حوليات آداب عين شمس المجلد ٤٦ (عدد يناير – مارس ٢٠١٨)
http://www.aafu.journals.ekb.eg
(دورية علمية محكمة)
جامعة عين شمس
كلية الآداب

معادلة اختبارات الذكاء المتقدمة باستعمال الدرجة الحقيقية في النظرية الكلاسيكية

شيماء صلاح العبيدي *

جامعة بغداد / كلية التربية للبنات

المستخلص

يرمي هذا البحث إلى معادلة اختبارات الذكاء المتقدمة الاربعة لواضعها فليب كارتر(٢٠٠٨) باستعمال الدرجة الحقيقية في النظرية الكلاسيكية ونظرية السمات الكامنة والمفاضلة فيما بين طرائق المعادلة، ولتحقيق ذلك طُبِّقَتْ أختبارين على عينة من طلبة المرحلة الإعدادية بواقع (٥٦٠) طالبا وطالبة ، وقد فاضل هذا البحث بين ثلاث من طرائق المعادلة في النظرية التقليدية وهي :(المتوسطات، الخطية، المئينيات) وخمس من طرق المعادلة في نظرية السمات الكامنة على وفق النموذج ثنائي المعلم (2PL) والثلاثي المعلم (3PL) ، وقد استخدم الباحث البرامج الإحصائية الآتية على الترتيب في معالجة البيانات وتحليلها وهي(Excel)،(SPSS17)،(Winsteps)،(Bilog–MG3)،(Xcalibre,4.1.6.0)،(ST)،(CPIE)،(PIE).

وبناءً على نتائج التحليلات الإحصائية ، تم التوصل إلى الاستنتاجات الآتية:

١.تناسب الأختباران والمستوى العقليّ لطلبة المرحلة الأعداديةفي بغداد، وتجانس مفرداتها، واعتدال التوزيع التكراريّ لدرجات المستجيبين عليها.

٢.تتعادل الأختباران بدرجة كبيرة ، وقد امكن من واقعها إعداد جداول تحويل الدرجات على الاختبار الأول إلى الدرجات المعادلة على الاختبارات الثلاث الاخرى، وذلك على وفق كلٍّ طريقة من طرائق المعادلة.

© جميع حقوق الطبع والنشر محفوظة لحولية كلية الآداب - جامعة عين شمس ٢٠١٧.

مشكلة البحث :

إنَّ لخطورة القرارات المترتبة على البيانات المأخوذة من اختبارات الذكاء، فقد حظيت باهتمام عددٍ من العلماء وعنايتهم ممّا أدى إلى ظهور الاهتمام في هذه الاختبارات وتطويرها ، إذ نشطت حركة القياس العقليّ في عددٍ من الدول العربية ومنها العراق ، ولكن نلحظ أنَّ المجتمع العراقي مازال بحاجة إلى تطوير كثيرٍ من الاختبارات ، لاسيما اختبارات الذكاء الحديثة ، والتي تعتمد على أكثر من صورة للاختبار الواحد ، والتي قد تكون هذه الاختبارات متوازية (Parallel tests)، أو غير متوازية (Non parallel tests)، وحتى نستطيع مقارنة نتائج الأفراد في هذه الاختبارات المختلفة ، واتخاذ القرارات ذات العلاقة ، فلابُدَّ من إجراء عملية معادلة يتم بموجبها وضع اقتران تناظري بين نتائج الاختبارات بحيث يتم في هذا الاقتران ربط كلّ نتيجة في الاختبار الواحد بما يعادلها في الاختبار الآخر.

فمعادلة تحويل نظام الوحدات في إحدى صور الاختبار إلى نظام الوحدات في الصور الأخرى للاختبار، حتى تكون الدرجات المشتقة من صور الاختبار بعد عملية التحويل متكافئة تماماً (Kolen , 1981: 4) ، وبذلك تحد الصور المتكافئة للاختبارات من مشكلة تطبيق الاختبار أكثر من مرة على المفحوص الواحد بوساطة استخدام أساليب إحصائية لمعادلة درجات صور تلك الاختبارات (Kolen & Whitney, 1982: 280) .

وتعتمد عملية معادلة الاختبارات على درجات الأفراد ، والدرجة التي نحصل عليها من تطبيق الاختبارات يطلق عليها الدرجة الخام (Obtained Score) ، وهذه الدرجة خليط من الدرجة الحقيقية (True Score)، وخطأ القياس (Measurement Error)، وبطرح درجة خطأ القياس من الدرجة الخام نحصل على الدرجة الحقيقية ، أي أنَّ الدرجة الحقيقية = الدرجة الخام – خطأ القياس ، وقد يكون خطأ القياس موجباً، أو سالباً، أو صفراً، ومن ثم فإنَّ الدرجة الخام في أي اختبار، أو قياس عقليّ قد تكون أكبر من، أو أقل من الدرجة الحقيقية، أومساوية لها (Anthony, 1983: 38).

وإنَّ أدبيات القياس النفسيّ تميز بين مدخلين رئيسين في تحليل بيانات الاختبارات ومعادلتها وحساب درجاتها الحقيقية ، وهذان المدخلان هما المدخل الكلاسيكيّ للاختبارات (Classical Test Theory) وما ينطوي عليه من مفاهيم ومبادئ ، ومدخل نظرية السمات الكامنة (Latent Tests Theory) ، التي يفضل بعض علماء القياس تسميتها نظرية الاستجابة للمفردة (Item Response Theory) لاهتمامها بالربط بين استجابة الفرد على مفردة اختبار ما، وخصائص هذه المفردة (علام ،١٩٩٠ : ١٢٩).

ولكلٍّ من المدخلين طرائقه الخاصة في معادلة الاختبارات باستعمال قدرة الفرد الحقيقية من الدرجات الخام (Obtain Score)، وهذه الطرائق تختلف في المدخلين الكلاسيكيّ أو التقليديّ ، والمعاصر أو نظرية السمات الكامنة، وممّا سبق تنبع مشكلة البحث في الإجابة عن الأسئلة الآتية: (ما صورتي اختبارات الذكاء لفليب كارتر وكن راسل (٢٠٠٧) على وفق النظرية الكلاسيكية في القياس ؟ ما المؤشرات الإحصائية

العامة لصور اختبارات الذكاء المتقدمة الأربعة على وفق النظرية التقليدية في القياس ؟ إلى أيٍّ حدٍّ تتعادل صور اختبارات الذكاء المتقدمة الأربعة لفليب كارتر (٢٠٠٧) على وفق طرائق المعادلة وباستعمال الدرجة الحقيقية للأفراد في نظريتي القياس التقليديّ والحديث ؟) .

أهمية البحث:

على الرغم من أنَّ موضوع معادلة الاختبارات قد بحثه المتخصصون في القياس والتقويم التربويّ والنفسيّ بنحوٍ مستفيضٍ بوصفه جزءاً من مهماتهم البحثية والمهنية إلاّ أنَّ عدد من المتخصصين في القياس العقليّ قد تجاهلوا هذا الموضوع المهم فخلال القرن العشرين سنة المنصرمة بدأ الاعتراف بأهمية معادلة الاختبارات من قطاع كبير من المتخصصين والمسؤولين عن بناء الاختبارات (Kolen & Brennan , 2004 :9) .

ويعزو أيلا Ayala (2009) الاهتمام بموضوع معادلة الاختبارات إلى ثلاثة تطورات مهمة خلال السنين الماضية ؛ التطور الأول : يتعلق بزيادة عدد برامج الاختبارات ذات الصور المتعددة وتنوعها ، فضلاً عن إدراك المسؤولين والمتخصصين عن بناء هذه الاختبارات والبرامج المستقاة من استمارات متعددة للاختبار الواحد، أولاختبارات متعددة تقيس الموضوع نفسه ، أهمية معادلة درجات الاختبارات المتوازية بعد الوصول إلى حقيقة صعوبة تحقيق تكافوء كامل للاختبارات التي تقيس الموضوع نفسه؛ لأنَّه عملية المعادلة هي خطوة أكثر واقعة وقابلة للتطبيق في ميدان القياس ، والتطورالثاني: هو تأكيد المتخصصين من مطوري الاختبارات في مجال القياس النفسيّ أهمية معادلة الاختبارات عند تحليل درجاتها ، وذلك لقدرة المعادلة على معالجة كثير من القضايا الفنية التي يثيرها النقاد في مجال القياس التي تظهر على الاختبارات بعد عملية بناء الصور المتكافئة للاختبارات، واستخدامها في عملية تصنيف الأفراد ، أما التطور الثالث : فهو قضية قدرة الاختبارات على تحقيق العدل في قياس ما يراد قياسه لمجموعات من الأفراد المتباينين في الجنس والأصل العرقيّ والثقافة والتحرر من آثارها، ودور عملية المعادلة في اشتقاق درجات من اختبارات طبقت في ظروف مختلفة (Ayala , 2009:310 - 311).

إنْ نلخص أهمية معادلة الاختبارات بأننا نحتاج إلى معادلة الاختبارات عندما يكون لدينا أكثر من صورة للاختبار الواحد ، ويراد تطبيقها على عدد من الأفراد ، ومقارنة أدائهم عليها بما لا يخل بالعدالة بينهم ، وهنا إذا ما تمت معادلة هذه الصور أمكننا تطبيقها على أفراد مختلفين ، فضلاً عن التعامل مع درجات التطبيق الناتجة ، كما لو كان الحكم على أداء الأفراد في التطبيق لاختبار واحد ، كما نحتاج إلى معادلة الاختبارات في بعض الحالات التي لا نستطيع فيها إعادة تطبيق الاختبار نفسه مرة أخرى أو مرات عدة ، بسبب إفشاء سرية الاختبار، أو التدرب عليه في المرات التالية ، وغير ذلك من آثار التطبيق ، لذلك نطبق أكثر من صورة للاختبار بما يضمن تكافؤها ومقارنة الأداء. نحتاج إلى معادلة الاختبارات؛ لتحقيق العدالة بين الأفراد ، إذا ما طبق عليهم صورعدة للاختبار الواحد، لاسيما في المواقف الحياتية التي تحتاج إلى إصدار أحكام بشان المنافسة بين الأفراد وتصنيفهم على أساس أدائهم ، وتعدُّ معادلة الاختبارات ضرورية عند الحاجة إلى الربط بين درجات إختبار ما بدرجات اختبار آخر، وهذا ما نحتاج إليه عند بناء اختبارات متكافئة ، إذ يستلزم الأمر أحدث تعادل بين مفردات الاختبارات، ووضع هذه المفردات الاختبارية على تدريج واحد مشترك ، ومن ثم يمكن أنْ نحدد الدرجة على الاختبار الواحد

بدلالة الدرجة على الاختبار الآخر وبالعكس ، أي أنَّ معادلة الاختبار هنا نحتاج إليها في بناء اختبارات متعادلة ، كما نحتاج إلى معادلة الاختبارات كثيرٍ من مواقف القياس والتقويم التربويّ والنفسيّ، مثل : تجنب تحيز المفردة الاختبارية ، وفي بناء بنوك الأسئلة ، وفي بناء الاختبارات المحبوكة بالحاسوب ، وبتطبيق صور متكافئة ومتعددة للاختبار الواحد ، كذلك نحتاج إلى معادلة الاختبارات في الحالات التي يراد فيها قياس التقدم، أو النمو في قدرة ما من القدرات عبر مُددٍ زمنية متتابعة ، إذ تكون لدينا عدة اختبارات ذات مستويات صعوبة متفاوتة ومتدرجة ، بهدف قياس القدرة عند هذا المستوى ومقارنته بما كانت عليه في المستوى السابق ، وما ستصير إليه في المستوى الآتي ، وذلك باعادة التطبيق عبر هذه المُدد الزمنية .

ويمكن أنْ نضيف بأنَّ مفهوم المعادلة للاختبارات يُعدَّ من المفاهيم الحديثة المهمة في مجال القياس والتقويم، التي لم يسلط الضوء عليها بنحو كافٍ ومازالت هناك جوانب لم تستوفِ البحث ولم تصل إلى إجابات وافية عن تساؤلات البحث العلميّ في هذا المجال، لاسيما تلك المتعلقة في مجال طرائق معادلة الاختبارات .

وبذلك تتجلى أهمية البحث الحالي في مجال معادلة الاختبارات من خلال أهمية مفهوم المعادلة فضلاً عن معرفة أي طرائق المعادلة أكثر دقة واستقراراً من خلال عملية المقارنة بين بعض طرائق المعادلة على وفق النظرية التقليدية و نظرية السمات الكامنة باستعمال الدرجات الحقيقة ، ومن هذه النقطة وممّا سبق تتجلى أهمية الجانب الأول لهذا البحث وهو مفهوم معادلة الاختبارات .

كلما تقدمت أساليب القياس ودقة التقدير الكميّ في أي علم من العلوم كان ذلك مظهراً من مظاهر رقي ذلك العلم وتقدمه وارتفاع مكانته بين العلوم الأخرى ، لذلك تُعدُّ عملية تكميم الظواهر من خصائص المنهج العلميّ (بركات ،١٩٨٣: ١٣).

وأنَّ أهداف العلم، هي : (الفهم ، والتنبؤ، والتحكم في ظاهرة ما)، ومعناه اكتشاف العلاقة الوظيفية بينها وبين غيرها من الظواهر. إنَّ معرفة شكل العلاقة وتقديرها تقديراً كمياً دقيقاً يستتبعه دقة في التنبؤ بحدوث الظاهرة، ومن ثم التحكم الدقيق في إحداثها. وقد كان التقدم الهائل في العلوم الطبيعية يرجع بدرجة كبيرة إلى اكتشاف الطرائق والوسائل التي تساعد على فهم الظواهر الطبيعية فهماً كمياً ، بلغ من دقته أنْ أوصل الإنسان إلى اكتشافات كانت إلى عهد قريب محض خيال (كاظم، ١٩٨١: ٣٧).

وتُعدُّ الأرقام التي أوجدها الإنسان أسلوباً للتعبير، وتصورالأشياء، وأبعاد المكان والزمان من خصائص المنهج العلميّ لدراسة الظواهر في أي علم من العلوم ، فالعلم يستحيل من دون استعمال شبكة معقدة ومستقرة من القياسات : carver ,1974) (12 .

في الوقت الذي نجد فيه أنَّ القياس الماديّ من الدقة والوضوح، وتوافر الوحدة والأداة أو الطريقة لقياس كلّ منها ، نجد أنَّ هناك مشكلة في مجال القياس النفسيّ ، إذ يكون أكثر تعقيداً من القياس الماديّ ، وإنَّ موضوع القياس في العلوم الإنسانية هو الإنسان ، وبذلك فهو يتأثر بالعوامل الداخلية والخارجية ، وتتمثل العوامل الخارجية بالبيئة التي يعيش فيها ويتفاعل معها ،أما العوامل الداخلية، فتتمثل بالعوامل البيولوجية والنفسية ، وعلى وفق طبيعة الظواهر والمتغيرات النفسية ، وإنَّ القياس النفسي قياسٌ نسبيٌّ وليس مطلق ، أي أنَّ الظاهرة لا تنعدم فيه ، لذا فالصفر يُعدُّ افتراضياً وليس حقيقياً، أي

أنَّ الصفر لا يعني انعدام الظاهرة تماماً (Nunnally & Bernstein, 1994 :12) .

إنَّ تلك الظواهر هي تكوينات افتراضية غير ملموسة ، ويصعب تحديدها بوضوح ودقة ، لذا إنَّ قياسها يكون غير مباشر، ومن خلال ما تُظهره من سلوكيات دالة عليها ، وأنْ يكون لعينة من الظاهرة، وليس لكامل الظاهرة (Aiken , 1988: 15-16).

