معادلة صورتي اختبار "توني " للذكاء غير اللفظي باستخدام طرق مختلفة للمعادلة في ضوء بعض المتغيرات المؤثرة على نتائجها

دكتوره/ مايسة فاضل أبو مسلم أحمد

باحث بقسم التقويم التربوي

المركز القومي للامتحانات والتقويم التربوي

ملخص البحث

هدفت الدراسة الى استخدام بعض طرق معادلة الاختبار (test equating) في اجراء معادلة بين صيغتي اختبار توني (Toni) للذكاء ، وهو اختبار غير لفظي له صيغتان متكافئتان ، وذلك في ضوء بعض المتغيرات المؤثرة على نتائج طرق حساب معادلة صيغتي الاختبار ومن خلال مقارنة نتائج طرق معادلة الاختبار المستخدمة والتي يتبع بعضها نظرية القياس التقليدي (Classical Test Theory (CTT، ويتبع بعضها الآخر نظرية القياس المعاصرة المعروفة باسم نظرية الاستجابة للبند أو المفردة Item Response Theory (IRT.

وتكونت عينة الدراسة من مجموعة من الطلاب من كليتين جامعيتين وأيضا طلاب المرحلة الثانوية بخمس مدارس ثانوية عامة في ثلاث محافظات من محافظات الجمهورية للعام الدراسي ٢٠٠٨ / ٢٠٠٩

أما أدوات الدراسة فقد تم تطبيق اختبار الذكاء بصورتيه ، وكذلك اختبار الجذع المشترك على أفراد العينة في مجموعات بحيث كانت تطبق على كل مجموعة من المجموعات صورة من الصورتين ، بينما طبق اختبار الجذع المشترك على أفراد جميع المجموعات وكان ذلك خلال العام الدراسي ٢٠٠٨/ ٢٠٠٩.

وقد أسفرت نتائج الدراسة عن أن متغيري خط المعادلة (ميل الخط والجزء الذي يقطعه من الاحداثي الصادي) وفق كل طريقة من الطرق الخمس لمعادلة الاختبار، أمكن اعداد جداول تحويل الدرجات الخام على الصورة الأولي(أ- A) الى ما يعادلها على الصورة الثانية (ب – B)) ، كما يلاحظ ما يلي بشأن معادلة درجات صورتي اختبار توني للذكاء وفق طرق معادلة الاختبار المختلفة وهو مايدور حوله التساؤل الأول للدراسة:

(١) درجات التعادل وفق طرق القياس التقليدي تتباعد عن درجات التعادل وفق طرق القياس المعاصر ، فمثلا بينما تتعادل الدرجة صفر على الصورة A الدرجات ٢٤٨ر٣ و ٧٩٢ر٥ و ٧٠٩ر٦ على الصورة B وفقا لطرق المعادلة : الخطية، وتوكر، والمئينيات بالترتيب، فانها تعادل الدرجة ٣١ر٠ و ٢٨ر٠ وفقا لطريقتي (M/M) و (M/S) بالترتيب .

(٢) عند مقارنة نتائج تطبيق طرق معادلة الاختبار وفق القياس التقليدي بعضها ببعض ، يلاحظ أن الفروق تزيد في طريقة توكر عنها في الطريقة الخطية ، ولكنها في طريقة المئينيات تزداد عنها في طريقة توكر .

(٣) عند المقارنة بين طريقتي القياس المعاصر ، يلاحظ أن نتائج طريقة (M/S) تتقارب وتقل الفروق فيها عما هو بالنسبة لنتائج طريقة (M/M).

(٤) بينما تتباعد الدرجات المتعادلة بين طرق القياس التقليدي عند الدرجات الدنيا المتطرفة (مثل صفر ، ٥ ، ١٠) ، فانها تتقارب عند الدرجات الوسطي (مثل ٢٥ ، ٣٠ ، ٣٥).

- أما عن النتائج الخاصة بفاعلية طرق معادلة الاختبار :

للاجابة عن التساؤل الثاني الذي تم طرحه خلال الدراسة والخاص بفاعلية طرق معادلة الاختبار ، تم حساب قيم " مؤشر كولن للصدق التقاطعي "Kolen's Cross-Validation" بمكوناته الثلاثة ، لكل طريقة من طرق المعادلة المستخدمة ، وحساب الدلالة الاحصائية للفروق بين نتائج هذه الطرق الخمس ، وبوجه عام يمكن القول أن طريقة المتوسط / سيجما تفضل نتائجها نتائج الطرق الأخري لمعادلة الاختبار وفقا لمؤشر كولن للصدق التقاطعي بجوانبه الثلاثة (التحيز - عدم الدقة - مقارنة المئينيات)، وتعتبر هذه النتيجة متوقعة باعتبار طريقة المتوسط / سيجما هي احدى طرق القياس المعاصر وفق النظرية (IRT) حيث أنه من المفترض أن تتميز طرق المعادلة وفق هذه النظرية عن طرق المعادلة وفق القياس التقليدي ، وهو ما يتفق مع نتائج العديد من الدراسات التي أجريت بهدف المقارنة بين طرق المعادلة وفقا لنوعي القياس (التقليدي والمعاصر) وتوصلت الى أن طرق المعادلة وفق القياس المعاصر تفضل طرق المعادلة وفق القياس التقليدي بوجه عام

وقد انتهت الدراسة بتقديم بعض التوصيات الخاصة باستخدام :طرق أخرى لمعادلة الاختبار وهي كثيرة عديدة ومتنوعة سواء في ظل القياس التقليدي أو القياس المعاصر فى ضوء ماأسفرت عنه من نتائج البحث .

معادلة صورتي اختبار "توني " للذكاء غير اللفظي باستخدام طرق مختلفة للمعادلة في ضوء بعض المتغيرات المؤثرة على نتائجها

دكتوره/ مايسة فاضل أبو مسلم أحمد
باحث بقسم التقويم التربوي
المركز القومي للامتحانات والتقويم التربوي

مقدمة :

تحتل الاختبارات بمختلف أنواعها – بوجه عام – مكانة لها أهميتها في مجال أدب القياس النفسي التربوي لما لها من دور فاعل أو مؤثر حيث يتم اتخاذ العديد من القرارات في ضوء نتائج تطبيق هذه الاختبارات ، وقد يحتاج في كثير من المواقف إلى بناء أكثر من صورة للاختبار الواحد, ويكون من الضروري جعل هذه الصور مختلفة المحتوى لكنها متكافئة الصعوبة , ومن ثم قد يكون من الضروري اتخاذ بعض القرارات كتصنيف الطلاب وإصدار أحكام موحدة بشأنهم على الرغم من تطبيق صور مختلفة من الاختبار عليهم , وهنا يكون من اللازم إجراء معادلة بين صور الاختبار الواحد التي يفترض أنها تقيس نفس القدرة أو السمة ، ومن ثم يكون من الضروري أيضا إيجاد اقتران تناظر بين نتائج صور الاختبار الواحد , بمعنى الربط بين نتائج الأداء على صورة ما بنتائج الأداء على الصور الأخرى لنفس الاختبار الواحد, ومن ثم يعد موضوع معادلة الاختبار (Test equating) من الموضوعات التي أصبح يزداد.

الاهتمام بها في مجال القياس النفسي التربوي ، حيث قد تكون هناك حاجة ملّحة إلى إجراء معادلة للصور الاختبارية التي تقدم في مواقف معينة لضمان تحقيق تكافؤ الفرص بين أفراد المجموعات المختلفة الذين تطبق عليهم الصور الاختبارية المختلفة المحتوى والتي يفترض أنها – رغم اختلاف محتواها – متكافئة الصعوبة لضمان تحقيق تكافؤ الفرص ، فمثلا في اختبارات القبول بالجامعات حيث يكون للاختبار أكثر من صورة واحدة يتم بناؤها لتكون مختلفة المحتوي متكافئة الصعوبة ، وتطبق على الطلاب دون تمييز .

ومن ثم لابد أن تتعادل هذه الصور الاختبارية لضمان تحقيق تكافؤ الفرص بين الطلاب الذين تطبق عليهم هذه الصور الاختبارية بصرف النظر عن أية مجموعة أعطي أفرادها أية صورة . وهناك العديد من المواقف الاختبارية التي نحتاج فيها إلى الرجوع لمعادلة الاختبار لعل من أهمها اختبارات القبول الجامعية – على سبيل المثال – والتي يكون عادة الاختبار الواحد منها له عدة

صور مختلفة المحتوى ويتم تطبيقها على مجموعات مختلفة من الطلاب في وقت واحد أو أوقات مختلفة, وتتخذ قرارات تتعلق بقبول هؤلاء الطلاب طبقا لنتائج استجاباتهم على الاختبار .

ومن أشهر الأمثلة لهذه الاختبارات اختبار الاستعداد الدراسي المعروف باسم (SAT) , واختبار أيوا IOWA وغيرهما من الاختبارات المعمول بها في الجامعات الأميركية ، ومن أمثلتها في البيئة العربية اختبارات القدرات التي تجريها بعض الكليات الجامعية المصرية ، واختبار القدرات العامة الذي يطبق في بعض الكليات والجامعات السعودية . ولعل معادلة مثل هذه الصور الاختبارية لهذا النوع من الاختبارات يطمئن متخذي القرارات إلى تحقيق تكافؤ الفرص بين الطلاب الذين تطبق عليهم هذه الصور الاختبارية , هذا فضلا عن الحاجة إلى مثل هذه الاختبارات المتعادلة في بعض المواقف البحثية التي تقوم على الدراسات المقارنة عموما والدولية خصوصا مثل الدراسة الدولية لتوجهات الرياضيات والعلوم المعروفة باسم (TIMSS) وهي دراسة دولية يكون للاختبار الواحد ما يقرب من خمس عشرة صورة متكافئة – رغم اختلاف محتواها – وتستخدم للمقارنة بين الطلاب سواء داخليا (بين وحدات تعكس المتغيرات البحثية داخل الدولة الواحد) أو خارجيا بين الدول .

هذا فضلا عن الحاجة الشديدة إلى معادلة الاختبار التي يلجأ اليها مصممو الاختبارات عند قيامهم بتصميم الاختبار من أكثر من صورة حيث يكون عليهم– عند بناء هذه الاختبارات وفق هذه الصور وضع الاقتران التناظري بين الدرجات على هذه الصور بعضها البعض لتكون مرشدا وموجها بما تعادله أية درجة على أية صورة اختبارية بالدرجة المعادلة لها على الصور الاختبارية الأخرى . ولا تتوقف استخدامات معادلة الاختبار عند حدود الاختبارات المعرفية المتعلقة بالتحصيل الدراسي فقط, بل تمتد للاختبارات غير المعرفية التي تتعلق بقياس الجوانب الأخرى غير المعرفية كالجوانب الانفعالية أو العاطفية أو الوجدانية, بل والجوانب النفس حركية.

وانطلاقا مما سبق ، استشعرت الباحثة وجود حاجة إلى التعرف على معادلة أحد الاختبارات التي لها أهميتها واستخداماتها في المجال النفسي التربوي ، وهو اختبار " توني " للذكاء , وله صورتان متكافئتان من حيث الصعوبة , ومختلفتان من حيث المحتوى ، وهو ما تدور حوله مشكلة الدراسة الحالية من حيث الطرق التي تحسب بها معادلة صورتي الاختبار والحكم على فاعلية كل طريقة من هذه الطرق.

أهداف الدراسة :

تهدف الدراسة الحالية إلى استخدام بعض طرق معادلة الاختبار (test equating) في إجراء معادلة بين صورتي اختبار توني (Toni) للذكاء ، وهو اختبار غير لفظي له صورتان متكافئتان ، وذلك في ضوء بعض المتغيرات المؤثرة على نتائج طرق حساب معادلة صورتي الاختبارر ومن خلال ما تهدف الدراسة إلى تحقيقه مقارنة نتائج طرق معادلة الاختبار المستخدمة والتي يتبع بعضها نظرية القياس التقليدي Classical Test Theory (CTT)، ويتبع بعضها الآخر نظرية القياس المعاصرة المعروفة باسم نظرية الاستجابة للبند أو المفردة(Item Response Theory (IRT امبريقيا باستخدام يوضح أعداد أفراد العينة وقيم المتوسطات الحسابية والانحرافات المعيارية لأعمارهم موزعين حسب الكلية الجامعية أو المدرسةValidation كمؤشر للحكم على فاعلية طرق المعادلة في ضوء ماتسفر عنه نتائج تطبيقها.

مشكلة الدراسة :

تتمثل مشكلة الدراسة الحالية في محاولة معادلة صورتي اختبار توني للذكاء باستخدام بعض طرق معادلة الاختبار وفقا للنظريتين التقليدية والمعاصرة (CTT & IRT) امبريقيا في ضوء بعض المتغيرات التي تؤثر على طرق المعادلة طبقا لما تسفر عنه نتائج تطبيقها ، وذلك باستخدام بعض المحكات المستخدمة للمقارنة بين نتائج هذه الطرق كمؤشر على فاعليتها .

تساؤلات الدراسة :

تجيب الدراسة عن تساؤلين رئيسيين هما:

(١) كيف تتعادل درجات تطبيق صورتي اختبار توني للذكاء وفقا لكل طريقة من طرق المعادلة المستخدمة ؟

(٢) أي طرق المعادلة المستخدمة تعتبر أكثر فاعلية باستخدام مؤشر كولن للصدق التقاطعي وفقا للمدخل المعروف بمعادلة التطابق (Identity Equating) ؟

أهمية الدراسة :

على الرغم من أن معادلة الاختبار تمثل إجراءات لها أهميتها الكبيرة في مجال القياس والتقويم لما يفيد فيه عند بنا ء عدة صور متكافئة للاختبار الواحد (test construction) ،

وتصحيح الاستجابات على هذه الاختبارات ومعادلة الطرق.الناتجة عنها ، إلا أن معادلة الاختبار الذي تدور حوله الدراسة الحالية يمكن أن تفيد نتائجها في الآتي :

(١) يضع تحت أيدي الباحثين جداول لتحويل الدرجات الخام بما يفيدهم في استخدام وتناول طرق المعادلة المختلفة في دراساتهم، وكذلك الممارسين لدى استخدامهم للاختبارات ذات الصور المتعددة، وذلك في ضوء ما تسفر عنه نتائج مقارنة هذه الطرق .

