ثبات المقياس وصدقه بين الطرق التقليدية ونموذج راش: البديل الإحصائي في تطوير المقياس النفسي

ميكائيل إبراهيم و أسامه عمر العزابي
ibidun18@yahoo.com
جامعة العلوم الإسلامية الماليزية، ماليزيا

استلم بتاريخ: ٢٠١١\٢\١٦ عُدل بتاريخ: ٢٠١١\٥\٨ قُبل بتاريخ: ٢٠١١\٥\١٠

تناولت الدراسة الحالية موضوع ثبات المقياس وصدقه من خلال المقارنة بين الطُّرق التقليدية، وبين نموذج راش الذي ينظر إليه الإحصائيون على أنه البديل الإحصائي الأفضل في تطوير المقياس النفسي من حيث الخصائص السيكومترية، لما للنموذج من إمكانيات هائلة قلّما تجدها في الطرق التقليدية الأخرى. فلقد أثبت الباحثون فشل الطرق الإحصائية التقليدية في مراعاة سمات المفردات، والأشخاص عند اختبار ثبات المقاييس أو المفردات، وعادة ما يلجاؤون إلى معامل ألفا في إثبات الاتساق الداخلي للمقياس بدون مراعاة - في كثير من الأحيان - تأثر معامل كرونباخ بالعوامل الخارجية غير المقاسة. أما نموذج راش الذي يعتبرنمطاً جديداً في مجال الإحصاء فإنه لا يتأثر بالعوامل غير المقاسة مثل طول المقياس، حتى قد يكون المقياس القصيرمن حيث عدد فقراته أكثر ثباتاً من المقياس الطويل. علاوة على ذلك، فإن نموذج راش قد يستخدم في اختبار صدق المقياس مثل صدق المحتوى، وصدق البناء، وصدق المحك، كما لنموذج راش إمكانية هائلة في تحديد مساهمة كلّ فقرة من فقرات المقياس وكلّ فرد من أفراد العيّنة في جودته الكلية. فمثلاً، فإن خريطة التغير تُستخدَم لاختبار مدى استهداف مفردات مقياس المفحوصين بمراعاة كلٍّ من صعوبة المفردات وقدرة المعينين. أما الطرق التقليدية فإنها تركّز على مجموع المربعات Sum of square، كما لا تقدم هذه الطرق الإحصائية الأخطاء القياسية لكل المفردات لتحديد الدقة العامّة للقياس، لذا، أثبت الباحثان في هذه الورقة أفضلية نموذج راش في اختبار ثبات المقياس وصدقه ودراسة مساهمة كل فقرة أو فرد على حدة في جودته.

The reliability and validity of a scale through the comparison between the traditional method and the Rasch model

Mikail Ibrahim & Osama Omar M. Elazzabi
Universiti Sains Islam Malaysia, Malaysia

The present study addressed the issue of reliability and validity of scale through the comparison between the traditional method and the Rasch model, which is seen by statisticians as the best method to psychometrically validate the scale and test its properties. Researchers have demonstrated the failure of traditional statistical methods to take into account the characteristics of the items and people when testing the reliability of the scale as well as respondents. They usually turn to alpha Cronbanch to examine the internal consistency of the scale without taking into consideration that alpha Cronbanch has been affected by external factors such as the length of the scale. However, the Rasch model that is relatively considered to be modern statistics, is not affected by external factors especially the length of the scale, even a short scale might be more reliable than the long one. Moreover, the Rasch model can also be used to investigate various types of validities such as content validity, construct validity, and criterion validity. Interestingly, the Rasch model is also a powerful statistics tool used to determine the contribution of the items and people in the total reliability of the scale. For example, the variable map is used to examine the extent to which the items adequately target the respondents taking into account the difficulty of items and ability of the subjects. Nevertheless, the traditional methods normally calculate the uniqueness of the scale by focusing on the sum of squares and these methods do not offer standard errors for each item to determine the accuracy of the measurement. Hence, the researchers suggested in this paper that favorability of the Rasch model in testing scales' reliability and validity was recommended compared to conventional statistical methods.