وقد تركزت جهود العاملين في مجال القياس النفسيّ بالدرجة الأساس على أهمية التقليل والحد من تلك المشكلات من خلال إعداد وسائل قياسية تقل فيها نسبة الخطأ (Anastasi & Urbina, 1997:10).

وعلى الرغم من الجهود الكبيرة التي بذلها المتخصصون في مجال القياس النفسيّ والعقليّ وتركيزهم في الجانب الإحصائي للتقليل من نسبة الخطأ والوصول قدر الإمكان بالقياس النفسيّ إلى مستوى القياس الماديّ في قياس بعض المفاهيم النفسية والعقلية، إلاّ أنَّه مازالت هناك كثير من المعوقات التي تحول من دون تحقيق ذلك المستوى للقياس النفسيّ والعقليّ (Brennan, 1997 :19) .

ويُعدُّ مفهوم الذكاء من المفاهيم التي واجهت عملية قياسه كثيراً من المشكلات، فضلاً عن أنّه من أكثر المفاهيم السيكولوجية التي يدور حولها النقاش ، وامتد إلى خارج النطاق الأكاديميّ ، ووصل إلى المسؤولين عن اتخاذ القرار في المجالات المختلفة ، و يرجع السبب في ذلك إلى أنَّ الذكاء من المفاهيم المهمة التي تتنوع حولها وجهات النظر والآراء ، والتي لم يتفق على ماهيتها بنحوٍ محدد ، ممّا أدى إلى تنوع النظريات التي تناولت مفهوم الذكاء الإنسانيّ، وانعكس ذلك على تنوع أساليب القياس لمفهوم الذكاء الإنسانيّ (Anastasi & Urbina , 1997:271) .

ونظراً لتعدد نظريات الذكاء فقد تنوعت الاختبارات التي تقيس الذكاء ، وقد صنفت تلك الاختبارات تبعاً إلى طريقة إجراءها إلى اختبارات فردية ، وأخرى جماعية في حين صنفت بناء على محتواها إلى اختبارات لغوية وأدائية ، واختبارات شبه أدائية في حين صنفها البعض الآخر إلى اختبارات متحيزة ثقافياً ، وأخرى متحررة من أثر الثقافة (Osterlind,1998:162).

ممّا يؤدي إلى أنَّ هذا التنوع في النظريات والاختبارات فرض الحاجة المتزايدة إلى بذل المزيد من الجهود للارتقاء بحركة بناء المزيد من الأدوات وتطويرها التي تتوافق كلٌّ منها مع الخلفيات النظرية التي بنية عليها ، وتحقيق الأهداف الأساسية لعملية القياس، وتتجاوز المشكلات التي يفرضها واقع القياس النفسيّ والعقليّ .

ومحاولة للتغلب على المشكلات التي واجهت طبيعة القياس للظواهر النفسية بصورة عامة، والذكاء بصورة خاصة انطلقت مع بدء الدعوة؛ لتحقيق الموضوعية في القياس النفسيّ، كما هو الحال في القياس الماديّ الدعوات للافادة من التقدم التكنولوجيّ في مجال الحاسب الآليّ في التوصل إلى وسائل حديثة في القياس .

فأهمية البحث الحالي في هذا الجانب من أهمية اختبارات الذكاء وعملية تطويرها، وأنَّ أهميتها لا تكمن في إضافة رقم جديد إلى الأرقام المتوافرة من اختبارات لقياس الذكاء بقدر ما تكمن في تحسين نوعية هذه الاختبارات المتمثلة باختبارات الذكاء المتقدمة لمؤلفها فليب كارتر وكن راسل (٢٠٠٧).

وتُعدُّ الدرجة الحقيقة مؤشر لمدى دقة الاختبار في تقدير القدرة، أو الصفة التي يقيسها

الاختبار فكلما اقتربت ، الدرجة الحقيقة لقدرة الفرد من القيمة المقدرة في الاختبار زادت الثقة في هذا الاختبار وفي نتائجه وتصل ، هذه الثقة إلى ذروتها عندما تعبر الدرجة الخام للفرد عن قدرته الحقيقية ، أي عندما س = ت وتصبح درجة خطأ (خ) في هذه الحالة مساوية للصفر (سعيد ، ٢٠٠٠: ٥٤) .

وبذلك فإنَّ القياس النفسيّ إلى تحقيق الدقة، والموضوعية، وتقليل خطأ القياس Error of measurement ، الذي يرجع إلى أسباب عدة ، منها ما يتصل بالفاحص ، أو بظروف الموقف الذي تتم فيه عملية القياس ، أو ترجع إلى طبيعة أداة القياس ذاتها ، أو قد ترجع إلى المفحوص الذي يقوم بتقدير القدرة، أو الصفة لدى الأفراد (عمر وآخرون، ٢٠١٠ : ٧٩) .

وبذلك تظهر أهمية الدرجة الحقيقية بوصفها مؤشراً لمدى دقة عملية القياس، ولقدرة الأداة للسيطرة على مصادر أخطاء القياس، والتي تُعدُّ السيطرة عليها خطوة مهمة لتحقيق الدقة والموضوعية، ويشير Lehmnn & Mehrens (2003) إلى أنَّ أهمية الدرجة الحقيقية تكمن بأنَّها درجة ثابتة لا تتغير، وأنَّ تباين الدرجة الحقيقية يساوي صفراً ، أي أنَّ الدرجة الحقيقة للفرد لا تتغير إذا ما أعدنا الاختبار على الفرد نفسه لعدة مرات ، ولكن يكون هناك شيئاً من عدم التوافق (خطأ) للدرجة الملاحظة لذا إنَّ الدرجة الملاحظة للفرد ستتباين وستكون إما أكبر من الدرجة الحقيقية أو اصغر ، وبذلك فإنَّه يفترض بدرجات الأفراد الملاحظة أنْ تقع ضمن توزيع اعتيادي حول درجاتهم الحقيقية(Lehmnn & mehrens,2003:338) .

ومن هنا يتضح أهمية اعتماد الدرجة الحقيقية للأفراد بدلاً من درجاتهم الملاحظة، وذلك لما تتصف به من ثبات، وما تعبر عنه من دقة وموضوعية من خلال التخلص من خطأ القياس، أو الحد من آثاره على درجات الأفراد الملاحظة، ويصبح في ضوء ذلك الهدف الأساس للقياس هو الوصول إلى الغاية النهائية، وهي الدرجة الحقيقية التي يتمتع بها كلّ فرد في أية عملية قياس إلى أية صفة، أو خاصية .

وممّا يضيف أهمية جديدة إلى أهمية هذا البحث هو أنَّ الدرجة الحقيقية من المفاهيم التي أخذت حيز من النقاش، ودار حولها كثيرٌ من الآراء ، وما اختلاف النظريات في مجال القياس النفسيّ إلاّ دليل على تنوع وتعدد تلك الآراء في كيفية الوصول إلى التقدير الحقيقيّ لدرجة الفرد في اختبار ما ، وعدم الوصول إلى نقطة القطع في صحة إحدى النظريات على بقية النظريات الأخرى ، ممّا يثري أهمية هذا البحث في تناوله للدرجة الحقيقة؛ وطرائق حسابها في النظرية التقليدية، ونظرية السمات الكامنة في تحديد الدرجة الحقيقية للأفراد وأيهما أكثر دقة واستقراراً في التقدير لها .

ثالثاً: أهداف البحث يرمي البحث إلى:

١- إعداد اختبارات الذكاء المتقدمة للبيئة العراقية على وفق النظرية التقليدية ، واستخراج الخصائص السيكومترية للاختبارات .
٢- حساب الدرجات الحقيقية للأفراد على وفق النظرية التقليدية في القياس.
٣- معادلة اختبارات الذكاء المتقدمة باستعمال الدرجات الحقيقية للأفراد على وفق هذه النظرية الكلاسيكية.

رابعا: حدود البحث : يتكون مجتمع البحث الحالي من طلبة المرحلة الإعدادية (الرابع ، والخامس ، والسادس) بفرعيها العلمي والأدبي في المديريات العامة للتربية بمحافظة بغداد (الكرخ / الرصافة) كافة للعام الدراسي ٢٠١٥ /٢٠١٦.

خامساً: تحديد المصطلحات : يشمل البحث الحالي على المصطلحات الآتية:-

معادلة الاختبارات عرَّفها كلٌّ من :

ليفنجستون Livingston (2004) : الدرجات التي تحصل عليها مجموعة من الأفراد على أحد صور الاختبار، والدرجات التي تحصل عليها المجموعة نفسها على صورة أخرى من الاختبار تمثل تقديرات القدرة نفسها داخل المجموعة (Livingston, 2004, 12).

ستوكنج Stocking (1997) :عملية إيجاد درجات متناظرة للصيغ المختلفة من الاختبار (Stocking,1997:18).

كولن وبرانان Kolen & Brennan (1995) : إجراء لإزالة الآثار التي يتركها الفرق بين متوسط مستويات الصعوبة لفقرات الاختبار، وكذالك الفروق بين متوسط مستويات التمييز لفقرات الاختبار، وذلك بين صورتين من الاختبار نفسه لجعلهما متكافئتين (Kolen & Brennan, 2004: 5).

دورانز Dorans (1990): عملية إجراء تعديل إحصائيّ على درجات صورة واحدة من الاختبار لجعل تلك الدرجات مكافئة بطريقة ما لدرجات صورة أخرى من الاختبار نفسه (Dorans, 1990, 3).

الاختبار Test عرّفه كلٌّ من :

عودة (١٩٩٨) : طريقة لتقدير درجة امتلاك الفرد لسمة معينة من خلال إجابات الفرد عن عينة من المثيرات التي تمثل السمة (عودة، ٢٠١٠: ٣٦).

انستازي ويوربينا Ana stasi & Urbina (1997) : قياس موضوعي ومقنن لعينة من السلوك (Anastasi & Urbina, 1997:4).

نانلي و بيرنستن Nunnally & Bernstein (1994) :استخدام الأعداد بحيث تدلُّ على الأشياء بأسلوب يشير إلى كميات من الخصائص (Nunnally & Bernstein, 1994:3) .

ومن خلال التعريفات السابقة jحدد الباحثة تعريف الاختبار، (وسيلة منظمة تستعمل لقياس درجة امتلاك الفرد للذكاء باستعمال عينة من الفقرات التي تمثل النطاق الشامل للذكاء العام على وفق نظرية سبيرمان للذكاء) .**ويعرف إجرائياً** (بأنَّه مجموعة من الأسئلة من نوع الاختيار من متعدد تقيس الذكاء لدى طلبة جامعات بغداد الحكومية)

الذكاء Intelligence عرَّفه كلٌّ من :

كارتر Carter (2008): قدرة عامة على استيعاب واستنباط المعلومات، وتحليلها بدقة وسرعة، واستعماله على نحو صحيح ، في المواقف التي تتطلب اكتشافات فكرية في الجوانب العددية والمكانية واللغوية والتفكير المجرد(Carter, 2008:6).

ستيرنبرغ Sternberg (2002):القدرة على التعلم من الخبرة، وضبط عمليات التفكير(Sternberg& Williams,2002:16)

ويتبنى الباحثان تعريف فليب كارتر للذكاء تعريفاً نظرياً للبحث، لأنَّه الباحثة تبنت اختباري كارتر وراسل Carter & Rassell (2008)، والتي بنيت على أساس

نظرية الذكاء العام ، وعرف إجرائيا بالدرجة التي يحصل عليها الطالب من الإجابة الصحيحة عن فقرات كلٍّ اختبار من اختباري كارتر وراسل للذكاء العام .

الدرجة الحقيقية عرَّفها كلٌّ من :

غنيم (١٩٩١): تقدير حقيقيّ لما لدى الفرد من قدرة معينة، أو صفة محددة(غنيم ،١٩٩١: ٢٨)

كروكر والجاينا (1986) : هو مفهوم إحصائي يعتمد على القيمة المتوقعة التي يحصل عليها الفرد من عملية قياس معينة ، وتمثل متوسط درجات الفرد الملاحظة في عدد من الاختبارات المتكافئة ، أو من مواقف اختبارية متكررة للاختبار نفسه من تكرار الاختبار نفسه بأوقات مختلفة .(Crocker & Algina, 1986: 153).

أما التعريف النظريّ فيتبنى الباحثان التعريف السابق تعريفاً نظرياً للدرجة الحقيقية ، والتعريف الاجرائيّ : (متوسط الدرجات الملاحظة للمستجيب والمعادلة باستخدام طريقة المتوسطات وبأسلوب التدوير لكلّ إختبار من اختبارات الذكاء الأربعة).

النظرية الكلاسيكية : واحدة من نظريات القياس التي تستخدم بغرض تحديد العوامل التي تؤثر في الدرجة التي يحصل عليها الفرد في الاختبار ، وترتكز هذه النظرية على مفهوم الدرجة الحقيقية والدرجة الخطأ ، الذي يفترض أنَّه لو أمكن أن نجرى الاختبارمرات عدة على الفرد بعناصر جديدة وتحت ظروف مختلفة ، فإننا نحصل على درجات ملحوظة مختلفة متوسطها هو أقرب تقدير غير متحيز لقدرة الفرد أو درجته الحقيقية (Randall,1998:4)

خلفية نظرية ودراسات سابقة

نظريات التكوين العقلي :إن ظهور نظريات متعددة حاولت تفسير النشاط العقلي للإنسان، وتحديد ماهيته وحقيقته ، وذلك لأن كل عالم له رأيه الخاص ، أو نظرية خاصة في تفسير النشاط العقلي وتحديد جوانبه ، وابتدأت هذه النظريات بالاعتماد على التفكير والتامل في النشاط العقلي لتنتهي بنظريات اعتمدت على التجربة، ومن هذه النظريات:-

١- (نظرية الملكات Faculty Theory): تعد نظرية الملكات من أقدم النظريات التي حاولت تفسير طبيعة النشاط العقلي التي ازدهرت في القرن التاسع عشر ، إذ ترى أن العقل مكون من ملكات (قدرات) مختلفة، مثل (التذكر، والسببية، والتمييز، والتصور ..الخ)، وهذه الملكات، أو القدرات كل منها مستقل عن الآخر ، ويمكن تطويرها من خلال التدريب الشديد، وحل المسائل الصعبة، غير ان علماء النفس التجريبيين أثبتوا عدم وجود ملكات منفصلة بعضها عن بعض في الدماغ يمكن تدريبها بشكل منفصل (Nunnally , 1970 : 101) .

٢- النظريات العاملية: التي اتبعت أسلوب التحليل العاملي الإحصائي في تطوير نظرياتها، ومنها: النظرية أحادية العامل (One Factor Theory) التي تصف القدرة العقلية (الذكاء) على انها قابلية عامة مفردة ، وانتقلت هذه النظرية من علم الاحياء إلى علم النفس التي ظهرت معالمها بوضوح في المعادلات المبكرة لبناء اختبارات الذكاء من أمثال مقياسي بينيه ويكسلر(Gronland , 1976 : 258). ونظرية العاملين لسبيرمان (The Two Factor Theory).التي تعد أول نظرية طورت على أسس

التحليل الإحصائي لدرجات الاختبارات وبنيت في شكلها الاساسي على ان النشاطات الفكرية جميعها تشترك في عامل وحيد مشترك يسمى العامل العام (General Factor) (S) يختص كل منها بنشاط فكري وحيد ، وينتسب كل ارتباط ايجابي بين أي مهتمين إلى العامل العام، وكلما زاد تشبع المهمتين بالعامل (G)ازدادت درجة الارتباط بينهما ، وأدى وجود العوامل الخاصة إلى تقليل الارتباط بينهما ، وعلى الرغم من وجود نوعين من العوامل المفترضة" العامة والخاصة" في هذه النظرية ، ان العامل العام هو الذي يعتمد عليه في الارتباط، وهذا ما ابرز خصوصية هذه النظرية ، لذلك سمت انستازي (Anastasi) هذه النظرية بنظرية العامل العام (Anastasi , 1990 : 380).