(٢) التعرف امبريقيا على أكثر طرق معادلة الاختبار فاعلية بما يفيد الباحثين عند استخدام وتناول طرق معادلة الاختبار المختلفة في دراساتهم ، وكذلك الممارسين لدى استخدامهم للاختبارات ذات الصور المتعددة وذلك في ضوء ما تسفر عنه نتائج مقارنة طرق حساب معادلة الاختبار.

حدود الدراسة :

يقتصر تعميم نتائج الدراسة على ما يلي كحدود للدراسة :

١) الاقتصار على عدد محدود من طرق معادلة الاختبار وهذه الطرق هي الخطية ، وتوكر ، والمئينيات (وفق النظرية vertical.) وطرق المتوسط / المتوسط ، والمتوسط / سيجما للنموذج أحادي المعلمة فقط (وفق النظرية المعاصرة) .

٢) الاقتصار على نمط المعادلة الأفقية (horizontal) فقط دون النمط الآخر للمعادلة (الرأسية أو العمودية vertical) .

٣) الاقتصار على تصميم المجموعات المتكافئة (العشوائية) كتصميم لجمع البيانات مع وجود اختبار جذع مشترك دون إتباع التصميمات الأخرى لجمع البيانات.

٤) حجم عينة الدراسة التي اقتصرت على عينات ليست كبيرة العدد من طلاب المدارس الثانوية بمحافظات ثلاث فقط من محافظات الجمهورية (هي القاهرة والجيزة و حلوان، وطلاب السنة الأولى بكليتين جامعيتين (كلية البنات – معهد الدراسات والبحوث التربوية) بجامعتين فقط من الجامعات المصرية (جامعة عين شمس وجامعة القاهرة).

٥) الاقتصار على مؤشر كولن كأسلوب لتقييم نتائج طرق معادلة الاختبار والحكم على كفاءتها أو فاعليتها دون غيره من المؤشرات الأخرى التي تستخدم كمؤشر للحكم على هذه الطرق.

٦) استخدام البرنامج الحاسوبي BILOG -MG لحساب معادلة الاختبار وفق الطرق المختلفة، وتقييم نتائجها للحكم على فاعليتها .

الإطار النظري والدراسات السابقة:

فيما يلي تتعرض الباحثة بالمناقشة لبعض قضايا معادلة الاختبار المرتبة بموضوع الدراسة الحالية من خلال البحوث والدراسات السابقة التي أمكنها الاطلاع عليها في هذا المجال سواء ما يتعلق منها بمفهوم معادلة الاختبار ومعناها , مع التركيز على المتغيرات المؤثرة على نتائج تطبيقها والطرق المختلفة التي يمكن إتباعها لحسابها , بما قد يؤثر بالتالي على فاعليتها , أو ما يتعلق بالمؤشرات المختلفة التي يمكن استخدامها في الحكم على فاعلية طرق المعادلة , وكذلك ما يتعلق بالبرامج الحاسوبية (software) المتخصصة والتي يمكن استخدامها لمعالجة وتحليل البيانات بهدف حساب نتائج تطبيق معادلة الاختبار بطرقها المختلفة وحساب قيم المؤشرات المستخدمة في المقارنة بين الطرق والحكم على فاعليتها :

أولاً – يقصد بمعادلة الاختبار (Test Equation) تلك الصيغة الرياضية (formula) التي يمكن بموجبها معرفة الدرجة على اختبار ما المعادلة للدرجة على اختبار آخر يقيس نفس القدرة أو السمة ، بمعنى أنه إذا كان لدينا اختباران للذكاء مثلا س و ص لمرحلة عمرية أو سنية معينة وكانت درجة فرد ما علي الاختبار س هي ٣٥ بينما حصل على الدرجة ٩٨ في الاختبار الثاني ص ، فانه من المفترض في هذه الحالة أن تكون درجتا الاختبارين س و ص(وهما ٣٥ ، ٩٨) متعادلتان لأنهما يقيسان مستوى ذكاء نفس الفرد الواحد ، والاختلاف هنا بين قيمتي الدرجتين إنما يرجع إلى اختلاف مستوى القياس (scaling) الذي يتبعه كل اختبار من الاختبارين س ، ص فقد يكون الاختبار س سقف درجاته ٥٠ بينما سقف درجات الاختبار الآخر ص هو ١٤٠مثلا (Dorans,Hammond 2008) ، ولتوضيح فكرة معادلة الاختبار بصورة أوضح لنضرب مثالا بمقاييس درجات حرارة الطقس المختلفة من حيث أنظمة قياسها والتي تتبعها مختلف دول العالم فبينما تتبع بعض الدول القياس الفهرنهايتى (Fahrenheit) تتبع دول أخرى ما يعرف بالقياس " السيليزي" (Celsius) ، ولنفترض أن سقف الدرجة في القياس الأول ٥٢ درجة وأنه في القياس الثاني ١٠ درجات ، أي أن الدرجة ٥٢ في القياس الأول تعادل الدرجة ١٠ في القياس الثاني ، وبافتراض أن س = درجة القياس السيليزي ، وأن ص = درجة القياس الفهرنهايتي ، فانه للتحويل من أحد القياسين إلى الآخر لأية درجة نتبع الصيغة الرياضية الآتية : ص = ٢ س + ٣٢

وهذه الصيغة الرياضية تسمي هنا معادلة تحويل درجات الحرارة من أحد القياسين إلى الآخر وبالعكس (الفهرنهايتى والسيليزي) ، فمثلا إذا كانت درجة حرارة الجو في إحدى

البلدان هي ٧ سيليزي ويراد معرفة الدرجة المعادلة لها وفق القياس الفهرنهايتي ، فإننا نعوّض في المعادلة أعلاه كالتالي : ص = (٢ x ٧) + ٣٢

= ١٤ + ٣٢ = ٤٦ درجة

والعكس أيضا صحيح ، فإذا كان المعلوم لدينا هو درجة الحرارة الفهرنهايتية ٤٦ فإنها تعادل درجة الحرارة السيليزية بالتعويض في المعادلة الموضحة أعلاه كالتالي :

٤٦ = ٢ س + ٣٢

٤٦ - ٣٢ = ٢ س أي أن س = ٧ درجات

وتقوم الفكرة الرئيسية لحساب معادلة الاختبار على تحديد قيمتين من واقع الدرجات الناتجة عن تطبيق صورتي الاختبار ، وهما الميل (slope)، والجزء المقطوع (intercept) ، فمثلا إذا كانت درجة فرد ما على الاختبار س مثلا ٧٠ ويراد معرفة الدرجة المعادلة لها على الاختبار ص بافتراض أن قيمة الميل = ٥٠ر٠ وقيمة الجزء المقطوع = ٢٥ر١ ، فان معادلة هذين الاختباريين (والتي بالتعويض فيها يمكن إيجاد أية درجة على الاختبار ص المعادلة لدرجة ما علي الاختبار س) تكون على النحو التالي : ص = ا س + ب أي أن ص = ٥ر٠ س + ٢٥ر١ (حيث ا = الميل ، ب = الجزء المقطوع).

وبالتعويض في المعادلة السابقة ينتج :

ص = (٥ر٠ x ٧٠) + ٢٥ر١

= ٢٥ ر ٣٦

أي أن الدرجة ٢٥ر٣٦ علي الاختبار ص تعادل الدرجة ٧٠ على الاختبار س ، ومن ثم يتوقع أن يحصل الفرد الذي في الاختبار س علي الدرجة ٢٥ر٣٦ في الاختبار ص .

ويلاحظ أنه كلما اقتربت قيمة الميل من الواحد الصحيح ، واقتربت قيمة الجزء المقطوع من الصفر كلما اقتربت قيمتا الدرجتين من بعضهما البعض وأصبحتا متساويتين ، بمعنى آخر فان الدرجة على الاختبار س تكون هي نفسها الدرجة المعادلة لها على الاختبار ص حيث أن المعادلة في هذه الحالة (عند الميل = 1، والجزء المقطوع = صفر) تكون : ص = س

وتسمى المعادلة في هذه الحالة بمعادلة التطابق (identity equating) أي المعادلة التي تحول الدرجة إلى نفسها (Kolen & Brennan, 2004)٠

وقد ينظر البعض الى معادلة الاختبار على أنها مجرد إجراء إحصائي يتم بموجبه تحويل الدرجة الخام على اختبار ما الى درجة مناظرة على اختبار آخر على النحو المذكور أعلاه ، إلا أنه في الواقع ، فان معادلة الاختبار ليست مجرد إجراء تحويلي إحصائي ، بل هي تعّد تكنيكا يستخدم في مجالات عديدة للقياس مثل بناء الاختبارات وصورها المتعددة ، وبناء بنوك الأسئلة التي نحتاج فيها الى التوصل الى صور متكافئة لاختبار واحد يقيس نفس السمة أو القدرة ، وذلك بالإضافة الى ما قد يتحقق من تكافؤ فرص بين مجموعات الأفراد المستجيبين على الصور المختلفة للاختبار (Kim & Hanson, 2002) ، ومما يجدر ذكره أن هناك مصطلحات عديدة تتشابه كثيرا في مجال معادلة الاختبار قد يراها البعض أنها مترادفة ، مثل مصطلح " المعادلة equating " ، و"التقايس scaling" و"التدريج calibration" و"التجميع assembly" و"الربط أو الوصل linking" وجميعها – وان كانت على علاقة وثيقة الصلة بين بعضها البعض – إلا أنها ليست مترادفة في معناها ، وان كانت الفروق بين معانيها هي فروق دقيقة الى حد بعيد (Van der Linden, 2006)

ثانياً – لمعادلة الاختبار أهميتها نظرا لاستخداماتها المتعددة في مجال القياس التربوي النفسي، فكثيرا ما نحتاج الى معرفة الدرجات المتعادلة على الاختبارات المختلفة القياس (scalin) التي تقيس نفس القدرة أو السمة الواحدة، كذلك ترجع أهمية معادلة الاختبار الى الحاجة إليها في معادلة التي يكون فيها لدينا صورة (أو أكثر) للاختبار الواحد ويكون من الضروري معادلة هذه الصور الاختبارية لنحقق تكافؤ الفرص بين الأفراد الذين تطبق عليهم هذه الصور بشكل تبادلي ، أو بشكل عشوائي بصرف النظر عن أية صورة تطبق على أفراد أية مجموعة من مجموعات الأفراد دون الأخرى التي تطبق عليها صورة أخرى ،(Sinharay,Holland 2007)

ومن ثم فانه إذا ما كانت الصور متعادلة فانه ليس من المهم علي أي المجموعات يطبق أي صورة من الصور الاختبارية ، أي أنه بصرف النظر عن الصورة الاختبارية التي تطبق على أفراد كل مجموعة ، فانه سيكون هناك تكافؤ فرص بين المجموعات (رغم اختلاف محتوي الصور الاختبارية) لتعادل صور الاختبار.

ويرى البعض أن لمعادلة الاختبار أهميتها الكبيرة التي تكمن في أنه مهما حاولنا بناء اختبارات متكافئة تماما (مختلفة من حيث المحتوي) ، فان ذلك لن يتحقق مما يضطرنا الى اللجوء لمعادلة صور الاختبار الواحد لمعرفة الدرجة علي الصورة ب المعادلة للدرجة على

- الصورة أ . ومما يؤكد أهمية معادلة الاختبار- بالإضافة الى ما سبق بشأن تكافؤ الفرص وصعوبة بناء اختبارات متكافئة تكافؤا تاما – أن كثيرا من المواقف الحياتية التي تصادفنا في مجال القياس النفسي تتطلب مثل هذه المعادلة، ومن هذه المواقف تجنب ما يعرف بتحيز المفردة (DIF)، وعندما يراد إنشاء أو بناء بنوك أسئلة item bank ، وعند بناء اختبارات تدار بالكمبيوتر (CAT) ، وكذلك عندما يراد بناء وتطوير أكثر من صورة واحدة لنفس الاختبار الواحد لقياس نفس السمة أو القدرة لأي سبب من الأسباب .

[Dorans, Hammond 2008; Tsai, et al, 2001; Moses,et al, 2007; Sinharay,Holland 2007 & Tong, Kolen, 2005]

ثالثا – هناك شروط ينبغي مراعاتها في معادلة الاختبار ، كأن تقيس صور الاختبار المراد معادلتها نفس السمة أو القدرة ، واعتبارات أخرى تتعلق بتطبيق وإدارة الصور الاختبارية ، وكذلك بعض الاعتبارات التي تراعى بالنسبة للأفراد الذين تطبق عليهم الصور الاختبارية من حيث الكم (الأعداد) والخصائص، ، كذلك بعض الشروط الخاصة بالمجتمع الأصلي (تباين المجتمع) وطبيعة المفردات المشتركة (common items) بين الصور الاختبارية سواء من حيث نوعها (داخلية/ خارجية، أو حجمها (عددها) ، كذلك من أهم هذه الشروط أوالاعتبارات تلك المتعلقة بإجراءات المعادلة ذاتها مثل نوع تصميم جمع البيانات المتبع (مجموعة واحدة – مجموعات متكافئة – مجموعات غير متكافئة مع مفردات مشتركة) ، ونمط المعادلة المستخدم (رأسية أو عمودية – أفقية) ومستويات أو أحجام عينات الأفراد المستجيبين (كبيرة – صغيرة) ونوع البيانات الحقيقية التي يتم المقارنة بها (مولدة أو محاكية – دائرية – عند حجم عينة كبيرة) (Kolen & Brennan, 2004) ومما يجدر ذكره أن هناك شروطا لمعادلة الاختبار تعتبر بمثابة شروط ينبغي مراعاتها عند إجراء معادلة الاختبار وفقا لنظرية القياس المعاصرة المعروفة باسم نظرية الاستجابة للمفردة (IRT)السالف الإشارة إليها وهذه الشروط هي على النحو التالي وذلك كما أشار إليها (Lord, F. & Wingresky, M. (1983) :

(١) **التساوي (equity)** , ويقصد بهذا الشرط أنه إذا ما كانت هناك مجموعتان من الأفراد لهما نفس مستوى القدرة , وأعطى أفراد كل مجموعة من المجموعتين صورة من صورتي الاختبار , فانه يتوقع أن يكون توزيع درجات أفراد المجموعتين على صورتي الاختبار بعد التحويل (conversion) متطابقا (identical) , أي أنه بصرف النظر عن أية صورة من الصورتين تقدم لأية مجموعة من المجموعتين , حيث تتحول كل درجة على إحدى الصورتين الى نفسها على الصورة الأخرى.