المقدمة:

اهتم علم النفس والتربويون واللغويون بالقياس منذ قرنين من الزمن تقريبا، إلا أن هذا الاهتمام ازداد ازديادا ملحوظا منذ بداية القرن المنصرم (١٩٠٠) عندما صمم بينيت (Binet) القياس لاختبار القدرة العقلية والمعرفية (الذكاء) للطلبة في فرنسا (كوهين سوردليك، ٢٠٠٧). وكان الهدف من هذا الاهتمام هو محاولة تحقيق الدقة في المعايير وإصدار الأحكام الموضوعية دون تعسف أو انحياز. وقد استهدف بينيت (Binet) اختبار القدرة العقلية للأطفال (نسبة الذكاء) بغية تقسيمهم إلى مستويات مختلفة على حسب قدراتهم العقلية وخلفياتهم المعرفية. ويعد اختراع الأرقام من أعظم الاختراعات بعد اللغة عبر العصور؛ لأن الأرقام تستخدم للقياس والتعداد معا، والتطورات العلمية والتكنولوجية المذهلة التي يشهدها العالم اليوم ما كان يمكن إنجازها لولا وجود الأرقام. وبدون القياس يصبح العلم مقتصرا على الملاحظات الانطباعية العابرة، والتصنيف الكيفي، والوصف اللفظيي، والأحكام الذاتية (علام، ٢٠٠٥). تاريخيا، فإن القياس لم يكن من اختراع علماء الرياضيات أو العلوم، وإنما بدأ في مجال البيوع والهندسة المعمارية، قبل بروز العلوم كالتخصص أو المهنة. فالقياس هو عملية تكميم الظواهر السلوكية ووصفها وصفا كميا بغية الحصول على قوانين يمكن تعميمها على الظواهر السلوكية المشابهة.

إلا أنالاهتمام بالقياس وجد تطور اكبير ا في الرياضيات والعلوم خاصة بعد الثورة المعلوماتية، مما دعا الخبراء في العلوم الإنسانية خاصة في علم النفس، والتربية، واللغة، والاقتصاد لتبني طريقة العلوم في تقعيد القياس وتصميمه، واختبار صلاحيته، ومصداقيته.

وقد أشار بوند و فوكس (Bond & Fox) (٢٠٠٧) إلى أن تعارض نتائج الدراسات العلمية في العلوم الإنسانية وخاصة في التربية وعلم النفس يرجع أساسا إلى مشكلة القياس، لذا لا بد من زيادة الاهتمام بكيفية تصميم القياس لتفادي هذه المشكلات التي يعاني منها المقياس، أو الاختبار في هذه العلوم.

يقول علام ٢٠٠٦: فالتقدم العلمي يعتمد إلى حد كبير على تمثيل الظواهر والأحداث وصياغتها بأساليب موضوعية دقيقة، بحيث تمكن الدارسين والباحثين والممارسين من التواصل فيما بينهم بلغة مشتركة متفق عليها، كما تمكنهم من التقييم الموضوعي للنتائج التي يتوصل إليها العلماء، وبدون هذه الأساليب يصبح تعريف المفاهيم والمصطلحات والقواعد المتعلقة بالظواهر المختلفة خاضعا للآراء الذاتية، والأحكام الفردية، وبذلك لا تكون هناك أسس متفق عليها في دراسة وبحث هذه الظواهر" (صـ١٣).

وبخلاف العلوم الإنسانية فقد تقدم القياس بصورة ملحوظة في العلوم الطبيعية نظر ا لأنها تهتم بقياس عناصر الطبيعية المشخصة القابلة للتقييم بصورة دقيقة ومباشرة بسبب توفر الأدوات والقواعد المقننة مثل قياس الحالة الجوية وقياس الوزن، والطول أو غير ذلك من العناصر الطبيعية التي لها أدوات قياس مقننة متفق عليها. لكن ليس معنى ذلك أن جميع العناصر الطبعية سهلة القياس إلا أنه مهما كان الأمر فإن الصعوبة التي قد تواجه الباحث عند قياس العناصر الطبيعية قليلة نسبيا مقارنة بالصعوبات التي قد تواجهه عند قياس النفس البشرية. فالإنسان كائن حي متحرك ونفسيته متشابكة ومعقدة للغاية، وأقل قابلية للتحكم، ومن

الصعوبة بمكان إخضاعها للقياس والعملية التجريبية. "أما عالم النفس فمن الصعب عليه إخضاع الإنسان للتجارب المعملية، وتثبيت جميع المتغيرات التي يود عزل أثرها" (علام، ٢٠٠٦) بسبب سرعة التقلب الذي تتسم به النفس البشرية كان من الصعوبة بمكان قياسها بطريقة علمية معملية.

