الفصل السابع

معايرة الاختبار

Test calibration

افترضت كل الفصول السابقة وللأغراض التعليمية أن قياس مقياس القدرة معروفٌ . حيث كان هذا القياس منتصف نقطة الصفر، ووحدة قياس (١) ومدى من اللانهائية السلبية إلى اللانهائية الإيجابية . وقد تم التعبير عن القيم العددية لمعالم المفردة ومعالم قدرة المفحوص في هذا القياس . وبينما ساعد ذلك في إدخال المفاهيم الرئيسة لنظرية الاستجابة للمفردة إلا أنه يمثل موقف الاختبار الفعلي . وعندما يقوم واضعو الاختبار بكتابة المفردة، يعرفون بذلك السمة التي يرغبون أن تقوم المفردة بقياسها، سواء تم تخطيط المفردة للأداء بين المفحوصين ذوي القدرة المنخفضة أو المتوسطة أو القدرة المرتفعة. لكن من غير الممكن تحديد قيم معالم المفردات كشيء أولي، وبالإضافة إلى ذلك فعند تطبيق الاختبار على مجموعة من المفحوصين نجد من غير المعروف مقدماً مدى السمة الكامنة التي يمتلكها كل مفحوص. ونتيجة لذلك تكمن المهمة الرئيسة في تحديد قيم معالم المفردات وقدرات المفحوص في القياس للسمة الأساسية الكامنة . وتسمى هذه المهمة في نظرية الاستجابة على المفردة بمعايرة الاختبار كما أنها تعطي نطاقاً مرجعياً يفيد في تفسير نتائج الاختبار . ويتم إنجاز معايرة الاختبار من خلال تطبيقه على مجموعة (M) مفحوصين من المستوى المتوسط مع التسجيل الثنائي

لإجابات المفحوصين لمفردات N . ويتم بعد ذلك تطبيق الإجراءات الحسابية لبيانات الإجابة على المفردات لإيجاد مقياس القدرة الفريد لمجموعة خاصة من مفردات الاختبار والمفحوصين . ويتم بعد ذلك التعبير عن قيم تقديرات معالم المفردات وقدرات المفحوصين التي تم تقديرها في هذا القياس . وعند إتمام ذلك يكون الاختبار قد تم معايرته ويمكن تفسير نتائج الاختبار بواسطة تراكيب نظرية الاستجابة للمفردة.

عملية معايرة الاختبار

The Test Calibration Process

تم افتراض الأسلوب المستخدم لمعايرة الاختبار بواسطة آلن بيرنبوم Alan Birnbaum في عام ١٩٦٨ وتم تطبيقه في برامج الحاسب المستخدمة على نطاق واسع مثل LOGIST (Wingersky , Barton and Lord, 1982) و (BICAL Wright and Mead ,1976) ويعتبر نموذج بيرنبوم تكرار إجراءً بديلاً يقوم بتوظيف مرحلتين لأقصى تقدير ترجيحي . ويتم في المرحلة الأولى تقدير معالم مفردات N في الاختبار ، وفي المرحلة الثانية تقدر معالم القدرة من (M) للمفحوصين . ويتم أداء المرحلتين بشكل متكرر حتى يتم الحصول على مجموعة ثابتة من تقديرات المعلم ، ويتم في هذه النقطة معايرة الاختبار وتعريف مقياس القدرة .

وخلال المرحلة الأولى من نموذج بيرنبوم يتم معاملة القدرة التي تم تقديرها لكل مفحوص كما لو تم التعبير عنها من خلال القياس الصحيح للسمة الكامنة . ويتم بعد ذلك تقدير معالم كل مفردة من خلال إجراء أقصى ترجيح والذي تم مناقشته في الفصل الثالث . ويتم إجراء ذلك لمفردة واحدة في كل مرة لأن الافتراض الضمني يشير إلى كون تلك المفردات مستقلة عن بعضها بعضاً. وكانت النتيجة مجموعة من القيم لتقديرات معالم المفردات في الاختبار وتفترض المرحلة الثانية أن تقديرات معلم المفردة

التي تم الحصول عليها بواسطة المرحلة الأولى تعتبر في الواقع قيم لمعالم المفردات وبعد ذلك يتم تقدير قدرة كل مفحوص باستخدام إجراء الاحتمالية القصوى التي تم عرضها في الفصل الخامس . ويفترض أن تكون قدرة كل ممتحن مستقلة عن قدرات كل المفحوصين الآخرين . ولذلك يتم الحصول على تقديرات القدرة لمفحوص واحد في مرة واحدة .

ويتم تكرار العملية ذات المرحلتين حتى يتم الحصول على محك التقارب المناسب. ويشير التأثير الكلي أنه تم تقدير معالم مفردات اختبار N ومستويات القدرة للمفحوصين ذوي المستوى المتوسط التي تم تقديرها في نفس الوقت على الرغم من إجرائها مرة واحدة . ويساعد هذا النموذج الماهر على خفض مشكلة التقدير المعقدة جداً إلى نموذج يمكن تطبيقه على الحاسب.

المشكلة القياسية

The Metric Problem

من الخصائص السيئة لنموذج بيرنبوم أنه لا يعطي قياساً واحداً لمقياس القدرة . ولذلك تعتبر النقطة المتوسطة ووحدة قياس القدرة الذي تم الحصول عليها غير محددة، أي أن كثيراً من القيم المختلفة تعمل جيداً بصورة متساوية. وبتعبير فني أو تقني يعتبر القياس فريداً وحتى مستوى التحويل الخطي. ونتيجة لذلك نجد من الضروري " إرساء " القياس من خلال القواعد الملزمة لتحديد نقطة الوسط ووحدة قياس القدرة ويرجع أمر كيفية فعل ذلك إلى الأشخاص الذين يقومون بتطبيق نموذج بيرنبوم في برنامج الحاسب . ومن خلال برنامج الحاسب BICAL يتم أداء هذه العملية للإرساء بعد إتمام المرحلة الأولى. ولذلك يتم أداء كل من المرحلتين في التكرار باستخدام قياس القدرة المختلف بدرجة طفيفة. وكلما تباينت العملية التكرارية الكلية كلما تباين أيضاً

قياس القدرة لنقطة متوسطة خاصة ولوحدة القياس. وكانت الخاصية الرئيسة لهذه العملية في اعتماد قياس القدرة الناتجة على مجموعة خاصة من المفردات التي تشكل الاختبار وإجابات مجموعة خاصة من المفحوصين على هذا الاختبار . ومن غير الممكن الحصول على تقديرات لقدرة المفحوصين ومعالم المفردات في القياس الصحيح للسمة الكامنة الأساسية . وأفضل ما يمكننا فعله الحصول على قياس يعتمد على تجمع خاص بين المفحوصين ومفردات الاختبار.