(٢) اللاتباين (invariance) , ويقصد بهذا الشرط عدم تباين أفراد المجموعتين من حيث القدرة , ومن ثم فانه بصرف النظر عن الصورة الاختبارية التي تعطى للمجموعة , فان نتائج التطبيق تكون متعادلة وذلك في ظل اللاتباين بين مستويات قدرات أفراد المجموعتين , وإلا فانه لن تتعادل نتائج التطبيق إذا ما كان هناك تباين بين مستويات قدرات أفراد المجموعتين.

(٣) التماثل (symmetry) ، ويعني ذلك أنه لا فرق بين أن يتم معادلة الصورة الاختبارية الأولى بالصورة الثانية أو العكس معادلة الصورة الثانية بالصورة الأولى لافتراض تماثل صورتي الاختبار، وتعتبر هذه الشروط الثلاثة بمثابة افتراضات ينبغي إتباعها والحرص على توافرها لضمان دقة نتائج المعادلة, وبالإضافة الى الشروط الثلاثة السابقة تفترض النظرية (IRT)ضرورة توافر شرطي أحادية البعد (unidimensionality) والاستقلال الموضعي أو المكاني (local independence) في صورتي الاختبار المراد معادلتهما , ويقصد بأحادية البعد أن يقيس الاختبار قدرة واحدة فقط , إلا أنه يلاحظ عدم الاتفاق حول وحدة القدرة , حيث يرى البعض أن القدرة الواحدة قد تتضمن أحيانا عدة قدرات فرعية مكونة لها مثل القدرة العقلية العامة التي قد تتكون من عدة قدرات فرعية (كالقدرة اللفظية والحسابية والاستدلالية) , وكذلك القدرة الرياضية التي قد تتكون من عدة قدرات فرعية (حسابية – هندسية – جبرية..الخ) , ومن ثم قد يصعب الحكم على الأحادية في مثل هذه الحالات , وعلى الرغم من أنه قد لا توجد إجراءات كافية بشكل كامل للتحقق من أحادية البعد , إلا أن أغلب الباحثين يستخدم التحليل العاملي المعروف باسم : (factor analysis of the inter-item tetrachoric correlation).

أما بالنسبة للاستقلال المكاني أو الموضعي فيقصد به ألا تعتمد الاستجابة على مفردة ما في الاختبار على مفردة أخرى , بمعنى أن تكون كل مفردة في الاختبار مستقلة تماما – من حيث القدرة على الاستجابة – عن باقي مفردات الاختبار , وقد يثير ذلك قضية المفردات التي تشترك معا في مثير واحد , كما هو الحال في اختبارات اللغات التي قد تعتمد عدة مفردات فيها على مثير واحد يمثل قصة أو موقفا واحدا , وقد يرجع احتمال الإجابة عليها معا الى فهم الموقف أو القصة ، فتأتي جميعها إما صائبة أو جميعها خاطئة بسبب الموقف الواحد . ومما يجدر ذكره أنه ليس هناك مقياسا أو مؤشرا إحصائيا للحكم على الاستقلال المكاني أو الموضعي , وهناك حل مقترح لمشكلة المثير ذي الاستجابات المتعددة يقوم على معاملة مجموعة المفردات التي على نفس الجذع المشترك (القصة أو الموقف) وكأنها مفردة واحدة , على أن يتبع في تصحيح الاستجابات على هذه المفردات بنظام تقدير الدرجات الجزئي المعروف باسم: (Kolen & Brennan, 2004) (Partial credit / Polytomus).

رابعاً – فيما يتعلق بالمتغيرات المؤثرة على نتائج معادلة الاختبار :

نناقش فيما يلي المتغيرات التي تناولها الباحثون في بحوثهم مما يؤثر على نتائج معادلة الاختبار، ومن ثم بالتالي يكون لها علاقة وثيقة – من ناحية أخرى – بمعايير الحكم على كفاءة معادلة الاختبار:

(١) **نمط المعادلة** ، نجد أن هناك نمطين أحدهما هو المعادلة الأفقية (horizontal equating) ، والآخر هو المعادلة العمودية أو الرأسية (vertical equating)، والنمط الأول يستخدم عندما يكون الأفراد الذين تطبق عليهم الاختبارات من مستويات قدرة متشابهة من حيث ما يقيسه الاختبار ، أي أن يكونوا من مستوى عمري أو تحصيلي دراسي واحد ، أما النمط الثاني فانه يستخدم عند وجود مستويات متباينة لقدرات الأفراد من حيث ما تقيسه الاختبارات ، كأن تكون الاختبارات لتلاميذ صفوف دراسية مختلفة (أول – ثان – ثالث ...الخ) ويكون لكل مستوى اختبار خاص به ويراد معادلة الاختبارات للصفوف المختلفة .

وقد قارنت بعض الدراسات بين نمطي المعادلة (الأفقية مقابل الرأسية) ، وبينما توصلت نتائج بعض هذه الدراسات إلي أن هناك بعض الطرق لا تكون نتائجها دقيقة إلا في ظل نمط ما من نمطي المعادلة دون الآخر [مثل دراسة (Kolen, 1981) التي توصلت الى أن نموذج المعلمة الواحدة "راش " يعتبر أكثر دقة إذا ما كان نمط المعادلة المستخدم هو المعادلة الرأسية " العمودية"] ، فان هناك العديد من الدراسات الأخرى التي توصلت نتائجها الى أنه لايوجد لنمط المعادلة المتبع تأثير على نتائج طرق المعادلة المستخدمة ، ومن هذه الدراسات دراسة كل من(Tong & Kolen, 2005) ، ودراسة(Wiley, 1999) .

---

(٢) **التصميم المتبع** لجمع البيانات لأغراض معادلة الاختبار ، توجد عدة تصميمات لعل أشهرها ثلاثة هي تصميم المجموعة الواحدة، وتصميم المجموعات المتكافئة ، وتصميم المجموعات غير المتكافئة مع وجود اختبار جذعي مشترك ، ويختار الباحث أحد هذه التصميمات الثلاثة حسب الموقف البحثي ، ففي التصميم الأول يتم تطبيق صورتي الاختبار على نفس المجموعة الواحدة صورة بعد أخرى، وهو ما يضمن التكافؤ التام لمجموعات العينة لأن التطبيق هنا يتم مرتين على نفس المجموعة الواحدة . أما التصميم الثاني فيتم التطبيق على مجموعتين يختارا عشوائيا ، أو تعد مجموعات متكافئة إحصائيا بحيث تطبق على كل مجموعة صورة من الصور الاختبارية عشوائيا بصرف النظر عن أية صورة تلك التي تطبق على أية مجموعة وفي التصميم الثالث يتم

تطبيق الصور الاختبارية على مجموعات غير متكافئة بشرط أن تكون هناك مفردات مشتركة يتم تطبيقها على جميع أفراد المجموعات. وهذه المفردات المشتركة هنا لها وظيفة ذات أهمية قصوى، ففي ضوئها يتم الحكم على صور الاختبارات – فمثلا إذا اختلفت نتيجة أفراد المجموعة الأولى على الاختبار (س مثلا) عن نتيجة أفراد المجموعة الثانية على الاختبار (ص مثلا) : فإلى ما ذا يرجع الاختلاف ؟ ، هل يرجع الى صعوبة أحد الاختبارين عن الآخر ؟ ، أم يرجع إلي اختلاف مستوى قدرة أفراد إحدى المجموعتين عن الأخرى ؟ إن من يحسم هذا هو نتيجة الاختبار المشترك الذي يطبق على أفراد كلتا المجموعتين والاختبار الجذعي المشترك قد يكون داخليا ، أي يكون ضمن مفردات الاختبارات الأصلية ، وقد يكون خارجيا بمعنى أن يكون اختبارا مستقلا يطبق على أفراد المجموعتين بالإضافة الى الاختبارات الأصلية ، وفي كل الأحوال من المفترض ألا تقل نسبة عدد المفردات المشتركة عن نسبة محددة (٢٠ %) من جملة عدد مفردات الاختبار الأصلي .

وإذا كان لكل تصميم من التصميمات الثلاثة لجمع البيانات سالفة الذكر مزاياه وعيوبه ، إلا أن الموقف البحثي هو الذي يفرض على الباحث اختيار نوع التصميم المناسب لموقف البحث ، فمثلا عندما لا يكون من الواقعي أو العملي إعادة تطبيق أكثر من صورة للاختبار الواحد على نفس الأفراد ، هنا يستبعد تصميم المجموعة الواحدة ، كذلك فانه حينما لا تكون المجموعات متكافئة ، أو يصعب ضمان تكافؤ المجموعات ، يستبعد تصميم المجموعات المتكافئة .

ويضطر الباحث الى إتباع تصميم المجموعات غير المتكافئة بشرط أن يكون هناك اختبار جذعي مشترك أو مفردات اختباريه مشتركة ، وإلا كان على الباحث تكوين مجموعات متكافئة إذا لم تتوافر مثل هذه المفردات المشتركة، وقد توصلت بعض الدراسات في نتائجها الى أن هناك بعض طرق المعادلة تفضل غيرها عند إتباع تصميم المجموعات غير المتكافئة ،

وذلك مثل دراسة (Petersen, Cook & Stocking, 2002) التى توصلت الى أن طريقة توكر الخطية (Tucker- Linear equating method) هي أفضل طرق المعادلة إذا ما كان التصميم المتبع هو تصميم المجموعات غير المتكافئة مع وجود اختبار جذعي مشترك (NEAT)، وكذلك دراسة (Quenette, 2006) التي توصلت نتائجها الى أن طريقة المئينيات تعد أفضل طرق معادلة الاختبار إذا ما كان التصميم المتبع هو تصميم المجموعات المتكافئة.

إلا أن هناك دراسات أخرى أوضحت نتائجها أنه لم توجد فروق جوهرية بين طرق

المعادلة عند إتباع تصميمات مختلفة ، بمعنى أن كفاءة طرق المعادلة لم تختلف باختلاف نوع التصميم المتبع ، ومن هذه الدراسات دراسة كل من (Tsai, et. Al., 2001) و (Wiley, 1999) اللتان أوضحتا أنه لا تتأثر نتائج طريقة المعادلة بنوع التصميم المستخدم لجمع البيانات ولكنها تتأثر بتباين مستويات قدرة المستجيبين، ويتفق مع هذه النتيجة نتائج بعض الدراسات الأخرى مثل دراسة كل من :

(Wright & Dorans, 1993 & Harris & Kolen, 1986)

(٣) معيار المقارنة الذي تتم وفقا له – أو في ضوئه – مقارنة نتائج طرق المعادلة للحكم على كفاءتها مما يؤثر على الاختيار من بين الطرق،ويقصد بذلك المعيار الحقيقي الذي تتم المقارنة وفقا له ، بمعنى أنه لابد وأن يتوافر لدينا معيار محدد لما ينبغي أن تكون عليه الدرجات الحقيقية من الناحية النظرية (طبقا لمستوى صعوبة المفردات ومستوى قدرة الأفراد المستجيبين) حتى يتسنى مقارنة الدرجات التي نحصل عليها من التطبيق الميداني الفعلي بها ، أي أنه تتم مقارنة الدرجات المشاهدة (observed) بتلك الدرجات التي كان ينبغي أن تكون عليها الدرجات الحقيقية، وهذا فانه بقدر كم الفروق بين نوعي الدرجات المعادلة (المشاهدة والحقيقية) تقدر كفاءة طريقة المعادلة، حيث تكون الطريقة التي تغل أو تنتج أقل قيمة من حيث مجموع هذه الفروق بين نوعي الدرجات هي الطريقة الأكفأ ، لأنه كلما اقتربت قيمتا النوعين من الدرجات (المشاهدة والحقيقية) قلت قيمة مجموع الفروق وبالتالي فان ذلك يعد دليلا على كفاءة الطريقة .

وللتوصل الى الدرجات الحقيقية (true score)تتبع عدة أساليب منها :

أ – الدرجات الناتجة عن معادلة التطابق (identity equating) أو ما يسمى المعادلة الدائرية (circular equating) وهي التي تتحول بموجبها الدرجة على الصورة الأولى الى نفسها على الصورة الثانية ، ويتم ذلك بطريقة مباشرة ، أو غير مباشرة عن طريق سلسلة من التعادلات كأن تعادل الصورة أ بالصورة ب ، وتعادل الصورة ب بالصورة ج التي تعادل بدورها بالصورة أ . وهذا المعيار لا بأس به كمؤشر للحكم على كفاءة الطريقة من خلال مقارنة نتائجها عن طريق المعادلة الدائرية أو معادلة التطابق.

ب – توليد بيانات (generated or simulated data) لما ينبغي أن تكون عليه الدرجات نظريا طبقا لكل من مستويات صعوبة المفردات ومستويات قدرة الأفراد المستجيبين. وتعتمد كفاءة الطريقة على مقدار ما ينتج عنها من فروق بين الدرجات المشاهدة أو

الملاحظة والبيانات الحقيقية التي تم توليدها ، وكلما قلت قيم هذه الفروق دل ذلك على كفاءة أكبر وبالعكس.