ويمثل الباحث عملية تصميم القياس وتقنينه بإجراءات الخياط عندما يحاول أن يفصل الملابس لشخص ما، عليه أن يكون دقيقا في قياس الحجم أو الطول لذلك الشخص. فإذا أخطأ في تحديد الحجم أو الطول زيادةة أو نقصانا، فإن ذلك الخطأ سيقود إلى إنتاج عينة من الملابس غير مناسبة للشخص الذي تم قياسه.

أجاز العلماء تبني القياس المقنن الجاهز للاستخدام في البحث العلمي سواء بالتعديل حتى يتمشى مع الظروف الجديدة التي سيستخدم فيها، أو بدون تعديل شريطة أن يكون مطابقا تمام التطابق للبيئة الجديدة التي سيستخدم فيها، وألا يزيد عمر المقياس عشر سنوات لدى بعضهم، وأن يستوفى الشروط العلمية اللازمة. إلا أن تبني قياس رديء وبصورة عشوائية واعتباطية قد يؤدي إلى نتائج عكسية بحيث يحصل الباحث على بيانات غير دقيقة، وبالتالي كل النتائج المستخلصة منه تكون مضللة وبدون مغزى أو معنى.

وتفاديا لتبني قياس خاطيء، فإن الخيار الوحيد المتبقي للباحث هو تصميم قياس خاص به يراعي كلالجوانب العلمية التي يريد دراستها بناءً على النظرية العلمية التي تبناها في دراسته. ولكن هناك بعض المعايير التي لا بد لمصمم القياس أن يراعيها كي يكون المقياس صالحا للاستخدام في البحث العلمي الهادف. ومع صعوبة قياس النفس البشرية فإنه يفضل على تبني المقياس الجاهز خاصة إذا استورد من ثقافة مغايرة للثقافة التي يراد أن يستخدم فيها أو المجتمع الغريب عنها. كذلك يستحسن ألا يكون عمر المقياس المتبنى قد تجاوز عشر سنوات مع مرعاة ما قد يطرأ على النظرية المعتمد عليها في المقياس من تعديل وتطوير وغير ذلك. "وذلك باتباع خطوات مماثلة للخطوات المتبعة في القياس الفيزيائي، وابتكار طرق وأساليب وأدوات قياس متنوعة تتميز بخصائص محددة" (علام، ٢٠٠٦، صـ١٧).

سيحاول الباحثان من خلال هذه الدراسة المشاركة في مناقشة قضايا الثبات والصدق في القياس النفسي وأهميتهما في تقنين القياس. وقد قارن الباحثان بين الطرق التقليدية خاصة اختبار معامل ألفا وبين نموذج راش في مسألة ثبات الاختبار وصدقه بغية إيجاد الخصائص السيكومترية التي تساعد الطلبة والباحثين على حد سواء في تصميم القياس المتسم بالمصداقية والثبات, والقادر على الفحص العلمي الموضوعي الدقيق.

الدراسات السابقة:

يعد الثبات قضية أساسية ومهمة في القياس (Ghiselli, Campbell & Zedbell, 1981) حيث يعمل على تجانس مفردات المقياس. وهو عبارة عن الاتساق الداخلي بين بنود الاختبار أو فقرات المقياس. وعادة ما ينوي الميزان على قياس ظاهرة واحدة أي هو أحادي البعد. وتظهر العلاقة بين مفردات المقياس دليلا على أنه يقيس السمات الكامنة المشتركة بين هذه المفردات. وتدل قوة الارتباط بين المفردات على شدة قواسم مشتركة بينها، أي أنها تقيس شيئا واحدا مشتركا، وبالتالي يمكن القول إن شدة الارتباط بين المفرادات يدلعلى الربط بينها أي (المفردات) و العامل الكامن. لذا اشترط كي يحقق المقياس خاصية أحادية البعد لا

بد أ تكون العلاقة بين مفرداته قوية، وارتباطاتها متينة.