٣- نظرية العوامل المتعددة Mullti – Factor Theory: رأت وجهة النظر الخاصة بتنظيم الصفات التي سادت في الثلث الاخير من القرن العشرين وجود عوامل طائفية(جمعية)(Group Factor) ، وقد يتداخل كل منهما باوزان مختلفة في اختبارات مختلفة ، وان الفرد في النشاط العقلي المعرفي يعود إلى عدد الارتباطات في الجهاز العصبي (Nunnally , 1970 : 105).

ومن دعاة نظرية العوامل المتعددة : ثورندايك (Thorndike 1926) ، الذي قدم تصنيفا ثلاثيا للذكاء يمثل المجموعات النوعية للذكاء ، وهي :
الذكاء الاجتماعي أو التعامل مع الناس ، والذكاء المادي أو التعامل مع الأشياء، والذكاء المجرد أو التعامل مع الرموز اللفظية والرياضية (السيد، ١٩٨٦ : ٢٦٧).
وثيرستون (Thurston 1938) ، فهو يعد من ابرز دعاة نظرية العوامل المتعددة، وقد استعمل الطريقة المركزية في التحليل العاملي، وتوصل من خلال أبحاثه إلى سبعة عوامل طائفية (جمعية) اطلق عليها القدرات العقلية الأولية (Primary Mental Abilities) وهي القدرات : اللغوية، والعددية، والاستدلالية، (الاستقرائية، والاستنباطية) ،والسرعة الادراكية، والعلاقات المكانية، والتذكر، والطلاقة اللفظية .

٤-نظريات النماذج الهرمية Hierarchical Models Theory :أثارت نظرية العوامل المتجمعة اهتمام العلماء في ميدان القياس العقلي بصورة عامة، والقدرات العقلية بصورة خاصة من ناحيتين ، الأولى تناولت التحقق من العوامل (القدرات) الأولية في حين تناولت الاخرى تحليل هذه القدرات إلى مكوناتها الأساسية ، مما تطلب هذا ظهور بعض النماذج النظرية التي سعت إلى توضيح العلاقة بين القدرات ، إذ تستند النماذج الهرمية إلى الفئات التي تقع داخل هذه الفئات، ثم تصنيفها إلى فئات أصغر يصبح التصنيف كالشجرة المعكوسة الذي يفترض وجود مستويات متعددة من العوامل ، وكلما ازداد المستوى الذ يوجد فيه العامل عاما ، كانت طبيعته أكثر إتساعا ومدى الأداء الذي يتضمنه أكثر شمولاً (ابو حطب ، ٢٠٠٠ : ١٦٥-١٦٧). ومن أهم هذه النماذج الهرمية، هي: الأنموذج الهرمي لـ بيرت Burt 1940 :الذي توصل إلى أنموذج هرمي للقدرات الذي يضع فيه عمليات الاحساس والنشاط الحركي البسيط في أدنى مستوى ، ثم يلي ذلك عمليات أكثر تعقيدا كالإدراك والتآزر الحركي ، وتقع في المستوى الثالث (المستوى الإرتباطي) عوامل الذاكرة ، والعادات التي اكتسبها وكونها اما المستوى الرابع، فيتضمن العلاقات التي تنقسم على الفهم والاستعمال، ويظهر الذكاء العام الذي يسميه "بيرت" الإمكانية التكاملية للعقل " في كل مستوى من حيث الدرجة والنوع) .Burt,1940 : 81)

٥- نظريات النماذج ذات الابعاد الثلاثية لتصنيف القدرات: لقد واجه علماء القياس العقلي بعد استخدام الاساليب الإحصائية المتمثلة بالتحليل العاملي مشكلة تنظيم العوامل، أو القدرات العقلية في نسق متكامل، فقد وصل عدد القدرات التي كشف عنها في القياس النفسي إلى (٥٠) قدرة حتى بداية الخمسينيات من القرن العشرين، ولم تكن النماذج النظرية التي طرحها بيرت Burt، وفرنون Vernon وكاتل Cattel في كل من بريطانيا والولايات المتحدة الامريكية متكاملة ، فقد لوحظ ان معظم البحوث العاملية لم تؤيد فرض العامل العام ، وانما كان الاتجاه هو تحديد كل عامل أو قدرة بعدد من الاختبارات في مستوى العوامل الطائفية الصغرى، وبناء عليه طرحت نماذج ثلاثية التصنيف مثل أنموذج ايزنك "Eysenk" ، وجلفورد "Guilford"، وفيما يأتي توضيح لهذين الأنموذجين.

أ- أنموذج ايزنك Eysenk 1953 :

اقترح ايزنك تصنيفا للقدرات العقلية المعرفية ، ميز فيه بين ثلاثة أسس متفاعلة، هي:-

١- العمليات العقلية ومن امثلتها ، الإدراك، والتذكر، والاستدلال .

٢- مواد الاختبار ، ويعني بها محتوى مادة الاختبار ، كالاعداد، والكلمات، والرسوم .

٣- كيفية الأداء ، الذي ميز فيه بين السرعة والدقة (Eysenk , 1967 : 50-55)

ب- أنموذج جيلفورد Guilford 1967:

فهو قدم تصنيفا ثلاثيا للقدرات العقلية يعرف باسم " بنية العقل " ، وقد طور هذا التصنيف حتى خرج بصورته النهائية عام ١٩٦٧ ، إذ لاحظ " جيلفورد " بعض نواحي التشابه والاختلاف بين القدرات، ومن ثم بدا محاولاته لتصنيفها فوجد بالإمكان تصنيف القدرات على أساس نوع العمليات العقلية المتضمنة فيها، مثل الفهم، والتذكر، وغيرها التي يمكن تصنيفها أيضا على وفق نوع المعلومات التي تكون عليها مفردات المعلومات، مثل الفئات، والعلاقات، والنظم، وهكذا تشكل التصنيف الثلاثي الابعاد الذي يشار إليه بالأنموذج المورفولوجي من العمليات ومحتوى ونتاجات (الشيخ ، ١٩٨٨ : ٧٦).

تقدير الدرجة الحقيقية في النظرية الكلاسيكية للاختبارات : تؤكد أدبيات النظرية التقليدية في القياس أنَّ درجة الفرد على الاختبار تتكون من جزئين أحدهما ثابت لا يتغير يمثل الدرجة الحقيقية للفرد، والآخر هو درجة الخطأ وهي جزء غير ثابت (متغير)، لذا فالتباين الكليّ للاختبار يتكون من تباين حقيقيّ مضافاً إليه تباين الخطأ، وكلما زاد تباين الخطأ قلَّ التباين الحقيقيّ، ممّا يؤدي إلى تذبذب الاختبار فيما يقيسه وعدم اتساقه، وبالعكس إذا زاد التباين الحقيقيّ يقلُّ تباين الخطأ، ممّا يؤدي إلى زيادة استقرار الاختبار واتساقه (Anastasi & Urbina,1997:110-111).

إنَّ قلة تباين الخطأ تعني أنَّ الدرجة الحقيقية للاختبار تقترب من الدرجة الملاحظة، وكذلك زيادة معامل ثبات الاختبار، فأحد مفاهيم الثبات، ويشير كلٌّ من (Nunnally&Bernstein) (1994) إلى أنَّهُ قدرة الاختبار على قياس الدرجة الحقيقية، ومن ثم فإنَّ درجات الاختبار التي ترتفع فيها نسبة الدرجة الحقيقية للتباين الكليّ تكون أكثر ثباتاً من تلك الدرجات التي تقل فيها هذه النسبة ، وأنَّه يمكن تقدير الدرجة الحقيقية بمعلومية معامل ثبات الاختبار، وذلك لأنَّه معامل الثبات الذي نحصل عليه من تطبيق اختبارين متكافئين هو مربع معامل الارتباط بين الدرجات الملاحظة والحقيقية، وإنَّ الجذر التربيعيّ لمعامل الثبات، هو معامل الارتباط بين الدرجات الملاحظة

والحقيقية (Nunnally&Bernstein,1994:248-249).

وهناك يوجد نوعان من التقدير للدرجة الحقيقية، هما: التقدير بالفترة (Interval Estimation):يتم تقدير الدرجة الحقيقية بالاعتماد على الخطأ المعياري من خلال حساب فترات الثقة التي يمكن أن تقع ضمنها الدرجة الحقيقية، إذ يمكن وبنسبة احتمالية معينة الإشارة إلى أنَّ الدرجة الحقيقية تقع بين قيمتين من الخطأ المعياريّ، وعليه فالدرجة الحقيقية يمكن تحديدها في مدى معين يعرف بفترة الثقة، ولتحديد الفترة التي يتوقع وجود الدرجة الحقيقية خلالها فإنّنا نحتاج إلى تحديد الخطأ المعياري للقياس، ومعامل الثبات (Urbina.2004:121-122) والتقدير بنقطة (Point Estimation): ان التقدير بالفترة يعبر عن مدى متوقع تقع داخله الدرجة الحقيقية، ولكي يتم اصدار احكام على أداء الأفراد يحتاج الفاحص إلى إعطاء قيمة محددة، ومن أجل إعطاء قيمة تقديرية للدرجة الحقيقية هناك طريقتان أساسيتان ، هما :

استخدام معادلات الانحدار الخطيّ، وقد وجد أنَّ الاختلاف بين الدرجات الخام (X) والدرجات الحقيقية (T) دالة لكلٍّ من ثبات الدرجات الخام للاختبار (r_{tt}) ،ومتوسط الدرجات الخام ($\bar{X}$) وانطلاقاً من مفاهيم النظرية التقليدية، فإنَّهُ يمكن معرفتنا لمقدار ثبات الاختبار معرفة مقدار الدرجة الحقيقية من الدرجة الكلية التي حصل عليها المستجيب. وعليه يمكن تقدير الدرجة الحقيقية بالاعتماد على الدرجة الخام (Row Score) في أي اختبار.

لذا إن عملية تقدير الدرجة الحقيقية تعتمد بنحوٍ كبير على ثبات أداة القياس، ومن خلال معامل الثبات يتم الحكم على أداة القياس المستخدمة في استخراج الدرجة الملاحظة ،التي يعتمد على تقدير الدرجة الحقيقية للمستجيب عليها (Torndike&Others,1986:376-377).

وتعرف الدرجة الحقيقية على أنَّها متوسط عدد من الدرجات الملاحظة التي حصل عليها المفحوص في مواقف اختبارية متكررة للاختبار نفسه ، أو في عدد من الاختبارات المتوازية أو المتكافئة ، وتواجه هذا الطريقة مشكلة صعوبة تكرار القياسات على الاختبار نفسه ،أو على صور مكافئة للاختبار وانتقال أثر الخبرة وعوامل الملل والتعب، والتي تُعدُّ عائق في هذا المجال (Crocker & Algina,1986:122)، ويرى (Kolen & Brennan (2004) أنَّه يمكن تجاوز هذه الصعوبات باستخدام طرائق المعادلة على وفق النظرية الكلاسيكية كطريقة المتوسط الحسابي بأنواعها أو الطريقة الخطية بأنواعها المختلفة شرط إجراء عملية تدوير كاملة للاختبارات ، أي أنَّ تتم عملية معادلة الاختبار الأول في بقية الاختبارات ، واستخراج الدرجات المعادلة له ، وهكذا لجميع الصور المستخدمة ، بعد ذلك يمكن الحصول على الدرجات المعادلة لكلِّ اختبار من دون إعادة تطبيق الاختبار نفسه، أو اختبار آخر مكافئ على الأفراد أنفسهم ، ولا يمكن استخدام هذه الطريقة ألأبعد التحقق من مدى تكافؤ الاختبارات، أو توازيها قبل إجراء عملية المعادلة (Kolen&Brennan,2004:293-294) ، ويمكن أنْ نستنتج أنَّ عملية المعادلة تمثل ابرز الحلول لايجاد تقديرات الدرجة الحقيقية ، وذلك لما توافره من قدرة على إيجاد تقديرات الأفراد على صور متكافئة، أو متوازية من دون الحاجة إلى الاختبار بجميع الاختبارت والاكتفاء بتطبيق صورة واحدة على كلِّ مجموعة ، لذا سيعتمد الباحث على هذه الطريقة في حساب الدرجات الحقيقية للأفراد على وفق النظرية الكلاسيكية

باستخدام التقدير بالنقطة، واعتماد المعادلة باستخدام اسلوب التدوير بطريقة المتوسطات ، فضلاً عن تحديد فترات الثقة التي تقع ضمنها كلُّ درجة حقيقة للاطمئنان بعدم خروج هذه الطريقة في التقديرللدرجة الحقيقية عن فترات الثقة.

الدرجة الحقيقية وعلاقتها بمفهوم معامل الثبات في النظرية التقليدية: يستند مفهوم معامل الثبات في القياس إلى فكرة استقرار الدرجة التي تم قياسها لسمة، أو قدرة معينة، ويشير أبسط تعبير لمفهوم الثبات إلى مدى دقة أداة القياس، وإنَّ الاختبار الثابت يعني إعطاء تقديرات ثابتة، أي أنَّ تكرار عملية القياس يعطينا نتائج متسقة عن الفــرد ، وإنَّ الثبات يشير إلى قدر الثقة التي يمكننا أنْ نضعها في نتائج اختباراتنا، ومن الحقائق الإحصائية والمنطقية أنْ معرفة الثبات لأختبار ما تسمح لنا بمعرفة الخطأ المحيط بالاختبار، أية درجة الدقة التي يمكن في ضوئها أن نفسر الدرجة التي حصلنا عليها ، وكما هو معروف أنَّ معامل الثبات يستخدم لأحد غرضين ، الأول : أنَّه يعبر عن دقة الاختبار نفسه كأداة للقياس ، والآخر: أنه يقدم تقديراً للأتساق (Consistency) في أداء المفحوص على الاختبار (عمر، وآخرون، ٢٠١٠: ٢١٥).

فالاختبار الذي يحقق صفة الثبات تتسم نتائجه بالاتساق ، وأننا سنحصل على التقديرات نفسها، أو تقديرات قريبة منها في كلِّ مرة نجري فيها هذا الاختبار على هذا الفرد، أو من يماثله من عينة التقنين.

ويشير كي وآخرون (Gay &Others) (2010) إلى أنَّه يقصد بثبات درجـــات اخـــتبار ما مدى خلوها من الأخـــطاء غير المنتظمة التي تشـــوب عملية القياس، ويعرفه بأنَّه نســـبة تباين الدرجة الحقيقية إلى تباين الدرجة الملاحـــظة (Gay &Others,2011:232-233).

وتشير (Crocker & Algina) (1986) إلى أن معامل الثبات لدرجات مجموعة من المستجيبين بأنَّه معامل الارتباط بين مجموعة تلك الدرجات ومجموعة درجات أخرى في اختبار مكافئ حصل عليها بنحوٍ مستقل أفراد مجموعة المستجيبين ذاتها، (Crocker & Algina, 1986:112) ، ويرى (Ayala) (2009) أنَّ الثبات يعني دلالة الاختبار على الأداء الحقيقيّ (الفعليّ) للفرد (الدرجة الحقيقية Ture Score) (Ayala,2009:6-7).

ويمكن من خلال معامل الثبات تقدير الدرجة الحقيقية للفرد، التي هي جزء من الدرجة الملاحظة، لتظهر العلاقة التي ترتبط بين الدرجة الحقيقية، والدرجة الملاحظة (Crocker & Algina,1986:114-115).