ج – الاعتماد على عينات كبيرة الحجم (ذات أعداد كبيرة) كبديل للمجتمع الأصلي، بحيث تكون البيانات الناتجة عن تطبيق هذه العينات الكبير بمثابة البيانات الحقيقية التي تتم المقارنة وفقا لها.

د – المقارنة مع دوال الاستجابة علي الاختبار:

(Test Response Functions : TRF's) أي أنه تحسب دوال الاستجابة على الاختبار طبقا لمستويات قدرات المستجيبين "سيتا" ومستويات صعوبة المفردات من الناحية النظرية (ما ينبغي أن يكون) ومقارنتها بدوال الاستجابة الامبريقية (الناتجة عن التطبيق على العينة) وتعتبر الفروق بين هذه الدوال بمثابة مؤشر على كفاءة الطريقة.

ومن الدراسات التي استخدمت أسلوب معادلة التطابق كمعيار لحساب الدرجات الحقيقية التي يتم المقارنة بها، دراسات كل من:

[Von Davier & Wilson, 2008 ; Wiley, 1999,]

أما الدراسات التي استخدمت أسلوب توليد البيانات أو ما يعرف بالبيانات المحاكية فمن أمثلتها دراسة كل من: (Kim, Brennan & Kolen, 2005)

أما الدراسات التي استخدمت العينات الكبيرة الحجم فلعل أشهرها دراسة (Kolen, 1982)، والدراسات التي استخدمت أسلوب المقارنة بين دوال الاستجابة على الاختبار دراسة (Kolen & Brennan, 2004)

(٤) طريقة حساب المعادلة : لاشك أنه من بين أهم المتغيرات التي تؤثر في معادلة الاختبار الطريقة التي تتبع لحساب الصيغة الرياضية المعبرة عن المعادلة ، والتي بتطبيقها تنتج الدرجات المتعادلة ، وتتعدد الطرق كثيرا وتتنوع بحيث قد يصعب حصرها ، إلا أن بعضها يندرج تحت النظرية التقليدية (الكلاسيكية) للقياس (CTT) : Classical Test Theory وبندرج بعضها الآخر تحت النظرية المعاصرة (الحديثة) للقياس المعروفة باسم " نظرية الاستجابة للمفردة : " Item Response Theory "، وتحت كل نظرية من النظريتين توجد العديد من طرق حساب المعادلة ، ولكل طريقة منطقها وفكرها الرياضي الذي تستند إليه الصيغة الرياضية ، ومما تجدر الإشارة إليه ذكر أمثلة لبعض هذه الطرق التي تكرر استخدام الباحثين لها في بحوثهم ، وهي

موضحة في الشكل (١) ، وعلى الرغم من أن هناك بعض الدراسات التي توصلت نتائجها الى أنه لا توجد فروق بين طرق معادلة الاختبار ، إما لأن النتائج تكاد تكون متشابهة أو متقاربة ، أو أن كل طرق المعادلة المستخدمة خلال الدراسة ينتج عنها نتائج معقولة أو مقبولة ، وهي دراسات ليست بالقليلة ، إلا أن هناك دراسات أخرى كثيرة أيضا أوضحت نتائجها تفضيل طريقة ما – أو عدة طرق – على طرق أخرى ، ومن أمثلة دراسات النوع الأول دراسة كل من :

[Tong & Kolen, 2005; Tsai, et al., 2001; Wiley, 1999 & Harris & Kolen,1986]

أما بالنسبة للنوع الثاني ، فهناك كثير من الدراسات في مجال المقارنة بين طرق معادلة الاختبار تؤكد نتائجها على تفوق طرق معادلة الاختبار وفق نظرية الاستجابة للمفردة (IRT) على طرق المعادلة وفق النظرية الكلاسيكية أو التقليدية ، ومن أمثلة هذه الدراسات دراسة كل من :

[Quenette, 2006; Kim, Hanson, 2002; Wright & Dorans, 1993; Skagges & Lissitz, 1986 & Cook, 1983] .

وفيما يلي طرق المعادلة التي يشيع استخدامها من قبل الباحثين في بحوثهم طبقا لكل نظرية من نظريتي القياس على حده:

أ – طرق النظرية التقليدية :

(١) طرق المتوسط الحسابي(mean equating) : وتعتمد هذه الطرق على المتوسط الحسابي لدرجات الصور الاختبارية المختلفة ، حيث تعتبر هذه الصور متعادلة إذا ما تساوت المتوسطات الحسابية لدرجاتها ، ومن ثم تحسب الدرجات المعادلة وفقا لذلك ، فمثلا إذا كانت قيمة المتوسط الحسابي لدرجات الصورة الاختبارية الأولى هي ٦٢ درجة بينما قيمة المتوسط الحسابي لدرجات الصورة الاختبارية الثانية هي ٦٦ درجة ، فان ذلك معناه أن الدرجة ٦٦ على الصورة الثانية هي الدرجة المعادلة للدرجة ٦٢ على الصورة الأولى ، وتبعا لذلك تكون مثلا الدرجة ٥٠ على الصورة الأولى معادلة للدرجة ٥٤ على الصورة الثانية ٠٠٠٠٠٠ وهكذا٠ وتعد طرق المتوسط الحسابي أبسط الطرق لاعتمادها على المتوسط الحسابي للدرجات فقط .

(٢) الطرق الخطية (linear equating) :وتعتمد هذه الطرق على كل من المتوسط الحسابي والانحراف المعياري لدرجات الصور الاختبارية وتعتبر الصور متعادلة إذا ما تساويت قيمة

كل من المتوسط الحسابي والانحراف المعياري لدرجاتها ، وتعتمد الصيغة الرياضية لطرق المعادلة الخطية على معادلة الخط المستقيم بالصورة الرياضية العامة ص = أ س + ب حيث تمثل ص الدرجة على الاختبار الثاني المعادلة للدرجة على الاختبار الأول س ، أما أ فتمثل ميل الخط المستقيم (slope) ، وتعبر ب عن قيمة الجزء المقطوع (intercept)

ومن ثم تعتمد الطرق الخطية لمعادلة الاختبار على كل من الميل والجزء المقطوع ، ففي المثال السابق الخاصة بمعادلة درجات الحرارة الفهرنهاتية بالدرجات السيليزية نجد أن المعادلة هي: ص = ٢ س + ٣٢ أي أن الميل = ٢ والقاطع (ويسمى الثابت constant) = ٣٢ وبموجب هاتين القيمتين (الميل والقاطع) يمكن معرفة الدرجة على اختبار ما المعادلة لدرجة ما على الاختبار الآخر . ومن أشهر أمثلة الطرق الخطية طريقة توكر الخطية Tucker Linear وطريقة ليفين الخطية Levin Linear ، وغيرهما الكثير .

(٣) طرق المئينيات (equipercentile equating) ، ولا تعتمد طرق المئينيات في معادلة الاختبار على المتوسط الحسابي والانحراف المعياري فقط ، وإنما تعتمد بالإضافة الى ذلك على شكل التوزيع التكراري للدرجات حيث تعتبر الصورتين الاختباريتين متعادلتين إذا تساوت الرتب المئينية (percentiles) لدرجاتيهما ، فإذا كان المئيني ٣٢ مثلا في الاختبار الأول تقابله الدرجة ٦٧ ، فان الدرجة المعادلة لها على الاختبار الثاني تكون هي الدرجة المقابلة للمئيني ٣٢ في هذا الاختبار الثاني ، وهكذا لبقية الدرجات عند جميع الرتب المئينية . وتتعدد كذلك طرق المئينيات لعل من أشهرها طريقة توكر المئينية وطريقة ليفين المئينية وغيرهما ، ويوجد نوعان من الطرق المئينية هما : المئينيات الممهدة (smoothed) ، وغير الممهدة (unsmoothed) ، وقد تعتمد طريقة معادلة المئينيات على التمثيل البياني (graph) كما قد تعتمد على الطريقة التحليلية الرياضية التي هي أدق في نتائجها من طريقة التمثيل البياني.

ب – طرق نظرية الاستجابة للمفردة (IRT) :

هناك العديد من الطرق التي تندرج تحت هذه النظرية ، وهناك ثلاثة نماذج لوغارتمية (لوجستية) لتقدير قيم المعلمات (parameters)هي النموذج أحادي المعلمة ويشتهر باسم " نموذج راش : Model ، والنموذج ثنائي المعلمة ، والنموذج ثلاثي المعلمة :

(One-,two-and three-logistic parameter model: 1PL, 2PL & 3PL)

أما الطرق التي تندرج تحت نموذج واحد أو أكثر من هذه النماذج اللوغاريتمية الثلاثة فهي عديدة ومتنوعة لعل أشهرها طريقة المتوسط / المتوسط (mean/mean) وطريقة المتوسط/سيجما(mean /sigma) وطريقة المتوسط المتزامن (concurrent mean) وغيرها من الطرق العديدة ، ولكل منها صيغته الرياضية الأكثر تخصصا .

ويوضح الشكل التالي ملخصا لأهم طرق معادلة الاختبار سالفة الذكر :

طرق معادلة الاختبار (الشهيرة)

وفق النظرية (CTT)	وفق النظرية (IRT)
▪ طرق المتوسط الحسابي:	طرق تتبع نموذجا أو أكثر
طريقة المتوسط الحسابي	من النماذج اللوغاريتمية الثلاثة:
▪ الطرق الخطية:	(أحادي – ثنائي – ثلاثي
الطريقة الخطية العادية	المعلمة1PL, 2PL & 3PL) :
طريقة توكر الخطية	. طريقة المتوسط / المتوسط
طريقة ليفين الخطية.	طريقة المتوسط / سيجما
▪ طرق المئينيات :	. طريقة المتوسط المتلاق أو
المئينيات غير الممهدة	المتزامن .
المئينيات الممهدة	[mean/mean, mean/sigma,
	Concurrent mean & Post-Hok]

شكل (١)

طرق معادلة الاختبار(Test equation methods)

(٥) **حجم العينة :** يعتبر حجم العينة المستخدمة بمثابة محدد له آثاره على اختيار الطريقة التي تستخدم في المعادلة ، فمثلا توصلت نتائج بعض الدراسات الى أن بعض طرق معادلة الاختبار لا يصلح استخدامها ، أو هي تعطي نتائج أقل دقة وموثوقية إذا ما كانت العينات

المستخدمة صغيرة الحجم (ذات أعداد قليلة) ، حيث أن من إشتراطات استخدام مثل هذه الطرق كبر حجم العينات المستخدمة ، وذلك مثل طرق المعادلة وفقا لنظرية الاستجابة للمفردة (IRT) للنموذج اللوغاريتمي ذي المعلمات الثلاث (three- (3PL parameter logistic modelومن أمثلة هذه الدراسات دراسة كل من مراد والشافعي (١٩٩٨) ، وعيد (Eid, 2004) وبارشال ورفاقه (Parshall, et al., 1995) إلا أنه على العكس من ذلك تماما فقد توصلت درسات أخرى في نتائجها الى أن بعض الطرق يصلح استخدامها عندما تكون العينات صغيرة العدد حيث تكون النتائج أكثر دقة ، ومن أمثلة هذه الدراسات دراسة كل من عبابنة (٢٠٠٤) ودراسة كولن (Kolen, 1981). إلا أنه مع ذلك ، فقد توصلت نتائج دراسات عديدة الى أنه لم يكن لحجم العينة أثره الدال على نتائج المعادلة ، أي أن النتائج التي نتجت عن تطبيق بعض طرق المعادلة المختلفة لم تختلف باختلاف أحجام العينات المستخدمة (كبيرة – صغيرة) ، ومن أمثلة هذه الدراسات دراسة كل من :

[Dorans, et al,2008; Petersen, 2008; Cook & Petersen, 1986 & Kolen & Whitney, 1982]

(٦) المؤشر الاحصائي المستخدم في الحكم : تتعدد المؤشرات الاحصائية التي يستخدمها الباحثون في دراساتهم لتقييم نتائج تطبيق طرق المعادلة والحكم على كفاءتها ، ولكل مؤشر من هذه المؤشرات منطقه ، فقد استخدم بعض الباحثين مؤشر " مربعات فروق المتوسطات الجذري"(root-square-mean differences : RSMD) ، واستخدم آخرون " خطأ مربعات فروق المتوسطات " (root-square-mean error : RSME) ، كما استخدم بعض الباحثين ما يعرف بالخطأ المعياري (CMSE) ، وهناك من الباحثين من استخدم ما يعرف ب (Bootstrap Standard error) ، واستخدم آخرون مؤشرا يعرف بالصدق التقاطعي (cross validation) وغير ذلك من المؤشرات الإحصائية التي تستخدم كمعايير للحكم على نتائج طرق المعادلة وكفاءتها ولكل مؤشر من هذه المؤشرات صيغته الرياضية التي يحسب وفقا لها والتي لا يتسع المقام هنا لعرضها ،

ومن الدراسات التي استخدمت مؤشر RSMD دراسة (Von Davier, 2007) أما الدراسات التي استخدمت مؤشر RSME فمن أهمها دراسة (Kolen & Whitney,1982) وغيرها الكثير من الدراسات الأخرى ،ومن الدراسات التي استخدمت مؤشر (CMSE) دراسة (Kim, et al., 2002)، والتي استخدمت مؤشر (Bootstrap Standard Error) دراسة (Tsai, et al., 2001)، أما الدراسات التي

استخدمت مؤشر الصدق التقاطعي (cross validation) فمن أمثلتها دراسة كل من : [Petersen, Cook & Stocking, 2002; Slindr & Linn, 1988 Skages & Lissitz, 1986] . ولا شك أن هناك الكثير من المؤشرات الأخرى التي استخدمت كمعايير للحكم على نتائج طرق المعادلة وكفاءتها والتي بلا شك لها تأثيرها باعتبارها من المتغيرات التي تؤثر على اختيار طريقة دون أخرى من طرق معادلة الاختبار.