ويختص الثبات بمدى الثقة بالدرجات التي يحصل عليها الباحث من تطبيق الاختبار أو المقياس، بمعنى أن هذه الدرجات أو النتائج يجب ألا تتأثر بالعوامل التي تعود إلى أخطاء الصدفة. فالدرجة التي يحصل عليها المجيب في مقياس معين تسمى الدرجة الملاحظة، ولكن غالبا ما تشوب هذه الدرجةَ الملاحظة الأخطاء؛ لأن لكل عملية قياس أخطاء تمنع - في الغالب - الحصول على الدرجة الحقيقية التي حصل عليها المجيب في الاختبار. وهناك مصادر كثيرة للأخطاء في القياس، فقد يأتي الخطأ من بناء المقياس، وقد يأتي من إدارته، أوقد تكون الأخطاء مجهولة المصدر. ولكي يعرف الباحث دقة المقياس عليه أن يطرح درجة الخطأ من الدرجة الملاحظة وبالتالي يحصل على درجة خالية من الأخطاء العشوائية. وتسمى هذه الدرجة المتحصل عليها الدرجة الحقيقية لشخص معين، هي عبارة عن الدرجة الملاحظة مطروحا منها درجة الأخطاء العشوائية. ومن ثم فالثبات يحدد الدقة التي تقيس بها الأداة أي شيء تقيسه. والجدير بالذكر أن الثبات ليس بضاعة المقياس، بل البيانات التي جمعت من خلاله؛ لأن الاختبار والمقياس قد يتسم بالثبات في حالة ولا يتسم به في حالة أخرى.

طُرق اختبار ثبات القياس:

طريقة إعادة تطبيق الاختبار

(Test retest method)

طريقة الصور البديلة: المتكافئة

(Parallel Form Method)

طريقة التجزئة النصفية (Split-half Method)

طريقة الاتساق الداخلي.

وليس القصد من هذه الورقة مناقشة هذه الطُرق بل سنعتني فقط بدراسة الاتساق الداخلي وتمايز طريقة نموذج راش في تقديره.

العينة:

تم اختيار العينة من أربع جامعات ماليزيا ومعهد تدريب المعلمين، وهي جامعة العلوم الإسلامية الماليزية، وجامعة ملايا، وجامعة فترا، والجامعة الإسلامية العالمية في ماليزيا. وتشمل هذه العينة ٧٥٢ طالبا وطالبة بمختلف أعمارهم وتخصصاتهم، وقد تم اختيارهم عشوائيا من هذه الجامعات الواقعة بين ولايتي سيلانجور، ونجري سنمبلان. وقد اعتمد الباحثان على مقياس مارش (Marsh, 1995) لتقييم الطلبة في جودة التعليم. والمقياس عبارة عن استبانة مكونة من ٢٩ فقرة موزعة على ثمانية عوامل رئيسة باستخدام التحليل العاملي. هذه العوامل هي التعلم/ أو القيمة التحصيلية، الحماسة التدريسية، التنظيم، التفاعل بين الأقران، الألفة الفردية، اتساع نطاق التغطية، التعمق، الاختبارات، عبء الواجبات. وقد أشادت دراسات كثيرة بالمقياس، وتم استخدامه في مقالات علمية متنوعة، كما ترجم إلى عدد من اللغات بسبب ما تميز به من الصدق التكويني، والثبات الداخلي العالي. ولقد قدمت هذه الاستبانة للمفحوصين مستهلة بالمعلومات أو المتغيرات الديموغرافية التي تتعلق بالجنس، والعمر، والسنة الدراسية، الجامعة، الكلية، أو العرق.

اختبار ثبات الاستبانة:

لقد استخدم الباحثان عدة طرق لاختبار صدق الاستبانة وثباتها منها التأكد من خلو الأسئلة من الأخطاء اللغوية والأسلوبية، وتسهيل مفرداتها لسهولة الفهم والاستيعاب. وتم استخدام معامل ألفا

لاختبار الثبات. وأظهرت النتيجة أن الاستبانة أوفت بجميع الشروط اللازمة بوصفها صالحة للاستخدام لوضوحها ودقتها، وأنها قاست ما صيغت من أجله حيث إن معامل ألفا تراوح ما بين ٩٣. إلى ٩٧. لكل فقرة من فقرات الاستبانة. اعتمد الباحثان على معامل ألفا في التحقق من ثبات المقياس لأن اختبار ثبات المقياس دلالة على مصداقيته وأنه يؤثر إيجابا أو سلبا على نتائج البحث والاستنباطات المستخلصة منه. كما تم الاعتماد على مقياس ليكرت (Likerts) في إعطاء الوزن النقطي (درجات) لكل بند من بنود الأسئلة. وسبب استخدام مقياس (ليكرت) يعود إلى ما يتيحه من إمكانية منهجية القراءة الشمولية للمعطيات الإحصائية (علي وطفة، بدون تاريخ).

