالث الثانوي. والتي لا تختلف فيما بينها إلا من حيث مستوى صعوبتها، والمدرجة على تدريج خطى بوحدة قياس معرفة بحيث تمثل مقياساً يغطي مدى واسعاً من المتغير وهو التحصيل الدراسي.

- **نموذج التقدير الجزئي** Partial Credit Model

هو أحد نماذج الاستجابة للمفردة المنبثق عن نموذج راش ثنائي التدريج والذي يهدف إلى تحديد معالم الصعوبة الفئوية لكل خطوة من خطوات الاستجابة للمفردة ذات الفئات المتعددة.

حدود البحث:

(1) الحدود الموضوعية:

تم استخدام نموذج التقدير الجزئي كأحد نماذج نظرية الاستجابة للمفردة أحادية البعد في تقدير معالم مفردات أداة البحث، وتقدير معالم قدرة عينة البحث. وقد اقتصرت مفردات بنك الأسئلة في هذا البحث على المفردات التي تكون الإجابة عنها مكونة من خطوة واحدة أو ثلاث خطوات أو أربع خطوات، والتي تقيس تمكن الطلاب من المفاهيم الكيميائية وحل المشكلات المتضمنة في الوحدتين الثانية والثالثة من كتاب الكيمياء للصف الثالث الثانوي.

(2) الحدود المكانية:

المدارس الثانوية التابعة لمديريات التربية والتعليم بالقاهرة والقليوبية ومنها مدرسة العبور الثانوية للبنين والعبور الثانوية للبنات.

(3) الحدود الزمانية:

الفصل الدراسي الثاني من العام الدراسي (2023 – 2024).

(4) الحدود البشرية:

تلاميذ الصف الثالث الثانوي الذكور والإناث وهم التلاميذ الذين يبلغون سن الثامنة عشر وقد يزيد العمر عن ذلك بسبب التأخر الدراسي أو الرسوب.

الإطار النظري والدراسات السابقة:

سوف توضح الباحثة في الجزء التالي المفاهيم النظرية ذات الصلة بموضوع البحث الحالي وتتضمن:

نماذج نظرية الاستجابة للمفردة:

يصنف مكدونالد (Macdonaled, 2002) نماذج الاستجابة للمفردة إلى نوعين وفقًا لنوع الاستجابة وهما:

(أ) نماذج الاستجابة ثنائية التدريج Dichotomous Response Models**:**

وهذه النماذج تحدد الاستجابة على المفردات بشكل ثنائي مثل (صواب – خطأ)، (موافق، غير موافق)، (نعم، لا) وبالتالي تأخذ استجابات التلاميذ فيها الدرجة (1) للإجابة الصحيحة والدرجة (صفر) للإجابة الخاطئة. ومن أشهر النماذج ثنائية التدريج: نموذج راش أحادي المَعْلم، ونموذج بيرنيوم ثنائي المَعْلم، ونموذج لورد ثلاثي المَعْلم. وسوف تتحدث الباحثة في الجزء التالي عن نموذج راش بشيء من التفصيل لكون نموذج التقدير الجزئي المستخدم في هذا البحث امتداد لنموذج راش.

النموذج أحادي المَعْلم One-Parameter Logistic Model**:**

يُعَّــد هذا النموذج من أشهر وأبسط نماذج الاستجابة للمفردة، وأكثرها استعمالاً، والذي اقترحه العالم الرياضي الدنماركي جورج راش (1960) لدراسة العلاقة بين احتمال استجابة الفرد على المفردة استجابة صحيحة، وبين ما يمتلكه الفرد من سمة أو قدرة مقاسة عن طريق مفردات الاختبار، وكذلك لتحقيق متطلبات القياس الموضوعي.

الأساس النظري الذي يتبناه "راش" في بناء هذا النموذج:

عندما يستجيب الفرد (v) عن المفردة (i) فإن قدرة هذا الفرد (θv) تحدد موقع هذا الفرد على متصل المتغير المقاس، وأيضًا صعوبة المفردة (bi) تحدد موقع المفردة على متصل المتغير المقاس، ورغم وجود عوامل داخلية كثيرة تؤثر في استجابة

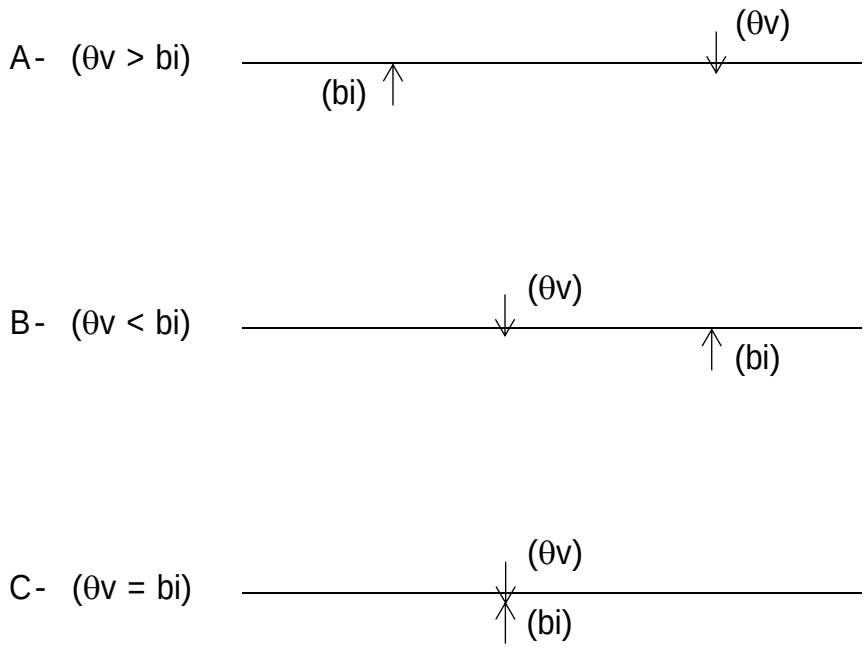
الأفراد، لكن عند بذل معد الاختبار جهد للسيطرة على العوامل المؤثرة في استجابة الفرد، فإن استجابة الفرد (v) على المفردة (i) هي (xvi) والتي تتحدد وفقًا لمتغيرين أساسيين هما:

- متغير قدرة الفرد (θv).
- متغير صعوبة المفردة (bi).

ومن المنطقي أنه إذا زادت قدرة قدرة الفرد (θv) عن صعوبة المفردة (bi) فإن احتمال حدوث الاستجابة الصحيحة هو الأرجح، أي أن (xvi=1) أما إذا انخفضت قدرة الفرد (θv) عن صعوبة المفردة (bi) فإن احتمال حدوث الاستجابة هو الأرجح أي أن (xvi=o). وقد يحدث أحيانًا أن تكون قدرة الفرد (θv) أكبر من صعوبة المفردة (bi) ومع هذا تكون استجابة الفرد لهذه المفردة السهلة استجابة خاطئة (xvi=o)، وقد يحدث عكس ذلك وهو أن تكون قدرة الفرد (الممتحن أو التلميذ) θv أقل من صعوبة المفردة (bi) ومع هذا تكون استجابة الفرد لهذه المفردة الصعبة استجابة صحيحة (xvi=1) ***(Wright & Stone.1979)***.

وحيث أن قدرة الفرد (θv) وصعوبة المفردة (bi) يمثلان وضعين على متصل متغير واحد يشتركان فيه، لذا فإن الفرق (θv - bi) هو الصيغة المناسبة للعلاقة بينهما، ولكن ليست هذه العلاقة قاطعة للفرق بين قدرة الفرد وصعوبة المفردة في تحديد قيمة الاستجابة (xvi)، بل من المنطقي أن يكون تأثير هذا الفرق (θv - bi) على شكل الاستجابة احتمالي ***(أمينة كاظم، 1988ب)***. والشكل (2 – 1) يوضح ثلاث حالات (A.B.C) لتأثير الفرق بين قدرة الفرد وصعوبة المفردة (θv - bi) على احتمالية الاستجابة الصحيحة للمفردة، فإذا كانت قدرة الفرد (θv) أكبر من صعوبة المفردة (bi) فإن الفرق (θv - bi) يكون أكبر من الصفر، وبالتالي يكون احتمال حدوث الاستجابة الصحيحة أكبر من (0.5) (الحالة A). وإذا كانت

قدرة الفرد (θv) أقل من صعوبة المفردة (bi) يكون أصغر من الصفر، وبالتالي يكون احتمال حدوث الاستجابة الصحيحة أقل من (0.5) (الحالة B). أما إذا تساوت قدرة الفرد (θv) مع صعوبة المفردة (bi) فإن احتمال حدوث الاستجابة الصحيحة للمفردة يساوي احتمال حدوث الاستجابة الخطأ ويساوي كل منهما (0.5) (الحالة 2) وهنا يكون (θv – bi =o). وذلك كما هو موضح بالشكل التالي (1–1):



شكل (1–1)

- تأثير الفرق بين مستوى قدرة الفرد ومستوى صعوبة المفردة في احتمال حدوث الاستجابة الصحيحة للمفردة.

خواص نموذج راش:

1– يشتمل النموذج على عدد أقل من متغيرات المفردة من السهل التعامل معها.

2– تعتبر مشاكل تقدير المتغيرات قليلة في عددها مقارنة مع النماذج الأخرى.

3- تمتلك جميع المفردات قوة تمييز متساوية.

4- يكون التخمين في حدوده الدنيا ***أمينة كاظم (1988 أ، 1988 ب)***.

- **النموذج ثنائي المَعْلم** Tow Parameter Logistic Model

هذا النموذج هو امتداد للنموذج أحادي المَعْلم، والذي اقترحه ***بيرنبوم (Birnbaum, 1968)***، حيث يقوم بتقدير صعوبة المفردات (bi) ومعاملات التمييز (ai)، ويفترض انعدام التخمين (C)، والصيغة الرياضية لهذا النموذج هي:

$$Pi\ (\theta) = \frac{{}_{e}Dai\ (\theta v - bi)}{1+{}_{e}Dai\ (\theta v - bi)}$$

حيث (ai) مؤشر تمييز المفردة.

(θ) Pi هو احتمال إجابة المفحوص الذي قدرته (θ) على المفردة i التي صعوبتها (bi) وتمييزها (ai) *(Ferrando, Loyenzo & Moline 2001)* وإضافة مَعْلم التمييز في هذا النموذج ترجع إلى صعوبة إيجاد مجموعة من المفردات تميز بدرجة واحدة مستويات القدرة أو السمة التي يقيسها الاختبار وهو الافتراض الذي استند إليه نموذج راش ***(صلاح الدين علام، 2005)***.

- **النموذج ثلاثي المَعْلم** Three – Parameter Logistic Model**:**

وهذا النموذج هو امتداد للنموذج أحادي المَعْلم والنموذج ثنائي المَعْلم حيث يضيف مَعْلم ثالث للمفردة هو الخط التقاربي الأدنى للمنحنى المميز للمفردة Lower Asymptote، والذي يعبر عن احتمالية توصل المفحوصين من ذوي القدرات المتدنية للإجابة الصحيحة للمفردة وبالتالي لا يصل هذا المنحنى إلى الصفر ***(صلاح الدين علام، 2005)***.