(٧) البرنامج الحاسوبي المستخدم في تحليل النتائج ، توجد العديد من برامج التحليل الإحصائي الحاسوبية (software) التي تستخدم في حساب المعادلة بالطرق المختلفة ، ومن الطبيعي ألا تظهر هذه الطرق نفس النتائج باستخدام البرامج الحاسوبية المختلفة، حيث تختلف نتائج تطبيق الطريقة الواحدة من برنامج الى آخر ، ومن ثم يعتبر البرنامج الحاسوبي المستخدم في التحليل الإحصائي وحساب معادلة الاختبار من المتغيرات المؤثرة على النتائج ، ومما يجدر ذكره أنه توجد غزارة الآن في البرامج الحاسوبية الجاهزة الممكن استخدامها في تحليل البيانات وحساب معادلة الاختبار وفقا لكل من نظريتي القياس (الكلاسيكية والاستجابة للمفردة) وبوجه عام فانه من أمثلة البرامج الحاسوبية للتحليل الإحصائي عموما وفق النظرية الكلاسيكي البرامج التالية :

SPSS; ITEMAN; SAS; LISREL; TESTFACT & LERTAP وغيرها أما أشهر برامج التحليل الإحصائي طبقا لنظرية الاستجابة للمفردة (IRT) فمن أمثلتها برامج النموذج اللوغاريتمي ذو المعلمة الواحدة الذي يشتهر باسم نموذج راش "Rasch" وهي : WINSTEPS; BIGSTEPS; RASCH; RUMM; ConQuest; RASCAL FACETS وغيرها ، أما برامج النماذج اللوغاريتمية الثلاثة (أحادي – ثنائي – ثلاثي المعلمات one-, two- and three-parameter Models : 1PL, 2PL & 3PL) فمن أمثلتها برامج : BILOG; BILOG-MG; MULTILOG & PARSCAL وبعض ما سبق هو برامج تحليل إحصائي عامة ، والبرامج المتخصصة في مجال معادلة الاختبار منها على وجه التحديد هي (Kolen & Brennan, 2004) : [CIPE; ST; PIE; POLYQUATE; POLYSCEM; LEGS; PIE] وغيرها. أما البرامج الحاسوبية التي تختص بتقييم طرق معادلة الاختبار والحكم على نتائجها وفق معايير ومؤشرات إحصائية مختلفة طبقا لبعض الجوانب فيما يتعلق بالنتائج التي تظهرها كل طريقة من طرق المعادلة فمن أمثلتها ما يلي (Zimowski, et al., 2008) :

[BILOG; MULTILOG; BILOG-MG; RUMM; FACETS; ConQuest; RASCAL; PARSCALE & WINSTEPS/BIGSTEPS]

خامساً – أساليب تقييم المعادلة للحكم على فاعليتها :

بالنسبة لأساليب تقييم المعادلة، فقد استخدم الباحثون في دراساتهم أساليب وطرقا عديدة لتقييم المعادلة والحكم علي فاعليتها adequacy of the equating should how the be assessed)، وفيما يلي توضيح لأشهر الأساليب التي شاع استخدامها من قبل الباحثين في دراساتهم لتقييم معادلة الاختبار(Harris & Crouse, 1993):

(١) ضعف التساوي(weak equity) وهو حالة خاصة من مفهوم المساواة عند "لورد" Lord's equity، ويقوم هذا الأسلوب في الحكم على المعادلة من خلال درجة تساوي التوزيعات التكرارية لدرجات صورتي الاختبار ، فبقدر تساوي هذه التوزيعات يتم الحكم على المعادلة. وممن استخدموا هذا الأسلوب كل من ديفجي (Divgi, 1981) و "ين" (Yen, 2008) ، ولعل من مزاياه أنه يطبق في إطار مفهوم محدد للمعادلة التي تتم من خلال تساوى التوزيع التكراري لدرجات الصور الاختبارية. أما من عيوب هذا الأسلوب صعوبة الشرح والتفسير للنتائج .

(٢) المؤشرات (indices) ويقوم هذا الأسلوب على استخدام بعض المؤشرات لمقارنة مجموعتين من الدرجات المحولة وفق نفس المعادلة (conversion)، ومن أمثلة هذه المؤشرات ما يعرف بخطأ مربع المتوسط الجذري : (root mean squared error RMSE) ويعد أشهر من استخدم هذا الأسلوب في دراساته أنجوف(Angoff, 1984) ومن أهم مزاياه سهولة تفسير النتائج ، إلا أنه يعاب عليه عدم تحديد درجة جوهرية الفروق الناتجة.

(٣) الخطأ المعياري (standard error) ويقوم هذا الأسلوب على تحديد مقدار كم الخطأ الذي ينتج عن التطبيق على عينات مختلفة , وممن استخدم هذا الأسلوب أيضا في دراساته في السبعينيات أنجوف (Angof, 1984) ومن أهم مزاياه سهولة تطبيقه وتفسير نتائجه, إلا أنه يعاب عليه تجاهله الخطأ النظامي (systematic error) الذي ينتج عن أسباب أخرى غير خطأ المعاينة (sampling error) .

(٤) البيانات المحاكية أو المولّدة (generated or simulated data) ويقوم على مقارنة نتائج المعادلة التي نتجت عن التطبيق العملي الفعلي بما كان يجب أن تكون عليه النتائج نظريا وفق بيانات يتم توليدها لما ينبغي أن تكون عليه النتائج ، وممن استخدم هذا الأسلوب لورد (Lord, 1983) ومن مزاياه أنه يحاول إيجاد علاقة تعادل حقيقية لما ينبغي أن تكون

عليه المعادلة نظريا لمطابقة النتائج الواقعية بها ، ومن ثم يسهل تقييم المعادلة طبقا لدرجة التطابق أو التقارب بين نوعي النتائج (النظرية والواقعية) ، أما عيوب هذا الأسلوب فتتمثل في التحيز الذي قد يشوب عملية توليد البيانات وعلاقتها بالبيانات الواقعية.

(٥) معادلة الاختبار بنفسه (equating a test to itself) ، ويتم هذا الأسلوب بطريقة مباشرة أو غير مباشرة من خلال إجراء سلسلة من المعادلات لعدة صور اختبارية متداخلة ، وممن استخدم هذا الأسلوب لورد (Lord, 1983) ، ومن مزاياه أن العلاقة التعادلية الحقيقية تكون معروفة حيث يمكن مقارنة نتائج التعادل الإمبريقي بها ، إلا أن من عيوبه أن التعادل قد لا يتم بشكل جيد دائما

(٦) استخدام عينات كبيرة (large sample) ، ويقوم هذا الأسلوب على استخدام عينات ذات أحجام كبيرة العدد بديلا عن المجتمع الأصل (population) تتخذ كأساس لمقارنة نتائج التعادل بها ومن ثم تقارن التحويلات من خلال العينات الكبيرة بالتحويلات الناتجة عن التعادل المطبقة على عينات البحث ، وممن استخدم هذا الأسلوب في دراساته أنجوف (Angoff, 1984) ، ومن مزايا هذا الأسلوب سهولة التطبيق والتفسير أما من عيوبه فهو أنه قد لا تتاح لدينا دائما عينات ذات أحجام كبيرة يسهل ا لحصول عليها في الواقع العملي.

(٧) الاتساق (consistency)، ويقوم هذا الأسلوب على مقارنة النتائج التي يسفر عنها تطبيق عدة طرق مختلفة للمعادلة لمعرفة درجة الاتساق بينها ، وبقدر هذه الدرجة يكون الحكم على المعادلة ، ويشيع استخدام هذا الأسلوب لدى العديد من الباحثين في بحوثهم حول معادلة الاختبار ، ومن أهم مزاياها سهولة تطبيقها ، إلا أن من أهم عيوبها عدم الاهتمام بدقة النتائج.

(٨) الاستقرار (stability) ، ويقوم هذا الأسلوب على إعادة التطبيق للصور الاختبارية على عينات مختلفة في أوقات متعددة ومقارنة النتائج ببعضها البعض للحكم على درجة استقرار النتائج ويستخدم كمؤشر للحكم وفق هذا الأسلوب مؤشر يعرف ب "التصدق التقاطعي : cross validation " ، ومن أشهر من استخدم هذا الأسلوب من الباحثين أنجوف (Angoff, 1984)، ومن مزاياه سهولة التطبيق والتفسير، إلا أن من أهم عيوبه عدم قياس الدقة لاقتصاره على قياس الاستقرار دون الدقة.

(٩) درجة الواقعية أو القابلية للتطبيق (practical) وهذا الأسلوب يقوم على فكرة إمكانية

التطبيق في الواقع العملي وان كان ليس له طريقة واحدة محددة ، فهناك مثلا طريقة تسمى rounded scale scores ، وأخري تسمى conversion tables for gaps ، ومن مزايا هذا الأسلوب سهولة التطبيق، أما عيوبه فهو قلة الموضوعية وقد استخدمه قلة من الباحثين لعل من أشهرهم أنجوف Angoff, 1984.

إجراءات تطبيق الدراسة :

أولاً – أدوات الدراسة :

فيما يلي الأدوات التي تم استخدامها خلال الدراسة:

(١) اختبار "توني" للذكاء:

تقوم الدراسة الحالية على استخدام أداة لقياس الذكاء تتمثل في اختبار توني للذكاء غير اللفظي ، وقد قام بتصميم هذا الاختبار توني (Toni) عام ١٩٨٣ كل من براون وشيربينو ودولار (Brown, Sherbenou & Dollar) ويعتبر من الاختبارات الجيدة للذكاء المتحررة من العوامل الثقافية، وعدم تقيده بزمن محدد للإجابة عن مفرداته (فهو ليس اختبار سرعة) ، وسهولة تطبيقه ، وكذلك صلاحيته للتطبيق الجمعي والفردي . ومما يجعل هذا الاختبار مناسبا للدراسة الحالية أن له صورتين (صورة أ وصورة ب) بحيث تمثل كل صورة منهما اختبارا مستقلا (مختلفة من حيث المحتوى ومتكافئة من حيث الصعوبة) بما يفيد في إجراء دراسة لإجراء المعادلة (equating) بين صورتيه .

وتتكون كل صورة من صورتي الاختبار من (٥٠) مفردة مرتبة بطريقة متدرجة حسب مستوى الصعوبة ، كما تضم هذه المفردات ستة محاور لكل منها مثال يوضح شكلا من الأشكال التي يحويها الاختبار بمحاوره الستة . ويقوم اختبار توني على أسلوب حل المشكلات (problem solving). حيث يطلب من المستجيب حل المشكلات المتضمنة في الاختبار عن طريق إدراك العلاقات بين الأشكال المجردة التي تشتمل عليها مفردات الاختبار. والمفردة الواحدة تتضمن عدة أشكال منها شكل واحد ناقص ، وعلى المستجيب استكمال النقص باختيار الشكل الصحيح الذي يكمل هذا النقص من خلال عدد من البدائل أو المشتتات المتاحة لكل مفردة من المفردات . والأشكال التي تضمها مفردات الاختبار تقوم على خاصية أو أكثر من خصائص الأشكال مثل خواص: الموقع المكاني – الاتجاه – التدوير – التماس – التظليل – الحجم – الطول – الحركة. وعلى المستجيب إدراك بعض العلاقات طبقا لهذه الخصائص لتحديد الإجابة الصحيحة

من خلال مجموعة من العمليات أو الحركات، ولعل أهم هذه العلاقات التي يجب على المستجيب إدراكها كالتناظر أو التقابل البسيط (simple matching).

والتناظر أو التقابل (analogies) الذي يقوم على عمليات متعددة كالتجانس، والإضافة addition والطرح subtraction والتعديل أو التغيير alteration والتوالي progression والتصنيف classification والتقاطعات intersections والمتواليات أو المتتابعات progressions (محمد شوكت, ١٩٨٧).

وهذا الاختبار مناسب للأفراد من الفئة العمرية ١٥ – ٢٤ عاما (طلاب المرحلة الثانوية والمرحلة الجامعية) ويطبق بصورة فردية أو جمعية دون التقيد بزمن (أو وقت) محدد للإجابة (فهو ليس اختبار سرعة) . ويبدأ الاختبار بعدد من الأمثلة يتولى القائم بالتطبيق تقديمها للمستجيب (المستجيبين) يتعرفون من خلال حلها على طريقة الإجابة فيما بعد عن مفردات الاختبار.

وكل مثال عبارة عن رسم به عدة أشكال وفيه شكل ناقص ، وعلى المستجيب أن يختار شكلا لتكملة هذا الرسم ووضعه مكان الشكل الناقص من بين عدة أشكال معطاة له لكي يصبح الشكل الكلي متسقا في وحداته أو أشكاله الفرعية طبقا لخاصية أو أكثر من الخصائص السابق ذكرها (كالموقع أو الاتجاه أو التدوير أو التماس أو التظليل أو الحجم أو الطول أو الحركة ..... الخ)، وبناء على ما يقوم به المستجيب من عمليات لإدراك العلاقة بين الأشكال (عن طريق التقابل أو التناظر أو التصنيف أو التوالي وما يستلزمه ذلك من عمليات كالتجانس أو الإضافة أو الطرح أو التعديل على النحو السالف ذكره)يصل المستجيب الى الإجابة التي يختارها .

وبعد دراسة الأمثلة يبدأ المستجيب في الإجابة عن مفردات الاختبار التي يبلغ عددها (٥٠) مفردة الواحدة تلو الأخرى، وذلك بوضع رمز الشكل الذي يختاره (أ أو ب أو ج أو د أو ه ..... الخ) ليوضع مكان الشكل الناقص في كل مفردة . وينصح المستجيب بعدم ترك أية مفردة دون إجابة ، ويترك دون تحديد وقت أو زمن محدد يقيده للانتهاء من الإجابة عن جميع مفردات الاختبار . ويتم تصحيح الاستجابات طبقا لمفتاح تصحيح يحدد الإجابة الصحيحة لكل مفردة من المفردات الخمسين وتعطى الإجابة الصحيحة الدرجة ١ بينما تعطى الإجابة الخاطئة صفرا , ومن ثم يتراوح المدى الكلي لدرجات المستجيب ما بين صفر الى (٥٠) درجة (محمد شوكت , ١٩٨٧).