منهج البحث:

يعتمد الباحثان في تحليل هذه البيانات على نموذج راش الإحصائي (Rasch Model) بغية اختبار ثبات هذه الاستبانة وصدقها التكويني. كما يساعد أيضا التعرف على أبعاد المقياس المعتمد وصلاحيته للقياس النفسي. وقد استهدفا (الباحثان) دراسة التميز الذي يتسم به تحليل راش في دقة اختبار المقياس وصدقه، وتحديد مساهمة كل من فقرات المقياس والأشخاص المشاركين في الإجابة عليها.

إن نموذج راش الإحصائي طريقة إحصائية حديثة تستخدم في المقابل الطرق الإحصائية التقليدية كرد فعل لإفراط الطرق التقليدية في الاعتماد على النسبة الفائية لقبول الفرضية، أو رفضها. لقد انتقد الإحصائيون طريقة فيشر الإحصائية التقليدية بسبب إفراطها في الاعتماد على النسبة الفائية (p-value) في قبول الفرضية، ورفضها بدون مراعاة العناصر الأخرى المؤثرة في التحليل الإحصائي، مما قد يؤدي في بعض الأحيان إلى التبني، والاعتماد على الحكم الخاطئ. وبخلاف الطرق الإحصائية التقليدية فإن نموذج راش ينبعث من مبدأ الاتحاد التام بين المفردات والأشخاص في تحليل الظواهر. وهذا يعني أن لنموذج راش إمكانية هائلة في تقييم قدرة الأشخاص، وتحديدها مع مراعاة صعوبة المفردات آنيا ومتزامنا.

لقد أثبت الباحثون (أويل Harwell، و غتي Gatti ٢٠٠١، وبوند Bond، و فوذ، ٢٠٠١) فشل الطرق الإحصائية التقليدية في مراعاة سمات المفرادات، والأشخاص عند اختبار ثبات المقاييس أو المفردات. تقليديا، فإن الباحثين أو الإحصائيين يركزون على مجموع المربعات Sum of square ويصدرون الحكم الحاسم بناء على العدد الكمي. كما لا تقدم هذه الطرق الإحصائية الأخطاء القياسية لكل المفردة، والشخص لتقدير الثبات، والدقة العامة للقياس. علاوة على ذلك، فقد اعتبر الخبراء أن تعارض نتائج الدراسات التجريبية في العلوم الإنسانية، وخاصة في التربية وتعليم اللغات وتعلمها قد يحل بزيادة العناية على المفردات والمقاييس.

كما يحاول الباحثان من خلال استخدام هذا النموذج تحقيق عدة أهداف منها أولا: تصميم المقياس المستوى الفتري (Interval Scale) للقياس حيث يستطيع المعادلة بين المفردات والأشخاص. ثانيا، تحقيق الصدق البنائي بصورة شاملة بحيث يراعي قدرة الأشخاص، وصعوبة المفردات تزامنيا. إضافة إلى ذلك فإن هذه الدراسة تقدم تحليلات مناسبة للخطأ المعياري جنبا إلى جنب مع الخطأ الإجمالي لثبات المفردات والأشخاص.

لذا، يحاول الباحثان في هذه الدراسة الاعتماد على المنهج الحديث والطريقة الإحصائية الجديدة للفحص ما إذا كانت العوامل الكامنة المخفية تظهر نفس السمات لكلٍ من المفردات، والأشخاص، وذلك

بمراعاة الصدق البنائي، وموافقة الأشخاص (المفحوصين)إ قرارهم على كل المفردات الموجودة في الدراسة.