أي أن المفردات تختلف فيما بينها في معاملات الصعوبة والتمييز والتخمين، ويرمز لهذا المَعْلم بالرمز Ci. والصيغة الرياضية لهذا النموذج هي:

$$Pi\ (\theta) = Ci + (1-Ci)\ \frac{{}_{e}Dai\ (\theta v - bi)}{1+{}_{e}Dai\ (\theta v - bi)}$$

$$i = 1, 2, 3\ldots\ldots, n$$

حيث (Ci) مَعْلم التخمين ***(Hambleton & Swaminathan)***

(ب) نماذج الاستجابة متعددة التدريج Polytomous Response Models

انتشرت الاختبارات الموضوعية من نوع (الخطأ والصواب، والاختيار من متعدد) بشكل كبير جداً، وذلك في عدد من الاختبارات، كاختبارات القبول للجامعات، والاختبارات التحصيلية التي يعدها المعلمون، وبنوك الأسئلة التي تعدها مراكز متخصصة. ويرجع ذلك إلى سهولة استخدام النماذج ثنائية التدريج والممثلة بنموذج "راش" في تحليل الاستجابات لمثل هذه الاختبارات، والنماذج ثنائية التدريج تفترض أن المفحوص أو التلميذ إما أن تكون إجابته صحيحة بشكل تام، أو تكون خاطئة بشكل تام. والاستخدام المفرط لهذا النوع من الاختبارات يؤدي إلى فقدان كثير من البيانات التشخيصية المهمة التي يمكن الحصول عليها من الإجابات الجزئية عن المفردة، حيث أن نماذج الاستجابة ثنائية التدريج تهمل مثل هذه الاستجابات الجزئية. لذلك ظهرت نماذج الاستجابة متعددة التدريج والتي تسمح بخيارات متعددة في الإجابة والتي يظهر فيها الاهتمام بالإجابات المتدرجة. وبالتالي تقدير قدرة الفرد قد تكون أقرب إلى الدقة عند الأخذ بالحسبان المعرفة الجزئية في عملية التقدير (De Ayala, 1989)، ويرى ماسترز (Masters, 1982) أن الهدف من استخدام مفردات مكونة من أكثر من خطوتين في الحل هو الحصول على معلومات أكثر حول السمة المراد قياسها وبالتالي الوصول إلى أكبر قدر من الدقة في تقدير قدرة الفرد، ويرى إمبرستون ورايس (Embrestone & Reise,

2000) أن هناك العديد من المواقف الاختبارية التي لا يمكن تدريجها بشكل ثنائي (صحيح، خاطئ)، كما في مفردات مقاييس الشخصية ومقاييس الاتجاهات، والاختبارات المقالية حيث تحتوي على مفردات ذات استجابات متعددة، والمفردة التي يكون لها أكثر من فئتين تصنف أنها متعددة الاستجابة (De Mars, 2010). ويصنف زيكار (Zickar, 1998) نماذج الاستجابة متعددة التدريج إلى: نموذج الاستجابة الاسمى (NRM)، ونموذج الاستجابة المتدرجة (GRM)، ونموذج التقدير الجزئي المعمم (GPCM)، ونموذج التقدير الجزئي (PCM) وسوف تتحدث الباحثة بشيء من التفصيل والتوضيح عن نموذج التقدير الجزئي لكونه النموذج المستخدم في البحث الحالي:

- **نموذج التقدير الجزئي** Partial Credit Model (PCM)**:**

يُعَّــد نموذج التقدير الجزئي لماسترز (Masters, 1982) توسيعًا لنموذج راش أحادي المَعْلم ثنائي التدريج (Davier & Yamamoto, 2004) ويتعامل نموذج التقدير الجزئي مع المفردات متعددة الاستجابة، والتي تتطلب الإجابة عنها المرور بعدة خطوات، بحيث تمثل كل خطوة نجاحًا جزئيًا في الإجابة عن المفردة، وبالتالي يتم تقدير كل خطوة تقديراً جزئيًا في ضوء الدرجة الكلية للمفردة، ومن ثم يهدف لدراسة النجاحات الجزئية للأفراد في الإجابة عن مفردات الاختبار، ولذا يصلح استخدامه مع الأسئلة الموضوعية والأسئلة المقالية، التي تتطلب الإجابة عن مفرداتها المرور بعدة خطوات متسلسلة ومتتابعة (Ferrare, 1989)، ولإيضاح طريقة إعطاء الفرد تقديرات جزئية على النجاح الجزئي نأخذ المثال التالي:

السؤال: ذرة عددها الذري (200) وضح باستخدام التوزيع الإلكتروني هل يحدث لها أكسدة أم اختزال؟

خطوات الحل:

الخطوة الأولى: يعطي الفرد الدرجة (1) بدلاً من (صفر) إذا قام بكتابة التوزيع الإلكتروني للذرة.

التوزيع الإلكتروني $IS^2 2S^2 2P^6 3S^2 3P^6 4S^2$

الخطوة الثانية: يعطي الفرد أو التلميذ الدرجة (2) بدلاً من الدرجة (1) إذا ذكر أن عدد الإلكترونات في المدار الخارجي ألكترونين.

الخطوة الثالثة: يعطي الفرد الدرجة (3) بدلاً من الدرجة (2) إذا ذكر أن الذرة أثناء التفاعل تفقد إلكترونيين لتصل إلى حالة الاستقرار.

الخطوة الرابعة: يعطي الفرد أو التلميذ الدرجة (4) بدلاً من (3) إذا ذكر أن الذرة حدثت لها عملية أكسدة. ويوضح الجدول التالي العلاقة بين خطوات الاستجابة وفئات الاستجابة.

جدول (1)

العلاقة بين خطوات الاستجابة

وفئات الاستجابة

خطوات الاستجابة	فئات الاستجابة صفر–1–2–3–4
$IS^2 2S^2 2P^6 3S^2 3P^6 4S^2$	الخطوة الأولى ← (1)
عدد الإلكترونات في المدار الخارجي إلكترونين	الخطوة الثانية ← (2)
الذرة أثناء التفاعل تفقد إلكترونين لتصل إلى حالة الاستقرار	الخطوة الثالثة ← (3)
الذرة حدثت لها عملية أكسدة	الخطوة الرابعة ← (4)

يتضح من الجدول السابق العلاقة بين عدد خطوات الاستجابة وفئات الاستجابة، حيث إن خطوات الحل للمثال السابق أربع خطوات، بينما فئات الاستجابة خمس فئات استجابة هي (صفر 1، 2، 3، 4)، والتلميذ إذا أتم الخطوة الأولى بنجاح حصل على الدرجة (1)، وإذا لم يتم الخطوة الأولى بنجاح حصل على الدرجة (صفر)، وإذا أتم الخطوة الثانية بنجاح حصل

على الدرجة (2)، وإذا فشل حصل على الدرجة (1)، وهكذا حتى يحصل الفرد أو التلميذ على الدرجة (4) بعد أن يقوم بالخطوة الرابعة، لذلك نجد أنه لحصول التلميذ على الدرجة الكاملة للمفردة، يجب أن يتم الخطوات الأربع بنجاح، وبشرط أن تكون الإجابات لجميع الخطوات متتابعة، أي يجب أن تتم بالترتيب ***(زياد العبدالله، 2009)***.

والدرجات التي تفصل بين خطوات الحل الأربع تسمى بفئات الاستجابة، وبشكل عام فإن وجود (mi + l) فئة استجابة لمفردة ما (i) يعني أن إتمام حل هذه المفردة يتطلب mi خطوة (زياد الخولي، 2006). ومن المهم أن نوضح أنه يتم ترتيب خطوات الاستجابة بشكل متتابع ابتداءً من الخطوة الأولى ثم الثانية حتى الخطوة الأخيرة، حيث يتوقف إعطاء التلميذ درجة عن المفردة عند أول خطوة خاطئة. وفي نموذج التقدير الجزئي ليس بالضرورة أن تكون صعوبة خطوات الاستجابة مرتبة تصاعدياً كما في نموذج الاستجابة المتدرجة، بل قد تكون صعوبة الخطوة الثانية أعلى من صعوبة الخطوة الثالثة في مفردة ما، وقد يكون العكس في مفردة أخرى.

نموذج التقدير الجزئي يتعامل مع هذا النوع من المفردات متعددة الاستجابة على أساس أن كل خطوة من خطوات المفردة عبارة عن سؤال ثنائي التدريج وتكون الإجابة عنها إما (1) للإجابة الصحيحة، أو (0) للإجابة الخاطئة، ويتعامل مع صعوبة كل خطوة من خطوات الاستجابة للمفردة الواحدة بشكل منفصل مقارنة بالخطوة السابقة لها. واعتماداً على ما سبق فقد قام ماسترز (Masters, 1982) باستنتاج العلاقة الرياضية العامة لنموذج التقدير الجزئي وهي:

$$Pijk = \frac{\exp \sum_{k=o}^{x} (\theta j_n - bik)}{\sum_{n=o}^{mi} \exp \sum_{k=o} (\theta j - bik)}$$

وهذه العلاقة تعبر عن احتمالية حصول الفرد (j) على الدرجة (x) في المفردة (i) كدالة لقدرة الفرد (θ j) وصعوبة الخطوة (mi) على المفردة (i) وحيث إن: bik تمثل صعوبة الخطوة k في المفردة i.

θ j تمثل قدرة الفرد j. (الممتحن أو المفحوص الذي يطبق عليه الاختبار).

وحيث أن موضوع البحث الحالي يهتم بتدريج مفردات بنك الأسئلة، لذا كان من المهم أن نتحدث عن بنوك الأسئلة بشيء من التوضيح في الجزء التالي:

مفهوم بنك الأسئلة Question Bank:

تشير أدبيات القياس إلى وجود مسميات متعددة أطلقت على بنك الأسئلة، منها مستودع المفردات Items Pool، وتجمع المفردات Items Collection، ومكتبة المفردات الاختبارية Test Item Library، ويتعامل ملمان وآرتر (1984) مع جميع هذه المسميات على أنها مترادفة. ويعرفان بنك الأسئلة بأنه مجموعة كبيرة نسبياً من الأسئلة الاختبارية التي يسهل استخدامها، والمقصود بـــ "مجموعة كبيرة نسبياً" أن عدد مفردات البنك يفوق العدد المستخدم في الاختبار بأضعاف كثيرة، أما سهولة الاستخدام فتعني أن المعلومات الخاصة بكل مفردة مفهرسة وموصوفة بدرجة عالية من الدقة، والوضوح، حيث يجب توفر معلومات حول المفردات تتعلق بخصائصها كمحتواها، ونوعها ومصدرها، ومستواها المعرفي، وكذلك معلومات حول خصائصها السيكومترية لمَعْلم الصعوبة ومَعْلم التمييز (*Millman & Arter, 1984*). ويعرف ***(صلاح الدين علام 1986)*** بنك الأسئلة بأنه: مجموعة من مفردات الاختبار يكون لها خصائص سيكومترية مميزة ومعلومة، وهذه المفردات موضوعة في البنك، بحيث يمكن سحب مجموعات منها تناسب أهداف معينة للقياس.