معايرة الاختبار في نموذج راش The Calibration Under the Rasch Model

توجد ثلاثة نماذج مختلفة للمنحنى المميز للمفردة يمكن الاختيار من بينهم ، بالإضافة إلى وجود طرق عديدة لتطبيق نموذج بيرنبوم . ولقد اختار المؤلف من خلال ذلك عرض الأسلوب المتمركز على النموذج الرياضي أحادي المعلم (راش) كما تم تطبيقه بواسطة بينجامين رايت Benjamin Wright و زملائه في برنامج BICAL . وفي ضوء هذا النموذج كانت مفردات معلم واحد فقط ليتم تقديره . وتعمل الإجراءات بصورة جيدة مع الأعداد الصغيرة لمفردات الاختبار والأعداد الصغيرة للمفحوصين . ويعتبر إجراء إرساء القياس بسيطاً ويعتبر من السهل عرض الأفكار الرئيسة لمعايرة الاختبار .

وسيتم استخدام معايرة اختبار من عشر مفردات ثم تطبيقه على مجموعة قوامها (١٦) مفحوصاً وفيما يلي توضيح لتلك العملية . وتعتمد المعلومات التي تم عرضها على تحليل بيانات المجموعة الأولى التي تم عرضها في جلسة الحاسب CALIBARATE A TEST في موقع الانترنت المصاحب . وقد يتم اختبار العمل في هذا الجزء خلال جلسة الحاسب الآلي لكنه لا يعتبر أمراً ضرورياً لأنه سيتم عرض إجراءات الحاسب الآلي فيما بعد .

ويعتبر الاختبار ذو العشر مفردات اختباراً يتوافق مع القدرة المتوسطة لمجموعة قوامها (١٦) مفحوصاً. وقد تم التسجيل الثنائي لإجابات المفحوصين على المفردة (١) للإجابات الصحيحة ، و(صفر) للإجابات الخاطئة . وقد كان الهدف استخدام هذه البيانات للإجابة على المفردات لمعايرة الاختبار . وفيما يلي عرض القيم الموجهة للإجابة الواقعية عن المفردات ويمثل كل صف إجابات المفردات التي تم إجراؤها بواسطة مفحوص بعينه .

الجدول رقم (٧,١). إجابات المفحوصين على المفردات .

المفحوصين	مفردات الاختبار المواقفة										المجموع
	١	٢	٣	٤	٥	٦	٧	٨	٩	١٠	
١	١		١						١		٣
٢	١		١								٢
٣	١	١	١		١					١	٥
٤	١	١	١		١						٤
٥					١						١
٦	١	١		١							٣
٧	١					١			١	١	٤
٨	١				١	١		١			٤
٩	١		١		١			١			٤
١٠	١				١		١				٣
١١	١	١		١	١	١	١	١	١	١	٩
١٢	١	١	١	١	١	١		١	١	١	٩
١٣	١	١	١		١		١			١	٦
١٤	١	١	١	١	١	١		١	١	١	٩
١٥	١	١		١	١	١	١	١	١	١	٩
١٦	١	١	١	١	١	١	١	١	١	١	١٠

وقد لوحظ في الفصل الخامس أنه من المستحيل تقدير قدرة المفحوص إذا لم يحصل على أي أو كل مفردات الاختبار الصحيحة . ويوضح الجدول رقم (٧,١) أن (١٦) مفحوصاً قد أجابوا على المفردات بطريقة صحيحة ويجب استبعادها من مجموعة البيانات وبطريقة مماثلة . فإذا تمت الإجابة على المفردات بطريقة صحيحة بواسطة كل المفحوصين أو لم يجب عليها أي مفحوص فلا يمكن تقدير معلم صعوبة مفرداتها . وبذلك يجب استبعاد هذه المفردة من مجموعة البيانات . وفي هذا المثال الخاص لم يتم استبعاد أي مفردة لهذا السبب . ومن إحدى الخصائص الفريدة لمعايرة الاختبار في نموذج راش أن لكل المفحوصين نفس عدد المفردات الصحيحة (نفس الدرجة الخام) التي ستحصل على نفس القدرة التي تم تقديرها . ونتيجة لذلك نجد من غير الضروري التمييز بين العديد من المفحوصين الذين حصلوا على نفس الدرجة الخام في الاختبار . وبذلك بدلاً من استخدام إجابات مفردة فردية كان كل ما نحتاجه عدداً من المفحوصين في كل درجة خام يقومون بالإجابة على المفردة إجابة صحيحة . وبسبب أيضاً ذلك وبسبب استبعاد المفردات يتم استخدام البيانات المحورة كنقطة بدء أولية لإجراءات معايرة الاختبار في نموذج راش . ويتم فيما يلي عرض مجموعة البيانات المحورة لهذا المثال.

الجدول رقم (٧,٢). حساب التكرارات للبيانات المحورة .

الدرجات	المفردات										المجموع
	١	٢	٣	٤	٥	٦	٧	٨	٩	١٠	
١					١						١
٢	١		٢					١			٤
٣	٢	١		١	١				١		٦
٤	٤	١	٢		٢	٣	١	١	٢		١٦
٥	١	١	١		١		١				٥
٦	١	١	١		١		١			١	٦
٩	٤	٤	٢	٤	٤	٤	٤	٤	٤	٥	٣٦
المجموع	١٣	٨	٨	٥	١٠	٧	٧	٦	٧	٣	٧٤