التحليلات الأولية للبيانات:

يعد التحليل الأولي للبيانات المستخدمة للإحصاء الاستدلالي من المتطلبات الأساسية. وهو عبارة عن اختبار مدى ملاءمة البيانات للطريقة الإحصائية التي يتبناها الباحثان في تحقيق أغراضهما البحثية. ويمكن اختبار مدى ملاءمة البيانات التي تستخدم في إجراء التحليل الكمي من خلال الاتساق الداخلي للأداة المستخدمة وصلاحيتها. وللتحقق من ثبات البيانات واتساقها الداخلي، أجرى الباحثان اختبار معامل كرومباخ ألفا Cronbanch Alpha، وتتراوح درجة الثبات لكل فقرة من فقرات المقياس كما ذكرنا سابقا ما بين 93.-97. وتشير هذه النتيجة إلى قوة ثبات البيانات، وأنها صالحة للاستخدام في البحث العلمي.وعلاو ة على ذلك، فقد تحقق الباحثان من قيمة الالتواء والتفرطح اللذين يدلان على التوز يع الاعتدالي والطبيعي للبيانات المستخدمة، وتشير نتيجة التحليل بعدم وجود أي انحرافات تذكر في البيانات، حيث إن درجة الالتواء والتفرطح تتراوح ما بين 2- و 2+ كما يفترض أي بين مساحة محددة للاختبار. بينما تم استخدام Mahalanobis للتحقق من افتراض تعدد المتغيرات. وعندما تم إجراء مزيد من التجارب للتحقق من جودة البيانات وصلاحيتها باستخدام كولموغوروف-سميرونوف (Kolmogorov-Smironov) أظهرت النتيجة أن الاختبار غير دال إحصائيا، أي أن النسبة الفائية كانت أكبر من (p > .05)، فيما عدا حالات بسيطة، في حين (P > .05) يعني التوز يع الاعتدالي للبيانات. فضلا عن ذلك، فقد أكدت نتيجة اختبار شابيروا ويلك (Shapiro-Wilk) هذه المزاعم, حيث أشارت إلى صلاحية البيانات، وعدم تأثرها بالعوامل الخارجية مثل أخطاء الصدفة. واستنادا إلى هذه النتائج، يمكن أن نقول إن هناك توذيعا اعتداليا للبيانات فضلا عن صلاحيتها وتبرير استخدام الطريقة البارامترية في هذه الدراسة.

نتائج تحليل الدراسة:

يتضح من جدول رقم (١) لتحليل البيانات الأساسية للمشاركين في هذه الدراسة، بأن عدد المشاركين٧٥٢، علما بأن ٣٠١ مشاركا (٤٠%) من أفراد العينة كانوا من الذكور، يقابلهم ٤٥١ مشاركة (٦٠%) من الإناث. ويمثل هذا التباين في العينة المجموعة الأصلية (population) في مجتمع الدراسة حيث إن عدد الإناث يفوق بكثير عدد الذكور الذين يدرسون في الجامعات الماليزية وكلياتها. أما فيما يتعلق بالعمر، فإن الغالبية الساحقة من المفحوصين (٥٥٠، ٧٣%) تتراوح أعمارهم ما بين 20-18 سنة، مقابل ٢٠٢ (٢٧%) تتراوح أعمارهم ما بين 25-22 سنة.

أما الجامعات التي ينتمي إليها أفراد عينة هذه الدراسة، فقد أظهرت نتائج التحليل أن ٢٢٠ طالبا يمثلون الجامعة الإسلامية العالمية (IIUM) (29.3%), في المقابل نجد 180 طالبا بنسبة (23.9%) من معهد تدريب المعلمين (IPI). أما طلبة جامعة العلوم الإسلامية (USIM) فيشكلون نسبة 13.4% أي 101طالبٍ ، وتشكل الجامعة الوطنية (UKM) 12% من أفراد العينة أي ٩٠ طالبا، وجامعة ملايا(UM) ٩٩ طالبا، أي بمعدل 13.2%، وجامعة فترا (UPM) كانت نسبتها 6.9%، أي ٥١ طالبا، أما بقية الجامعات فتمثل 1.46%، أي ١١ طالبا من عدد المشاركين في هذه الدراسة.