خصائص بنك الأسئلة:

يتميز بنك الأسئلة المصمم باستخدام نماذج الاستجابة للمفردة بالخصائص التالية والتي ذكرها صلاح علام (2005) وهي: (الاقتصاد، المرونة، الاتساق، السرية).

- **خطوات تصميم بنك الأسئلة المصمم باستخدام نموذج التقدير الجزئي:**

تمر مرحلة إعداد بنك الأسئلة باستخدام نموذج التقدير الجزئي بالخطوات التالية:

(أ) كتابة قائمة من المفردات التي تقيس المحتوى نفسه وذلك بإتباع الخطوات التالية:

1- تحديد العناصر الرئيسية المراد قياسها.
2- تحليل العناصر الرئيسية إلى عناصرها الفرعية، ثم صياغتها في صورة أهداف سلوكية.
3- إعداد جدول مواصفات للمفردات (أسئلة الاختبار).
4- صياغة المفردات، وتجميعها في قائمة (اختبار واحد) يسمى الاختبار المرجعي، ويتضمن نوعين من المفردات (الاختيار من متعدد، أسئلة المقال).
5- التحقق من صدق الاختبار المرجعي وثباته (الخصائص السيكومترية للاختبار).

(ب) تدريج قائمة المفردات في الاختبار المرجعي على مقياس واحد مشترك باتباع ما يلي:

1- تطبيق مفردات الاختبار المرجعي على عينة الدراسة.
2- التحقق من فروض نموذج التقدير الجزئي المستخدم في الدراسة.
3- استخدام نموذج التقدير الجزئي في تقدير معالم الصعوبة الفئوية لمفردات الاختبار المرجعي، وتقدير معالم القدرة لجميع أفراد العينة باستخدام وحدة التقدير اللوجيت Logit.
4- استبعاد الأفراد والمفردات غير المطابقين للنموذج.

5- الحصول على بنك الأسئلة بعد تدريج مفردات الاختبار المطابقة لنموذج التقدير الجزئي (زياد الخولي، 2006).

وسوف تتحدث الباحثة في الجزء التالي عن الدراسات السابقة ذات الصلة بمتغيرات البحث الحالي وذلك كالآتي:

الدراسات السابقة ذات الصلة بمتغيرات البحث الحالي:

(1) **دراسة (**Deayala & Bolesta, 1999**)**: وهي دراسة بعنوان أثر نوع توزيع البيانات ودالة المعلومات للمفردة، وعدد فئات المفردات على دقة تقدير معالم نموذج الاستجابة الأسمية".

وقد هدفت إلى دراسة تأثير كل من نوع توزيع البيانات، ودالة المعلومات للمفردة، وعدد فئات المفردات على دقة تقدير معالم نموذج الاستجابة الأسمية (NRM). واستخدمت الدراسة طولاً ثابتًا للاختبار من خلال توليد بيانات لـــــ (28) مفردة متعددة الاستجابة ذات ثلاث فئات، وأربع فئات بالتساوي، كما استخدمت برنامج الحاسوب (MultiloG) والذي يعتمد على طريقة الأرجحية العظمى الهامشية في تقدير معالم الأفراد والمفردات.

وتوصلت نتائج الدراسة إلى أن هناك تأثيراً واضحًا لتقدير معالم المفردات عند تغير العوامل المذكورة، وكانت أفضل النتائج عند استخدام التوزيع الطبيعي للبيانات، وعند عدد الفئات الثلاث.

(2) **دراسة (**Demars, 2003**)**: وهي دراسة بعنوان "أثر عدد المفردات، وعدد فئات المفردات على تقدير معالم المفردات باستخدام نموذج الاستجابة الأسمى".

هدفت الدراسة إلى الكشف عن أثر عدد المفردات، وعدد فئات المفردات على تقدير معالم المفردات. ولتحقيق هدف الدراسة ثم توليد بيانات أربعة اختبارات، حيث تكون الاختبار الأول من (20) مفردة ذات ثلاث فئات، والاختبار الثاني من

(40) مفردة ذات ثلاث فئات، وعدد فئات الاختبار الثالث (20) مفردة ذات ست فئات، وعدد فئات الاختبار الرابع (10) مفردات ذات ست فئات، وذلك لعينات مختلفة مكونة من (600) فرداً و(2400) فرداً واستخدمت الدراسة برنامج الحاسوب (Multilog) الذي يعتمد على طريقة الأرجحية العظمى الهامشية في تقدير معالم المفردات. وتوصلت الدراسة إلى أن:

- زيادة حجم العينة له تأثير قليل منخفض على دقة تقدير معالم المفردات.
- زيادة عدد فئات المفردات يؤدي إلى زيادة أخطاء التباين، وبالتالي يؤدي إلى انخفاض دقة تقدير معالم المفردات.

(3) **دراسة محمد العلى (2004):**

بعنوان "فاعلية نموذج التقدير الجزئي في تقدير قدرة المفحوصين، ومعايرة أسئلة متعددة الخطوات في مادة الكيمياء للصف الثاني الثانوي العلمي".

هدفت الدراسة إلى استقصاء فاعلية نموذج التقدير الجزئي في تقدير قدرة المفحوصين ومعايرة أسئلة متعددة الخطوات في مادة الكيمياء للصف الثاني الثانوي العلمي. وتم تطبيق اختبار تحصيلي في مادة الكيمياء للصف الثاني الثانوي العلمي، مكون من (20) مفردة متعددة الاستجابة، ومختلفة في عدد خطواتها، على (771) من طلبة الصف الثاني الثانوي العلمي، في مديرية عمان الكبرى. استخدمت الدراسة برنامج الحاسوب (Bigsteps) الذي يعتمد طريقة الأرجحية العظمى غير المشروطة في تقدير معالم الأفراد، والمفردات متعددة الخطوات، وتوصلت الدراسة إلى أن:

- درجة صعوبة الأسئلة تزداد بزيادة عدد خطوات الاستجابة وبالتالي يزداد الخطأ المعياري في القياس.

- الكفاية النسبية للاختبار تزداد بزيادة عدد خطوات الاستجابة، وتتساوى الكفاءة النسبية للاختبار الصعب مع الاختبار الكلي عند أصحاب القدرات العالية، بينما تزداد الكفاية النسبية للاختبار السهل عند أصحاب القدرات المتدنية جداً.

(4) دراسة زياد الخولي (2006):

بعنوان "أثر طرق معادلة درجات الاختبار في تدريج مفردات بنك الأسئلة باستخدام نموذج التقدير الجزئي".

هدفت الدراسة إلى معرفة أثر استخدام طرق المعادلة وهي: (طريقة رايت وستون الموسعة على نموذج التقدير الجزئي، وطريقة المتوسط، والانحراف المعياري الموسعة، والطريقة المقترحة) في تدريج مفردات بنك الأسئلة ثنائية الاستجابة ومتعددة الاستجابة باستخدام نموذج التقدير الجزئي.

وتم تطبيق اختبار تحصيلي في الرياضيات مكون من (47) مفردة متعددة الاستجابة، ومختلفة في عدد فئاتها على (1130) من طلاب الصف الأول الثانوي العام، وتم تقسيم مفردات بنك الأسئلة إلى اختبارين إحداهما سهل والآخر صعب لإجراء عمليات المعادلة باستخدام تصميم المفردات المشتركة. واستخدمت الدراسة برنامج الحاسوب (Parscale) الذي يعتمد على طريقة الأرجحية العظمى الهامشية في تقدير معالم الأفراد والمفردات.

- **وتوصلت الدراسة إلى أن:** تدريج مفردات بنك الأسئلة في المقاييس المشتركة تختلف فيما بينها من جهة، وفيما بينها وبين التدريج المرجعي لبنك الأسئلة باختلاف طرق المعادلة الثلاث المستخدمة في هذه الدراسة من جهة أخرى.
- طريقة معادلة رايت وستون (Wright & stone) الموسعة، كانت الطريقة الأكثر ملائمة للدراسة مقارنة بالطرق الأخرى.

(5) دراسة (ye, 2007):

بعنوان "تأثير خصائص مفردات الرباط، وتعدد أبعاد الاختبار في المعادلة الرأسية". هدفت الدراسة إلى بحث تأثير مستوى صعوبة مفردات الربط وبنية الأبعاد على دقة المعادلة الرأسية. ولدراسة أثر المتغيرين في دقة المعادلة الرأسية تم استخدام بيانات محاكاة معتمدة على النموذج ثلاثي المَعْلم، لثلاث صفوف دراسية هي (3، 4، 5) لبيانات الرياضيات مكونة من (40) مفردة، و(10) مفردات رابطة في كل صف دراسي، وطبقت (10) مفردات سهلة و(10) مفردات متوسطة الصعوبة و(10) مفردات صعبة، وافترضت رياضيات الصفين الرابع والخامس مكونة من بعدين، ورياضيات الصف الثالث أحادية البعد، والارتباط بين الأبعاد افترض أنه مساوٍ (80%) للصف الرابع و(60%) للصف الخامس.

واستخدمت الدراسة برنامج الحاسوب (MICROSCALE) الذي يعتمد على طريقة الأرجحية العظمى الهامشية في إجراء عملية تقدير معالم القدرة والمفردة. وتوصلت الدراسة إلى أن المفردات المتوسطة الصعوبة هي الأفضل في إجراء المعادلة الرأسية.

(6) دراسة نسرين إسماعيل (2010):

بعنوان "تأثير بعض المتغيرات على دقة القياس باستخدام نظرية الاستجابة للمفردة". هدفت الدراسة إلى اكتشاف العلاقة بين بعض المتغيرات المتعلقة بنماذج الاستجابة للمفردة والتي تتمثل في (النموذج المستخدم في التدريج، وحجم العينة المستخدمة في التدريج، وعدد المفردات المشتركة، ومستوى صعوبة المفردات المشتركة) وبين دقة القياس.

وتم تطبيق بنك أسئلة في موضوع المعادلات الجبرية من كتاب الجبر للصف الثاني الإعدادي والمكون من (79) مفردة على عينة مكونة من (1045) طالبًا وطالبة أخذت من (8) مدارس تابعة لإدارة مدينة نصر التعليمية بالقاهرة. واستخدمت

الدراسة برنامج الحاسوب (parscale v4.1) المعتمد على طريقة الأرجحية العظمى الهامشية في تقدير معالم الأفراد، والمفردات، وفقًا للنماذج اللوغاريتمية الثلاثة (الأحادي المَعْلم، والثنائي المَعْلم، والثلاثي المَعْلم).