وتمثل العلاقة بين الدرجة الحقيقية، والدرجة الملاحظة في النظرية التقليدية (مؤشر الثبات)، ومن المفيد في هذا الجانب توضيح العلاقة بين الدرجات الحقيقية (غير المعلومة)، والدرجات الملاحظة (المعلومة) لاختبار ما في ضوء المفهوم الإحصائي لمعامل الثبات، وهو عبارة عن جذر معامل الثبات (معامل التحديد)، إذ إن قيمة معامل الثبات هي القيمة التي يذكرها المتخصصون في مجال القياس، وإعداد الاختبارات، والمقاييس في أدلة الاختبارات، فمعامل الثبات يدلُّ على العلاقة بين مجموعتين من الدرجات الملاحظة (المعلومة)، كدرجات اختبارين متكافئين يقيسان القدرة نفسها، في حين يدلُّ مؤشر الثبات على العلاقة بين الدرجات الحقيقية (غير المعلومة)، والدرجات الملاحظة (المعلومة)، لذلك فإنَّ مؤشر الثبات يعطي قيمة تقديرية لهذه العلاقة بين

الدرجات الحقيقية والدرجات الملاحظة، ويمكن أنْ نصل ممّا سبق إلى حقيقة أنَّ معامل الثبات يساوي مربع مؤشر الثبات ، أي مربع معامل الارتباط بين الدرجات الحقيقية، والدرجات الملاحظة، لذا إنَّ معامل الثبات يُعدُّ بمنزلة معامل التحديد، إذ يحدد نسبة تباين الدرجات الملاحظة التي تعزى إلى التباين الحقيقيّ للدرجات فيما يقيسه أي اختبار، وإذا طرحنا قيمته من الواحد الصحيح فسنحصل على نسبة التباين في الدرجات الملاحظة التي تعزى إلى الأخـــطاء العشوائية (علام، ٢٠٠٠: ١٣٤-١٣٦) .

على وفق ما تم استعراضه في الاطار النظري سيتم مناقشته مناقشة تحليلية ، وعلى ما ياتي:

لقد تبنت الباحثة نظرية العاملين لسبيرمان في هذا البحث للأسباب الآتية :-

١-لأن الاختبار الذي اعتمدت عليه الباحثة في البحث الحالي لم يكن هدفه قياس قدرة عقلية خاصة معينة، وانما قياس الذكاء العام بوصفه قدرة عقلية عامة يبدو أثرها في شتى نواحي النشاط العقلي.

٢- لأن هذه النظرية هي المنطلق الذي نشـــأت عنه اختبارات الذكاء العام التي تضم في العادة فحوصا لقـــدرات متنوعة وتعـــطي من ثم قدراً من الحرية في تنويع الاخـــتبارات للحصول على المستوى العام للذكاء عند الأفراد(توق وعدس ، ١٩٨٤ :٣٠٢) .

٣- لأن معظم النظريات الأخرى للذكاء لم تستطع ان تتجاوز المفهوم الأساس لنظرية سبيرمان، وهو العامل العام (G) على الرغم من اعترافها الضمني بوجوده انها حاولت ان تتميز منه، أو تفسيره بطريقة مختلفة نوعا ما عن تفسير سبيرمان ، ومن هذه النظريات نظرية تومسون، وثورندايك، وثيرستون، وكاتل، وبيرت، وفيرنون ، ثم افترض وجوده في أعمال بينيه وترمان (Binet & Terman)، وهناك دليل على أن الأداء في اختبارات ستانفورد وبينيه والاختبارات المشابهة يمكن تفسيره بشكل كبير بمفهوم العامل العام(G) (Aiken , 1988 :153) .

٤- لأن نظرية سبيرمان تركز على العامل العام ، وهو العامل الذي يعتمد عليه في أكثر المجالات العلمية، فقد توصل سبيرمان إلى أن نتائج بعض البحوث تؤكد اعتماد الدراسات العلمية كالفلسفة، والرياضيات، واللغات، والعلوم الطبيعية على العامل العام أكثر من اعتمادها على القدرات الخاصة، واعتماد الدراسات الفنية في مجالات الرسم، والموسيقى، والاشغال اليدوية، والزراعة، والصناعة، على القدرات الخاصة أكثر من العامل العام (ابو حطب ، ١٩٧٨ : ١٤٨) ، فضلا عن أن فكرة العامل العام اصبحت قضية مركزية في بحوث الذكاء وظلت حيه حتى يومنا هذا (Gregory , 1996 : 157).

٥- لأن العامل العام هو في الجوهر العامل المشترك للقدرة المعرفية العامة المسؤولة بشكل أولي عن ثبات الدرجات، أو استقرارها في أكثر الاختبارات المعرفية (Aiken , 1988 :153).

دراسات سابقة:

١-دراسة بالخيور (٢٠٠٩): رمت الدراسة إلى مقارنة طرائق معادلة النماذج الاختبارية الأربعة لاختبارات القدرات الذي أعدّه المركز الوطنيّ للقياس والتقويم في وزارة التعليم العاليّ السعودية ، الذي يطبق على طلاب الصف الثالث الثانويّ (السادس الاعداديّ) ، توصلت الدراسة إلى تناسب الاختبار مع المستوى المعرفي (العقليّ) لطلاب هذه المرحلة، وكذلك إلى تعادل نماذج الاختبارات الأربعة بدرجة كبيرة.

٢- دراسة هولاند وآخرون (Holland & Others,2008): رمت الدراسة إلى المقارنة

بين طرائق المعادلة المتبعة عندما يكون التصميم هو تصميم المجموعات غير المتكافئة مع وجود اختبار رابط، وتمثل هذه الطرائق نمطين من طرائق معادلة الدرجات الملاحظة ،وتوصلت الدراسة إلى أنَّ في مواقف المعادلة حينما تكون دالة الربط غير خطية ، فإنَّه يلحظ أنَّ كلا النوعين من التوقعات في هذه الحالة يكونان متشابهين(Holland &Others,2008:17-18).

٣- دراسة سادي وهتروي (Saida & Hattori,2008): هدفت الدراسة إلى مقارنة طرائق المعادلة عند تطبيق اختبارات في اللغة الانجليزية باستخدام أسلوب التعادل (Post. Hok IRT Equating)، واستخدمت نتائج تطبيق اختبار يتكون من (٨) صور، تكونت عينة الدراسة من (١٤٠.٠٠٠) مستجيب يتوزعون على (٧) مجموعات تغطي كلّ مجموعة اختبار من الاختبارات السبعة ،توصلت الدراسة إلى أنَّ متوسط درجات طلاب الصف الأول بالمرحلة الثانوية كان اتجاهها في تناقص مستمر خلال السنوات الثمان، وأنَّه من خلال مقارنة أداء الطلاب اليابانيين في اللغة الانجليزية خلال ثمان سنوات على مقياس تقدير مشترك (Saida & Hattori,2008:187-188).

٤- دراسة فان دايفر وولسن (Von Davier&Wilson,2008): تُعدُّ الدراسة الحالية امتداد لدراسة سابقة قام بها الباحثان حول مدى تأثير المجتمع الأصل للدراسة الذي يراد سحب عينات منه على دقة نتائج عملية معادلة الاختبار باستعمال طريقة الدرجات الملاحظة (Observed Scores)، إذ تتناول الدراسة الحالية الأهداف نفسها، ولكن تستعمل طريقة الدرجات الحقيقية (True Score) ، توصلت الدراسة إلى أنَّ أكثر الطرائق حساسية للمجتمع الأصل المحسوب منه العينات، ومدى دقة التمثل لهذا المجتمع طبقاً للتوزيعات المتمثلة للدرجات هي طريقة الدرجات الحقيقية للمعادلة(Von Davier&Wilson,2008: 12-13)

٥- دراسة كوينتي (Quenette,2006):رمت الدراسة إلى مقارنة طريقتين من طرائق المعادلة هي طريقة المرتكزة على أنموذج (Model Based Equation)، والطريقة الامبريقية (Empirical Equating) وقد استخدمت هذه الدراسة بطارية اختبارات الاستعداد المهني للجيش Armed Services Vocational Aptitude Test Battery ((ASVAB)) ، توصلت الدراسة إلى أنَّ تحليل البيانات الناتجة عن تطبيق البطارية (ASVAB) في طريقتي معادلة الاختبار تتشابهان بنحوٍ عام، وأنَّ الطريقة الامبريقية تتضمن معياراً للحكم على دقة نتائج المعادلة(Quenette,2006 :167-168)

٦-دراسة فان دار لاين (Vander Linden,2006): رمت هذه الدراسة إلى إجراء عملية المقارنة بين طريقتين من طرائق تقدير خطأ المعادلة للاختبارات، وذلك من حساب خطأ المعادلة بالاعتماد على الدرجات الملاحظة، لحلِّ مشكلة معادلة الاختبارات الخطية والمكيّفة (Adaptive)،توصلت الدراسة إلى أنَّ كلا المدخلين جيدان من حيث الدقة في تقدير خطأ المعادلة (Vander Linden,2006:493-494).

دراسة الشريفين (٢٠٠٣) : رمت الدراسة إلى الكشف عن مدى تحقق معايير الفاعلية في معادلة اختبارين توصلت الدراسة إلى أنَّ نتائج المعادلة الأفقية التي أجريت على وفق تصميم المجموعات المتكافئة عن أن الأنموذج أحادي المعلم كان أكثر فاعلية من طريقتي

المعادلة الخطية والمئينية (الشريفين، ٢٠٠٣:ص ـــــ ر)

٧- دراسة أيوب (١٩٩٤): رمت الدراسة إلى المقارنة بين أربع طرائق لمعادلة الاختبارات هي الطريقة الخطية، والطريقة المئينية المنبثقتان عن النظرية التقليدية في القياس، ونماذج نظرية السمات الكامنة أحادي المعلم وثنائي المعلم ، توصلت الدراسة إلى أنّ نماذج النظرية الحديثة كانت أكثر فاعلية من طريقتي المعادلة الخطية والمئينية، وفيما يتعلق بالمعادلة الأفقية في تصميم المجموعات المتكافئة أو العشوائية، إذ كانت قيمة معاملات الصدق التقاطعيّ أقل ما يمكن في الأنموذج ثنائي المعلم(ايوب ،١٩٩٤ :٢٤).

٨-دراسة رايت ودورانس (Wright&Dorans,1993): رمت الدراسة إلى مقارنة طريقتين من طرائق معادلة الاختبارات هما الطريقة الخطية، والرتب المئينية المتساوية ، وقد استخدم الباحثان درجات اختبارين من مجتمع دراسة حقيقي وآخر افتراضي، وتوصلت الدراسة إلى أنَّ كلاً من الطريقة الخطية والمئينية لهما التأثير نفسه فيما يتعلق بدقة نتائج معادلة الاختبارات، كما أن استخدام طرائق نظرية السمات الكامنة في معادلة الاختبارات كان أكثر دقة من طرائق النظرية التقليدية الكلاسيكية (Wright&Dorans,1993:3-4).

إنّ عملية التعقيب على الدراسات واستقرائها التي تناولت مفهوم معادلة الاختبارات، واستخدام الطرائق المختلفة للمعادلة سواء على وفق نظرية التقليدية خلص إلى ما يأتي:-

١- لا توجد طريقة واحدة تتفوق على جميع الطرائق الأخرى لمعادلة الاختبارات، ولكن يمكن أنْ تكون نتائج الطرائق التي تعتمد على نظرية السمات الكامنة أدق من نتائج الطرائق التي تعتمد على النظرية التقليدية (الكلاسيكية) في القياس، وذلك عندما تكون العينات المستخدمة في عملية المعادلة عشوائية وذات أحجام كبيرة، مع تحقيق كامل افتراضات الأنموذج المستخدم في عملية المعادلة.

٢-تحقق طرائق المعادلة الخطية والمئينية نتائج جيدة وبمستوى دقة مقبول شرط أن تكون البيانات المستخدمة في عملية المعادلة لها درجة مقبولة من الثبات مع تقارب مستويات صعوبة الاختبارات المستخدمة في عملية المعادلة، مع تقارب مستويات قدرة الأفراد في جميع العينات المستخدمة في المعادلة.

٣-إنَّ التباين في المجتمع الأصليّ للأفراد المستجيبين قد يؤدي إلى تحيزٍ في نتائج عملية المعادلة لاسيما إذا كانت عملية المعادلة تمتد على فقرات زمنية طويلة، إذ إنَّ تباين المجتمع (Population Invariance) له تأثير كبير في نتائج في عملية معادلة الاختبارات والمفاضلة بين طرائق المعادلة المختلفة بعد ذلك.

٤-لم تتفق الدراسات السابقة للمحور الأول في نتائجها حول أفضل الطرائق التي ينبغي استخدامها في عملية حساب تقديرات الدرجة الحقيقية لكلِّ فرد.

٥- لم تتفق الدراسات السابقة للمحور الثاني في نتائجها حول أفضل الطرائق التي ينبغي استخدامها ، بل لم يكن هناك أتفاق حول الجوانب التي يمكن استخدامها في عملية المقارنة والمفاضلة بين الطرائق للحكم على الأفضلية إلى أحداها على بقية الطرائق الأخرى، فبعضها أعتمد على الدقة (Accuracy)، وبعضها الآخر (Stability)، وبعضها اعتمد على حجم الخطأ المعياري (Standard Error) ، فضلاً عن هذا النوع في معايير الحكم وتطبيقها إحصائياً، وان كان اشهرها كما طبقت الدراسات السابقة هو: الصدق التقاطعيّ، وجذر متوسط مربعات الخطأ (RMSE).

٦- توصلت الباحثة إلى أنَّه قد تم استخدام بعض البرامج الإحصائية كالبرنامج الإحصائي

(SPSS) لاستخراج بعض المعاملات، مثل: التحليل العاملي، وحساب معاملات الصدق والثبات، ومؤشرات الصعوبة والتمييز قبل تطبيق البرامج الخاصة بعملية معادلة الاختبارات.

الجوانب التي أفادت الباحثة من الدراسات السابقة:

خلص البحث الحالي إلى جملة من الفوائد التي تفيد وتعزز هذا البحث من تعقبه لسلسلة الدراسات السابقة التي توزعت على المحورين السابقين ، التي أفادت الباحث في الوقوف على حقائق تفيده في هذا البحث ويصب مراعاتها في خطوات الدراسة الحالية، وهي كالآتي:

١- عدم استخدام عينة من الأفراد قليلة ؛ ذلك لتوصل الدراسات إلى نتائج تؤكد عدم غمكانية المقارنة بين الطرائق باستخدام عينات صغيرة عن الأفراد.

٢- عدم وجود دراسة استخدمت اختبارات ذكاء في عملية المعادلة، وبذلك تعدُّ الدراسة الحالية الأولى في استخدامها لاختبارات ذكاء حديثة.

٣- عدم وجود دراسة استخدمت تقديرات الدرجة الحقيقة لدرجات مستمدة من تطبيق حقيقي لنتائج اختبارات حقيقية غير مولدة (محاكات).

٤- مراعات مستوى تباين العينات واختلافها قبل الشروع في عملية المعادلة والمفضلة بين طرائقها المختلفة، وذلك لما أجمعت عليه نتائج الدراسات السابقة من أثره في نتائج عملية المعادلة.

٥- قلة عدد الدراسات العربية التي أجريت حول معادلة الاختبارات، والدرجة الحقيقية بنحوٍ ملحوظ ، إذ لم يبلغ عدد الدراسات العربية للمحور الأول سوى دراستين فقط وللمحورالثاني، وهو مفهوم المعادلة، فقد بلغ عدد الدراسات السابقة العربية (٣) دراسات، ولم توجد أية دراسة عراقية للمطورين.