وفي الجدول رقم (٢,٧) يتم تسجيل الصفوف بواسطة درجات الاختبار الخام التي تتراوح من (١) إلى (٩) . ويعتبر صف الهوامش العدد الإجمالي للإجابات الصحيحة التي تم الحصول عليها بواسطة المفحوصين ذوي درجة الاختبار الخام . ويتم تسجيل الأعمدة بواسطة عدد المفردات من (١) إلى (١٠) . ويعتبر صف الهوامش العدد الإجمالي للإجابات الصحيحة التي تم وضعها لمفردة خاصة بواسطة المفحوصين الباقين . (وقد كان الصف المزدوج لأعمدة العد الإجمالي ضرورياً للعمل حول شروط المسافة لشاشة العرض) . وفي نموذج راش تمثل المعلومات الوحيدة المستخدمة في بيرنبوم إجمالي التكرارات التي تم احتواؤها في صف وأعمدة الهوامش . ويعتبر هذا نسقاً فريداً لهذا النموذج ويؤدي إلى حسابات بسيطة من خلال إجراءات الاحتمال الأقصى التي تم توظيفها في كل مرحلة من العملية الكلية . وبافتراض وجود اثنين من القيم الموجهة للتكرار يمكن حينئذ تطبيق عملية التقدير ويتم الحصول على التقديرات الأولية لمعالم صعوبة المفردات في المرحلة الأولى ويجب إرساء قياس القدرة . وفي نموذج راش تتخذ إجراءات الإرساء مزايا حقيقية كون معلم تمييز المفردة ثابتة في قيمة (١) لكل المفردات في الاختبار . وبسبب ذلك تكون وحدة قياس القدرات التي تم تقديرها ثابتة في قيمة (١) . وما ينبغي بعد كل ذلك تعريف النقطة المتوسطة للمقياس . وفي برنامج الحاسب BICAL يتم تعريف النقطة المتوسطة بأنها متوسط صعوبات المفردات التي تم تقديرها . ومن أجل الحصول على قيمة النقطة المتوسطة الملائمة يتم طرح متوسط صعوبة المفردات من قيمة تقدير صعوبة كل مفردة وينتج عنه متوسط إعادة القياس لصعوبة المفردة ذات القيمة صفر . وحيث يتم التعبير عن صعوبات المفردة بنفس القياس للقدرة فقد تم الآن تحديد النقطة المتوسطة ووحدة قياس الأخير . وحيث تم إجراء ذلك بين المراحل ستكون بذلك القدرات التي تم تقديرها في المرحلة الثانية في القياس المعروف بواسطة تقديرات معالم المفردات التي أعيد قياسها والتي تم الحصول عليها في المرحلة الأولى . ويتم الحصول على تقدير القدرة المناظر لكل درجة خام للاختبار في المرحلة الثانية باستخدام صعوبات المفردة المعاد قياسها كما لو كانت معالم

الصعوبة والقيم الموجهة لإجمالي صفوف الهوامش . وقد كانت نتيجة هذه المرحلة تقدير قدرة كل درجة خام للاختبار في مجموعة البيانات . وفي هذه النقطة يتم فحص التقارب لعملية التكرار الكلية . ومن خلال برنامج BICAL قام رايت Wright بتلخيص الفروق المطلقة بين قيم تقديرات معلم صعوبة المفردات لتكرارين متواليين في النموذج . وإذا كان هذا الحاصل أقل من (٠,٠١) فحينئذ يتم توقف عملية التقدير . وإذا كان ذلك أكبر من (٠,٠١) فحينئذ يتم أداء تكرار آخر ويتم القيام بالمرحلتين مرة أخرى . ولذلك يتم تكرار عملية المرحلة الأولى ، وإرساء القياس ، عملية المرحلة الثانية وفحص التقارب حتى يتم تطابق المعيار . وعندما يحدث ذلك يتم تقبل القيم الحالية للمفردة وتقديرات معلم القدرة ويتم حينئذ تعريف قياس القدرة . و فيما يلي يتم عرض تقديرات معالم صعوبة المفردة للمثال الحالي .

مجموعة بيانات رقم (١)

تقديرات معلم المفردة

الجدول رقم (٧,٣). معالم صعوبة المفردات التي تم تقديرها .

المفردة	الصعوبة
١	٢,٣٧-
٢	٠,٢٧-
٣	٠,٢٧-
٤	٠,٩٨-
٥	١,٠٠-
٦	٠,١١+
٧	٠,١١+
٨	٠,٥٢+
٩	٠,١١+
١٠	٢,٠٦+

ويمكنك البرهنة على أن حاصل صعوبات المفردة يساوي صفراً (في الأخطاء الدائرية) . حيث كان تفسير قيم تقديرات معلم المفردات بالضبط هو ما تم عرضه في الفصل الثاني . فعلى سبيل المثال كان لمفردة (١) صعوبة مفردة (-٢,٣٧) والتي تضعها في النهاية المنخفضة لمقياس القدرة . وكان لمفردة (٦) صعوبة (+٠,١١) والتي يضعها قريباً من منتصف مقياس القدرة . وكان لمفردة (١٠) صعوبة (+٢,٠٦) التي تضعها في أعلى نهاية لمقياس القدرة . ولذلك فالتمسك بالتفسير المعتاد لصعوبة المفردات على أنه وضع للمفردات في مقياس القدرة . وبسبب استخدام إجراءات الاعتماد تعتبر هذه القيمة نسبية في الواقع لصعوبة مفردات الاختبار لهؤلاء المفحوصين .

وعلى الرغم من تقدير القدرة في الجدول رقم (٧,٤) لكل مفحوص، كل المفحوصين ذوى نفس درجة الاختبار الخام حصلوا على نفس تقدير القدرة . وعلى سبيل المثال فقد كان لكل من المفحوص الأول والثاني درجات خام تصل إلى (٢) وحصلوا على قدرة تم تقديرها (-١,٥) وقد كان للمفحوص السابع والثامن والتاسع درجات خام (٤) بحيث أن قدرتهم تم تقديرها بـ (-٠,٤٢) . وتعتبر هذه السمة الفريدة نتيجة مباشرة لحقيقة أنه في نموذج راش كان معلم التمييز ثابتاً (١) لكل المفردات في الاختبار . ويروق هذا الجانب من نموذج راش لكل الممارسين لأنهم يشعرون بطريقة بديهية أنه ينبغي على المفحوصين الذين حصلوا على نفس الدرجة الخام في الاختبار الحصول على نفس تقدير القدرة . وعند استخدام نماذج المنحنى المميز في النماذج الثنائي والثلاثي ، يعتمد تقدير قدرة الممتحن على شكل خاص لإجابات المفردات بدلاً من الدرجات الخام . وفي هذه النماذج سيحصل المفحوصون ذوو نفس شكل الإجابات على المفردات على نفس تقدير القدرة . ولذلك يمكن أن يحصل المفحوصون ذوو نفس الدرجة الخام على تقديرات مختلفة للقدرة حتى ولو أجابوا على المفردات المختلفة بطريقة صحيحة .

مجموعة بيانات رقم (١)

تقديرات القدرة

الجدول رقم (٧,٤)- تقديرات القدرة التي تم الحصول عليها.