وتوصلت الدراسة إلى أن:

- أفضل دقة قياس عند استخدام النموذج الثلاثي المَعْلم.
- أفضل دقة قياس عندما كان الحجم (13) ضعف عدد مفردات الاختبار وباستخدام النموذج الثلاثي.
- أفضل دقة قياس عندما كانت عدد المفردات المشتركة (9) مفردات، أي بنسبة (23%) من عدد مفردات الاختبارين الذي يتم الربط بينهما، كذلك أوضحت النتائج أن أفضل دقة قياس عندما كان مستوى صعوبة المفردات المشتركة ممثل للاختبار، كما أعطت أفضل النتائج من حيث مؤشرات دقة القياس بعد المفردات الممثلة للاختبار لصالح المفردات المشتركة المتوسطة المستوى، حيث تفوقت على المستويين الآخرين (سهلة وصعبة).

تعقيب الباحثة على الدراسات السابقة:

يتضح من العرض السابق لنتائج الدراسات السابقة ذات الصلة بمتغيرات البحث الحالي الاختلاف بينها وبين الدراسة الحالية في المتغيرات المستقلة والتابعة ومجتمع الدراسة والعينة ومنهج الدراسة وأدوات القياس، وكذلك البرامج الإحصائية المستخدمة في تحليل البيانات، حيث يتضح أن للمفردات المشتركة متوسطة الصعوبة دقة عالية في تدريج المفردات، وهذا ما أشارت إليه دراسات كل من ***(Ye, 2007، نسرين إسماعيل 2010)***.

كذلك أوضحت نتائج دراسات كل من (محمد العلى 2004، Demars, Deayala 1999, 2003) أن زيادة عدد المفردات يؤدي إلى تناقص ملحوظ في دقة تقدير معالم المفردات والأفراد (المفحوصين).

كذلك اتضح للباحثة من الدراسات التي اهتمت ببناء بنوك الأسئلة، والاختبارات المختلفة، أن هناك ندرة في الدراسات التي استخدمت نموذج التقدير الجزئي في بناء بنوك الأسئلة.

لذلك صممت الباحثة بنك الأسئلة الخاص بالدراسة الحالية بالاستفادة من دراسة كل من (زياد الخولي، 2006)، (زياد العبدالله، 2009) حيث اتفقت هذه الدراسات في بناء بنك الأسئلة على استخدام مفردات متعددة الاستجابة، ومختلفة في فئات المفردات، فمنها ما يتكون من فئتين (0، 1) وثلاث فئات (0، 1، 2)، وأربع فئات (0، 1، 2، 3) وخمس فئات (0، 1، 2، 3، 4، 5) وذلك من خلال إجراء المعادلة بين مجموعة من الصور الاختبارية التي تقيس نفس المحتوى باستخدام نموذج التقدير الجزئي، وباستخدام تصميم المفردات المشتركة. وقد اتفق البحث الحالي مع جميع الدراسات السابقة من خلال تصميمها بنك أسئلة باستخدامها نموذج التقدير الجزئي، واستخدامها تصميم المفردات المشتركة.

كما اتفقت مع دراسة (زياد الخولي 2006) باستخدام طريقة التقدير، وهي الأرجحية العظمى الهامشية، والبرنامج المستخدم هو (PARSCALE)، بالإضافة إلى طريقة معادلة رايت وستون الموسعة على نموذج التقدير الجزئي.

وفي ضوء ما سبق يمكن صياغة فروض البحث على النحو التالي:

فروض البحث:

الفرض الأول: لا يختلف تدريج مفردات بنك الأسئلة في التدريج المرجعي عنه في التدريجات الناتجة عن معادلة مفردات اختباري البنك متكافئ1، ومتكافئ2، في تدريج مشترك وذلك باختلاف عدد فئات المفردات المشتركة (فئتين، وأربع فئات، وخمس فئات).

الفرض الثاني: لا يختلف تدريج مفردات بنك الأسئلة في التدريج المرجعي عنه في التدريجات الناتجة عن معادلة مفردات اختباري البنك متكافئ1، ومتكافئ2 في تدريج

مشترك باختلاف مستوى صعوبة المفردات المشتركة (صعبة، ومتوسطة، وسهلة) المستوى.

إجراءات البحث:

أولاً: المنهج:

- استخدمت الباحثة المنهج التجريبي لدراسة أثر المتغيرات المستقلة، والمتمثلة بعدد فئات المفردات المشتركة، ومستوى صعوبتها، على المتغير التابع والمتمثل في تدريج مفردات بنك الأسئلة.

ثانيًا: عينة البحث:

1- **العينة الاستطلاعية:**

- تكونت من (87) طالبًا من خارج عينة البحث الأساسية تم تجريب أداة البحث عليها (الاختبار التحصيلي)، وتم حساب صدق وثبات الاختبار.

2- **العينة الأساسية:**

- تكونت من (1149) طالبًا وطالبة من الصف الثالث الثانوي القسم العلمي في مدارس التعليم العام الحكومي بمحافظة القاهرة والقليوبية، تم اختيارهم بطريقة المعاينة العشوائية العنقودية من (6) مدارس منها (3) للذكور، و(3) للإناث تابعة لمديرية التربية والتعليم بالقاهرة والقليوبية.

ثالثًا: أدوات البحث:

تكونت أداة البحث من اختبار تحصيلي محكى المرجع في الوحدتين الثانية، والثالثة من كتاب الكيمياء للصف الثالث الثانوي، نظراً لما تتمتع به الاختبارات المحكية من مميزات، حيث تستخدم في تقويم التعليم الاتقاني واتخاذ قرارات تشخيصية، كما أن الاختبارات المحكية تعتمد على التعريف الجيد للأهداف المقاسة؛ مما يجعل عملية بناء المفردات أسهل وأكثر صدقًا ***(منار طومان، 2006)*** ولذلك قامت الباحثة بإعداد اختبار تحصيلي محكى المرجع باتباع الخطوات التالية:

1- **تحديد المجال السلوكي الذي يقيسه الاختبار:**

- حيث حددت الباحثة المجال السلوكي المراد قياسه في هذه الدراسة بالوحدتين الثانية والثالثة من كتاب الكيمياء للصف الثالث الثانوي وتتكون الوحدتين من موضوعات رئيسية يتضمنها عدد من الموضوعات الفرعية، والتي يمكن قياسها بمفردات ثنائية الاستجابة، ومفردات متعددة الاستجابة، وهو ما يتطلبه نموذج التقدير الجزئي المستخدم في البحث وذلك كما يلي:

جدول (2)

الموضوعات المتضمنة في الوحدتين الثانية والثالثة

من كتاب الكيمياء للصف الثالث الثانوي

العنوان	الموضوعات الرئيسية	الموضوعات الفرعية
الطاقة الحرارية المصاحبة لتغيرات المادة	صور الطاقة، وتحولاتها	الطاقة الكيميائية – الكيمياء الحرارية – النظام – السعة الحرارية
	التفاعلات الكيميائية، والمعادلة الكيميائية الحرارية	تفاعلات طاردة للحرارة – تفاعلات ماصة للحرارة – المعادلة الكيميائية الحرارية
	أنواع التغيرات الحرارية	تغيرات حرارية فيزيائية – تغيرات حرارية كيميائية
	حساب حرارة التفاعل	باستخدام حرارة التكوين القياسية باستخدام قانون هس
الطاقة الكهربية وتفاعلات الأكسدة والاختزال	الأكسدة والاختزال	مفهوم الأكسدة والاختزال – التأكسد، والاختزال بمفهوم التطرية الإلكترونية – أعداد التأكسد – قواعد حساب أعداد التأكسد
	الخلايا الجلفانية	مفهوم الخلية الجلفانية – جهود الاختزال القطبية – بعض أنواع الخلايا الجلفانية
	خلايا التحليل الكهربية	– مفهوم الخلية الإلكترونية – تطبيقات على التحليل الكهربي – قانوناً فارادي للتحليل الكهربي

2- **تحليل المجال السلوكي إلى مكوناته (العناصر الرئيسية للاختبار):**

- حيث تم عرض هذه العناصر على مجموعة من المحكمين من ذوي الخبرة والاختصاص لإبداء آرائهم، وتم الأخذ في الاعتبار لهذه الآراء وقامت الباحثة بتعديل ما يلزم من تعديل، كما سيتضح في الجزء التالي.

3- **صياغة الأهداف السلوكية المنبثقة عن العناصر الرئيسية:**

- وفي هذه الخطوة تم صياغة هذه العناصر وتحويلها إلى مجموعة من الأهداف السلوكية التي يمكن ملاحظتها ملاحظة مباشرة من خلال سلوك الأفراد، وقياسها إجرائيًا بمفردات اختبارية، وقد تم صياغة (27) هدفًا إجرائيًا، حيث تم مراعاة الأسس العلمية لصياغة الأهداف السلوكية والتي حددها (صلاح الدين علام، 2007). والجدول (3) يوضح أهم العناصر الرئيسية والأهداف السلوكية المشتقة منها:

جدول (3)

العناصر الرئيسية والأهداف السلوكية المشتقة منها

العناصر الرئيسية للمحتوى	م	الأهداف السلوكية
صور الطاقة وتحولاتها	1	يحسب الطالب الحرارة النوعية لمادة
	2	يبين الطالب التفاعلات الماصة للحرارة من التفاعلات الطاردة للحرارة.
المعادلة الكيميائية الحرارية، والتفاعلات الكيميائية	3	يبين الطالب المعادلة الكيميائية الحرارية التي تعبر عن تفاعل معين.
التغيرات الحرارية	4	يحسب الطالب الحرارة الناتجة من بعض أنواع التغيرات الفيزيائية.
	5	يحسب الطالب الحرارة الناتجة من بعض أنواع التغيرات الكيميائية.
حرارة التفاعل	6	يوجد الطالب حرارة التكوين القياسية لمركب ما.
	7	يحسب الطالب حرارة التفاعل باستخدام حرارة التكوين القياسية لتفاعل ما.
الأكسدة والاختزال	8	يذكر الطالب الحالة التي تحدد التغير في المحتوى الحراري للتفاعل.
	9	يحدد الطالب الذرات والأيونات التي يحدث فيها تأكسد واختزال.
	10	يوضح الطالب العامل المؤكسد من العامل المختزل.

	11	يميز الطالب التفاعلات التي يحدث فيها أكسدة واختزال.
	12	يحسب الطالب عدد التأكسد لعنصر ما.
الخلايا الجلفانية	13	يتعرف الطالب على رمز الخلية الجلفانية.
	14	يذكر الطالب أقطاب الخلايا الجلفانية.
	15	يحسب الطالب القوة الدافعة لخلية جلفانية.
	16	يحسب الطالب جهد الاختزال بمعلومية القوة الدافعة للخلية.
	17	يوضح الطالب، إذا كان تفاعل الأكسدة، والاختزال يحدث بشكل تلقائي أم لا.
	18	يتعرف الطالب على تدرج العناصر حسب جهود الاختزال القطبية لها.
	19	يوضح الطالب التفاعلات التي تحدث في بعض أنواع الخلايا الجلفانية.
	20	يذكر الطالب نوع الإلكتروليت في بعض أنواع الخلايا الجلفانية.
	21	يذكر الطالب مميزات بعض أنواع الخلايا الجلفانية.