منهجية البحث :

إن المنهج الوصفي شكل من أشكال التحليل والذي يفسر ، ويصور الظاهرة أو المشكلة تفسيراً علمياً ومنظماً ، ويصنفها، ويحللها ويخضعها للدراسة بشكل دقيق (عبيدات ، وآخرون ، ٢٠٠٠ : ٢٤٧).

ونظراً لأن الدراسة الحالية تسعى إلى معادلة اختبارات فيليب كارتر وكين راسل للقدرة العقلية على طلبة المرحلة الإعدادية لإدارة مديريات تربية محافظة بغداد الكرخ والرصافة كافة ، والذي يتراوح أعمارهم بين ١٦ – ١٨ سنة ، ومن ثَمَ فإن الباحثة ستتبع إجراءات البحث الوصفي في البحث الحالي ، وهو المنهج الملائم لموضوع الدراسة الحالية، وذلك لوصف خصائص الاختبار، وفيما يأتي وصفاً لهذه الإجراءات :-

أولاً: مجتمع البحث :

يتكون مجتمع البحث الحالي من جميع طلاب المرحلة الإعدادية (الرابع ، والخامس ، والسادس) بفرعيها العلمي والأدبي في المديريات العامة للتربية بمحافظة بغداد (الكرخ / الرصافة) كافة للعام الدراسي ٢٠١٥ /٢٠١٦ ، وبلغ حجم مجتمع العينة (١٦٣٩١٣) طالباً وطالبة موزعين على مدارس المحافظة ، موزعين بحسب متغير التخصص العلمي بنسبة (٥٣%)، والأدبي بنسبة(٤٧%)، ويتوزعون بحسب متغير الجنس بنسبة(٤٥%) للذكور و(٥٥%) للإناث ، وبنسب (٣١% ، ٣٠% ، ٣٩%) للصفوف الرابع، والخامس، والسادس الأعدادي، بحسب إحصائية مديرية التخطيط والإحصاء المركزي التابعة لوزارة

التربية.

ثانياً : عينة البحث :إن معادلة اختبارات فيليب كارتر وكين راسل للقدرة العقلية يتضمن تطبيقات عدة على عينات مختارة من مجتمع البحث، لذلك ستقوم الباحثة بعرض حجم العينة، وأسلوب اختيارها ضمن الإجراء الذي سيتبعه في حينه

ثالثاً : أداة البحث : استخدمت الباحثة أثنان من سلسلة اختبارات فيليب كارتر وكين راسل "Philip Carter & Ken Russell " ، للقدرة العقلية والمعد في عام (٢٠٠٧) ، للأعمار من ١٦ – ١٨ سنة ؛ لكونها متماثلة في إعدادها من حيث البناء والصعوبة ، وهي أداة اختبار لقياس القدرة العقلية العامة ،والذي أعد على وفق نظرية سبيرمان للعامل العام.

رابعاً : إجراءات البحث والإعداد للاختبار :

إن ترجمة الاختبارين عند محاولة إعداده، أو تقنينه من ثقافة إلى أخرى ، وذلك لجعل مفرداته وفقراته تتفق مع ما هو شائع في ذلك البلد (Butcher & Rouse , 199 : 4) ، لذلك قامت الباحثة بترجمة الاختبارين من اللغة الإنكليزية إلى اللغة العربية بالاستعانة بمترجم متخصص*١، ولإجراء الترجمة العكسية أعيد ترجمة الاختبارين الذي أعد باللغة العربية إلى اللغة الإنكليزية مترجم آخر *٢ ، ثم عرضت النسخة الأصلية للاختبارين مع النسخة الإنكليزية المترجمة على خبير ثالث *٣ لمطابقتها ، وقد وجدت بأنهما متكافئتان فيما بعض الفقرات القليلة التي عدلت وترجمت أساساً بما يتناسب ، مع معناها الحقيقي تجنباً للترجمة الحرفية ، وبما يتلاءم مع البيئة العراقية، ثم عرض الاختبار بترجمته العربية على متخصصان باللغة العربية* ، للاطمئنان على سلامة لغته العربية ، وقد أبدا صلاحية تعليمات وفقرات الاختبارين ، فضلاً عن بعض التعديلات التي أجرياها على بعض الفقرات ، وبذلك تأكدت الباحثة من صلاحية تعليمات الاختبار وفقراته للتطبيقات المنطقية والإحصائية.

***١ الأستاذ المساعد الدكتور علي عارف *٢ الأستاذ المساعد الدكتور منصور كاظم هيجل *٣ الأستاذ الدكتورة فاتن خيري الرفاعي**

كما قامت الباحثة بالإعداد للاختبارين من خلال توفير الأدوات التي تسهل من إتمام التطبيق، وتوفير وتجهيز الأدوات اللازمة لتطبيق الاختبار على عينة الدراسة، والتي تمثلت في: (تهيئة نسخ كافية من الاختبار المترجم لعينة التطبيق ، ونسخ كافية من ورق الإجابة).

كما أعدت الباحثة: (١-البيانات الخاصة بأعداد الطلبة حيث حصل الباحثة على بيانات كافية عن أعداد الطلبة في محافظة بغداد ، وأعداد الطلبة في المدارس للدراسة الصباحية، إذ استخدم في تحديد عدد أفراد العينة على وفق طريقة وخطوات اختيار العينة ، ٢-تحديد المدارس التي سوف يطبق فيها الاختبار على وفق طريقة وخطوات اختيار العينة،٣-أعداد جدول زمني للتطبيق، حيث قامت الباحثة بإعداد جدول زمني لتطبيق الاختبار في المدارس التي سيطبق فيها الاختبار موزع على مدة زمنية امتدت من ١٥ /٣ /٢٠١٦ إلى ١٥/٤/ ٢٠١٦ .

التحليل المنطقي للفقرات وملائمتها للبيئة العراقية : للتثبت من صلاحية فقرات اختباري فيليب كارتر وكين راسل للقدرة العقلية في شكلها الظاهري وملائمتها للبيئة العراقية .

عرضت الباحثة تعليمات الاختبار وفقراته بصورته الأولية على أربعة متخصصين في علم النفس والقياس والتقويم ، لاســتطلاع آرائهم بالتعديلات التي أجريت من حيث

تأثيرها في قياس القدرة العقلية مع الفقرات الأخـــــرى.

التجربة الاستطلاعية: من الضروري التحقق من مدى فهم العينة المســـــتهدفة لتعليمات المقياس، ومدى وضوح فقراته لديهم (فرج، ١٩٩٧: ١٦٠)، وهدف الباحثان التعرف على وضوح التعليمات، وتدوين أسئلة الطلبة واستفساراتهم، في أثناء الإجابة عن فقرات الاختبار، فضلاً عن التعرف على طريقة الإجابة على ورقة الإجابة المنفصلة، ومن ثم تعميمه عند التطبيق النهائي ، التعرف على مدى فهم واستيعاب المفحوصين لتعليمات الاختبارين، التعرف على مدى إمكانية تطبيق الاختبارين في البيئة المدرسية في العراق على وفق الإمكانات المتاحة .

وكانت التجربة الاستطلاعية على عينة مؤلفة من (٦٠) طالباً وطالبة ، موزعين بصورة متساوية على متغير الجنس في مدرستين ، لمديرية تربية الكرخ الثانية / محافظة بغداد ، وتم اختيارهم بشكل عشوائي ، وقام الباحثان بتوضيح التعليمات، لكي تساعد على فهم واستيعاب المفحوصين للاختبار ، والإجابة عن أسئلة، واستفسارات الطلبة، وتسجيل جميع الملاحظات مما سهل التطبيق ، أما فقرات الاختبار كانت متنوعة ، وكل سؤال له بدائل ، واحدة منها صحيحة ، وهذا يساعد في تقدير الدرجة بسرعة وموضوعية ، وكان متوسط الزمن المستغرق (٢٨) دقيقة ، وتراوح الزمن الذي استغرقه الطلبة بين (١٩-٣٨) دقيقة، علما إن وقت قراءة التعليمات العامة لم يحسب ضمن الزمن المقرر .

التحليل الإحصائي لفقرات الاختبار : يعد التحليل الإحصائي جزءاً أساساً لا يتجزأ من عملية البحث ذاتها ، ويترتب عليها استخلاص النتائج التي يقومون بتفسيرها ، ويمكن باستخدام الحاسوب في تخزين كميات كبيرة من البيانات وتحليلها عن طريق البرامج الإحصائية، وذلك بهدف تيسير إمكانية التوصل إلى تفسيرات صادقة ومفيدة من مجموعات كبيرة من البيانات (علام ، ٢٠٠٣: ٣١- ٣٢) .

وعليه تألفت عينة الدراسة الحالية من طلبة الصف الرابع، والخامس، والسادس الإعدادي، والتي تم اختيارها بالأسلوب الطبقي العشوائي بأتباع الخطوات الآتية:-

١- تتوزع مديريات التربية في محافظة بغداد إلى ست مديريات في قضائي الكرخ والرصافة ، وقد تم اختيار مدرستين، أو أكثر من كل مديرية ولكلا الجنسين.

٢-لقد تم اختيار الطلبة بشكل عشوائي في كل مدرسة من المدارس (بطريقة السحب) التي وقع عليها الاختيار، إذ بلغ عدد الطلاب الذين اختيروا لعينة التحليل الإحصائي (٥٠٠) طالب وطالبة بواقع (٢٥٠) من الذكور، و (٢٥٠) للإناث ، وبحسب النسبة المئوية لمجتمع الدراسة والبالغة (٣١%) للصف الرابع ، و(٣٠%) للصف الخامس ، و(٩%) للصف السادس ، وبالاختصاصين العلمي والأدبي ، وتطوع (٦٠) طالب وطالبة للمشاركة بواقع (٣٤) من الذكور و(٢٦) من الإناث ، ولم تهمل هذه المشاركات، فقد أضيفت إلى عينة التحليل الإحصائي، وبذلك أصبح عدد مجتمع الدراسة (٥٦٠) طالباً وطالبة والجدول (٣) يوضح حجم عينة التحليل الإحصائي، وأسماء المدارس المشاركة، وعدد الطلاب، وقد توزعوا على عدد أفراد العينة الكلي ، وبذلك أصبح عدد مجتمع الدراسة المشاركين (٣٠٠) من ذكور و(٢٦٠) من الإناث ، ويعد هذا العدد مناسباً على وفق ما أشار إليه .

الجدول (١)
حجم عينة التحليل الإحصائي وأسماء المدارس المشاركة وأعداد الطلبة

مديريات التربية	المدرسة	الصف الرابع		الصف الخامس		الصف السادس		مجموع الذكور والإناث		المجموع الكلي للذكور والإناث
		ذكور	إناث	ذكور	إناث	ذكور	إناث	ذكور	إناث	
الرصافة الاولى	إعدادية السويس للبنين	٢٠		٢٠		٢٠		٦٠	٤٩	١٠٩
	إعدادية الشعب للبنات		١٧		١٧		١٥			
الرصافة الثانية	ثانوية الرواء للبنين	١٥		٢٠		٢١		٥٦	٥٤	١١٠
	إعدادية التسامح للبنات		٩		٩		١٠			
	ثانوية سومر للبنات		٩		٨		٩			
الرصافة الثالثة	إعدادية قتيبة للبنات	١٠		٩		١٢		٣١	٢٩	٦٠
	إعدادية الفضيلة للبنات		١٠		١٠		٩			
الكرخ الاولى	إعدادية المنصور للبنين	١٤		١٠		١٦		٦٩	٥٢	١٢١
	إعدادية ذي قار للبنين	٩		١٠		١٠				
	إعدادية الرسالة للبنات		٨		٩		١٠			
	ثانوية بغداد للبنات		٩		٩		٧			
الكرخ الثانية	إعدادية السيدية للبنين	٨		٩		٨		٤٩	٤٦	٩٥
	إعدادية المعارف للبنين	٨		٧		٩				
	إعدادية مريم العذراء للبنات		٧		٩		٧			
	إعدادية البياع للبنات		٧		٨		٨			
الكرخ الثالثة	إعدادية سيف الدولة للبنين	١٢		١٣		١٠		٣٥	٣٠	٦٥
	إعدادية أم عمارة للبنات		١٠		١٠		١٠			
المجموع	١٧ مدرسة	٩٦	٨٦	٩٨	٨٩	١٠٦	٨٥	٣٠٠	٢٦٠	٥٦٠

قامت الباحثة بتطبيق الاختبار بعد توضيح الهدف منه، وهو البحث العلمي، ويطلبان من أفراد العينة تثبيت المعلومات الموجودة في أعلى الورقة، وهي المعلومات العامة ، وصححت إجابات الطلبة عن فقرات الاختبار البالغة (٤٠) فقرة، وقد أعطيت الإجابة الصحيحة درجة واحدة (١)، والإجابة الخاطئة (صفراً) باستخدام مفتاح التصحيح، وحسبت الدرجة الكلية للطالب على فقرات الاختبار بجمع عدد الإجابات الصحيحة عن جميع فقرات الاختبار . ثم استخرجت الباحثة بعض المؤشرات الإحصائية لدرجات أفراد العينة.

وحسبت المؤشرات الإحصائية مقاييس النزعة المركزية والتشتت والتفرطح والالتواء لدرجات أفراد عينة التحليل الإحصائي على فقرات الاختبار وتأكد الباحث من مدى قرب درجات عينة التحليل الإحصائي من التوزيع الاعتدالي.

صعوبة الفقرات : لقد بلغ عدد الأفراد في كل مجموعة من المجموعتين المتطرفتين (١٥١) طالبا وطالبة ، وقد تراوحت درجات المجموعة العليا بين (٢٥ – ٣٤) درجة و درجات المجموعة الدنيا بين (١١ – ٢١) درجة ، ثم استخرج عدد الطلبة الذين أجابوا إجابة صحيحة في كل من المجموعتين المتطرفتين عن كل فقرة من فقرات الاختبار ، وبإضافة أعداد الطلبة التي نحصل عليها من الخطوة السابقة إلى بعض، ثم قسمتها على مجموع عدد الطلبة في المجموعتين المتطرفتين، فيكون الناتج معامل صعوبة الفقرة (١)،

كما يوضح ذلك الجدول (٢).

وقد تراوحت قيم معاملات الصعوبة للاختبار بين (٠.١٠ – ٠.٨٩)، فقد حصلت (٢) فقرات على معامل صعوبة من (٠.١٩–فاقل) ، وهي تمثل نسبة (٥ %)من فقرات الاختبار، وحصلت (٣) فقرات على معـامل صعوبة بين (٠.٢٠ – ٠.٢٩) ، وتمثل نسبة (٧.٥%) من فقرات الاخـتبار، وحصلت (٣٠) فقرة على معامل صعوبة بين (٠.٣٠ – ٠.٦٩)، وتمثل نسبة (٧٥ %)، وحصلت (٣) فقرات على معامل صعوبة بين (٠.٧ – ٠.٧٩) وتمثل نسبة (٧.٥%) وحصلت (٢) فقرة على معامل صعوبة (٠.٨٠ – فأكثر)، وتمثل نسبة (٥ %). ولغرض الحكم على معامل صـعوبة الفقرات للاختبار، لابد من الإشارة إلى آراء بعض المختصين في مجال القياس النفسي، فيرى بلوم وداوني ان مدى صعوبة الفقرات المقبولة يتراوح بين(٠.٢٠) و(٠.٨٠) (Downi,1967: 215)

أما براون، فيقترح ان تكون معاملات الصعوبة بين (٠.٥٠ – ٠.٧٠)، وبالإمكان تضمين الاختبار فقرات صعبة ، وفقرات سهلة يستطيع (٩٠ %) من الطلبة الإجابة عنها (Brown , 1983 : 279) ، ويفضل ثورندايك وهيجن أن تكون متوسط معاملات الصـــعوبة (٠.٥٠) وبمدى واســع على ان لا يتضمن فقرات صـــعبة جـداً أو ســهلة جـــداً (Thorondike & Hagen ,1981 : 214 – 215)، وقد اعتمد الباحث على المدى بين (٠.٢٠ – ٠.٨٠) في توزيع فقرات الاختبار على مستويات الصعوبة . والجدول (٢) يوضح ذلك .