المفحوص	تقدير القدرة	الدرجة الخام
١	١,٥٠-	٢
٢	١,٥-	٢
٣	٠,٠٢ +	٥
٤	٠,٤٢-	٤
٥	٢,٣٧-	١
٦	٠,٩١-	٣
٧	٠,٤٢-	٤
٨	٠,٤٢-	٤
٩	٠,٤٢-	٤
١٠	٠,٩١-	٣
١١	٢,٣٣+	٩
١٢	٢,٣٣+	٩
١٣	٠,٤٦+	٦
١٤	٢,٣٣+	٩
١٥	٢,٣٣+	٩
١٦	****	١٠

ولم يدخل المفحوص رقم (١٦) في الحسابات بسبب استبعاده لأنه حصل على درجة خام كاملة . ويتم تفسير تقدير القدرة التي تم الحصول عليها بواسطة مفحوص ما فيما يتعلق بمكان وضع المفحوص في مقياس القدرة .وعلى سبيل المثال فقد حصل المفحوص رقم (٧) على درجة (-٠,٤٢) من القدرة التي تم تقديرها والتي وضعته تحت النقطة المتوسطة للمقياس . ويمكن وضع توزيعهم في مقياس القدرة ويمكن حساب ملخص إحصائيات هذا التوزيع . وفي الحالة الحالية يحصل هذا على

متوسط (٠,٠٦) وكان الانحراف المعياري (١,٥٧) . ولذلك حصل المفحوص رقم (٧) على درجة قدرة تصل إلى حوالي (٠,٢٧) من الانحرافات المعيارية تحت متوسط قدرة المجموعة ومع ذلك لا يفسر الفرد بصورة نمطية درجة قدرة المفحوص فيما يتعلق بتوزيع الدرجات كمجموعة من المفحوصين . ولفعل ذلك يجب علينا تجاهل حقيقة إمكانية تفسير درجة القدرة بطريقة مباشرة كموقع المفحوص على مقياس القدرة .

ملخص عملية معايرة الاختبار Summary of the test Calibration Process

الناتج النهائي لعملية معايرة الاختبار يمثل تعريف تدريج مقياس القدرة . وفي نموذج راش نجد لهذا المقياس وحدة قياس (١) و نقطة المنتصف هي الصفر . ويبدو هذا بصورة سطحية نفس الشيء كمقياس القدرة المستخدم في الفصول السابقة . ومع ذلك لا يعتبر قياساً للسمة الكامنة الأساسية. ويعتمد القياس الذي تم الحصول عليه على إجابات المفردات التي تم الحصول عليها بواسطة الدمج بين المفحوصين ومفردات الاختبار التي اختبرت وفق نموذج بيرنبوم وحيث لا يمكن تحديد قياس السمة الكامنة الأساسية ، لذلك يتم استخدام القياس الذي نتج عن نموذج بيرنبوم كما لو كان القياس صحيحاً . ويتم تفسير قيم صعوبة المفردات التي تم الحصول عليها وقدرة المفحوص في هذا القياس . ولذلك تم معايرة الاختبار . وكانت نتيجة إجراء تقديم الاختبار وضع كل مفحوص ومفردة عبر مقياس القدرة الذي تم الحصول عليه . وكان للمفردة (٥) في المثال الحالي صعوبة (-١) وكان للمفحوص العاشر تقدير قدرة (-٠,٩١) . ولذلك كانت احتمالية إجابة المفحوص العاشر على المفردة (٥) بطريقة صحيحة (٠,٥) تقريباً . وتعتبر إمكانية وضع المفردات والمفحوصين في مقياس عام سمة قوية لنظرية الاستجابة للمفردة . وتسمح هذه السمة للفرد بتفسير نتائج معايرة الاختبار وفق مرجعية واحدة وتعطي معنى لقيم تقديرات المعالم .

جلسة الحاسب الآلي للفصل السابع

Computer Session for Chapter7

تعتبر هذه الجلسة للحاسب الآلي مختلفة قليلاً عن جلسات الفصول السابقة . حيث سيكون من الصعب عليك إيجاد مجموعات بيانات ليتم معايرتها لذا تم تخزين ثلاث مجموعات من البيانات في موقع الإنترنت . وسيتم استخدام كل مجموعة من هذه المجموعات لتقويم الاختبار وسيتم عرض النتائج على الشاشة وستنتقل ببساطة من كل مجموعة من مجموعات البيانات ونتائج المعايرة . وتوجد بعض الأهداف المحددة في هذه العملية . أولاً ستصبح معتاداً على بيانات المدخل وكيفية تحريرها . ثانياً يمكن تفسير تقديرات صعوبة المفردة وتقديرات قدرة المفحوص . ثالثاً سيتم إيضاح وتفسير المنحنى المميز للاختبار ودالة معلومات الاختبار .

وسيتم استخدام ثلاثة اختبارات مختلفة ذات عشر مفردات تقوم بقياس نفس السمة الكامنة . وستقوم مجموعة مشتركة قوامها (١٦) مفحوصاً بأداء ثلاثة اختبارات من هذه الاختبارات . وتم وضع الاختبارات بحيث يتوافق متوسط صعوبة الاختبار الأول مع متوسط قدرة المجموعة العادية من المفحوصين . وتم وضع الاختبار الثاني ليكون اختباراً سهلاً لهذه المجموعة . وتم بعد ذلك وضع الاختبار الثالث ليكون صعباً لهذه المجموعة . وسيتعرض كل اختبار من هذه الاختبارات لنموذج بيرنبوم والمعايرة بطريقة منفصلة . وهناك سببان لهذا الأسلوب : الأول أنها توضح أن كل معايرة للاختبار تنتج قياساً وحيداً لمقياس القدرة . ثانياً ، يمكن استخدام النتائج لإيضاح العملية والتي يمكن من خلالها وضع ثلاث مجموعات من نتائج الاختبار في مقياس القدرة العام .

إجراءات جلسة معايرة الاختبار Procedures for the test Calibration session

١- بيانات المجموعة الأولى

وقد كان لهذا الاختبار ذو العشر مفردات متوسط صعوبة يتوافق مع متوسط قدرة المجموعة ذات الـ (١٦) مفحوصاً .

١- اتبع الإجراءات الافتتاحية التي تم وصفها في المقدمة .

٢- استخدم الفأرة لإبراز جلسة CALIBRATE A TEST واضغط على [SELECT] .

٣- اقرأ الشاشات الشارحة واضغط على [CONTINUE] للانتقال من عرض للذى يليه .

٤- سيظهر جدول القيم الموجهة للإجابة على المفردات . وسيكون ذلك نفس الشيء مثلما كان في الجدول رقم (٧,٢) . ولاحظ أنه تم استبعاد المفحوص السادس عشر لأنه أجاب على المفردات كلها إجابة صحيحة . اضغط على [CONTINUE] .

٥- سيظهر عرض جدول البيانات المحررة . وستكون نفس الشيء مثلما كانت في الجدول رقم (٧,١) . ولاحظ أنه تم استبعاد المفحوص السادس عشر ولم يتم استبعاد المفردات . اضغط على [CONTINUE] .