العناصر الرئيسية للمحتوى	م	الأهداف السلوكية
خلايا التحليل الكهربي	22	يذكر الطالب أقطاب خلايا التحليل الكهربي.
	23	يبين الطالب التفاعل الذي يحدث عند المصعد عند تحضير مركب ما.
	24	يوضح الطالب مكونات خلية التحليل عند طلاء معدن معين.
	25	يحسب الطالب مسائل متعلقة بالقانون الأول لفارادي.

4- مفردات الاختبار:

- بعد الانتهاء من إعداد وتحكيم الأهداف السلوكية، ومواصفات الاختبار بدأ الباحث ببناء مفردات الاختبار، حيث تم وضعها بالاعتماد على الكتاب المدرسي المقرر، والتي تم صياغتها وفقًا للأسس العلمية لصياغة المفردات، ومطابقتها للهدف السلوكي الذي تقيسه من حيث المحتوى، والمستوى المعرفي. كما تم الأخذ في الاعتبار أن لا تعتمد إجابة إحدى المفردات على إجابة مفردة أخرى، والتي أشار إليها كل من (صلاح مراد، وأمين سليمان، 2012، رجاء أبو علام، 1987)، وتم صياغة (36) مفردة اختبارية لقياس (27) هدفًا سلوكيًا، تمثل (13) مفردة منهم المفردات ثنائية الاستجابة من نوع

الاختيار من متعدد، (23) مفردة متعددة الاستجابة (مقالية، وبحيث تحقق هذه المفردات أهداف البحث التي تتطلب مفردات ثنائية الاستجابة ومفردات متعددة الاستجابة، وبفئات استجابة مختلفة. كما هو موضح بالجدول التالي:

جدول (4)

توزيع مفردات الاختبار

المفردات ثنائية الاستجابة	المفردات متعددة الاستجابة		المجموع
الاختيار من متعدد (فئتين)	أربع فئات	خمس فئات	
13 – 1	24 – 14	36 – 25	
13	11	12	36

وقد تم عرض المفردات مع الأهداف السلوكية على مجموعة من المحكمين لاستطلاع آرائهم، كما طلبت الباحثة من كل مُحَكم إبداء الرأي في تعديل أو تغيير، أو حذف أية مفردة تحتاج إلى ذلك.

وقد اهتمت الباحثة بالتقييم الكيفي الذي قدمه المحكمين أكثر من اهتمامها بالنسب المئوية للاتفاق فيما بينهم، إذ أنهم قدموا اقتراحات ساعدت الباحثة في مراجعة عدد من المفردات التي كانت تحتاج إلى الدقة، أو الوضوح في الصياغة، وبهذا أصبحت مفردات الاختبار صالحة للتطبيق.

5- إعداد تعليمات الاختبار ونموذج ورقة الإجابة:

- أعدت الباحثة تعليمات الاختبار، وأرفقتها مع الاختبار، حيث اشتملت على حل مثالين يمثل كل منهما قسم، الأول فيهما يمثل تعليمات إجابة المفردات من نوع الاختيار من متعدد، وكيفية كتابة رمز الإجابة الصحيحة للمفردة في ورقة الإجابة، والثاني يمثل تعليمات إجابة المفردات متعددة الاستجابة (المقالية)، وأعدت الباحثة ورقة الإجابة التي اشتملت على قسمين أيضاً هما: الأول يمثل الإجابة على مفردات الاختيار من متعدد. والثاني يمثل الإجابة على المفردات المقالية.

6- الخصائص السيكومترية للاختبار محكي المرجع:

أ- صدق الاختبار:

- للتحقق من صدق مفردات الاختبار تم عرضه على مجموعة من خبراء المناهج والمتخصصين في المجال من معلمين ومشرفين، وذلك لاستطلاع آرائهم حول هل المفردة تقيس الهدف الذي أعدت لقياسه أم لا؟ وأجمع المحكمون على أن كل مفردة تقيس الهدف الذي أعدت لقياسه. وأن الاختبار صادق في المحتوى الذي يقيسه.

ب- ثبـــات الاختبـــار:

- الاختبار المرجعي يستخدم في تقدير درجات الأفراد أو التلاميذ إلى مجموعتين أو أكثر وفقًا لدرجة تمكنهم أي تصنيفهم إلى متقنين، وغير متقنين، ولحساب ثبات الاختبار قامت الباحثة بالإجراءات التالية:

(1) تطبيق الاختبار المرجعي مرتين على عينة مكونة من (87) طالبًا وذلك بفاصل زمني أسبوعين.

(2) تم استبعاد أوراق إجابة ستة طلاب نظراً لعدم حضورهم في مرتي التطبيق، وبذلك تكون حجم العينة الاستطلاعية (81) طالبًا.

(3) تصنيف الطلاب تبعًا لدرجات القطع التالية (50%، 80%) إلى مجموعتين هما مجموعة المتقنين، ومجموعة غير المتقنين.

(4) إيجاد قيمة معامل كابا لسوامينثان وهامبلتون وإلجانيا والذي يُعَّد من المعاملات المهمة في حساب معامل الثبات وتقدير اتساق تصنيف الطلاب، وذلك وفقًا للمعادلة التالية:

$$K = \frac{Lo - Le}{1 - Le}$$

حيث إن: Lo ترمز إلى نسبة الاتفاق الملاحظ في التصنيفات.

Le ترمز إلى نسبة الاتفاق المتوقع في التصنيفات.

- وقد بلغت قيمة معامل الثبات بالنسبة لدرجة القطع (50%) (98) وبالنسبة لدرجة القطع (80%) 99 وهي نسبة مرتفعة ومقبولة.

إجـراءات البحـث:

بعد التحقق من الخصائص السيكومترية للاختبار (صدق الاختبار وثباته)، والذي تم من خلال بيانات العينة الاستطلاعية التي بلغت (87) طالباً، وبعد الانتهاء من إعداد الصورة النهائية للاختبار، قامت الباحثة بالإجراءات التالية:

1- زيارة المدارس التي تم اختيارها، وذلك للاتفاق على مديري المدارس ومدرسي مادة الكيمياء حول إجراءات تطبيق الاختبار وتحديد موعده، وزيارة طلاب الصف الثالث في غرفهم الدراسية؛ لإطلاعهم بموعد الاختبار، ومحتواه.
2- تجهيز كراسات الاختبار النهائي وورقة التعليمات على عدد الطلاب في كل مدرسة.
3- تم تطبيق الاختبار المكون من (36) مفردة اختبارية على عينة الدراسة المكونة من (1149) طالباً وطالبة في الفصل الدراسي الثاني من العام الدراسي (2023 – 2024م) ابتداءً من تاريخ 2023/3/26م وحتى تاريخ 2023/12/31م أي بواقع يوم في كل مدرسة.
4- أشرفت الباحثة على عملية تطبيق الاختبار، وإيضاح تعليماته بنفسها، حيث تم تطبيق الاختبار في جميع المدارس ابتداءً من الحصة الأولى، وحتى الحصة الثالثة، وقد استعانت بمدرسي الحصص الأولى أثناء تأدية الاختبار، بالإضافة إلى عدد من المدرسين لملاحظة سير عملية الاختبار والذي سبق لهم المعرفة الكافية بإجراءات الاختبار، ومدى أهمية التقيد بها بالنسبة للدراسة. وكون الاختبار محكى المرجع ويطبق عليه نموذج التقدير الجزئي أحد نماذج الاستجابة للمفردة، والتي تتطلب التحرر من عامل السرعة، فإن الباحثة أتاحت زمناً كافياً لكل طالب للإجابة عن مفردات الاختبار، حيث استغرق زمن تطبيق الاختبار (120) دقيقة.
5- تم تصحيح أوراق الإجابة وفق نموذج الإجابة المعد مسبقًا من قِبل الباحث، حيث تعطي المفردة الدرجة (صفر) للإجابة الخاطئة، والدرجة (1) للإجابة الصحيحة، أي

إكمال الخطوة الأولى بنجاح، والدرجة (2) للإجابة الصحيحة عند إكمال الخطوة الثانية بنجاح، والدرجة (3) للإجابة الصحيحة عند إكمال الخطوة الثالثة بنجاح، والدرجة (4) للإجابة الصحيحة عند إكمال الخطوة الرابعة بنجاح، وبحيث يتوقف التصحيح عند أول خطوة خاطئة، وبذلك تكون درجة الطالب مساوية لمجموع الخطوات الصحيحة والمتتابعة والتي أكملها الطالب بنجاح.

6- تم إدخال درجات الطالب على ذاكرة الكمبيوتر من خلال برنامج SPSS، والذي يتكون من (36) عموداً رأسياً يمثل المفردات، و(1056) صفًا أفقياً تمثل الأفراد، بعد استبعاد (96) طالبًا، حيث حصل (11) طالبًا منهم على الدرجة النهائية للاختبار، وبالتالي يكون الاختبار أقل من مستوى تحصيلهم التعليمي، و(82) طالبًا حصلوا على الدرجة (صفر) وبالتالي يكون الاختبار أعلى من مستوى تحصيلهم التعليمي، كما أنه لم يتم استبعاد أي مفردة؛ حيث لا توجد أي مفردة أجاب عنها جميع الطلاب وبالمثل لا توجد أي مفردة لم يجب عنها أحد.

(أ) إجراءات اختبار فروض البحث المتعلقة بعدد فئات المفردات المشتركة:

لاختبار صحة فروض البحث المتعلق بعدد فئات المفردات المشتركة اتبعت الباحثة الإجراءات التالية:

1- اشتقاق ثلاث مجموعات من المفردات المشتركة حسب عدد فئات الاستجابة من بنك الأسئلة في التدريج المرجعي على النحو التالي:

- (10) مفردات ذات فئتين وهي: (1، 2، 3، 4، 5، 8، 9، 11، 12، 13).
- (11) مفردة ذات أربع فئات وهي: (14، 15، 16، 17، 18، 19، 20، 21، 22، 23، 24).
- (12) مفردة ذات خمس فئات وهي: (25، 26، 27، 28، 29، 30، 31، 32، 33، 34، 35، 36).

2- تقسيم مفردات بنك الأسئلة في التدريج المرجعي المتبقية عن كل مجموعة من مجموعات المفردات المشتركة السابقة إلى اختبارين متكافئين، حيث يكون المتوسط

الحسابي، والانحراف المعياري لهما متقارب جداً، وذلك كما أوصت دراسة منار طومان (2006) بأنه كلما اقتربت متوسطات صعوبة مفردات الاختبارات المعادلة من بعضها البعض، أدى ذلك إلى الحصول على نتائج معادلة أفضل. وقد سمى الباحث الاختبار الأول مع المفردات المشتركة بالاختبار متكافئ1، والاختبار الثاني مع المفردات المشتركة بالاختبار متكافئ2، وذلك كما هو موضح بالجدول (5).