الجدول (٢)
توزيع فقرات الاختبار على مستويات الصعوبة

ت	مستوى صعوبة الفقرات	أرقام الفقرات	معيار الصعوبة	عددها	نسبتها
١-	٠.١٩ – فاقل	٢٠،٢٢	صعبة جداً	٢	٥ %
٢-	٠.٢٠-٠.٢٩	٣١ ، ٣٥ ، ٣٧	صعبة	٣	٧.٥%
٣-	٠.٣٠-٠.٦٩	١،٢،٣،٥،٧،٨،٩،١٢، ١٣ ، ١٤ ، ١٥ ، ١٦ ، ١٧ ، ١٨ ، ١٩، ٢١ ، ٢٣ ، ٢٤ ، ٢٥ ، ٢٦ ، ٢٧، ٢٨ ، ٢٩ ، ٣٠ ، ٣٢ ،٣٣ ، ٣٤ ، ٣٦ ، ٣٩ ، ٤٠	متوسطة الصعوبة	٣٠	٧٥%
٤-	٠.٧٩ -٠.٠٧	٤،٦،٣٨	سهلة	٣	٧.٥%
٥-	٠.٨٠- فأكثر	١٠،١١	سهلة جداً	٢	٥%
	المجموع		–	٤٠	١٠٠

القوة التمييزية : لقد اتبعت الباحثة طريقة المجموعتين المتطرفتين في الدرجة الكلية في استخراج معامل التمييز من خلال الفرق بين عدد الذين أجابوا إجابة صحيحة في المجموعتين (العليا والدنيا) مقسوماً على أفراد إحدى المجموعتين(٢)، وتنحصر قيمتها

بين (+١) و(−١)(الجدول ٩) .

وفيما يتعلق بمعيار تحديد معامل التمييز للفقرات، فقد اعتمد الباحث على معيار ايبل ١٩٧٢ (Eble)، اذ قسم ايبل الفقرات بحسب قوة التمييز على النحو الآتي:

معامل التمييز	تقويم الفقرة
٠.٤٠ – فأكثر	فقرة جيدة جداً
٠.٣٠ – ٠.٣٩	فقرات جيدة ، ويفضل تحسينها
٠.٢٠ – ٠.٢٩	فقرات حدية ، تحذف أو تخضع إلى التحسين
٠.١٩ – فاقل	فقرات ضعيفة ، وتحذف

(Eble , 1972 : 392)

وقد ظهر ان عدد الفقرات(الجيدة جدا) من حيث قوتها التمييزية من (٠.٤٠ – فأكثر) بلغ (٣١) فقرة بنسبة (٧٧.٥%)، وبلغ عدد الفقرات (الجيدة) من حيث قوتها التمييزية من (٠.٣٠ – ٠.٣٩) (٩) فقرة بنسبة (٢٢.٥%)، ولم تحصل أية فقرة على مستوى(حدي) بنسبة (٠.٠٠%)، ولم تحصل أية فقرة على مستوى(الضعيف) (٧) فقرات ونسبتها (٨.٧٥%). والجدول (٣) يوضح ذلك .

الجدول (٣)

توزيع فقرات اختبار القدرة العقلية على مستويات دليل التميز

معاملات تمييز الفقرات	أرقام الفقرات	عددها	نسبتها
٠.٤٠ – فأكثر جيدة جداً	٢ ، ٤ ، ٧ ، ٨ ، ٩ ، ١٠ ، ١١ ، ١٢ ، ١٣ ، ١٤ ، ١٧ ، ١٨ ، ١٩، ٢٠ ، ٢١ ، ٢٢ ، ٢٣ ، ٢٤ ، ٢٥ ، ٢٦ ، ٢٨ ، ٢٩ ، ٣١ ، ٣٢ ، ٣٣ ، ٣٤ ، ٣٥ ، ٣٧ ، ٣٨ ، ٣٩ ، ٤٠	٣١	% ٧٧.٥
من ٠.٣٠ – ٠.٣٩ جيدة	١،٣،٥،٦،١٥،١٦،٢٧، ٣٠،٣٦،	٩	% ٢٢.٥
من ٠.٢٠ – ٠.٢٩ حدية	لا يوجد	٠	% ٠
من ٠.١٩ – فاقل ضعيفة	لا يوجد	٠	% ٠

معامل صدق الفقرة :

إن المتخصصين في القياس النفسي مثل جيزلي وآخرون (Ghiselli etal , 1981) يشيرون إلى أهمية اختيار الفقرات التي لها ارتباط عالٍ مع محك خارجي، أو مع الدرجة الكلية للمقياس (Ghiselli etal,1981:436)، وحساب معامل صدق الفقرة استخدم الباحثان معامل ارتباط بوينت بايسيريال (Point – Biserial) لحساب الارتباط بين الدرجة الكلية للاختبار (المتصلة)، وبين الدرجة الثنائية (المتقطعة) لكل فقرة، بواسطة برنامج الحاسوب الآلي (SPSS)، واتضح ان معاملات صدق الفقرات جميعها كانت بدلالة إحصائية عند مستوى (٠.٠٥).

وبعد الانتهاء من استخراج أهم الخصائص السيكومترية(القياسية) لفقرات إختباري القدرة العقلية ، اقدم الباحثان على خطوة اتخاذ القرار النهائي في استبعاد الفقرات أو

إبقائها، إذ عُدت الفقرة جيدة عندما يتراوح معامل صعوبتها بين (٠.٢٠ – ٠.٨٠)، ويكون معامل تمييزها (٠.٣) فأكثر، وذات علاقة موجبة دالة مع الدرجة الكلية (Eble, 1972 : 399 – 401). وعلى هذا الأساس تبين ان جميع الفقرات تمتاز بخصائص جيده لذلك لم تحذف أية فقرة من فقرات الأختبار.

فعالية البدائل الخاطئة : يذكر عودة (١٩٨٥) ان البديل الخاطئ يكون فعالاً إذا اختاره (٠.٥) فأكثر من المجموعتين العليا والدنيا، ونسبة أكبر من المجموعة الدنيا (عودة، ١٩٨٥: ١١٥)، ولفحص اجابات الطلبة على كل بديل من بدائل الفقرة ، نتبع الخطوات نفسها المتبعة في إيجاد قوة تمييز الفقرة ، وباستخدام المعادلة نفسها تم استخراج فعالية كل بديل من البدائل الخاطئة، ولكل فقرة من فقرات الاختبار ، وتبين ان جميع بدائل الفقرات فعالة، وجذبت أفراد المجموعة الدنيا أكثر من أفراد المجموعة العليا.

صدق الاختبار:

يشير جينكنس (Jenkins) إلى أن الصدق مفهوم واحد مهما تعددت أنواعه، لذا فهي مؤشرات وطرائق للدلالة عليه، وكلما توافرت للمقياس مؤشرات أكثر زادت الثقة في قياس ما اعد لقياسه (Jenkins, 1966 : 33)، وقد حددت جمعية السيكولوجيين الأمريكية (A.P.A) ثلاثة أنواع من الصدق، هي (صدق المحتوى ، والصدق المرتبط بمحك ، وصدق البناء) (A.P.A , 1985 : 9) .

وقد استخرجت الباحثة في البحث الحالي نوعين من الصدق ،وفيما يأتي توضيح لكيفية الحصول على كل منها: (أ- صدق الظاهري عندما قام بعرض الاختبار بصيغته الأولية على مجموعة من الخبراء المتخصصين في القياس النفسي، ليحكموا على مدى صلاحيته في البيئة العراقية ، وحصلت جميع الفقرات على موافقة الخبراء المختصين على صلاحيتها لقياس ما وضعت لقياسه وبنسبة(٨٠%).(ب- صدق البناء(ConstructionValidity): ان القياس النفسي السيكومتري يستند إلى قياس الفروق الفردية ، وتقارن فيه درجة الفرد بدرجات المجموعة التي ينتمي اليها ، وتصبح درجات هذه المجموعة معياراً للحكم على درجته، وتحديد موقعه بالنسبة لهم، وهي تعرف بالاختبارات المرجعية المعيار (الشرقاوي، وآخرون ، ١٩٩٦: ٨٢ – ١٠٠).

ويذكر كرونباخ وميهل (Cronbach & Meehl1965و) ان الفروق بين الجماعات والأفراد في الخصيصة المقيسة هي دلائل ومؤشرات لصدق البناء . وقد تحقق الباحث من هذا الافتراض باستبعاد الفقرات غير المميزة واستبقاء الفقرات المميزة.

أما تجانس الفقرات في قياس ما أعدت لقياسه ، أي من خلال ارتباطها بالدرجة الكلية للاختبار، والذي يعد احد مؤشرات صدق البناء . فقد تم التحقق منه من خلال إبقاء الفقرات ذات العلاقة الدالة إحصائيا ، كما تحققت الباحثة أيضاً من الاتساق الداخلي لإجابات الأفراد عن فقرات الاختبار، وذلك باستخراج معاملات صدق الفقرة .

ثبات الاختبار : يعد الثبات من الخصائص القياسية الأساسية للمقاييس النفسية على الرغم من أن الصدق أكثر أهمية منه ، لأن المقياس الصادق يعد ثابتاً ، فيما قد لا يكون المقياس الثابت صادقاً ، إلا أنه ينبغي التحقق من ثبات المقياس على الرغم من مؤشرات صدقه؛ لأنه لا يوجد مقياس يتسم بالصدق التام(Brown, 1983: 72) .

ولحساب ثبات الاختبار اختار الباحث عينة مكونة من (١٥٠) طالباً وطالبة اختيروا عشوائيا من عينة التحليل الإحصائي البالغ عددها (٥٦٠) طالبا وطالبة.

وقد تم حساب الثبات بطريقة التجزئة النصفية، وتعد هذه الطريقة من الطرائق الشائعة في حساب معامل ثبات المقاييس والاختبارات النفسية، ويمثل معامل الثبات المحسوب بهذه الطريقة معامل التجانس الداخلي (homogeneity) بين فقرات المقياس(Weiner & Stewart, 1984 : 61)، ولحساب معامل الثبات بهذه الطريقة ، فبلغت قيمة معامل الثبات المستخرج بهذه الطريقة (٠.٧٨) ، وبعد تصحيحه باستخدام معادلة التصحيح لـ(سبيرمان – براون) بلغت قيمته (٠.٨٧) وهو مؤشر إضافي على أن معامل ثبات الاختبار جيد في مقارنة الكثير من الدراسات السابقة .

الخطأ المعياري للاختبار : ان خطأ القياس هو المشوه الأساس للدرجة الحقيقية ، لذلك يحاول المختصون في القياس تقدير الخطأ، لكي يحصلوا على اكبر قدر من الدقة في القياس من خلال تقدير المدى الذي تقع ضمنه الدرجة الحقيقية (عودة ،١٩٨٥: ١٥٥)، ويقصد بالخطأ المعياري للاختبار ، انحراف معياري متوقع لنتائج أي شخص يجري اختباره (Nunnaly, 1981: 218).

و يذكر أيبل (Ebel, 1972) ان الخطأ المعياري للقياس مؤشر من مؤشرات دقة المقياس ، لأنه يوضح مدى اقتراب درجات الفرد على المقياس من الدرجة الحقيقية (Ebel , 1972 : 429) ، فالدرجة التي نحصل عليها من القياس قد لا تكون معبره بدقة عن السمة، أو القدرة المراد قياسها، فتتضمن دائماً قدراً من الخطأ (فرج، ١٩٨٠ : ٣٢٢). وبعد تطبيق معادلة الخطأ المعياري للاختبار ، بلغت قيمة الخطأ (٤.٦) عندما كان معامل الثبات (٠.٨٧) الذي استخرج بطريقة التجزئة النصفية.

نتائج البحث :لتحقيق الهدف الأول قامت الباحثة بجميع الخطوات اللازمة من ترجمة الاختبارين واستخراج مؤشرات صدق الترجمة للاختبارين ، وحساب الخصائص القياسية للاختبارين ككلّ وللفقرات الاختبارية ، وقد تبيّن أنَّ الاختبارين يتمتعان بمؤشرات جيدة للفقرات وللاختبارات، وبذلك حققت الباحثة الهدف الأول من أهداف البحث .

ولتحقيق الهدف الثاني من أهداف هذا البحث اعتمدت الباحثة على النظرية الكلاسيكية والاتجاه النظريّ الذي يتوافق مع تعريف الباحثة للدرجة الحقيقية والذي يشير إلى أنَّ الدرجة الحقيقية متوسط أداء الفرد في عددٍ من الاختبارات المتكافئة، أو متوسط أداء الفروق في اختبار ما أعيد تطبيقه عليه عدداً من المرات، ونتيجة لصعوبة إعادة تطبيق الاختبار نفسه عدد كبير من المرات على الفرد نفسه، أو على صورة مكافئة للاختبار نتيجة لمشكلة انتقال أثر الخبرة عند تكرار تطبيق الاختبار. فأصبح هذا الاتجاه يواجه صعوبة في التطبيق في ميدان تقدير الدرجة الحقيقية للأفراد في القياس النفسيّ والتربويّ، وقد أوجدت عملية معادلة الاختبار على وفق النظرية الكلاسيكية حلٌّ لمشكلة إعادة التطبيق للاختبار نفسه، أو لصورة أخرى ، وذلك من استخراج درجة الفرد في كلِّ صورة من صور الاختبار من دون الحاجة إلى إعادة تطبيق الاختبار مرة أخرى (Von Daviera, 2007:960).

وقد اعتمدت الباحثة على عملية المعادلة باستخدام الدرجات الخام لاستخراج الدرجات المقابلة لها، أو ما تمثلها من الدرجات في كلِّ اختبار، وبالاعتماد على برنامج (CIPE) وتبيّن تقديرات الدرجات لكلِّ درجة من درجات الاختبار في بقية الاختبارات

الأخرى، وقام الباحثان بحساب متوسط أداء الفرد في تلك الاختبارات ليمثل أقرب تقدير للدرجة الحقيقية لكلِّ درجة من درجات الاختبارات الأربعة ، وللاطمئنان بصورة أكبر قام الباحثان بحساب الفترة التي تقع فيها الدرجة الحقيقية لكلِّ درجة من الدرجات بالاعتماد على الخطأ المعياريّ ومؤشر الثبات، إذ بلغ الخطأ المعياريّ للاختبارات على التوالي الأول { (١,٠٣)∓ ، (١,٢١)∓}، ومؤشرات الثبات بالاعتماد على تحليل التباين وباســـــتخدام معادلة (هويت) على التوالي للاختبار الأول (٠.٨٦، ٠.٨١) بانحراف معياري (٢.٨٠، ٢.٨٢) على التوالي لكلِّ من الاختبارين.