٦- الشاشة تشير إلى نموذج بيرنبوم لمعايرة الاختبار الذي سيتم عرضه . اضغط على [CONTINUE] .

٧- سيتم عرض جدول تقديرات صعوبة المفردة للاختبار الأول . ويعتبر هذا نفس الشيء كما في الجدول رقم (٧,٣) . اضغط على [CONTINUE] .

٨- سيتم عرض القدرات التي تم تقديرها لـ (١٦) مفحوصاً ودرجاتهم الخام . وستكون الشاشة نفس الشيء كما في الجدول رقم (٧,٤) . وقد كان لتقديرات القدرة

متوسط (٠,٠٦٢) والانحراف (١,٥٧) . ولاحظ أن المفحوص السادس عشر لم يحصل على تقدير قدرة .

٩- تظهر بعد ذلك رسالة DO YOU WANT TO REVIEW DATA SET 1 ? RESULTS AGAING وإذا ضغطت على زر YES ، ستعود إلى الخطوة الرابعة . وإذا ضغطت على زر NO ستظهر الشاشة التالية . اضغط على زر NO .

١٠- ستؤدي إجابة No إلى عرض المنحنى المميز للاختبار. لاحظ حقيقة أن الدرجة نصف حقيقية (الدرجة الحقيقية تساوي نصف عدد المفردات) تناظر مستوى قدرة الصفر . ويعكس ذلك إجراء الاعتماد الذي يضع متوسط صعوبة المفردات عند الصفر . اضغط [CONTINUE] .

١١- وسيتم بعد ذلك عرض دالة معلومات الاختبار . ويعتبر المنحنى متماثلاً بطريقة مقبولة وذا ارتفاع واضح جداً في الوسط . ويشير شكل المنحنى إلى أنه يتم تقدير القدرة بأعظم دقة بمصاحبة متوسط مقياس القدرة . وتحدث ذروة دالة معلومات الاختبار في نقطة أعلى بدرجة طفيفة من النقطة المتوسطة لمقياس القدرة . ويعكس ذلك توزيع صعوبات المفردات ووجود ست مفردات ذات قيم إيجابية وأربع فقط ذات قيم سلبية . ولذلك يوجد تأكيد طفيف جداً على مستويات القدرة الإيجابية .

١٢- ويساعد الضغط على [DISPLAY FIRST CURVE] على إعادة ظهور الشكل البياني للمنحنى المميز للمفردات . وسيسمح لك ذلك بالتبادل بين المنحنى المميز للاختبار و دالة معلومات الاختبار .

١٣- ولاستمرار الجلسة أجب على سؤال DO NEXT DATA SET ? بالضغط على زر YES .

٢- بيانات المجموعة الثانية

وقد تم تشكيل الاختبار ذي العشر مفردات ليكون اختباراً سهلاً لمجموعة مكونة من (١٦) مفحوصاً . وحيث ستكون إجراءات الحاسب لهذه المجموعة من البيانات نفس الشيء بالضبط مثل بيانات المجموعة الأولى ، لذلك لن يتم تكرارها بالتفصيل . وسيتم تدوين النتائج ذات الدلالة فقط .

١- ومن خلال عرض البيانات المحورة تم استبعاد المفحوص الخامس عشر والسادس عشر لحصولهم على درجات خام كاملة .

٢- وكان متوسط صعوبات المفردة (٠,٠٩٨) والذي يقترب من الصفر . وقد حصلت ست من هذه المفردات على صعوبات إيجابية للمفردات وكان توزيع الصعوبات على شكل U إلى حد ما .

٣- وكان لتقديرات القدرة متوسط (٠,٤٤) والانحراف معياري (١,٣٥) . ومن المهم ملاحظة أنه كان للمفحوص التاسع درجة خام (٤) في الاختبار الأول وحصل على تقدير قدرة (-٠,٤٢) وكانت الدرجة الخام في الاختبار الثاني (٧) وكان تقدير القدرة (١,٠٢) . وبذلك تعتبر القدرة الحقيقية للمفحوص نفس الشيء في كلا الحالتين .

٤- وتناظر الدرجة نصف الحقيقية للمنحنى المميز للمفردة لمستوى القدرة صفر. ويعتبر شكل المنحنى المميز للاختبار متطابقاً تقريباً مع شكل منحنى الاختبار الأول .

٥- وتعتبر دالة معلومات الاختبار متماثلة . ولها مظهر دائري إلى حد ما . وأقصى قدر من المعلومات كان عند مستوى قدرة (٠,٥) تقريباً .

٦- أجب على رسالة DO NEXT DATA SET ? بواسطة الضغط على زر YES

٣- بيانات المجموعة الثالثة

وقد تم تشكيل هذا الاختبار ذي العشر مفردات ليكون اختباراً صعباً لمجموعة مشتركة قوامها (١٦) مفحوصاً . وحيث كانت إجراءات الحاسب نفس الشيء كما كانت للمثالين السابقين فسيتم مناقشة النتائج ذات الأهمية فقط .

١- يوضح استكشاف جدول القيم الموجهة للإجابة على المفردات أن المفحوص الأول والثالث حصلوا على درجات خام صفر وسيتم استبعادهم . ويوضح استكشاف الأعمدة عدم إجابة أي من المفحوصين على المفردة العاشرة بطريقة صحيحة . وسيتم استبعادها من مجموعة البيانات . وبالإضافة إلى ذلك وبعد استبعاد المفحوصين الاثنين تم الإجابة على مفردة (١) بطريقة صحيحة من قبل كل المفحوصين الباقين . ولذلك يجب أيضاً استبعاد هذه المفردة . وبناء على ذلك حصل المفحوص الثاني والسادس على درجات خام صفر لأن المفردة الوحيدة التي أجابوا عليها بطريقة صحيحة هي (١) . وبعد استبعاد هذين المفحوصين لا نجد من الضروري وجود تحرير إضافي . ويعتبر هذا التحرير المتعدد للمرحلة مألوفاً تماماً في معايرة الاختبار . وينبغي ملاحظة أنه بعد التحرير تعتبر مجموعة البيانات أصغر من المجموعتين السابقتين ويعتبر مدى الدرجات الخام الآن من (١) إلى (٧) .

٢- ومتوسط الثماني تقديرات لصعوبات المفردات كان (٠,٠٠١٣) . والذي يقترب مرة أخرى من الصفر . وقد كان لثلاث مفردات قيم إيجابية لتقديرات الصعوبة . وقد كان للمفردة الثامنة صعوبة تقدر بـ (١,٣٤) بينما تقع صعوبات المفردات السبع المتبقية من (-٠,٦٧) إلى (+٠,٧٩) .