جدول (5)

توزيع المفردات إلى اختبارين متكافئ1، ومتكافئ2

حسب عدد فئات المفردات المشتركة

	مفردات مشتركة ذات فئتين		مفردات مشتركة ذات أربع فئات		مفردات مشتركة ذات خمس فئات	
الاختبار	المفردات	المفردات المشتركة	المفردات	المفردات المشتركة	المفردات	المفردات المشتركة
متكافئ1	17-19 21-23 24-25 26-27 29-31 33-36	1-2-3 4-5-8	2-4-5 8-13 25-26 28-30 31-33	14-15 16-17 18-19	1-3-5 8-11 14-17 20-21 22-24	25-26 27-28 29-30
متكافئ2	14-15 16-18 20-22 28-30 32-34 35	9-11 12-13	1-3-9 11-12 27-29 32-34 35-36	20-21 22-23 24	2-4-9 12-13 15-16 18-19 23	31-32 33-34 35-36
المجموع	33		33		33	

3- تطبيق الاختبارين متكافئ1، ومتكافئ2، وذلك حسب كل فئة من فئات الاستجابة الثلاث (فئتين، وأربع فئات، وخمس فئات) على عينتين تم اختيارهما عشوائيًا من العينة الكلية، حجم كل منها (500) طالبًا وطالبة.

4- استخدام برنامج PARSCALE في تحليل بيانات كل اختبار من الاختبارات متكافئ1، ومتكافئ2، حسب كل فئة من الفئات الثلاث للحصول على تدريج الصعوبة لكل اختبار على حده من خلال نموذج التقدير الجزئي، وذلك تمهيداً لإجراء التدريجات التالية:

(ب) إجراءات اختبار فروض البحث المتعلقة بمستوى صعوبة المفردات المشتركة:

لاختبار صحة فروض البحث المتعلقة بمستوى صعوبة المفردات المشتركة اتبعت الباحثة الإجراءات التالية:

1- تقسيم مفردات بنك الأسئلة في التدريج المرجعي الموضح بالجدول (5) إلى ثلاث مجموعات متساوية كل مجموعة تتكون من: (11) مفردة تمثل مستوى من مستويات الصعوبة على النحو الآتي:

- (11) مفردة صعبة المستوى وهي: (1، 4، 5، 11، 12، 15، 17، 19، 22، 23، 25).
- (11) مفردة متوسطة المستوى وهي: (9، 14، 16، 18، 24، 27، 28، 30، 33، 34، 36).
- (11) مفردة سهلة المستوى وهي: (2، 3، 8، 13، 20، 21، 26، 29، 31، 32، 35).

2- تقسيم مفردات بنك الأسئلة في التدريج المرجعي المتبقية عن كل مجموعة من مجموعات المفردات المشتركة السابقة إلى اختبارين متكافئين، بحيث يكون المتوسط الحسابي، والانحراف المعياري للاختبارين متقاربين جداً. حيث استخدم محمد الشافعي (1996) الاختبارات المتكافئة في دراسة متغير مستوى صعوبة المفردات المشتركة باستخدام نموذج راش. وقد سمي الباحث الاختبار الأول مع المفردات

المشتركة بالاختبار متكافئ1، والاختبار الثاني مع المفردات المشتركة بالاختبار متكافئ2، وذلك كما هو موضح بالجدول (6).

جدول (6)

توزيع المفردات إلى اختبارين متكافئ1، ومتكافئ2

حسب مستوى صعوبة المفردات المشتركة

	مفردات مشتركة صعبة المستوى		مفردات مشتركة متوسطة المستوى		مفردات مشتركة سهلة المستوى	
الاختبار	المفردات	المفردات المشتركة	المفردات	المفردات المشتركة	المفردات	المفردات المشتركة
متكافئ1	3-8-9 16-18 20-21 26-27 30-31	1-4-5 11-12 15-17 19-23 25	2-4-5 13-15 17-20 21-22 29-35	9-14 16-18 24-27 28-30 33-34 36	4-11 15-16 17-18 23-24 25-27 28	2-3-8 13-20 21-26 29-31 32-35
متكافئ2	2-13 14-24 28-29 32-33 34-35 36		1-3-8 11-12 19-23 25-26 31-32		1-5-9 10-14 12-22 30-33 34-36	
المجموع	33		33		33	

3- تطبيق الاختبارين متكافئ1، ومتكافئ2، وذلك حسب كل مستوى من مستويات صعوبة المفردات المشتركة (صعبة، ومتوسطة، وسهلة) المستوى على عينتين تم اختيارهما عشوائيًا من العينة الكلية، حجم كل منها 500 طالبًا وطالبة.

4- استخدام برنامج PARSCALE في تحليل بيانات كل اختبار من الاختبارات متكافئ1، ومتكافئ2 للحصول على تدريج صعوبة كل اختبار على حده من خلال نموذج التقدير الجزئي؛ وذلك تمهيداً لإجراء التدريجات الآتية:

أولاً: مناقشة وتفسير نتائج الدراسة المتعلقة بعدد فئات المفردات المشتركة:

يناقش الباحث في هذا الجزء نتائج اختبار فروض الدراسة المتعلقة بعدد فئات المفردات المشتركة، وذلك تبعاً للمحكات المستخدمة. كما يقوم بتفسير نتائج الدراسة حسب نتائج الدراسات، والبحوث السابقة.

لخصت الباحثة نتائج اختبار الفرضين الأول والثاني والمتعلقين بعدد فئات المفردات المشتركة في الجدول الآتي:

جدول (7)

نتائج الإجابة عن الأسئلة المتعلقة بعدد فئات المفردات المشتركة

التدريج	Q	قيمة (ت)	الدلالة الإحصائية	الدلالة اللفظية
المرجعي – والفئتين	0.010	2.71	0.018	دال
المرجعي – والأربع فئات	0.008	1.112	0.275	غير دال
المرجعي – والخمس فئات	0.006	0.393	0.697	غير دال
الفئتين – والأربع فئات	0.013	2.609	0.014	دال
الفئتين – والخمس فئات	0.011	1.929	0.063	غير دال
الأربع فئات – والخمس فئات	0.009	1.198	0.240	غير دال

- **مناقشة نتائج الفرض الأول:** لا يختلف تدريج مفردات بنك الأسئلة في التدريج المرجعي عنه في التدريجات الناتجة عن معادلة مفردات اختباري البنك متكافئ1، ومتكافئ2 في تدريج مشترك وذلك باختلاف عدد فئات المفردات المشتركة (فئتين، وأربع فئات، وخمس فئات).

من الجدول (7) يبين المحك الإحصائي (ت) من خلال قيمة الدلالة الإحصائية أنه:

(1) يوجد اختلاف ذو دلالة إحصائية بين التدريجين المرجعي، والمشترك الناتج عن استخدام مفردات مشتركة ذات فئتين؛ وذلك لأن الدلالة الإحصائية تساوي 0.018، وهي أقل من مستوى الدلالة 0.05، لكن الدلالة العملية لها متوسطة.

(2) لا يوجد اختلاف ذو دلالة إحصائية بين التدريجين المرجعي، والمشترك الناتج عن استخدام مفردات مشتركة ذات أربع فئات؛ وذلك لأن الدلالة الإحصائية تساوي 0.275، وهي أكبر من مستوى الدلالة 0.05.

(3) لا يوجد اختلاف ذو دلالة إحصائية بين التدريجين المرجعي، والمشترك الناتج عن استخدام مفردات مشتركة ذات خمس فئات؛ وذلك لأن الدلالة الإحصائية تساوي 0.697، وهي أكبر من مستوى الدلالة 0.05.

ويعزي الباحث النتائج السابقة إلى أن المفردات ثنائية الاستجابة ذات الفئتين تكون الاستجابة عنها ذات خيار واحد، وهو إما أن تكون صحيحة أو خاطئة فقط، ولا تسمح للفرد بالتعبير عن معرفته الجزئية للمفردات، وبالتالي تكون استجابة الفرد على المفردات لا تعبر عن قدرته الحقيقية؛ مما يؤد يإلى انخفاض دقة تقدير المعالم، وبالتالي يحدث اختلاف بين التدريج المرجعي، والتدريج المشترك عند استخدام مفردات مشتركة ذات فئتين. بينما زيادة عدد فئات المفردات المشتركة تسمح للفرد بالتعبير عن معرفته الجزئية للمفردات، وبالتالي تكون استجابة الفرد على المفردات أكثر واقعية، وتعبر عن قدرته الحقيقية؛ مما يؤدي إلى زيادة دقة تقدير المعالم، وبالتالي يقترب التدريج المشترك من التدريج المرجعي. وهذا ما يؤكده المحك الإحصائي Q بأن أدق التدريجات كانت بين التدريجين المرجعي، والمشترك الناتج عن استخدام مفردات مشتركة ذات خمس فئات، يليه التدريجين المرجعي، والمشترك الناتج عن استخدام مفردات مشتركة ذات أربع فئات. وهذه النتيجة تتفق مع دراسة محمد شاهين (1993) التي تؤكد أن زيادة عدد فئات المفردات (خطوات الاستجابة) تؤدي إلى انخفاض الأخطاء المعيارية، وزيادة سهولة الأسئلة، وبالتالي تزداد دقة تقدير معالم المفردات، والأفراد. كما تختلف نتائج هذه الدراسة مع دراسة (DeMars, 2003, Choi, Cook & Dodd, 1996, DeAyala

(Bolesta, 1999 & ودراسة محمد العلي (2004) التي تؤكد أنه كلما زادت عدد فئات المفردات، فإن هذا يؤدي إلى تناقص ملحوظ في دقة تقدير معالم المفردات والأفراد.

ثانيًا: مناقشة نتائج الدراسة المتعلقة بمستوى صعوبة المفردات المشتركة:

لخصت الباحثة نتائج اختبار الفرض الثاني والمتعلق بمستوى صعوبة المفردات المشتركة في الجدول (8) التالي:

جدول (8)

نتائج الإجابة عن الأسئلة المتعلقة بمستوى صعوبة المفردات المشتركة

التدريج	Q	قيمة (ت)	الدلالة الإحصائية	الدلالة اللفظية
المرجعي – والصعبة	0.008	0.619	0.56	غير دال
المرجعي – والمتوسطة	0.006	1.054	0.30	غير دال
المرجعي – والسهلة	0.0079	1.210	0.235	غير دال
الصعبة – والمتوسطة	0.009	1.240	0.224	غير دال
الصعبة – والسهلة	0.011	1.335	0.191	غير دال
المتوسطة – والسهلة	0.010	0.298	0.768	غير دال

- **مناقشة نتائج الفرض الثاني**: لا يختلف تدريج مفردات بنك الأسئلة في التدريج المرجعي عنه في التدريجات الناتجة عن معادلة مفردات اختباري هذا البنك متكافئ1، ومتكافئ2 في تدريج مشترك باختلاف مستوى صعوبة المفردات المشتركة (صعبة، ومتوسطة، وسهلة).