وقد توصل الباحثان إلى أنَّ جميع التقديرات للدرجة الحقيقية تقع بنسبة (١٠٠%) داخل الفترات الحقيقية لكلِّ درجة من الدرجات ، وهي في الوقت نفسه لا تتطابق مع الدرجات الملاحظة بنسبة (١٠٠%)، وهذا يؤشر إلى عدم وجود أي فردٍ من الأفراد المستجيبين تتطابق درجته الحقيقية مع درجته الملاحظة وذلك يتوافق مع الاتجاه النظريّ للنظرية الكلاسيكية ، والذي يشير إلى أنَّ تطابق درجة الفرد الملاحظة مع الحـــــقيقية يتحقق في حالة واحدة ، وهي أنْ يكون الخطأ مساوياً للصفر، أي في حالة انعدام الخطأ ، وذلك شيءً صعب التحقيق في القياس النفسيّ ، لذا فقد خلصت الباحثة من جملة تلك المؤشرات إلى أن تقديرات الدرجات الحقيقية للاختبارات الأربعة هي أقرب تقدير للدرجة الحقيقية بالاعتماد على الاتجاهين النظريين المعتمدين في تقدير الدرجة الحقيقية على وفق النظرية الكلاسيكية ، وبذلك حقق الباحثان الهدف الثالث من اهداف بحثها ، والجداول (٤)، و(٥) تشير إلى نتائج معادلة كل اختبار من الاختبارات ، كما يظهرها برنامج (CIPE)، فضلاً عن تقديرات الدرجة الحقيقية، والفترات التي تقع فيها تلك الدرجات.

الجدول (٤)

نتائج معادلة الاختبار الأول كما يظهرها برنامج (CIPE) وتقديرات الدرجة الحقيقية على وفتراتها

الدرجة الخام في الاختبار الاول	الدرجات المعادلة لها في الاختبارات الاختبار الثاني	تقديرات الدرجة الحقيقية للاختبار الاول	الفترات التي تقع ضمنها الدرجات الحقيقية للاختبار الاول ∓(١,٠٣)
الثابت	٠.٥٤		
١	١.٥٤	١.٢١	(٢.٠٣-٠.٠٣-)
٢	٢.٥٤	٢.٢١	(٣.٠٣-٠.٩٧)
٣	٣.٥٤	٣.٢١	(٤.٠٣-١.٩٧)
٤	٤.٥٤	٤.٢١	(٥.٠٣-٢.٩٧)
٥	٥.٥٤	٥.٢١	(٦.٠٣-٣.٩٧)
٦	٦.٥٤	٦.٢١	(٧.٠٣-٤.٩٧)
٧	٧.٥٤	٧.٢١	(٨.٠٣-٥.٩٧)
٨	٨.٥٤	٨.٢١	(٩.٠٣-٦.٩٧)
٩	٩.٥٤	٩.٢١	(١٠.٠٣-٧.٩٧)
١٠	١٠.٥٤	١٠.٢١	(١١.٠٣-٨.٩٧)
١١	١١.٥٤	١١.٢١	(١٢.٠٣-٩.٩٧)
١٢	١٢.٥٤	١٢.٢١	(١٣.٠٣-١٠.٩٧)

١٣	١٣.٥٤	١٣.٢١	(١٤.٠٣-١١.٩٧)
١٤	١٤.٥٤	١٤.٢١	(١٥.٠٣-١٢.٩٧)
١٥	١٥.٥٤	١٥.٢١	(١٦.٠٣-١٣.٩٧)
١٦	١٦.٥٤	١٦.٢١	(١٧.٠٣-١٤.٩٧)
١٧	١٧.٥٤	١٧.٢١	(١٨.٠٣-١٥.٩٧)
١٨	١٨.٥٤	١٨.٢١	(١٩.٠٣-١٦.٩٧)
١٩	١٩.٥٤	١٩.٢١	(٢٠.٠٣-١٧.٩٧)
٢٠	٢٠.٥٤	٢٠.٢١	(٢١.٠٣-١٨.٩٧)
٢١	٢١.٥٤	٢١.٢١	(٢٢.٠٣-١٩.٩٧)
٢٢	٢٢.٥٤	٢٢.٢١	(٢٣.٠٣-٢٠.٩٧)
٢٣	٢٣.٥٤	٢٣.٢١	(٢٤.٠٣-٢١.٩٧)
٢٤	٢٤.٥٤	٢٤.٢١	(٢٤.٠٣-٢٢.٩٧)
٢٥	٢٤.٥٤	٢٥.٢١	(٢٥.٠٣-٢٣.٩٧)
٢٦	٢٦.٥٤	٢٦.٢١	(٢٦.٠٣-٢٤.٩٧)
٢٧	٢٧.٥٤	٢٧.٢١	(٢٧.٠٣-٢٥.٩٧)
٢٨	٢٨.٥٤	٢٨.٢١	(٢٨.٠٣-٢٦.٩٧)
٢٩	٢٩.٥٤	٢٩.٢١	(٢٩.٠٣-٢٧.٩٧)

الجدول (٥)

نتائج معادلة الاختبار الثاني كما يظهرها برنامج (CIPE) وتقديرات الدرجة الحقيقية وفتراتها

الدرجة الخام في الاختبار الثاني	تقديرات الدرجة الحقيقية للاختبار الثاني الاختبار الاول	الفترات التي تقع ضمنها الدرجات الحقيقية للاختبار الثاني $\mp$ (١.٢١)	
الثابت	٠.٥٤ -		
١	٠.٤٦	٠.٦٧	(-٠.٢١-٢.٢١)
٢	١.٤٦	١.٦٧	(٠.٩٧-٣.٢١)
٣	٢.٤٦	٢.٦٧	(١.٧٩-٤.٢١)
٤	٣.٤٦	٣.٦٧	(٢.٧٩-٥.٢١)
٥	٤.٤٦	٤.٦٧	(٣.٧٩-٦.٢١)
٦	٥.٤٦	٥.٦٧	(٤.٧٩-٧.٢١)
٧	٦.٤٦	٦.٦٧	(٥.٧٩-٨.٢١)
٨	٧.٤٦	٧.٦٧	(٦.٧٩-٩.٢١)
٩	٨.٤٦	٨.٦٧	(٧.٧٩-١٠.٢١)
١٠	٩.٤٦	٩.٦٧	(٨.٧٩-١١.٢١)
١١	١٠.٤٦	١٠.٦٧	(٩.٧٩-١٢.٢١)
١٢	١١.٤٦	١١.٦٧	(١٠.٧٩-١٣.٢١)
١٣	١٢.٤٦	١٢.٦٧	(١١.٧٩-١٤.٢١)
١٤	١٣.٤٦	١٣.٦٧	(١٢.٧٩-١٥.٢١)
١٥	١٤.٤٦	١٤.٦٧	(١٣.٧٩-١٦.٢١)
١٦	١٥.٤٦	١٥.٦٧	(١٤.٧٩-١٧.٢١)
١٧	١٦.٤٦	١٦.٦٧	(١٥.٧٩-١٨.٢١)
١٨	١٧.٤٦	١٧.٦٧	(١٦.٧٩-١٩.٢١)
١٩	١٨.٤٦	١٨.٦٧	(١٧.٧٩-٢٠.٢١)
٢٠	١٩.٤٦	١٩.٦٧	(١٨.٧٩-٢١.٢١)

٢١	٢٠.٤٦	٢٠.٦٧	(٢٢.٢١-١٩.٧٩)
٢٢	٢١.٤٦	٢١.٦٧	(٢٣.٢١-٢٠.٧٩)
٢٣	٢٢.٤٦	٢٢.٦٧	(٢٤.٢١-٢١.٧٩)
٢٤	٢٣.٤٦	٢٣.٦٧	(٢٥.٢١-٢٢.٧٩)
٢٥	٢٤.٤٦	٢٤.٦٧	(٢٦.٢١-٢٣.٧٩)
٢٦	٢٥.٤٦	٢٥.٦٧	(٢٧.٢١-٢٤.٧٩)
٢٧	٢٦.٤٦	٢٦.٦٧	(٢٨.٢١-٢٥.٧٩)
٢٨	٢٧.٤٦	٢٧.٦٧	(٢٩.٢١-٢٦.٧٩)
٢٩	٢٨.٤٦	٢٨.٦٧	(٣٠-٢٧.٧٩)

وقد تحققت من الهدف الثالث، إذ جميع شروط معادلة الاختبارات على وفق النظرية الكلاسيكية، وكذلك شروط تصميم المجموعات العشوائية ، إذ اختارت الباحثة منها ثلاث طرائق هي : طريقة المتوسط الحسابيّ (Mean Equating)، وطريقة المعادلة الخطية (Linear Equating)، وطريقة المعادلة المئينية، وفيما يأتي ستبيّن الباحثة الخطوات المتعلقة في كلٍّ طريقة من الطرائق:-

أ- معادلة المتوسط الحسابي (Mean Equating): تعتمد هذه الطريقة لحساب المعادلة للاختبارات على قيم المتوسط الحسابي للدرجات، ويكون الفرق بين المتوسط الحسابي للاختبار الأول الذي يتخذ كأسس لعملية المعادلة، والمتوسط الحسابي للاختبار الأول والثاني.

فعلى وفق هذه الطريقة عند أية درجة على الاختبار الأول ، ويمكن معرفة الدرجة المعادلة لـــــها على أي اختبار ، وذلك بإضافة مقدار ما إلى المتوسط الحسابيّ للدرجات على الاختبار الأول، وهو (١٧,٥٠٩) للاختبار الثاني(١٨,٠٤٤)، وقد بلغ المتوسط الحسابي للتقديرات الحقيقية للأختبار الأول (١٧,٧١٩)، وللاختبار الثاني (١٧,٧١٤)، وهي لم تختلف كثيراً عن متوسطات الدرجات الخام ، وعلى وفق ذلك فقد تم حساب جميع الدرجات الحقيقية للاختبارين المعادلة للدرجات الحقيقية على الاختبار الأول، وتم ذلك باستخدام برنامج (CIPE) الخاص بطرائق المعادلة على وفق النظرية الكلاسيكية ، يتضمن نتائج تحويل الدرجات للاختبارين.

الجدول (٦)

معادلة الدرجات الحقيقية في الاختبار الأول مع الدرجات الحقيقية للأختبار الثاني بطريقة المتوسط الحسابي، وكما يظهرها برنامج (CIPE)

الدرجة الحقيقية في الاختبار الأول	الدرجات المعادلة لها في الاختبار الثاني
١,٢١	٠,٦٦٥
٢,٢١	١,٦٦٥
٣,٢١	٢,٦٦٥
٤,٢١	٣,٦٦٥
٥,٢١	٤,٦٦٥
٦,٢١	٥,٦٦٥
٧,٢١	٦,٦٦٥
٨,٢١	٧,٦٦٥
٩,٢١	٨,٦٦٥

١٠,٢١	٩,٦٦٥
١١,٢١	١٠,٦٦٥
١٢,٢١	١١,٦٦٥
١٣,٢١	١٢,٦٦٥
١٤,٢١	١٣,٦٦٥
١٥,٢١	١٤,٦٦٥
١٦,٢١	١٥,٦٦٥
١٧,٢١	١٦,٦٦٥
١٨,٢١	١٧,٦٦٥
١٩,٢١	١٨,٦٦٥
٢٠,٢١	١٩,٦٦٥
٢١,٢١	٢٠,٦٦٥
٢٢,٢١	٢١,٦٦٥
٢٣,٢١	٢٢,٦٦٥
٢٤,٢١	٢٣,٦٦٥
٢٥,٢١	٢٤,٦٦٥
٢٦,٢١	٢٥,٦٦٥
٢٧,٢١	٢٦,٦٦٥
٢٨,٢١	٢٧,٦٦٥
٢٩,٢١	٢٨,٦٦٥

ومن قراءة الجداول التفصيلية لمعادلة الدرجة الحقيقية في الاختبارين، ومن التدقيق يتضح أنَّ الدرجات المعادلة للدرجات على الاختبار الأول تزداد بمقدار ثابت، يمثل الفرق بين قيمتي المتوسط الحسابيّ للدرجات على كلٍّ اختبار من الاختبارات الثلاثة، والاختبار الأول.

ب–المعادلة الخطية (Linear Equating):

تعتمد هذه الطريقة على حساب قيم المتوسط الحسابيّ، والانحراف المعياريّ لدرجات الاختبارات المراد معادلتها ، وممّا يجب ذكره أنَّ هناك طرائق خطية أخرى كطريقة (توكر)، وطريقة (هولاند) الخطية ، واللتان تختلفان عن الطريقة الخطية في أنَّهما تستخدمان في الحالات التي لما قد يشوب مجموعات العينة من عدم التكافؤ، ولتجنب تفسير الفروق الكاملة بأنَّها تعود إلى مستويات صعوبة صور الاختبارات بأنها تعتمد في حسابها لعملية المعادلة على مفردات مشتركة(Parshall &Other,1995:38) .

الجدول (٧)

قيم المتوسطات الحسابية والانحرافات المعيارية للتقديرات الحقيقية للاختبارات الأربعة

الاختبار	المتوسط الحسابي للتقديرات الحقيقية	الانحراف المعياري للتقديرات الحقيقية
الاختبار الأول	١٧,٧١٩	٢,٨٠٢
الاختبار الثاني	١٧,٧١٤	٢,٨٠٢

قامت الباحثة بحساب قيم الميل، والجزء المقطوع كخطوة مهمة لعملية المعادلة، وبالاعتماد على الصيغة الرياضية السابقة ، والجدول (٨) يبيّن قيم الميل، والجزء المقطوع الخاصة بعملية المعادلة.

الجدول (٨)
قيم الميل والجزء المقطوع لمعادلة الاختبارات باستخدام الطريقة الخطية

الاختبارات	قيم الميل	قيم الجزء المقطوع
الاختبار الاول- الاختبار الثاني	١	-٠,٠٠٥

وبالاعتماد على قيم الميل، والجزء المقطوع في الجدول (٨)، وباستخدام برنامج (CIPE) أكملت الباحثة عملية المعادلة الخطية، وتبيّن نتائج عملية المعادلة كما أظهرها برنامج (CIPE)، وبنحوٍ مفصل للتقديرات الحقيقية لاستجابات الأفراد للاختبارين.

المعادلة المئينية (centile Equating Equiper): تعتمد عملية معادلة الاختبار بطريقة المئينيات على التوزيع التكراريّ للدرجات وحساب الرتب المئينية (Percentile Ranks) للدرجات لكلٍّ اختبارين ، والجدول (٩) تبيّن تفاصيل الرتب المئينية لكلٍّ اختبار.

وتدلُّ الرتب المئينية المتساوية لدرجات الأفراد في صورتي الاختبار على تساوي مستوى الأداء في كلتا الصورتين،وبذلك تكون درجات الأفراد في صورتي الاختبار متكافئة إذا تساوت رتبهما المئينية(Petersen&Other,2002:140-141).