٣- وقد كان لاثني عشر مفحوصاً تم استخدامهم في معايرة الاختبار متوسط (-٠,١١) وانحراف معياري (١,٤١) .

٤- ويعتبر المنحنى المميز للاختبار مماثلاً للمنحنين السابقين إذ توجد منتصف الدرجة الحقيقية مرة أخرى عند مستوى القدرة الصفر . لكن يقترب الجزء الأعلى من المنحنى من قيمة (٨) بدلاً من (١٠) .

٥- وتعتبر دالة معلومات الاختبار متماثلة تقريباً عند مستوى قدرة الصفر تقريباً . ويعتبر المنحنى أقل في ذروته من العميلتين السابقتين لمعلومات الاختبار وكان حدها الأقصى أقل بدرجة طفيفة .

٦- أجب على رسالة DO NEXT DATA SET ? بواسطة الضغط على زر NO وسيؤدي ذلك إلى إنهاء الجلسة وستظهر القائمة الرئيسة على الشاشة مرة أخرى .

وينبغي على القارئ أن يتأمل قليلاً في سبب عدم كون متوسط قدرة المجموعة المألوفة من المفحوصين نفس الشيء لكل المعايير الثلاثة . ومبدأ ثبات المفردة ينص على أنه ينبغي أن تكون كلها نفس الشيء . فهل هذا المبدأ خاطئ أم أن شيئاً آخر له دور في هذا المقام ؟ ويتم عرض التمثيل البياني لعدم الثبات بعد جزء " ملاحظات " .

ملاحظات Things To Notice

١- وفي كل المعايير الثلاثة تم استبعاد المفحوصين في عملية التحرير . ونتيجة لذلك لا تعتبر المجموعة المألوفة نفس الشيء تماماً في كل معايير الاختبار .

٢- وعلى الرغم من تصميم الاختبارات لكي تمثل الاختبارات السهلة ، الصعبة ، والمتوافقة نسبياً مع متوسط قدرة المجموعة المشتركة ، إلا أن النتائج لا تعكس ذلك . وبسبب عملية الاعتماد حصلت كل المعايير الثلاثة للاختبار على متوسط صعوبة للمفردة(صفر).

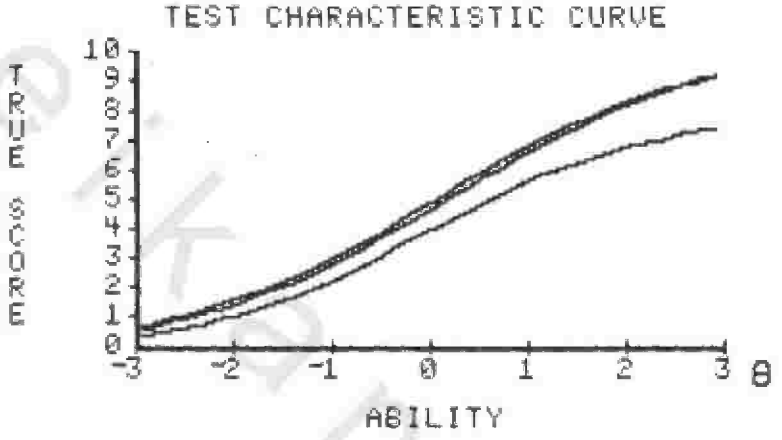
٣- وفي كل معيار حصل المفحوصون ذوو نفس درجة الاختبار الخام على نفس التقدير للقدرة ومع ذلك لم تنتج الدرجة الخام المفترضة على نفس القدرة المقدرة عبر المعايير الثلاثة .

٤- وعلى الرغم من إعطاء نفس مجموعة المفحوصين كل الاختبارات الثلاثة إلا أن المتوسطات والانحرافات المعيارية لتقديرات قدراتهم كانت مختلفة لكل معيار. ويمكن عزو ذلك إلى عدد من الأسباب. ويشير السبب الأولى إلى أنه من خلال عملية الاعتماد يتم التعبير عن قيمة متوسط تقدير القدرات بطريقة نسبية لمتوسط صعوبة المفردة الاختبار. ولذلك ينبغي أن يؤدي متوسط الصعوبة في اختبار سهل إلى متوسط قدرة إيجابية. وينبغي أن يكون لمتوسط القدرة في الاختبار الصعب قيم سلبية. وينبغي أن يقترب متوسط القدرة في الاختبار المتوافق من الصفر. ويفسر التغير في أفراد المجموعة بعض الاختلافات أيضاً وخاصة عندما تكون المجموعة صغيرة للبدء معها. وفي النهاية يعتبر القدر الكلي للمعلومات صغيراً إلى حد ما في كل دوال معلومات الاختبار الثلاثة. ولذلك يتم تقدير مستوى القدرة لأي من المفحوصين بدقة شديدة. ونتيجة لذلك ليس من الضروري أن تكون تقدير القدرة لمفحوص ما قريبة من قدرته الحقيقية.

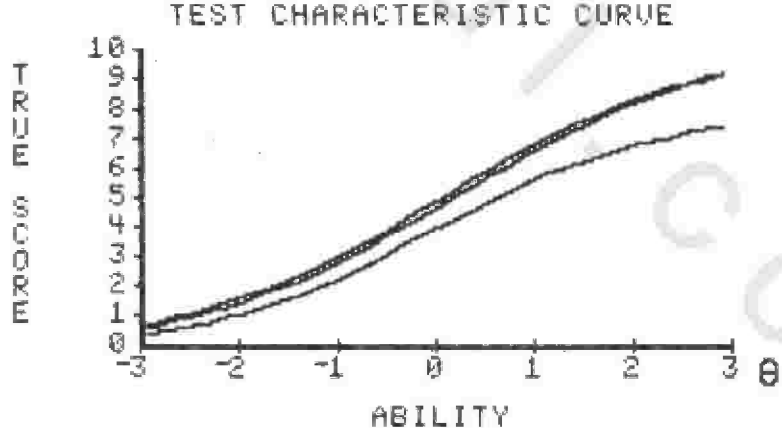
٥- ويضع إجراء الاعتماد متوسط صعوبة المفردة يساوي الصفر وكذلك نقطة المنتصف لمقياس القدرة صفر. وكانت النتيجة المباشرة لذلك وجود منتصف الدرجة الحقيقية لكل المنحنيات المميزة للاختبارات الثلاثة عند مستوى قدرة الصفر. ومرد التشابه في أشكال المنحنيات لأول مجموعتين من البيانات بسبب صعوبات المفردات التي يتم توزيعها بأسلوب متطابق تقريباً حول نقطة الصفر. وتساهم حقيقة أنه كان لكل المفردات نفس قيمة المعلم التمييزي (١,٠) في جعل الحذار أول منحنيين متماثلين. ويقع منحنى بيانات المجموعة (٣) تحت منحنيات مجموعة (٢، ١) لأنها اعتمدت على ثمان مفردات فقط. ومع ذلك شكلها العام متماثلاً مع المنحنين السابقين و منتصف الدرجة الحقيقية تقع عند مستوى القدرة صفر.