من الجدول (8) يتبين من المحك الإحصائي (ت) من خلال قيمة الدلالة الإحصائية أنه:

- لا يوجد اختلاف ذو دلالة إحصائية بين التدريج المرجعي، والتدريجات المشتركة الثلاثة الناتجة عن مستوى صعوبة المفردات المشتركة (سهلة، ومتوسطة، وصعبة). وتتفق الدراسة الحالية في هذه النتيجة مع دراسة محمد الشافعي (1996)، ودراسة

نسرين إسماعيل (2010)، من حيث استخدام ثلاث مستويات لصعوبة المفردات المشتركة (صعبة، ومتوسطة، وسهلة) لضم اختبارين متكافئين في تدريج واحد، وبينت النتائج عدم وجود فروق جوهرية عند استخدام أي مستوى للمفردات المشتركة (صعبة، ومتوسطة، وسهلة) على الربط بين الاختبارين، واعتبار المفردات متوسطة الصعوبة أفضل من (السهلة، والصعبة) في إجراء عملية الربط. وهو ما وجده الباحث من خلال المحك Q الذي أعطى أقل قيمة مع التدريجين المرجعي، والمشترك الناتج عن استخدام المفردات المشتركة متوسطة المستوى عند عملية المعادلة.

ثالثًا: التطبيقات التربوية للدراسة:

1- إنشاء وحدة لبنوك الأسئلة تابعة لإدارة الاختبارات في وزارة التربية والتعليم، وتكوين مجموعات من متخصصي المواد الدراسية، ثم تدريب هذه المجموعات على طريقة بناء المفردات الاختبارية الجيدة، وطريقة تجريبها، وتنقيحها، وإضافتها إلى بنك الأسئلة، حيث تسهم هذه البنوك في تحقيق الدقة، والموضوعية في عملية القياس والتقويم التربوي.

2- الاعتماد في بناء بنوك الأسئلة على جميع أنواع المفردات ثنائية الاستجابة، ومتعددة الاستجابة حتى تناسب قياس كل الأهداف التربوية.

3- توفير الحاسبات الآلية، والبرامج الإحصائية المختلفة، والمتخصصة في بنوك الأسئلة، وتدريب كوادر بشرية متعددة عليها، وذلك للاستعانة بهم في تدريج وتخزين المفردات واستدعاء المفردات المطلوبة.

4- تدريب المعلمين على استخدام اختبارات متعددة من بنك الأسئلة تناسب الأهداف التي يريدون قياسها.

5- استخدام نماذج الاستجابة متعددة التدريج في تطوير، وتدريج اختبارات القدرات العقلية.

رابعًا: أبحاث مقترحة:

1- إعادة الدراسة الحالية باستخدام نموذج الاستجابة المتدرجة ومقارنة نتائجها مع نتائج الدراسة الحالية التي استخدمت نموذج التقدير الجزئي.

2- إعادة الدراسة الحالية باستخدام نموذج التقدير الجزئي المعمم، ومقارنة نتائجها مع نتائج الدراسة الحالية التي استخدمت نموذج التقدير الجزئي.

3- إعادة الدراسة الحالية باستخدام بيانات مولدة ثم مقارنة نتائجها مع الدراسة الحالية، وذلك لمعرفة تأثير نوع البيانات على دقة تدريج مفردات بنك الأسئلة.

إعادة الدراسة الحالية باستخدام اختبارات مختلفة الصعوبة عند إجراء المعادلة، ومقارنة نتائجها مع نتائج الدراسة الحالية؛ لمعرفة تأثير صعوبة الاختبارات المعادلة على تدريج مفردات بنك الأسئلة.

مراجـع البحـث

أولاً: المراجع العربية:

- أحمد محمد التقي. (2009). **النظرية الحديثة في القياس**. عمان: دار المسيرة.
- اعتدال غازي عبدالله. (2003). استخدام نموذج راش في تدريج مقياس القدرة العقلية لدراسة بعض العوامل المؤثرة على دقة القياس. **رسالة دكتوراه:** كلية البنات. جامعة عين شمس.
- المعلم في الكيمياء الصف الأول الثانوي. (2011). إعداد نخبة من متخصصي الكيمياء المؤسسة العربية الحديثة – القاهرة.
- أمينة محمد كاظم. (1988أ). دراسة نظرية نقدية حول القياس الموضوعي للسلوك (نموذج راش). **مطبوعات جامعة الكويت.**
- أنور الشرقاوي، وسليمان الخضري، وأمينة كاظم، ونادية عبدالسلام. (1996). **اتجاهات معاصرة في القياس والتقويم النفسي والتربوي**. القاهرة: الأنجلو المصرية.
- بندر بن حمدان الزهراني. (2008). أثر اختلاف حجم العينة واتساع مدى القدرة على دقة الدرجة الحقيقة المقدرة باستخدام النظرية التقليدية والنماذج أحادية البعد في النظرية الحديثة في القياس. **رسالة دكتوراه:** جامعة أم القرى، السعودية.

- حمزة محمد دودين. (2004). استخدام نظرية الاستجابة على الفقرة في تحليل الاستجابات على اختبار المهارات الكمية لدى طالب الجامعة. **المجلة التربوية:** 18، (721).
- راشد حماد الدوسري. (2001). معادلة الاختبارات مفهومها، وطرقها، ومشكلات تطبيقها. **مجلة العلوم التربوية والنفسية.** مجلة جامعة البحرين: (2) 4، 141–106.

رجاء محمود أبو علام. (1987). **قياس وتقويم التحصيل الدراسي.** الكويت: دار القلم

- رجاء محمود أبو علام. (2009). **مناهج البحث في العلوم النفسية والتربوية.** ط6. القاهرة: دار النشر للجامعات.
- رضا سمير عمر. (2008). بعض المتغيرات المؤثرة على دقة كشف الدالة المميزة للمفردة في ضوء نظرية الاستجابة للمفردة. **رسالة دكتوراه:** معهد الدراسات التربوية، جامعة القاهرة.
- زياد أحمد العبدالله. (2009). أثر حجم العينة وعدد المفردات المشتركة على تدريج مفردات بنك الأسئلة باستخدام نموذج التقدير الجزئي. **رسالة ماجستير:** معهد الدراسات التربوية، جامعة القاهرة.
- زياد عبدالحسيب الخولي. (2006). أثر طرق معادلة درجات الاختبار في تدريج مفردات بنك الأسئلة باستخدام نموذج التقدير الجزئي. **رسالة دكتوراه:** معهد الدراسات التربوية، جامعة القاهرة.
- شفاء بنت عبدالله بالخيور. (2009). فاعلية طرق معادلة نماذج اختبار القدرات العامة بالمركز الوطني للقياس والتقويم وفق نظريتي القياس التقليدية والحديثة في ضوء بعض المتغيرات. **رسالة دكتوراه:** جامعة أم القرى.
- صابر حكيم. (2011). **الامتحان في الكيمياء أسئلة ومسائل ونماذج وامتحانات دورية.** الدولية للطباعة والنشر، القاهرة.
- صبحي حمدان أبو جلالة. (1999). اتجاهات معاصرة في التقويم التربوي وبناء الاختبارات وبنوك الأسئلة. الكويت: مكتبة الفلاح للنشر.

- صلاح الدين محمود علام. (1986). تطورات معاصرة في القياس النفسي والتربوي. الكويت: جامعة الكويت، إدارة التأليف والترجمة والنشر.
- صلاح الدين محمود علام. (2000). القياس والتقويم التربوي والنفسي **أساسياته وتطبيقاته وتوجهاته المعاصر**. القاهرة: دار الفكر العربي.
- صلاح الدين محمود علام. (2005). **نماذج الاستجابة للمفردة أحادية البعد ومتعددة الأبعاد وتطبيقاتها في القياس النفسي والتربوي.** القاهرة: دار الفكر العربي.
- صلاح الدين محمود علام. (2007). **الاختبارات التشخيصية محكية المراجع في المجالات التربوية والنفسية والتدريبية.** القاهرة: دار الفكر العربي.
- صلاح مراد وأمين سليمان. (2012). **الاختبارات والمقاييس في العلوم التربوية والنفسية – خطوات إعدادها وخصائصها.** القاهرة: دار الكتاب الحديث.
- عادل سعد خضر. (2007). **بنوك الأسئلة بين النظرية والتطبيق.** القاهرة: دار السحاب.
- عبدالله زيد الكيلاني ونضال كمال الشريفين. (2007). **مدخل إلى البحث في العلوم التربوية والاجتماعية: أساسياته – مناهجه – تصاميمه – أساليبه الإحصائية.** عمان: دار المسيرة.
- عماد غصاب عبابنة. (2009). **الاختبارات المرجعية – فسلفتها وأسس تطويرها.** عمان: دار المسيرة.
- محمد محمود العلي. (2004). فاعلية نموذج التقدير الجزئي في بناء أسئلة من فقرات متعددة الخطوات في مادة الكيمياء في الصف الثاني الثانوي العلمي. **رسالة دكتوراه:** جامعة عمان العربية.
- محمد محمود شاهين. (1993). أثر عدد خطوات تصحيح الأسئلة المقالية في تقدير قدرات الأفراد وقيم صعوبة الأسئلة باستخدام نموذج التقدير الجزئي. **رسالة دكتوراه:** الجامعة الأردنية.
- محمد منصور الشافعي. (1996). أثر طرق معادلة درجات الاختبار وضوابط اختيار العينة على تدريج بنود بنك الأسئلة باستخدام نموذج راش. **رسالة دكتوراه:** الجامعة الأردنية.

- محمد منصور الشافعي. (1996). أثر طرق معادلة درجات الاختبار وضوابط اختيار العينة على تدريج بنود بنك الأسئلة باستخدام نموذج راش. **رسالة دكتوراه:** كلية التربية، جامعة المنصورة.

محمد منصور الشافعي. (2008). تأثير انتهاك افتراضي أحادية البعد واستقلالية المحل في تدرج بنك الأسئلة ودقة معادلة درجات الاختبارات البنكية المسحوبة. **بحث منشور**.

- مصطفى أحمد طيفور. (2007). دراسة مقارنة لنماذج نظرية الاستجابة للمفردة في معادلة درجات الاختبارات. **رسالة دكتوراه:** معهد الدراسات التربوية، جامعة القاهرة.
- منار أحمد طومان. (2006). فاعلية استخدام نظرية الاستجابة للمفردة في عملية معادلة درجات الاختبارات متعددة الأبعاد والمتغيرات المؤثرة فيها. **رسالة دكتوراه:** معهد الدراسات التربوية، جامعة القاهرة.
- نسرين إسماعيل السيد. (2010). تأثير بعض المتغيرات على دقة القياس باستخدام نظرية الاستجابة للمفردة. **رسالة دكتوراه:** كلية البنات، جامعة عين شمس.