الجدول (٩)
الدرجات الحقيقية على الاختبارين والرتب المئينية المقابلة لها وكما أظهرها برنامج (CIPE)

الاختبار الاول		الاختبار الثاني	
الدرجة الحقيقية	الرتبة المئينية	الدرجة الحقيقية	الرتبة المئينية
١,٢١	١,٢	١,٦٧	١,٦
٢,٢١	١,٥	٢,٦٧	٢,٠
٣,٢١	٢,٠	٣,٦٧	٢,٤
٤,٢١	٣,٠	٤,٦٧	٢,٥
٥,٢١	٣,٤	٥,٦٧	٢,٦
٦,٢١	٣,٦	٦,٦٧	٢,٧
٧,٢١	٤,١	٧,٦٧	٢,٨
٨,٢١	٤,٣	٨,٦٧	٣,٠
٩,٢١	٤,٦	٩,٦٧	٣,٢
١٠,٢١	٤,٨	١٠,٦٧	٣,٥
١١,٢١	٥,٥	١١,٦٧	٤,١
١٢,٢١	٧,٧	١٢,٦٧	٦,٠
١٣,٢١	١٠,٤	١٣,٦٧	٩,٥
١٤,٢١	١٦,٥	١٤,٦٧	١٧,٤
١٥,٢١	٢٥,٧	١٥,٦٧	٢٨,٥
١٦,٢١	٣٧,٤	١٦,٦٧	٤٠,٩
١٧,٢١	٥١,٨	١٧,٦٧	٥٧,٤
١٨,٢١	٦٥,٨	١٨,٦٧	٧٣.٠
١٩,٢١	٧٧,١	١٩,٦٧	٨٤,١
٢٠,٢١	٨٧,٧	٢٠,٦٧	٩٠,٨
٢١,٢١	٩٣,٥	٢١,٦٧	٩٥,٣
٢٢,٢١	٩٦,٧	٢٢,٦٧	٩٨,٠
٢٣,٢١	٩٨,٧	٢٣,٦٧	٩٨,٩
٢٤,٢١	٩٩,٥	٢٤,٦٧	٩٩,٣
٢٥,٢١	٩٩,٧	٢٥,٦٧	٩٩,٦

٢٦,٢١	٩٩,٨	٢٦,٦٧	٩٩,٨
٢٧,٢١	٩٩,٩	٢٧,٦٧	٩٩,٩
٢٨,٢١	١٠٠	٢٨,٦٧	١٠٠

ويعد هذا البحث إحدى المحاولات المتواضعة، لتسليط الضوء على عملية تقييم طرائق المعادلة للاختبارات ، وذلك لتحديد أفضل الطرائق لمعادلة الدرجات ، حتى يمكن وضع درجات الاختبارين على سلم وتدريج موّحد يضمن مصداقيتها، لاسيما أن هذه الدرجات تمثل معايير مهمة في تزويد متخذي القرارات بالمعلومات اللازمة لبناء قرارات موضوعية. وقد أكدت النتائج على ضرورة البحث في هذا المجال لتحديد أكثر الطرائق دقة واستقراراً في عملية معادلة الاختبارات. وبناءًعلى ما توصل إليه البحث الحالي من نتائج فإن الباحثة تؤكد على ضرورة موضوع معادلة الاختبارات، لما يترتب من خطورة في التشكيك في موثوقية نتائج درجات الاختبارات التي تعتمد صور متعددة ، لذا فيحسب إعطاء هذا المجال الأهمية التي تتناسب مع دورها في تحقيق العدالة بين من تطبق عليهم هذه الاختبارات ، وتمثل معادلة درجات الاختبار من نماذج متعددة قضية مهمة وملحة من أبرز قضايا الاختبارات التربوية والنفسية ، حيث تؤكد على تحقيق العدالة وتكافؤ الفرص بين الأفراد المطبق عليهم النماذج الاختبارية، كما تؤكد أدبيات البحث على أهمية ودور المعادلة في الحصول على تفسيرات ملائمة للدرجات التي نحصل عليها من تطبيق عدة صور للاختبار الواحد تقيس القدرة نفسها. وعلى الرغم من أهمية هذه القضية، إلا أنها لم تخضع كثيراً للدراسة والبحث من قبل الباحثين في مجال القياس والتقويم فمن خلال المراجعة المضنية والبحث الدقيق والتقصي المرهق للبحوث والدراسات لم يتوصل الباحث إلا إلى عدد قليل من الدراسات المتعلقة بهذا المجال البحثي.

التوصيات: وفيما يأتي يمكن للباحثة تقديم جملة من التوصيات الآتية:

١- عدم الاكتفاء بالاعتماد على طريقة واحدة من طرائق المعادلة للاختبارات، وإن تباين نتائج الطرائق لا يعني عدم كفاءتها، لأن لكل طريقة أهميتها، إن طرائق المعادلة تتكامل فيما بينها، إذ لا توجد هناك طريقة واحدة هي الأفضل في جميع الحالات أو الموافق.

٢- ضرورة إجراء معادلة للاختبارات وإعادة المعادلة بإتباع طرائق مختلفة عبرمدد زمنية متفاوتة؛ وبصورة مستمرة وبانتظام لضمان بقاء تعادل هذين الاختبارين، ومن ثم تحقيق العدالة والتكافؤ للفرص بين الأفراد المستجيبين على هذه الاختبارات.

٣- تطوير بعض المفردات في الاختبارات الأربعة تحقيق المزيد من الدقة في معادلة الاختبارات.

٤- استخدام تصاميم تعتمد مفردات مشتركة بين الاختبارات بعد عملية تطوير الاختبارات، على أن لا تقل نسبة عدد المفردات عن (٢٠%) من مجموع عدد المفردات للاختبار الواحد، عما هو متعارف شائع في مجال معادلة الاختبارات، وذلك عند بنائها على وفق نظرية القياس الحديثة.

المقترحات: يمكن للباحثة تقديم المقترحات الآتية:

١- إعادة البحث الحالي بتطبيق طرائق أخرى للمعادلة لم تستخدم في البحث الحالي، مثل طريقة المئينيات المتسلسلة المعدلة (Chained Equipercentile)، والتي تجري خلالها المعادلة مرات عدة متتابعة للتوصل إلى أفضل تعادل ممكن، وكذلك طرائق ليفين الخطية والمئينية (Levien Linear or Equioercentile)، فضلاً عن طرائق أخرى لتقدير ثوابت الموازين (Scaline Constant) المعتمدة على المتوسط الحسابي أو السيجما

كالطريقة المعروفة باسم (Robus Means)، وطريقة الانحدار (Recression) وغيرها من الطرائق الأخرى الكثيرة.

٢- إعادة البحث على اختبارات أخرى باستخدام إحدى تصميمات معادلة الاختبارات كتصميم المجموعة الواحدة، كأن يطبق أكثر من اختبار على نفس المجموعة الواحدة نفسها من الأفراد، ومقارنة النتائج في هذه الحالة بنتائج البحث الحالي.

٣- إجراء دراسة للمقارنة بين طرائق المعادلة للاختبارات إذا كان نمط المعادلة المستخدم هو المعادلة العمودية، وذلك إذا تم تقديم هذه الاختبارات إلى عينات من الطلاب ذات مستويات عمرية مختلفة بمديات واسعة.

٤- إجراء دراسة مشابهة إلى البحث الحالي باعتماد نفس الطرائق نفسها ولكن باستخدام درجات مولدة، وأخرى ملاحظة حقيقة، ومقارنة نتائجها بنتائج الدراسة الحالية.

Abstract

The Equation Of Intelligence Tests Developed Using The Real Degree In The Classical Theory

By Shimaa Salah

The research aims to current equation intelligence tests developed the Two Philip Carter (2008) to The test was applied to a sample of (560) students, and Fadel current research between the three methods of the equation in the classical theory, namely, (averages, linear, Almienja) and five of the methods of the equation in the theory of features inherent in such form binary parameter (2PL) and three-parameter (3PL), the researcher used the statistical programs of the following, respectively, in data processing and analysis of the (Excel), (SPSS17), (Winsteps), (Bilog-MG3), (Xcalibre, 4.1.6.0), (ST), (CPIE), (PIE); and the calculation methods of the equation and compared. and reached was the following conclusions:

1. Fit tests with the four-level students to mental Baghdad five-governmental organizations, and the homogeneity of its vocabulary, and mild degrees of frequency distribution of the respondents.

2. The four tests be equally significantly, it has been possible from being the preparation of conversion tables on the first test grades to the grades equation on the three other tests, according to each method of the equation

المصادر العربية:

١- أبو حطب ، فؤاد(1978) : (بحوث في تقنين الاختبارات النفسية)، المجلد الثاني، مكتبة الأنجلو المصرية، القاهرة – مصر .

٢- أبو حطب ، فؤاد(2000) : (القدرات العقلية) ، مكتبة الأنجلو المصرية ، القاهرة ، مصر.

٣- أيوب ، حسين محمد عبد القادر (١٩٩٤) : المقارنة بين أربع طرق للمعادلة عندما يكون التصميم من مجموعات متكافئة وغير متكافئة. رسالة دكتوراه، مقدمة إلى كلية الدراسات العليا، عمان، الجامعة الأردنية.

٤- بركات ، محمد خليفة (١٩٨٣): علم النفس التعليمي: القياس النفسي والتقويم التربوي، ج٢، الكويت، دار القلم.
٥- سعيد ، محمد حسين(٢٠٠٠): درجات امتحان الثانوية العامة: دراسة سيكومترية، رسالة ماجستير، غير منشورة كلية التربية، بني سويف، جامعة القاهرة.
٦- الشرقاوي ، انور محمد ، والشيخ ، سليمان الخضري ، كاظم ، امينة ،محمد عبدالسلام، نادية محمد ، (١٩٩٦) : (القياس والتقويم النفسي والتربوي) ، القاهرة ، مكتبة الانجلوالمصرية .
٧- الشريفين ، نضال كمال (٢٠٠٣) : مدى تحقق معايير الفاعلية في معادلة اختبارين احدهما ثنائي التدريج والأخر متعدد التدريج وفق نماذج النظرية الكلاسيكية والنظرية الحديثة في القياس. رسالة دكتوراه، مقدمة إلى كلية العلوم التربوية، عمان، جامعة عمان العربية للدراسات العليا.
٨- الشيخ ، سليمان الخضيري (١٩٩٠) : (الفروق الفردية في الذكاء). دار الثقافة للنشر والتوزيع ، القاهرة – مصر .
٩- عبيدات ، ذوقان و أبو السميد ، سهيله (١٩٩٦) : استراتيجيات التدريس في القرن الحادي والعشرين، ديبونو للطباعة والنشر، عمان – الاردن .
١٠- عدس ، عبد الرحمن ، ومحي الدين توق (1984) : (المدخل في علم النفس)، دار الفكر العربي ، عمان – الاردن .
١١- علام ، صلاح الدين محمود (٢٠٠٠): القياس والتقويم التربوي والنفسي : أساسياته وتطبيقاته وتوجيهاته المعاصرة، ط١، القاهرة، دار الفكر العربي.
١٢- علام ، صلاح الدين محمود. (٢٠٠٣)، نماذج الاستجابة للمفردة الاختبارية أحادية البعد ومتعدد الأبعاد وتطبيقاتها في القياس النفسي والتربوي، ط١، القاهرة، دار الفكر العربي.
١٣- عمر ، محمود احمد وفخرو ،حصه عبد الرحمن والسبيعي ، تركي و تركي ،امنة عبد الله (٢٠١٠) ، ط١ ، عمان ، الاردن ، دار المسيرة للنشر والتوزيع والطباعة .
١٤- عودة ، احمد سليمان (١٩٨٥): القياس والتقويم في العملية التدريسية ، ط٤ ـــــــــ ص١ ،الاردن ، دار الامل للنشر والتوزيع .
١٥- عودة ، احمد سليمان (٢٠١٠) ، القياس والتقويم في العملية التدريسية ، ط٤ ـــــــــ ص١ ،الاردن ، دار الامل للنشر والتوزيع .
١٦- غنيم ، أحمد (١٩٩١): تعميم معامل ألفا لحساب معامل ثبات المقاييس ذات المفردات غير المتجانسة ، مجلة كلية التربية ، العدد(١٥) ، جامعة القاهرة – مصر .
١٧- غنيم، احمد الرفاعي (١٩٩١) ، تعميم معامل ألفا لساب معامل ثبات المقاييس ذات المفردات غير المتجانسة، مجلة كلية التربية، ١٥، ٢٠٧–٢٣٨، جامعة عين الشمس.

المصادر الانكليزية

18- A.P.A. (American Psychological Association)(1985): ***Standard for Educational and Psychological Test***, Washington, D.C: Author.
19- Aiken, L.R. (1988): ***Psychological Testing and Assessment***, 6th ed, Boston, Allyn & Bacon.
20- Anastasi, A. (1990): ***Psychological Testing***, 6th ed, New York, Mac Milan.
21- Anastasi& Urbina, S. (1997): ***Psychology Testing***, (7th. Ed) New Yoek: Macmillan Company.
22- Ayala.R.J.de. (2009): ***The Theory and Practice of Item Response Theory***, (1th. Ed), New York ,The Guilford Press .
23- Brennan,L.R.(1997):A perspective on the History of Generalizability Theory, ***EducationalMeasurement Issues and Practice***, 16,4, 14-20.
24- Butcher, J,N & Rouse, S,V. (1994): ***Personality Individual Differences and Clinical Assessment***, University of Minnesota, Minnesota, U.S.A.
25-Carter, Philip J.(2008), ***Advanced IQ tests***, Kogan Page Limited, London and Philadelphia, United Kingdom and USA.

26- Carver, R.P., (1974) :***Two Dimensions of Test Psychometric and Edumetric***, American Psychologist, Vol.29, No.7, PP.512-518.
27- Crocker, L.& Algina, J. (1986): ***Introduction to Classical and Modern Test Theory***. NY: Holt, Rinehart and Winston.
28- Dorans, N.J. (1990): ***Equating methods and Sampling designs***,Applied Measurement in Education, 3(1), 3-18.
29- Ebel, R.L. (1972): ***Essentials of Educational Measurement***, New Jersey, Englewood Cliffs Prentice Hall.
30- Gay. L.R, Mills.Geoffreye & Airasian. Peter. (2011): ***Educational Research: Competencies For Analysis and Applications***, (9th.Ed) , Pearson Education.Inc .
31- Ghiselli, E.E.; Campbell, J.P.; & Zedeck, S., (1981***): Measurement Theory for the Behavioral Science***, San Francisco: Freeman and Company.
32- Gregory. R.L. (1996): ***Psychological testing. History, Principles, and applications***, 2nd ed Boston, Allyan and Bacon, U.S.A.
33- Gulliksen, H. (1967): ***Theory of Mental Test***, New York: John Wiley and Sons, Inc.
34- Holland. P, Sinharay. S, Von Davier, A & Han, N.(2008): ***An Approach to Evaluating the Missing Data Assumptions of the Chain and Post-Stratification Equating Methods For the NEAT Design*** , Journal of Educational Measurement, Vol.45,No.1,PP.17-43.
35- Kolen , M.J, & Whitney , D.R .(1982): ***Comparison of Procedures for Equating the Tests of General Educational Development*** , Journal of Education Measurement 9 (4), PP.279-293.
36- Kolen.M.J,Zeng.L,Hanson,Cui.Z & Chien.Y .(2004): ***Rage-Rgequate Manual***, The University of Iowa, USA.
37- Livingston, S. (2004): ***Equating Test Scores (Without IRT)***, Educational Testing Service (ETS), U.S.A.
38- Nunnally, J.C. (1978): ***Psychological Theory***, 2nd Ed., New York: Mc Graw-Hill.
39- Osterlind, Steven.J(1998):***Construccting Test Items: Multiple-Choice,Constructed-Response Performance and Other Formats***, (2th .Ed) , Boston .Dordrecht. London, Kluwer Academic Publishers .
40-Parshall, C.; Houghton, P. & Kromrey, J. (1995): ***Equating Error and StatisticalBias in Small Sample Linear Equating***, Journal of Educational Measurement, Vol. 32, No. 1, PP. 37-54.
41- Petersen,N., Cook ,L.& Stocking (2002): ***IRT VersusConventional EquatingMethod: A comparative Study of Scale Stability***. Journal of Educational Statistics, Vol. 8, No. 2, PP. 137-156.
42- Quenette, M. (2006): ***Model-Based Empirical Equating of TestForms***, Applied Psychological Measurement , Vol. 30, No. 3, PP. 167-182.
43- Saida, C. & Hattori, T. (2008): ***Post-Hok IRT Equating of Previously Administered English Tests For Comparison of Test Scores***. Language Testing, Vol. 25,No. 2, PP. 187-210.
44- Van der Linden, W.J.(2006): ***Equating Scores From Adaptive to Linear Tests***, Applied Psychological Measurement, Vol. 30, No. 6, PP. 493-508.
45- Von Davier, A.(2007):***IRT True-Score Test Equating***, Educational and Psychological Measurement. Vol. 67, No. 6, PP. 940-957.
46- Wright, N. & Dorans, N. (1993): ***Using the Selection Variable For Matching or Equating*** , Research Report 3-92**,** Princeton, NJ: ETS.