٦- وعلى الرغم من كون دوال معلومات الاختبار متشابهة إلا أننا نجد بعض الاختلافات الهامة. وقد كان لمنحنى الاختبار المتزاوج نفس المستوى العام كمنحنى الاختبار السهل لكنه كان أكثر سطحية وهذا يشير إلى احتفاظ الاختبار بمستوى دقته

وعلى مدى أوسع وبدرجة طفيفة . وقد كان لدالة معلومات الاختبار الصعب قدر أقل وبدرجة طفيفة من المعلومات في وسطه . ولذلك فقد كان له دقة أقل في هذه النقطة . ومع ذلك انخفض المنحنى بسرعة أقل من المنحنيين الآخرين مشيراً بذلك أن الاختبار لم يحتفظ بدقته على مدى أوسع من القدرة .



الشكل رقم (٧,١). المنحنيات المميزة للاختبار لثلاث مجموعات من البيانات.



الشكل رقم (٧,٢). دالة معلومات الاختبار لثلاث مجموعات من البيانات .

وضع الاختبارات الثلاثة على مقياس القدرة العام (معادلة الاختبار) Putting the Three Tests on a Common Ability Scale (Test Equating)

يشير مبدأ ثبات المفردات لقدرة المفحوص إلى أنه ينبغي على المفحوص الحصول على نفس تقدير القدرة بغض النظر عن مجموعة المفردات المستخدمة . ومع ذلك وفي معايرة الاختبارات الثلاثة التي تم إجراؤها فيما سبق ، لم يطبق ذلك . ولم تكن المشكلة في مبدأ الثبات لكن في معايير الاختبار . ويفترض مبدأ الثبات أن قيم معالم المفردة لمجموعات مختلفة من المفردات تعبر عن نفس قياس مقياس القدرة .

ويوجد في الموقف الحالي ثلاثة مقاييس مختلفة القدرة وتمثل كافة المعايير . وبسبب ذلك سيحصل نفس المفحوص على ثلاث قيم مختلفة بصورة واضحة لتقديرات القدرة بدلاً من قيمة مشتركة . وقد كان الهدف من الاختبارات الثلاثة أن يتوافق أحدها مع متوسط قدرة مجموعة مشتركة قوامها (١٦) مفحوصاً ، وأن يكون أحدها سهلاً لمجموعة والآخر صعباً لها. ومن الواضح أن متوسط الصعوبات للاختبارات الثلاثة مالت للاختلاف ، لكن عملية الاعتماد قيدت كل اختبار ليكون لديه متوسط صعوبة مفردات تصل إلى الصفر . ومع ذلك لم يتم فقدان كل شيء حيث إن جعل متوسط صعوبة مفردات الاختبار عند الصفر يؤدي إلى متوسط قدرة مقدرة للمجموعة تعكس متوسط صعوبات المفردات قبل إعادة القياس . ولذلك فما كان في الأصل اختلافات في متوسط صعوبة الاختبارات الثلاثة يصير الآن فروق في متوسط قدرة المجموعة المشتركة للمفحوصين. ومن خلال النتائج التي تم عرضها سابقاً نجد أن متوسط المجموعة المشتركة كان (٠,٠٦) للاختبار المشترك ، (٠,٤٤) للاختبار السهل ، (-٠,١١) للاختبار الصعب . ويدل ذلك على أن متوسط القدرة للاختبار المشترك يشير إلى ما ينبغي فعله . ويشير متوسط الاختبار السهل إلى أن متوسط القدرة يفوق متوسط صعوبة مفردات الاختبار ، وهذا ما ينبغي فعله تماماً وفي النهاية كان

متوسط القدرة للاختبار الصعب أقل من متوسط صعوبة المفردات ومرة أخرى يعتبر هذا ما يتوقعه الفرد . وحيث تم قياس صعوبة المفردة والقدرة بنفس القياس لذلك يمكننا استخدام متوسط القدرات لوضع الاختبارات في المقياس العام . ويصبح السؤال حينئذ (ما المقياس) ؟ ويصبح الاختيار انتقاء أي معيار مميز للاختبار ليتم استخدامه كخط قاعدي . وفي الحالة الحالية يعتبر المقياس الذي تم إنتاجه من خلال معايرة الاختبار المشترك والمجموعة المألوفة اختباراً أكثر منطقية لقياس الخط القاعدي . وهذا المعيار أنتج للاختبار، متوسط قدرة يصل إلى (٠,٠٦٢) ومتوسط صعوبة المفردات يصل إلى الصفر . وبالإضافة إلى ذلك نعرف أن اختباراً سهلاً والآخر صعباً ولذلك يبدو استخدام معيار الاختبار المتوافق كخط قاعدي أمراً ملائماً . وبسبب استخدام نموذج راش تعتبر وحدة القياس لكل المعايير الثلاثة (الوحدة) . ولجعل الاختبار السهل والصعب على مستوى قياس الخط القاعدي لابد من إجراء تعديل الفروق في النقاط المتوسطة . وفي الفقرات التالية سيتم تحويل نتائج الاختبار السهل والصعب إلى قياس قاعدي (أساسي).

اختبار سهل Easy test

إن عامل الإزاحة الذي نحن بحاجة إليه لعمل إزاحة للفرق بين متوسط تقدير القدرة في المجموعة المشتركة على الاختبار (٠,٤٤٤). وعلى الاختبار المتوافق(المتزاوج) (٠,٠٦٢) هو (٠,٣٨٢) . ولتحويل قيم صعوبات المفردة للاختبار السهل لقياس الخط القاعدي يقوم الفرد ببساطة بطرح (٠,٣٨٢) من كل صعوبة للمفردة ويتم إيضاح القيم الناتجة في الجدول رقم (٧,٥) . وبطريقة مماثلة يمكن التعبير عن كل قدرة للمفحوص في قياس الخط القاعدي بواسطة طرح (٠,٣٨٢) منها . ويتم فيما يلي إيضاح القيم المحولة في الجدول رقم (٧,٦) .