ثانيًا: المراجع الأجنبية:

Angoff, W.H. (1979). **Summary and Derivation of Equatinh Methods Used at Est**. in Holland & Rubin (Eds) Test Equating.

Baker, F.B. (1987). Methodology Review: Item Parameter Estimation Under the One-and Two-and Three- Parameter Logistic Models. **Applied Psychological Measurement**, Vol. 11, No. 2, pp. 111-141.

Bludworth, j. (2007). A Structural Investigation and Revision of the Inventory of Interpersonal Problems Using Item Response Theory. Dissertation Presented In Partial Fulfillment of the Requirements for the **Degree of Doctor of Phillosophy. Arezona UNIVERSITY**.

Bock, R.D. (1972). Estimating Item Parameters and latent Ability when Responses are Scored in two or more Nominal Categories. **Psychometrika**. Vol. 37, No. 1, pp. 29-51.

Bock, R.D., & Aitkin, M. (1981). Marginal Maximum likelihood Estimation of Item Parameters: Application of an EM Algorithm. Psychometrika, Vol. 46, pp. 433-459.

Candace, L. & Macken, R. (2008). A Comparison of Multi Stage and Computerized Adaptive Tests Based on The Generalized Partial Credit Model. **Doctor of Philosophy, University of Texas at Austin August**.

Chang, S.T & Bashaw, W.L. (1984). Characteristics of Anchor tests. Paper Presented at the Annul Meeting of the American Education Research Association. (68th, New Orleans, LA, April 23-27).

Chen, H., Cui, Z., Zhu, R. & Gao, X. (2010). **Evaluating the Effects of Differences in Group Abilities on the Tucker and the Levine Observed-Score Methods of Common-Item Nonequivalent Groups Equating**. ACT Research Report Series.

Choi, S.W., Cook, K.F., & Dodd, B.G. (1996). Parameter Recovery for the Partial Credit Model using MULTILOG. Paper presented at the Annual Meeting of the National Council on Measurement in Education, New York.

Cook, L.L. & Eignor, D.R. (1991). An NCME Instructional Module on IRT Equating Methods. **Educational Measurement:** Issues and Practice, 10, 37-45.

Crocker, I. & Algina, J. (1986). **Introduction to Classical and Modern Test Theory**. Orlando, Harcourt Brace Jovanovich, Inc.

Davier, M.V. & Yamamoto, K. (2004). Partially Observed Mixtures of IRT Models: An Extension of the Generalized Partial Credit Model. **Applied Psychological Measurement**, Vol. 28, No. 6, pp. 389-406.

De Ayala, R.J. (1989). A Comparison of the Nominal Response Model and the Three Parameter Logistic Model in Computerized Adaptive Testing. **Educational and Psychological Measurement**, Vol. 49, No, pp. 789-805.

De Ayala, R.J. & Bolesta, M.S. (1999). Item Parameter Recovery for the Nominal Response Model. Paper Presented at the Annul Meeting of the American Education Research Association, San Francisco. CA.

De Champlain, f. (1996). The Effects of Multidimensionality on IRT True-Score Equating for Subgroups of Examinees. **Journal of Educational Measurement**, Vol. 33, No. 2, pp. 181-201.

DeMars, C. (2003). Sample Size and the Recovery of Nominal Response Model Item Parameters. **Applied Psychological Measurement**. Vol. 27, No. 4, pp. 275-288.

DeMars, C. (2010). **Item Response Theory**. Published by Oxford University Press, Inc. New York.

Dodd, B.G., Kock, W.B. & De Ayala, R. J. (1989). Operational Characteristics. **Applied Psychological Measurement,** Vol. 13, No. 2, PP. 129-143.

Dorans, N. J. & Holland, P. W. (2000). Population Invariance and Equitability of Tests: Basic Theory and the Linear Case. **Journal of Education measurement,** Vol. 27, No. 3, pp. 281-306.

Embretsone, S. E. & Reise, S. P. (2000). Item Response Theory for Psychologists. Mahwah, NJ: Erlbaum.

Ferrando, P. J., Lorenzo, U. & Molina, G. (2001). An Item Response Theory Analysis of Response Stability in Personality Measurement. **Applied Psychological Measurement**, Vol. 25, No. 1. Pp. 3-17.

Ferrare, S. (1989). Constructing an Essay Prompt Bank Using the Partial Credit Model. Papar Presented at the Annual Meeting of the National Council on Measurement in Educational, Sanfranciso, available online: **www.eric.ed.gov. ED306249**.

Flowers, C.P., Oshima, T.C. & Raju, N.S. (1999). A Descriptior Demonstration of Polytomous- DFIT Framework. **Applied Psychological Measurement**, Vol. 23, No. 4, pp. 309-326.

Goldstien, H. & Wood, R. (1989). Five Decades of Item Response Modeling, Britsh. **Journal of Mathematical and Statistical Psychology**, Vol. 42, No. 2, pp. 139-167.

Hambleton, R.K. (1989). **Principles and Selected Applications of Item Response Theory**. In R. L, Linn (Ed), Educational Measurement third edition. New York, NY: American council on Education and Macmillan Publishing Company.

Hambleton, R.K. & Swaminathan, H. (1985). **Item Response Theory**: **Principles and Applications.** Kluwer-Nijh off Publishing, Boston. U.S.A.

Hambleton, R.K., Swaminathan, H. & Rogers, H.J. (1991). **Fundamentals of Item Response Theory**. Sage Publications, Inc., Teller Road, Newburypark, California, U.S.A.

Heinen, A.G. (1996). **Latent Class and Discrete Latent Trait Models**: Similarities and Differences. London, Sage.

Jurich, D. & Goodman, J. (2009). A Comparison of IRT Parameter Recovery In Mixed Format Examinations Using Parscale and ICL, Poster presented at the Annual meeting of Northeastern Educational Research Association, October.

Kline, P. (1993). **The Handbook of Psychological Testing**. Published London Routledge.

Kolen, M. (1988). Traditional Equating Methodology: Issues and Practices. **Educational Measurement,** Vol. 7, No. 4, pp. 29-36.

Kolen, M.J. & Brennan, R.L. (2004). **Test Equating Scaling and Linking Methods and Practices.** (2nd ed.). New York: Springer- Verlag.

Lee, G., Kolen, M., Frisbie, D. & Ankenmann, R. (2001). Comparison of Dichotomous and Polytomous Item Response Models in Equating Scores From Tests Coposed of Testlets. **Applied Psychological Measurement**, Vol. 25, No. 4, pp. 357-372.

Linacre, J.M. (2008). **Winsteps Rasch Measurement Computer Program**. Chicago. Winsteps.Com.

Lord, F.M. (1980). **Applications of Item Response Theory to Practical Testing Problems**. Hillsdale, NJ: Lawrence Erlbaum **Associates**.

Lord, F.M. & Stocking, M.L. (1990). **Item Response Theory**. Walberg, H.J., & Haertel, G.D. (Eds). The International in Cyclopedia of Educational Evaluation, Perman **Press LTD**. 281-284.

Macdonald, P. (2002). Computer Adaptive Test for Measuring Personality Factors Using Item Response Theory. **Submitted in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy. The University of Western Ontario**.

Masters, G.N. (1982). A rasch Model for Partial Credit Scoring. **Psychometrika**, Vol. 47, No. 2, pp. 149-174.

Masters, G.N. (1984a). Constructing an Item Bank Using Partial Credit Scoring. **Journal of Educational Measurement**, Vol. 21, No. 1, pp. 19-32.

Masters, G.G. & Evans, J. (1986). Banking Non-Dichotomously Scored Items. Applied Psychological Measurement, Vol. 10, No. 4, PP. 355-367.

Michaelides, M. (2006). **Effects of Misbehaving Common Item on Aggregate Scores an Application of the Mental-Haenszel Statistic in Test Equating**. Center for the Study of Evaluation (CSE). Los Angeles, CA.

Millman, J. & Arter, J.A. (1984). Issues in Item Banking. **Journal of Educational Measurement**, Vol. 21, No. 4, pp. 315-330.

Mislevy, R.J. & Stocking, M.L. (1989). A consumer's Guide to LOGIST and BILOG. **Applied Psychological Measurement,** Vol. 13, No. 1, pp. 57-75.

Muraki, E. (1992). A Generalized Partial Credit Model Application of an EM Algorithm. **Applied Psychological Measurement,** Vol. 16, No. 2, pp. 159-176.

Petersen, N., Kolen, M., & Hoover, H. (1989). **Scaling, Norming and Equating**, pp. 221-262. In Linn, R. L., ed. Educational Measurement, 3nd ed., New York: MacMillan.

Popham, W.J. (1990). **Modern Educational Measurement A practitioner's Perspective**. New Jersy: Prentice-Hall.

Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. **Psvchometrika Monograph Supplement 34 (4, Whole No. 17)**.

Suen, H.K. (1990). **Principles of Test the Ories.** Hillsdale, NJ: Lawrence Erlbaum.

Urbina, S. & Anastasi, A. (1997). **Psychological Testing**. (7th ed). New Gersey, Prentice Hall.

Van der Linden. W. & Hambleton, R. K. (1997). **Handbook of Modern Item Response Theory**. New York, Springer – Verlag,, Inc.

Van der Linden. W. & Wiberg, M. (2010) Local Observed-Score Equating With Anchor-Test Designs. **Applied Psychological Measurement,** Vol. 34, No. 8, pp. 620-640.

Walker – Bartnick, L. A. (1990). **An Investigation of Factors Affecting Invariance of Item Parameter Estimates for the Partial Credit Model**.

Wei, B. & Hua, M. (2010). Impact of Non-Representative Anchor Items on Scale Stability. Paper Presented at the Annual Meeting of the National Council on Measurement in Education.

Weiss, L. & Yoes. (1994) Item Response Theory, In Hambleton, R. K,. & Zaal. J. N. (EDS) **Advances in Educational and Psychological Testing: Theory and Applicayions.** Khrwer Academic Publishers, Boston.

Woldbeck, T. (1998). Basic Concepts in Modern Methods on Equating. Paper Presented at the Annual Meeting of Sout Psychological Association (New Orleans, LA, April).

Wright, B.D. & Liancre, J.m. (1992). **Bigsteps Rasch Analysis Computer Program.** Chicago: MESA Press.

Wright, B.D. & Stone, M.H. (1979). **Best Test Design**. Chicago: MESA Press.

Ye, T. (2007). Effects of Anchor Item Properties and Dimensionality of Test on Vertical Scaling. Papar Presented At The Annual Meeting Of The National Council On Measurement In Education. Chicago.

Yi, C. (2008). Mixed Format Test Equating: Effects of Test Dimensionality and Common Item Sets. **Doctor of Philosophy**.

Yu, F. & Nandakumar, R. (2001). Poly-Detect For Quantifying the Degree of Multidimensionality of Item Response Data. In **Journal of Educational Measurement,** Vol. 38, No. 2, pp. 99-120.

Zickar, M. J. (1998). Modeling Item – Level Data with Item Response Theory. **American Psychological Society,** Vol. 7, No. 4, pp. 104-110.