ولعل من المتعارف عليه بين رجال علم النفس بصفة عامة أنه لا يوجد اتفاق فيما بينهم حول ماهية الذكاء وتعريفه أو طبيعته , وربما يرجع ذلك الى أسباب عديدة لعل من أهمها أن مفهوم الذكاء نشأ – شأنه شأن علم النفس – في أحضان الفلسفة ثم بدأت تهتم به علوم أخرى مثل البيولوجي والفسيولوجي حيث استقر أخيرا في مجال علم النفس .

كذلك تتباين النظرة الى الذكاء فهناك من ينظر إليه من حيث الوظيفة في حين ينظر إليه البعض الآخر من حيث المكونات ولذلك تعددت مفاهيم الذكاء وتنوعت تعريفاته بتنوع أو تعدد النظرة إليه, فمن الباحثين من ينظر الى الذكاء على أنه قدرة على التكيف أو المواءمة مع البيئة , ولكن آخرين يرون أن في هذه النظرة الى الذكاء إغفال للكثير من النشاطات العقلية الأخرى التي تظهر بشكل واضح في حياة الإنسان بصورها المختلفة (أبو حطب , ١٩٨٣).

ولعل من ينظر الى التعاريف التي أوردها الباحثون للذكاء يدرك أن ما يوجد بينها من اختلاف أكثر مما بينها من اتفاق , وعلى الرغم من ظهور اتجاه ينظر الى الذكاء كقدرة عقلية عامة , أو كعامل عقلي عام يظهر في كل الأنشطة العقلية بحيث يدل على قدرة عامة مشتركة بين عدة قدرات عقلية فرعية أولية , وبحيث يكون الذكاء هو عامل العوامل أو القدرة العقلية العامة التي تتكون من عدة قدرات عقلية فرعية ,

ومن ثم يمكن – وفقا لهذه النظرة – النظر الى الذكاء باعتباره عاملا عاما يدخل في جميع العمليات العقلية المعرفية , إلا أنه ظهرت في الآونة الأخيرة نظريات أخري مغايرة لعل من أشهرها نظرية " جاردنر " التي ترى غير ذلك حيث تأخذ بفكرة تعدد القدرات لدى الفرد – أي فرد – واعتبار كل قدرة من هذه القدرات بمثابة ذكاء , وهو ما قد يعبر عنه البعض بالذكاءات المتعددة ، وباختصار فانه ليس من بين أهداف الدراسة الحالية الخوض في موضوع الذكاء كإشكالية جدلية حيث أن صلة الدراسة الحالية بموضوع الذكاء تكاد تنحصر في أداة الدراسة التي يتم معادلة صورتيها , ومن ثم فهي تتناولها كأداة لقياس الذكاء بصرف النظر عن وجهة النظر التي اعتمدت عليها في تعريفها للذكاء .

وقد قام بالتحقق من صدق وثبات الاختبار وتقنينه في البيئة المصرية الدكتور محمد شوكت عام ١٩٨٧ على عينة قوامها (٧٤٦) طالبا من طلاب المدارس الثانوية والمرحلة الجامعية ممن تراوحت أعمارهم ما بين ١٥ – ٢٤ عاما ،

واستخدم طريقة الاتساق الداخلي لحساب صدق الاختبار وسلامة التماسك الداخلي له , وقدرته على التمييز بين المجموعات المتضادة (في أدني وفي أعلى التوزيع) , ولحساب الصدق التلازمي تم حساب معاملات الارتباط بين درجات هذا الاختبار واختبارات مقننة أخرى للذكاء (اختبار الذكاء العالي للسيد محمد خيري , واختبار الذكاء المصور لأحمد زكي صالح , واختبار كاتل للذكاء لعبد السلام عبد الغفار) ,

كذلك تم حساب معاملات الارتباط بين الدرجات على الاختبار ودرجات التحصيل الدراسي العام لأفراد العينة , ومن ثم يشير الباحث الى أن الاختبار (بصورتيه) ولكل صورة على حده يتصف بقدر عال من الصدق سواء من حيث اتساقه الداخلي , أو قدرته على التمييز بين المجموعتين الطرفيتين أو الصدق التلازمي (محمد شوكت , ١٩٨٧) . أما من حيث الثبات فقد استخدم الباحث طريقتي إعادة تطبيق الاختبار (test-retest) والتجزئة النصفية , وذلك بتطبيقه على عينة أخرى قوامها ١٨٠ طالبا من المرحلة الثانوية و ٢١٩ طالبا من المرحلة الجامعية , وطبق مرتين بفاصل زمني قدره ٢١ يوما بين مرتي التطبيق, وتبين من قيم معاملات الارتباط الناتجة أن الاختبار (بصورتيه) ولكل صورة على حده يتسم بدرجة ثبات عالية (محمد شوكت , ١٩٨٧) .

ومن ثم فان الباحثة تطمئن الى توافر درجات صدق وثبات مناسبة للاختبار بصورتيه قبل البدء في استخدامه وتطبيقه خلال الدراسة الحالية وهو ما تحرص عليه الباحثة للوفاء بأحد أهم شروط معادلة الاختبار والتي سوف يتم الإشارة إليها تفصيلا فيما بعد .

(٢) اختبار الجذع المشترك (Anchor Test):

حتى يتسنى إجراء معادلة بين صورتي الاختبار ، كان لابد من توافر مفردات مشتركة (common items) بين الصورتين ، حيث تقوم مثل هذه المفردات بوظيفة هامة في المعادلة حيث أنه بافتراض أن أداء مجموعتين من المستجيبين على صورتي اختبار ما كان متباينا ، فإلى ما يرجع هذا الاختلاف ؟ هل الى اختلاف مستويات قدرة أفراد مجموعة ما عن الأخرى ؟ أم الى اختلاف مستوى صعوبة إحدى الصور الاختبارية على الأخرى ؟ ، وهنا يكون دور الاختبار المشترك ، فإذا ما كان أداء أفراد المجموعتين عليه متساويا ، فان التباين يرجع الى اختلاف

مستوى الصعوبة بين صورتي الاختبار ، أما إذا كان الأداء متباينا على الاختبار المشترك فان ذلك معناه تباين مستويات قدرة أفراد المجموعتين، ولذلك فقد تم بناء اختبار جذعي مشترك خارجي (anchor test) يتكون من ٢٥مفردة منها ١٣مفردة مأخوذة من مفردات الصورة (أ) للاختبار الأصلي بطريقة عشوائية و ١٢مفردة من مفردات الصورة (ب) أيضا بطريقة عشوائية ، وفيما يلي أرقام المفردات التي تكوّن منها الاختبار الجذعي المشترك الذي تم تكوينه ليطبق على جميع أفراد مجموعات الدراسة بجانب الاختبار الأصلي بصورتيه :

الصورة	أرقام المفردات المكوّنة للاختبار الجذعي المشترك
أ (A)	٣، ٤ ، ٧ ، ٩ ، ١١ ، ١٦ ، ٢٣ ، ٢٤ ، ٢٧ ، ٣٠ ، ٣٩ ، ٤٠ ، ٤٩.
ب (B)	٢ ، ٣ ، ٥ ، ٦ ، ٨ ، ١٥ ، ١٦ ، ٢٠ ، ٢٥ ، ٣٠ ، ٣٥ ، ٤٣ .

ثانياً – عينة الدراسة :

تمثل عينة الدراسة الطلاب الجامعيين من كليتين جامعيتين وطلاب المرحلة الثانوية بخمس مدارس ثانوية عامة في ثلاث محافظات من محافظات الجمهورية للعام الدراسي ٢٠٠٨ / ٢٠٠٩, ويوضح جدول (١) خصائص أفراد عينة الدراسة الذين طبق عليهم اختبار الذكاء الأصلي بصورتيه واختبار الجذع المشترك.

جدول (١)

يوضح أعداد أفراد العينة وقيم المتوسطات الحسابية والانحرافات المعيارية لأعمارهم موزعين حسب المرحلة التعليمية

الكلية الجامعية/ المدرسة	عدد الطلاب	المتوسط الحسابي للعمر	الانحراف المعياري للعمر
▪ المرحلة الجامعية:			
كلية التربية/ جامعة عين شمس	٣٨٩		
معهد الدراسات والبحوث التربوية /جامعة القاهرة	٣٩١		

▪ المرحلة الثانوية:			
مدرسة ناصر الثانوية بنات إدارة الساحل محافظة القاهرة	١٦١		
مدرسة شبرا الثانوية بنات إدارة شبرا محافظة القاهرة	١٥١		
مدرسة أم الأبطال الثانوية بنات إدارة.العمرانية	١٤٥		
مدرسة الهرم الثانوية بنات إدارة الهرم محافظة الجيزة	١٥٢		
مدرسة المعادى الثانوية بنات.			
ادارة المعادى محافظة حلوان	١٧١		
المجموع الكلي	١٥٦٠		

ثالثاً – خطوات جمع البيانات وتهيئتها للتحليل الإحصائي:

قامت الباحثة بالخطوات التالية لجمع البيانات وتهيئتها للتحليل الإحصائي لإمكان إجراء المعادلة وتحقيق أهداف الدراسة :

(١) تطبيق الاختبارات على أفراد العينة :

تم تطبيق اختبار الذكاء بصورتيه ، وكذلك اختبار الجذع المشترك على أفراد العينة في مجموعات بحيث كانت تطبق على كل مجموعة من المجموعات صورة من الصورتين ، بينما طبق اختبار الجذع المشترك على أفراد جميع المجموعات وكان ذلك خلال العام الدراسي ٢٠٠٨/ ٢٠٠٩ ، وقد قامت الباحثة بنفسها بتطبيق الاختبارات على طلاب المرحلة الجامعية ، وعاونها بعض المتخصصين في التطبيق على طلاب المرحلة الثانوية بعد تلقيهم تعليمات التطبيق من قبل الباحثة للاطمئنان الى استيعاب تعليمات التطبيق بدرجة كافية وتوحيدها والإحساس بأهمية الالتزام بها.

(٢) التحقق من توافر افتراضات النظرية IRT :

كان لابد قبل البدء في تحليل النتائج ضرورة الاطمئنان الى توافر افتراضات أو شروط معادلة الاختبار وفق نظرية القياس المعاصرة ولعل أهمها شرط أحادية البعد (unidimensionality) وشرط الاستقلال المكاني أو الموضعي (local independence) ، وللتحقق من أحادية البعد للاختبار تم إخضاعه للتحليل العاملي (tetrachoric) والذي أسفرت نتائجه عن القيم المميزة

(eignvalues) ونسبة التباين المفسر للعوامل الاثني عشرة الأولى ، وهو ما يوضحه جدول (٢) :

جدول (٢)

القيم المميزة والنسب المئوية للتباين المفسر للعوامل

(factors)لاختبار توني للذكاء طبقا للتحليل العاملي بين المفردات

العامل (factor)	القيم المميزة (eignvalues)	نسبة التباين المفسر %
١	٢٠ر٦٥	٢٥ر٨١
٢	٢ر٧٥	٣ر٤٤
٣	٢ر٢٧	٢ر٨٤
٤	٢ر١٠	٢ر٦٣
٥	١ر٨٣	٢ر٢٩
٦	١ر٤٢	١ر٧٨
٧	١ر٣٦	١ر٧٠
٨	١ر٣٥	١ر٦٩
٩	١ر٢٣	١ر٥٤
١٠	١ر٢٢	١ر٥٣
١١	١ر٢١	١ر٥١
١٢	١ر١٢	١ر٤٠

وكما يتضح من جدول (٢) الذي يوضح القيم المميزة(eignvalues) والنسبة المئوية للتباين المفسر بواسطة كل عامل من العوامل الاثني عشر الأول أن نسبة التباين المفسر بواسطة العامل الأول بلغت ٢٥ر٩٨ % ، ومن ثم فانه مادامت تزيد قيمة هذه النسبة عن ٢٠% من التباين المفسر الكلي ، فان ذلك يدل على توافر أحادية البعد للاختبار على النحو السالف ذكره من قبل .

(٣) تصحيح الاستجابات وإدخالها للحاسب :

تم إدخال الاستجابات الى الحاسب وتصحيحها لتقدير الدرجات سواء بالنسبة للاختبار الأصلي (بصورتيه) أو لاختبار الجذع المشترك وذلك بعد أن تم استبعاد الاستجابات غير الكاملة أو غير السليمة ، وقد تم تصنيف أفراد العينة حسب صورة الاختبار المطبقة عليهم الى مجموعتين متساويتي العدد لكل صورة من الصورتين ، ويشار الى إحداهما بالمجموعة الأولي ، وهي التي

طبق على أفراها الصورة (A) والى الأخرى بالمجموعة الثانية وهي التي طبق علـى أفرادهـا الصورة (B) وتبلغ جملة عدد أفراد كل مجموعة منهما (٧٨٠) طالبا وطالبـة مـن المـرحلتين التعليميتين الثانوية والجامعية .

(٤) تحديد طرق المعادلة ومؤشرات الحكم على فاعلية الطرق :

تم تحديد الطرق التي يتم حساب المعادلة وفقا لها بثلاث طرق تتبع القياس التقليدي (وهـي : الخطية – توكر – المئينيات) وطريقتين فقط وفق النموذج اللوغاريتمي أحادي المعلمة (1-PL) تتبع القياس المعاصر (وهما : المتوسط/المتوسط – المتوسط/سيجما)، أما مؤشرات الحكم على فاعلية الطرق فهو مؤشر كولن للصدق التقاطعي بجوانبه الثلاثة (التحيز – عدم الدقة – فـروق المئينيات) .