اختبار صعب Hard test

ويمكن التعبير عن نتائج الاختبار الصعب في قياس الخط القاعدي بواسطة استخدام الفروق في متوسط القدرة . وقد كان عامل الإزاحة (-١١١,٠ ، -٠٦٢,٠ أو -١٧٣,٠) ومرة أخرى يساعد طرح هذه القيم من كل تقدير من تقديرات صعوبة المفردات على وضعها في قياس الخط القاعدي . ويتم إيضاح القيم الانتقالية في الجدول رقم (٧,٥) . ويمكن تحويل تقديرات القدرة للمجموعة المشتركة التي تم الحصول عليها بواسطة الاختبار الصعب إلى قياس الخط القاعدي للاختبار المتوافق . وتم إنجاز ذلك من خلال استخدام نفس عامل الإزاحة كما تم توظيفه لإعادة قياس تقديرات صعوبة المفردات . وفي الجدول رقم (٧,٦) نتائج إعادة قياس كل تقدير لقدرة المفحوص لقياس الخط القاعدي .

الجدول رقم (٧,٥). صعوبة المفردات في الخط القاعدي .

المفردة	الاختبار السهل	الاختبار المتوافق	الاختبار الصعب
١	-١,٤٩٢	-٢,٣٧	******
٢	-١,٤٩٢	-٢,٢٧	-٠,٠٣٧
٣	-٢,١٢٢	-٠,٢٧	٠,٤٩٧
٤	-٠,١٨٢	٠,٩٨	٠,٤٩٧
٥	-٠,٥٦٢	-١,٠٠	٠,٩٦٣
٦	+٠,١٧٨	٠,١١	-٠,٤٩٧
٧	٠,٥٢٨	٠,١١	٠,٣٨٣
٨	٠,٥٨٢	٠,٥٢	١,٥٣٣
٩	٠,٨٨٠	٠,١١	٠,٤٤٣
١٠	٠,٨٨٠	٢,٠٦	******
المتوسط	-٠,٢٨٥	٠,٠٠٠	+٠,٢٢٤

وبعد التحويل يوضح متوسط صعوبة المفردة العلاقات المطلوبة في مقياس القدرة للخط القاعدي . وقد كان للاختبار المتوافق متوسط في نقطة الوسط لمقياس قدرة الخط القاعدي . وقد كان للاختبار السهل قيمة سالبة وللاختبار الصعب قيمة

موجبة . وكان متوسط صعوبة كلا الاختبارين حوالي نفس البعد عن وسط المقياس . ولدينا في المصطلحات الفنية (معادلة) الاختبارات التي وضعها في مقياس مشترك .

الجدول رقم (٧,٦). تقديرات القدرة للمجموعة العامة في الخط القاعدي.

المفردة	الاختبار السهل	الاختبار المتوافق	الاختبار الصعب
١	٢,٩٠٠-	١,٥٠-	******
٢	٠,٧٧٢-	١,٥٠-	******
٣	١,٩٦٢-	٠,٠٢	******
٤	٠,٢٩٢-	-٤٢.	٠,٨٧٧-
٥	٠,٢٩٢-	٢,٣٧-	٠,٨٧٧-
٦	٠,١٦٨	٠,٩١	******
٧	١,٩٦٨	٠,٤٢-	١,٦٣٧-
٨	٠,١٦٨	٠,٤٢-	٠,٨٧٧-
٩	٠,٦٣٨	٠,٤٢-	١,٦٣٧-
١٠	٠,٦٣٨	٠,٩١-	٠,٨٧٧-
١١	٠,٦٣٨	٢,٣٣	٠,١٥٣
١٢	١,١٨٨	٢,٣٣	٠,١٥٣
١٣	٠,٢٩٢-	٠,٤٦	٠,١٥٣
١٤	١,٩٦٨	٢,٣٣	٢,٠٠٣
١٥	******	٢,٣٣	١,٢١٣
١٦	******	******	٢,٠٠٣
المتوسط	٠,٠٦٢	٠,٠٦٢	٠,٠٦٢
الانحراف المعياري	١,٣٤٤	١,٥٦٦	١,٤١٣

ويمكن صياغة عدد من الملاحظات الهامة من هذه النتائج . فقد كان متوسط القدرة المقدرة للمجموعة العامة نفس الشيء لكل الاختبارات الثلاثة . وقد كانت الانحرافات المعيارية لتقديرات القدرة نفس الشيء تقريباً للاختبارات السهلة والصعبة ، وكان ذلك في الاختبار المتوافق تقريباً . وعلى الرغم من كون الإحصائيات النهائية متماثلة تماماً لكل المجموعات الثلاثة من النتائج إلا أن تقديرات القدرة لمفحوص بعينه قد تنوعت على نطاق واسع . ولم يشر مبدأ الثبات إلى المحراف وما تراه هو تغير في

المعاينة . وقد اشتملت مجموعات البيانات لكل معيار من معايير الاختبار الثلاثة على عدد صغير من المفردات (١٠) وعدد صغير من المفحوصين (١٦) . ونتيجة لذلك سيكون تغير المعاينة لبيانات الإجابة على المفردات كبيراً تماماً ولا يتوقع الفرد أن تكون التقديرات العديدة للقدرة نفس الشيء . وقد تم في الفصل الخامس تعريف القارئ بهذا المفهوم . وترى ذلك في هذا الفصل بصورة عملية . وبافتراض الحجم الصغير لمجموعات البيانات فهد من المدهش تماماً أن النتائج جاءت بطريقة جيدة .

ويوضح ذلك وجود إمكانيات قوية لنموذج راش ونموذج بيرنبوم لتقدير أكثر قدر من الترجيح كما تم تطبيقه في برنامج BICAL .

ويعتبر ما تم إنجازه فيما سبق معروفاً في مجال القياس النفسي كمعادلة الاختبار. وقد تم وضع كل الاختبارات الثلاثة في مقياس عام . وبعد المعادلة يمكن استخدام القيم العددية لمعالم المفردات لمقارنة مكان دالة المفردات المختلفة في مقياس القدرة . ويتم أيضاً التعبير عن القدرات المقدرة للمفحوصين ويمكن مقارنتها كذلك . وعلى الرغم من فعل ذلك ، إلا أنه من الممكن أيضاً حساب المنحنى المميز للاختبار ودالة معلومات الاختبار للاختبارات السهلة والصعبة في قياس الخط القاعدي . وبلغة فنية ، يتم معادلة الاختبارات باستخدام الأسلوب الجماعي العام مع اختبارات الصعوبة المختلفة ويعتبر إيجاد السهولة التي يتم من خلالها معادلة الاختبار من إحدى المزايا الرئيسة لنظرية الاستجابة للمفردة التي تتميز بها عن نظرية القياس التقليدية .