(٥) اختيار البرامج الحاسوبية التي يستعان بها في التحليل الإحصائي:

تم استخدام البرنامج الحاسوبي (RUMM) لحساب معادلة الاختبار بالطرق المختلفة ، والبرنامج الحاسوبي BILOG-MG لحساب مؤشرات الحكـم على فاعلية هذه الطرق ، بالإضافة الى البرنامج الحاسوبي المعـروف باسـم SPSS لحسـاب نتائـج التحليـل العاملـي (factor analysis) للمفردات للتحقق من توافر افتراضية أحادية البعد.

رابعاً – معالجة البيانات والتحليل الإحصائي:

تم إتباع ما يلي لمعالجة البيانات وتحليلها إحصائيا:

(١) حساب قيم كل من الميل (slope) والجزء المقطوع (intercept) لخط المعادلة ، وكذلك حساب قيم المئينيات المقابلة للدرجات الخام، وفق كل طريقة من طرق المعادلة المستخدمة ، والتي من واقعها يمكن إعداد جداول تحويل الدرجات من إحدى صورتي الاختبار الـى الصورة الأخرى، وقد اتبعت الإجراءات اللازمة لذلك بالتطبيق المباشر للصيغ الرياضـية لمعادلة الاختبار وفق القياس التقليدي ، أما بالنسبة لطرق القياس المعاصر فقـد تـم إتبـاع الخطوات المتضمنة في المراحل المختلفة من تقدير للمعلمات قبل الـربط أو المعـايرة ، وتقدير المعلمات بعد الربط (linked parameters) بين كل من معلمات صعوبة المفردات ومستويات قدرات الأفراد ومعايرتها (calibrated) ووضـعهما علـى قيـاس مـشترك (common scale).

(٢) حساب فاعلية نتائج الطرق طبقا لقيم مؤشر كولن للصدق التقـاطعي (kolen's Cross-

(Validation بجوانبه الثلاثة (التحيز – عدم الدقة –فروق المئينيات) ، وحساب دلالة الفروق بين نتائج الطرق لكل جانب من الجوانب الثلاثة على حده والتي هي على النحو التالي :

أ – التحيز (Bias):

تتحدد قيمة التحيز بحساب مربعات الفروق بين متوسطات الدرجات لكل من الاختبار الجذعي المشترك(anchor test) والدرجات المتعادلة لصورتي الاختبار الأصلي ، وهي تتضمن نوعين من الخطأ هما خطأ المعادلة (equating error) وخطأ المعاينة (sampling error) وطريقة المعادلة التي ينتج عن تطبيقها أكبر قيمة في هذا المؤشر تعد هي الطريقة الأكثر تحيزا من طرق المعادلة الأخرى ، أي أنه كلما قلت قيمة هذا المؤشر كلما كانت طريقة المعادلة أقل تحيزا وبالعكس .

ب– عدم الدقة (Imprecision):

تتحدد قيمة عدم الدقة بحساب مقدار التباين (variance) بين فروق الدرجات للاختبار الجذعي المشترك والدرجات المتعادلة للاختبار الأصلي ، ولا يتضمن هذا المؤشر الفرعي إلا خطأ المعادلة فقط دون خطأ المعاينة كسابقه ، ويفسر بنفس التفسير السابق بحيث أنه كلما زادت قيمة هذا المؤشر دل ذلك على عدم دقة بدرجة أكبر ، بمعنى أنه كلما زادت قيمة هذا المؤشر الناتج عن تطبيق طريقة ما من طرق معادلة الاختبار ، دل ذلك على درجة أكبر من عدم الدقة لهذه الطريقة بالنسبة لغيرها من طرق المعادلة الأخرى .

ج– فروق المئينيات (Percentile Comparison) :

تتحدد قيمة هذا المؤشر الفرعي بحساب الفروق بين الدرجات على الاختبار الجذعي المشترك والدرجات المعادلة المقابلة لها (التي لها نفس الرتب المئينية) على الاختبار الأصلي ، حيث يتم حساب هذه الفروق (وفقا للرتب المئينية) ، وتربيعها ثم جمع المربعات وإيجاد متوسطها. وتفضل طريقة المعادلة على غيرها إذ ما نتج عن تطبيقه لفروق مئينية ذات قيمة أقل مما ينتج عن تطبيق الطرق الأخرى ، أي أن الطريقة الأفضل هي الطريقة ذات القيمة الأقل من حيث فروق الدرجات التي لها نفس الرتب المئينية .

نتائج الدراسة :

تعرض الباحثة فيما يلي النتائج التي أسفرت عنها الدراسة مبوبة حسب التساؤلات التي تم طرحها خلال الدراسة وتهدف الإجابة عنها:

أولاً – النتائج الخاصة بمعادلة صورتي الاختبار وفق الطرق المختلفة:

للإجابة عن التساؤل الأول الذي تم طرحه خلال الدراسة الحالية فيما يتعلق بمعادلة الدرجات بين صورتي اختبار توني للذكاء باستخدام طرق المعادلة المختلفة وفق كل من القياس التقليدي والمعاصر ، كان لابد من حساب قيمة كل من الميل والجزء المقطوع لخط معادلة درجات صورتي الاختبار ، وهو ما يوضحه الجدول (٣) :جدول (٣)

يوضح قيم الميل والجزء المقطوع(slope & intercept)

لخط معادلة صورتي الاختبار وفق طرق المعادلة المستخدمة

طريقة المعادلة	الميل (slope)	الجزء المقطوع(intercept)
القياس التقليدي:		
الخطية	٠ر٩١٥٣	٣ر٢٤٧
توكر	٠ر٨٩٨٨	٥ر٧٨٩
المئينيات	٠ر٩٠٨٠	٦ر٧٠٩
القياس المعاصر:		
المتوسط/المتوسط (M/M)	٠ر٣٠٠	١ر١١٩
المتوسط/الانحراف (M/S)	-٠ر٥١٠	١ر١٣٨

وفي ضوء ما سبق بشأن متغيري خط المعادلة (ميل الخط والجزء الذي يقطعه من الاحداثي الصادي) وفق كل طريقة من الطرق الخمس لمعادلة الاختبار، أمكن إعداد جداول تحويل الدرجات الخام على الصورة الأولي (أ- A) الى ما يعادلها على الصورة الثانية (ب – B)، وهو ما يوضحه الملحق (١) لجميع الدرجات الخام من صفر الى (٥٠) ، والذي من واقعه ويوضح الجدول (٤) جزءا من هذه الدرجات :

جدول (٤)

تحويل بعض الدرجات الخام على الصورة (A) لاختبار توني للذكاء

الى الدرجات المعادلة لها على الصورة (B) وفقا لطرق المعادلة المستخدمة

الدرجة الخام على الصورة (A)	الدرجة المعادلة على الصورة (B)				
	الخطية	توكر	المئينيات	(M/M)	(M/S)
صفر	٣ر٢٤٨	٥ر٧٩٢	٦ر٧٠٩	٠ر٣١٠	٠ر٢٨٠
٥	٧ر٨٢٤	١٠ر٢٨٦	١١ر٢٤٩	٥ر٩٠٠	٥ر٧٠٠
١٠	١٢ر٤٠١	١٤ر٧٧٩	١٥ر٧٨٩	١١ر٤٩٠	١٠ر٨٧٠
١٥	١٦ر٩٧٧	١٩ر٢٧٣	٢٠ر٣٢٩	١٧ر٠٩٠	١٦ر١٦٠
٢٠	٢١ر٥٥٣	٢٣ر٧٦٧	٢٤ر٨٦٩	٢٢ر٦٨٠	٢١ر٤٥٠
٢٥	٢٦ر١٣٠	٢٨ر٢٦١	٢٩ر٤٠٩	٢٧ر١٥٠	٢٥ر٦٩٠
٣٠	٣٠ر٧٠٦	٣٢ر٧٥٥	٣٣ر٩٤٩	٣٣ر٧٦٠	٣١ر٥٤٠
٣٥	٣٥ر٢٨٢	٣٧ر٢٤٩	٣٨ر٤٨٩	٣٨ر٩١٠	٣٦ر٠٥٠
٤٠	٣٩ر٨٥٩	٤١ر٧٤٣	٤٣ر٠٢٩	٤٣ر٢٧٠	٤١ر٦٤٠
٤٥	٤٤ر٤٣٥	٤٦ر٢٣٧	٤٧ر٥٦٩	٤٧ر١٣٠	٤٦ر٦٨٠
٥٠	٤٩ر٠١١	٥٠ر٧٣١	٥٢ر١٠٩	٥١ر٠٣٠	٥٠ر٢٩٠

وبالنظر الى الجدول (٤) والملحق (١) السالف الإشارة إليهما أعلاه ، يلاحظ ما يلي بشأن معادلة درجات صورتي اختبار توني للذكاء وفق طرق معادلة الاختبار المختلفة وهو مايدور حوله التساؤل الأول للدراسة :

(١) درجات التعادل وفق طرق القياس التقليدي تتباعد عن درجات التعادل وفق طرق القياس المعاصر ، فمثلا بينما تتعادل الدرجة صفر على الصورة A الدرجات ٢٤٨ر٣ و ٧٩٢ر٥ و ٧٠٩ر٦ على الصورة B وفقا لطرق المعادلة : الخطية ، وتوكر ، والمئينيات بالترتيب، فإنها تعادل الدرجة ٣١ر٠ و ٢٨ر٠ وفقا لطريقتي (M/M) و (M/S) بالترتيب.

(٢) عند مقارنة نتائج تطبيق طرق معادلة الاختبار وفق القياس التقليدي بعضها ببعض ، يلاحظ أن الفروق تزيد في طريقة توكر عنها في الطريقة الخطية ، ولكنها في طريقة المئينيات تزداد عنها في طريقة توكر .

(٣) عند المقارنة بين طريقتي القياس المعاصر، يلاحظ أن نتائج طريقة (M/S) تتقارب وتقل الفروق فيها عما هو بالنسبة لنتائج طريقة (M/M).

(٤) بينما تتباعد الدرجات المتعادلة بين طرق القياس التقليدي عند الدرجات الدنيا المتطرفة (مثل صفر، ٥، ١٠)، فإنها تتقارب عند الدرجات الوسطي (مثل ٢٥، ٣، ٣٥).ومما يجدر ذكره أنه بعد تقدير المعلمات (parameters estimates) لمفردات صورتي الاختبار ، وهو ما يوضحه الملحق (٣) ، فانه تمت معادلة مستويات قدرات الأفراد " سيتا O "على الصورتين وفق كل طريقة من الطريقتين وهو ما يوضحه الملحق (٢) ، ومن ثم تم تقدير معلمات الربط linked parameters وتدريجها أو معايرتها(calibrated) بمستويات القدرة ووضعهما معا على قياس مشترك (common scale)، وحساب الدرجات المتعادلة باستخدام كل طريقة من طريقتي المعادلة ، وهو ما يوضحه الملحق (٤) ، والذي منه أمكن إعداد الجدول (٥) كجزء من هذا الملحق :

جدول (٥)

معادلة الدرجات ببعض مستويات القدرة"سيتا O" لصورتي الاختبار وفقا لطريقتي المعادلة (M/M) و (M/S)

مستوى القدرة O	(M/M)	(M/S)
-٤ر٠	١ر٦٣٣	١ر٥١٣
-٣ر٠	١ر٨٣٥	٢ر٠٠٧
-٣ر٠	٢ر١١٥	٣ر١١٢
-٢ر٠	٣ر٢٠٠	٣ر٧٨٥
-٢ر٠	٤ر١٤٨	٥ر١٩٨
-١ر٠	٨ر٢١٥	٩ر١١٩
-١ر٠	١٢ر١١١	١٣ر٠٠١
-٠ر٠	٢٠ر٧٨٧	٢١ر٠٩٧
صفر	٢١ر٢٢٥	٢٢ر١١٣
٠ر٠	٢٥ر٤١٥	٢٤ر١٦٤
١ر٠	٢٩ر١٠٨	٣٠ر١١٨
١ر٠	٣٢ر٨٧٨	٣٢ر١٧٩
٢ر٠	٣٣ر١٧٨	٣٤ر١١٥
٢ر٠	٣٥ر٦٩١	٣٥ر٨٧٥
٣ر٠	٣٨ر١١٧	٣٩ر١١٨
٣ر٠	٤٣ر٨٧٧	٤٤ر٧٨٥
٤ر٠	٤٩ر٦٨٧	٤٩ر٥٧٨

ويلاحظ من جدول (٥) أن مستويات قدرات الأفراد " سيتا O " تتراوح ما بين -٤ و + ٤ ، وتوضح الدرجة المعادلة لكل مستوى من مستويات القدرة التي بين هذين الحدين ، وبينما يلاحظ أنه عند مستوى القدرة صفر تكون الدرجة المعادلة هي ٢٢٥ر٢١ وفقا لطريقة (M/M) وهي

١١٣ر٢٢ وفقا لطريقة (M/S) لتصل الى ١٧٨ر٣٣ و ١١٥ر٣٤ عند مستوى القدرة ٢ ، ثم تصل الى ١١٧ر٣٨ و ١١٨ر٣٩ عند مستوى القدرة ٣ ولتصبح ٦٨٧ر٤٩ و٥٧٨ر٤٩ عند مستوى القدرة ٤.

وبوجه عام ، فان ما سبق عرضه من نتائج يجيب عن التساؤل الأول للدراسة بشأن معادلة درجات تطبيق صورتي اختبار "توني" للذكاء باستخدام طرق المعادلة المختلفة وفق كل من القياس التقليدي والمعاصر.

ثانياً – النتائج الخاصة بفاعلية طرق معادلة الاختبار :

للإجابة عن التساؤل الثاني الذي تم طرحه خلال الدراسة والخاص بفاعلية طرق معادلة الاختبار ، تم حساب قيم " مؤشر كولن للصدق التقاطعي Kolen's Cross-Validation" بمكوناته الثلاثة ، لكل طريقة من طرق المعادلة المستخدمة ، وحساب الدلالة الإحصائية للفروق بين نتائج هذه الطرق الخمس ، وهو ما يوضح نتائجه الجدول (٦) :

جدول (٦)

يوضح متوسطات قيم مؤشرات الصدق التقاطعي (cross-validation indices) وقيم توكي (Tukey) لدلالة الفروق بين المتوسطات وفقا لطرق معادلة الاختبار المختلفة

م	المؤشر(index)	طريقة المعادلة	قيمة المؤشر
١	التحيز(Bias)	الخطية	٦٣ر٤
		توكر	٥٢ر٣
		المئينيات	٧٢ر٢
		M/M	٧٣ر٣
		M/S	٥٧ر٢
		دلالة الفروق (توكي)	٦٢ر٠*
٢	عدم الدقة (imprecision)	الخطية	٢٢ر٣٤
		توكر	٠٣ر٣٢
		المئينيات	١١ر٣٥
		M/M	٠٧ر٣٣
		M/S	٤٤ر٣٠
		دلالة الفروق (توكي)	١١ر٢*
٣	مقارنة المئينيات (Percentile Comparison)	الخطية	٣٣ر١١
		توكر	٥٦ر١٢
		المئينيات	٣٩ر١٠
		M/M	٨٠ر١١
		M/S	٢٦ر٩
		دلالة الفروق (توكي)	٤٧ر٤*

* = الفرق دال على مستوى ٠٥ر٠ على الأقل لصالح الطريقة ذي المؤشر الأقل قيمة .

ويلاحظ من خلال قراءة الجدول (٥) ما يلي :

(١) تعتبر طريقة المعادلة الخطية هي أكثر الطرق تحيزا في نتائجها من الطرق الأخرى حيث بلغت قيمة مؤشر التحيز لها أكبر قيمة من نظائرها في الطرق الأخرى ، وعلى العكس من ذلك تعتبر طريقة المتوسط / سيجما (M/S) من أفضل الطرق ، حيث بلغت قيمة مؤشر التحيز لها أقل قيمة ، وقد كانت الفروق بين قيم التحيز – كأحد عناصر مؤشر كولن للصدق التقاطعي – دالة إحصائيا طبقا لمقياس توكي لصالح القيمة الأقل ، مما يعني أن طريقة المتوسط/سيجما تعد أفضل الطرق الخمس لمعادلة الاختبار من حيث التحيز .

(٢) تعتبر طريقة المعادلة المئينية أقل الطرق دقة في نتائجها من الطرق الأخرى حيث بلغت قيمة مؤشر عدم الدقة لها أكبر قيمة من القيم المناظرة في الطرق الأخرى ، وعلى العكس من ذلك ، فان طريقة المتوسط / سيجما أفضل الطرق من حيث هذا الجانب حيث بلغت قيمة مؤشر عدم الدقة لها أقل قيمة ، ولما كانت الفروق بين قيم عدم الدقة – كأحد عناصر مؤشر كولن للصدق التقاطعي دالة إحصائيا لصالح القيمة الأقل ، فان ذلك يعني أن طريقة المتوسط/ سيجما تعد أفضل الطرق من حيث الدقة.

(٣) تعد طريقة معادلة توكر من أقل الطرق تفضيلا من حيث مقارنة الرتب المئينية كأحد عناصر الصدق التقاطعي كما يعكسه مؤشر كولن ، حيث بلغت قيمة الفروق بين الرتب المئينية لها أكبر قيمة من القيم المناظرة في الطرق الأخرى، وعلى العكس من ذلك فان طريقة المتوسط / سيجما تعد أفضل الطرق من حيث هذا الجانب ، ويلاحظ أن الفروق بين نتائج الطرق كانت دالة إحصائيا لصالح القيمة الأقل ، وهذا يعني أن طريقة المتوسط/ سيجما تعتبر أفضل الطرق من حيث المقارنة بين المئينيات .

وبوجه عام يمكن القول أن طريقة المتوسط / سيجما تفضل نتائجها نتائج الطرق الأخرى لمعادلة الاختبار وفقا لمؤشر كولن للصدق التقاطعي بجوانبه الثلاثة (التحيز – عدم الدقة – مقارنة المئينيات) ، وتعتبر هذه النتيجة متوقعة باعتبار طريقة المتوسط / سيجما هي إحدى طرق القياس المعاصر وفق النظرية (IRT) حيث أنه من المفترض أن تتميز طرق المعادلة وفق هذه النظرية عن طرق المعادلة وفق القياس التقليدي ، وهو ما يتفق مع نتائج العديد من الدراسات التي أجريت بهدف المقارنة بين طرق المعادلة وفقا لنوعي القياس (التقليدي والمعاصر) وتوصلت الى أن طرق المعادلة وفق القياس المعاصر تفضل طرق المعادلة وفق القياس التقليدي بوجه عام ، و من هذه الدراسات دراسة كل من :

[(Felan, 2002) , (Petersen, et al., 2002) , (Cook & Petersen, 1986) (Skages & Lissitz, 1986) , (Cook, 1983) , (Kolen & Whitney, 1982) (Kolen, 1981)].

توصيات واقتراحات لبحوث ودراسات مستقبلية:

ربما لا تكون هناك توصيات محددة يمكن للباحثة اقتراحها في ضوء نتائج الدراسة الحالية ، حيث أن هدف مثل هذا النوع من الدراسات تنفيذي إجرائي أو تطبيقي يتمثل في التوصل الى واقع محدد ، وهو التعرف على معادلة الدرجات بين صورتي اختبار ما بإتباع عدة طرق مختلفة لمعادلة الاختبار ، مع الحكم على فاعلية النتائج التي تفعلها كل طريقة من هذه الطرق ، وغالبا ما تكون نتائج مثل هذه الدراسات محدودة بحدود تحكمها المتغيرات التي يتم تناولها خلال الدراسة ، وهو ما ذهبت إليه الدراسة الحالية حيث تعد نتائجها محدودة بما تم تناوله من متغيرات ، ومن ثم يوصي باقتراح تكرار البحث باستخدام :

(١) طرق أخرى لمعادلة الاختبار وهي كثيرة عديدة ومتنوعة سواء في ظل القياس التقليدي أو القياس المعاصر.

(٢) نماذج لوغاريتمية أخري كالنموذج ثنائي المعلمة والنموذج ثلاثي المعلمة (2- & 3-PL).

(٣) عينات ذات أحجام أكبر .

(٤) النمط الآخر غير نمط المعادلة الأفقية، (المعادلة الرأسية أو العمودية).

(٥) تصميم آخر من تصميمات جمع البيانات اللازمة لمعادلة الاختبار (مثل تصميم المجموعة الواحدة، وتصميم المجموعات غير المتكافئة) .

(٦) شكل آخر من أشكال الاختبار الجذعي المشترك ليكون داخليا أو متضمنا في الاختبار الأصلي بدلا من اختبار الجذع المشترك الخارجي .

المراجـــع

أولاً – باللغة العربية :

(١) صلاح مراد، ومحمد الشافعي (١٩٩٨). أثر حجم العينة في دقة وكفاءة ضم اختبارين في تدريج مشترك. مجلة البحوث النفسية والتربوية ، كلية التربية بجامعة المنوفية ، مصر ، ص ٩٧ – ١٤٨ .

(٢) عماد غصاب عبابنة (٢٠٠٤). أثر حجم العينة وطريقة انتقائها وعدد الفقرات وطريقة انتقائها على دقة تقدير معالم الفقرة والقدرة لاختبار قدرة عقلية باستخدام نظرية الاستجابة للفقرة ، رسالة دكتوراه ، كلية الدراسات العليا ، جامعة عمان ، الأردن .

(٣) فؤاد أبو حطب (١٩٨٣) . القدرات العقلية . القاهرة ، مكتبة الأنجلو المصرية .

(٤) محمد محمد شوكت (١٩٨٧/أ) . اختبار توني للذكاء (اختبار ذكاء غير لفظي) ، كراسة التعليمات .

(٥) – – – – – – – – – – (١٩٨٧/ب). اختبار توني للذكاء (اختبار ذكاء غير لفظي) ، كراسة الأسئلة : الصورة (أ : A) ، والصورة (ب : B).

ثانياً – باللغة الانجليزية :

1. Angoff, W. (1984). Scales Norms and Equivalent Scores. Educational Testing Services, Princeton.

2. Cook, L. (1983). An Investigation of the Feasibility of Applying Item Response Theory to Equate Achievement Tests. Paper Presented at the Annual Meeting of the American Research Association. (67th , Montreal, Quebec, April.

3. Cook, L. & Petersen, N. (1986). Problems Related to the Use of Conventional and Item Response Theory Equating Methods in Less than Optimal Circumstances. Paper Presented at the Annual Meeting of the National Council of Measurement in Education. (San Francisco, CA, April).

4. Divgi, D. (1981). Model-free Evaluation of Scaling. Applied Psychological Measurement, vol. 5, no.2, pp. 203-9.

5. Dorans, N.; Liu, J. & Hammond, S. (2008). Anchor Test Type and

Population Invariance: An Exploration across Subpopulations and Test Administration. Applied Psychological Measurement, vol. 32, no. 1, pp. 81 – 97.

6. Eid, G. (2005). The Effect of Sample Size on the Equating of Test Item. Education, vol. 126, no. 1, p. 165.

7. Felan, G. (2002). Test Equating: Mean, Linear, Equipercentile, and Item Response Theory. Paper Presented at the Annual Meeting of the Southwest Educational Research Association, (Austin, TX, Feb., 14 –16).

8. Harris, D. & Crouse, J. (1993). A Study of Criteria in Equating. Applied Measurement in Education, no. 6, pp. 195 – 240.

9. Harris, D. & Kolen, J. (1986). Effect of Examinee Group on Equating Relationships. Applied Psychological Measurement, vol. 10, pp.35 – 43.

10. Kim, D.; Brennan, R. & Kolen, M. (2005). A Comparison of IRT Equating and Beta 4 Equating. Journal of Educational Measurement, vol. 42, no.1, pp. 77 - 99.

11. Kim, J. & Hanson, B. (2002). Test Equating Under Multiple-Choice Model. Applied Psychological Measurement, vol. 26, no. 1, pp. 27 – 42.

12. Kolen, M. (1981). Comparison of Traditional and Item Response Theory Methods of Equating Tests. Journal of Educational Measurement, vol. 18, no. 1, pp. 1 – 11.

13. Kolen, M. & Brennan, R. (2004). Test Equating; Scaling and Linking : Methods and Practices. Second Ed., New York, Springer-Verlag.

14. Kolen, M. & Whitney, D. (1982). Comparison of Four Procedures for Equating the Tests of General Education Development. Journal of Educational Measurement, vol. 9, no.4, pp. 279 – 93.

15. Lord, F. & Wingresky, M. (1983). Comparison of IRT Observed Score and True Score Equating. Unpublished Academic Report.

16. Moses, T.; Yang, W. & Wilson, C. (2007). Using Kernel Equating to Assess Item Order Effects on Test Scores. Journal of Educational Measurement, vol. 44, no .2, pp. 157 - 78.

17. Parshall, C.; Houghton, P. & Kromery, J. (1995). Equating Error and Statistical Bias in Small Sample Linear Equating. Journal of Educational Measurement, vol.32, no.1, pp. 37 – 54.

18. Petersen, N. (2008). A Discussion of Population Invariance of Equating. Applied Psychological Measurement, vol. 3, no 1, pp, 98 – 101.

19. Petersen, N.; Cook, L. & Stocking, M. (2002). IRT Versus Conventional Equating Methods: A Comparative Study of Scale Stability. Journal of Educational Statistics, vol. 8, no. 2, pp. 137-56.

20. Quenette, M. (2006). Model-Based Empirical Equating of Test Forms. Applied Psychological Measurement, vol. 30, no. 3, pp.167-82.

21. Sinharay, S. & Holland, P. (2007). Is It Necessary to Make Anchor Test Mini-Versions of the Tests Being Equated of Can Some Restrictions Be Relaxed?. Journal of Educational Measurement, vol. 44, no.3, pp.249 – 75.

22. Skaggs, G. & Lissitz, R. (1986). An Exploration of the Robustness of Four Test Equating Models. Applied Psychological Measurement, vol. 56, no. 4, pp. 495 – 509.

23. Slinde, J. & Linn, R. (1988). An Exploration of the Adequacy of the Rasch Model for the Problem of Vertical Equating. Journal of Educational Measurement, vol. 15, no. 1, pp. 23 – 35.

24. Tong, Y. & Kolen, M. (2005). Assessing Equating Results on Different Equating Criteria. Applied Psychological Measurement, vol. 29, no. 6, pp. 418 - 32.

25. Tsai, T; Hanson, B.; Kolen, M. & Forsyth, A. (2001). A Comparison of Bootstrap Standard Errors of IRT Equating Methods for the Common-Item Nonequivalent Groups Design. Applied Measurement in Education, vol. 14, pp. 17 – 30.

26. Van der Linden, W. (2006). Equating Error in Observed-Scores Equating. Applied Psychological Measurement, vol.30, no 5, pp. 355-78.

27. Von Davier, A. (2007). IRT True-Scores Test Equating. Educational and Psychological Measurement, vol. 67, no. 6, pp. 940- 57.

28. Von Davier, A. & Wilson, C. (2008). Investigation the Population

Sensitivity Assumptions of Item Response Theory True-Scores Equating Across Two Subgroups of Examinees and Two Test Formats. Applied Psychological Measurement, vol. 32, no. 1, pp. 11 – 26.

29. Wiley, A. (1999). An Investigation into Two Models for Equating Examinations with Multiple Item Formats. Dissertation Abstract International, DAI, A-3351.

30. Wright, N. & Dorans, N. (1993). Using the Selection Variable for Matching or Equating. Research Report 92-3, Princeton, NJ: ETS.

31. Yen, W. (2002). Scaling and Equating. New York State Technical Conference, March: 13.

32. Zimowski, M. ; Muraki, e. ; Mislevy, R. & Rock, F. (2006). BILOG-MG 3 for Windows. Assessment Systems Corporation, ASC.