جدول (١)

التحليل الوصفي لمعلومات أساسية للمشاركين في الدراسة

معلومات أساسية		عدد	نسبة مئوية
الجنس	الذكور	301	301%
	الإناث	451	451%
العمر	18-20	550	74.0%
	21-23	202	27.0%
الجامعة	الجامعة الإسلامية العالمية	220	29.3%
	معهد تدريب المعلمين	180	23.9%
	جامعة العلوم الإسلامية الماليزية	101	13.4%
	الجامعة الوطنية	90	12.0%
	جامعة ملايا	99	13.2%
	غيرها	51	6.9%
		11	1.45%

إجراءات البحث:

استهل الباحثان هذه الدراسة بإجراء التحليل العاملي وذلك لقصد تلخيص المتغيرات واختصارها في عدد أقل من العوامل. فهذه الطريقة عبارة عن ارتباط عدد من المتغيرات ببعضها ارتباطا عاليا، وارتباطها بغيرها ارتباطا منخفضا. علاوة على ذلك، فقد لجأ الباحثان إلى التحليل العاملي لغرض دراسة أبعاد المقياس المستخدم والتأكد أنه استوفى شرط أحادية البعد والتي تعد شرطا أساسيا في تحليل راش. قد أثبتت نتائج التحليل العاملي أن مفردات الأسئلة محملة حسب المتوقعات.

وبناء على نتيجة التحليل العاملي أجرى الباحثان تحليل نموذج راش وذلك عن طريق استخدام برنامج الإحصاء (Winsteps) الذي صممه مارك ليناكار 1991-2004. يوفر هذا البرنامج نوعين من جودة التوفيق (١) جودة التوفيق الداخلية Infit Meansquare (٢) جودة التوفيق الخارجيةOutfit Meansquare، ولكل واحد منهما امتيازتها ومجالاتها. أما جودة التوفيق الداخلية فإنها مجموع الثقل للمعلومات في البيانات المجتمعة لدى المفحوصين، وأما جودة التوفيق الخارجية فإنها عبارة عن مجموع المربعات للخطأ المعياري standardized residual بطريقة تقليدية. وتعد كلا الجودتين (الداخلية والخارجية) مربع المتوسطات مقسم بالدرجة الحرية Degree of Freedom، وتتراوح هذه الدرجة الحرية بين ٠-١ (بوند وفوذ ٢٠٠١، سلفار، سميث و غرين، ٢٠٠١). وأكد غرين وزملاؤه أن مربع متوسط إحصائي أقرب من واحد يدل على التكرار والاتكاء والتقيد للمعلومات، كما أن مربع متوسط أكبر من واحد يدل على التفاوت غير المتوقع، والتناقض الداخلي للمعلومات. وتكون جودة التوفيق الداخلية والخارجية بدرجة مقاربتها من واحد، لأن درجة ١ تدل على دقة تلائم البيانات وا إصابتها للنموذج المدروس. وجدير بالذكر أن قلة الإنحراف بين المعلومات والنموذج هو المطلوب في نموذج راش.

وبناء على توصية الإحصائيين والخبراء التوبويين (رايت و ليناكار، Wright & Linacre, 1994، ١٩٩٤) فإن النطاق المقبول (acceptable

range) للجودة الداخلية والخارجية لهذه الدراسة تتراوح ما بين 67.-١.٤٠ حيث يعد هذا النطاق مقبولاباعتباره قريبا نسبيا من النطاق المثالي المتقن لنموذج راش. وكلما تبتعد جودة التوفيق الداخلية والخارجية من هذا النطاق كلما كانت عناصر أوى غير المقاسة تلعب دور ا حيويا في قرار أفراد العينة. علاوة على جودتي التوفيق: الداخلية والخارجية فإن نموذج راش يوفر الثبات الكلي، ومجموع الخطأ المحتمل للتحليل.

التعريف بنموذج راش:

لقد أثبت الباحثون (هارويل Harwell و غاتي Gatti ٢٠٠١، وبوند Bond، و فوكس، ٢٠٠١) فشل الطرق الإحصائية التقليدية في مراعاة سمات المفرادات، والأشخاص عند اختبار ثبات المقاييس أو المفردات. تقليديا، فإن الباحثين أو الإحصائيين يركزون على مجموع المربعات Sum of square ويصدرون الحكم الحاسم بناء على العدد الكمي. فضلا عن أن هذه الطرق الإحصائية لا تقدم الأخطاء القياسية لكل المفردات، والأشخاص لتقدير الثبات، والدقة العامة للقياس. ويعين نموذج (السمات الكامنة) العلاقة المتوقعة بين الاستجابات الملاحظة على اختبار ما، والسمات غير الملاحظة التي يفترض أنها تحدد هذه الاستجابات، وتوفر نماذج السمات الكامنة تقدير ا للقدرة مستقلة عن العينة، كما توفر أيضا مميزات ذات فئات متساوية (أمينة محمد كاظم، ١٩٨٨). وقد يحقق الباحثون من جراء تبنيهم لنموذج راش عدة أهداف، منها: تصميم مقياس الفترات (Interval Scale) للقياس الذي يستطيع المعادلة بين المفردات والأشخاص. ثانيا، تحقيق الصدق البنائي بصورة شاملة بحيث يراعي قدرة الأشخاص، وصعوبة المفردات تزامنيا. فضلا عن ذلك التقدير الدقيق للخطأ المعيار ي جنبا إلى جنب مع الخطأ الإجمالي لثبات المفردات والأشخاص.

ومن فروض نموذج راش (١) أحادية البعد (٢) استقلالية القياس (٣) توازي المنحنيات المميزة للبنود أو المفردات (٤) تفاعل قدرة الفرد مع صعوبة البند أو المفردة (بوند وفوكس ٢٠٠١م، كاظم، ١٩٨٨).

طريقة الاتساق الداخلي:

تعتمد هذه الطريقة على مدى ارتباط الوحدات أو البنود مع بعضها بعضا داخل الاختبار أو المقياس، وكذلك ارتباط كل بند مع الاختبار ككل. ومن أكثر المعادلات استخداما لقياس الاتساق الداخلي، هي: معادلة كودر-ريتشاردسون ومعامل ألفا. إلا أن جل الاهنمام سيركز على معامل ألفا في السطور التالية.

عادة ما يرادف الاتساق الداخلي بكرونباخ ألفا (α) لأسباب كثيرة، منها:

الشيوع: عادة ما يلجأ الباحث إلى معامل ألفا؛ لإثبات جودة المقياس، واتساقه الداخلي لتبرير استخدامه وتبنيه في إجراء البحث العلمي.

يعد ارتباطَ ألفا بثبات المقياس أشد وأوضح مقارنة بالطرق الأخرى.

ومع شيوع معامل ألفا وكثرة استخدامه في البحوث العلمية والمقالات المنشورة إلا أنه أُثبتَ علميا أنه غالبا ما يتأثر بطول المقياس بحيث تزداد قيمة معامل الثبات بزيادة عدد أسئلة المقياس أو المفردات.

فمعادلة معامل ألفا هي:

$$\alpha=(N)(r_{msan})/[1+(r_{msan})(N-1)]$$

حيث N يرمز إلى العدد الكلي لمفردات الاختبار أو المقياس و mean متوسط الارتباط بين المفردات. فمثلا، نفترض أن لدينا ٢٠ مفردة، والارتباط بين هذه المفردات = 0.30، يمكن حساب معامل ألفا كما يلي:

$$\alpha=\frac{(20)(0.30)}{[1+(0.30)(20-1)]}-\frac{6}{6.7}-0.896$$

وتظهر هذه المعادلة أن تقدير معامل ألفا ليس مرهونا ومعتمدا على الارتباط بين المفردات فقط، ولكن كذلك على عدد المفردات. فمثلا، إذا قلصنا عدد المفردات في المعادلة السابقة إلى ١٠ مفردات تكون النتيجة كالتالي:

$$\alpha=\frac{(10)(0.30)}{[1+(0.30)(10-1)]}-\frac{3}{3.7}-0.810$$

ويرجع ذلك إلى أن زيادة عدد الأسئلة أو المفردات يؤدي إلى شمول أكثر للمحتوى، ومن ثم صدق محتوى مرتفع. ومن هنا يمكن القول أنه كلما زاد عدد المفردات ارتفع معه معامل الثبات.

ومن جانب آخر، أثبت الإحصائيون أن الثبات في نموذج راش لا يتأثر بطول المقياس، بل حتى المقياس القصير (أي بالنسبة إلى عدد المفردات) قد يكون أكثر ثباتا وجودة من المقياس الطويل. وهذا الملخص لا يعني التقليل من قيمة معامل ألفا بقدر ما ينبه الباحثين إلى عدم الغرور بنتيجته (سميت Smith، ٢٠٠٤). وعلى الباحث التأكد من نتيجة معامل ألفا عن طريق استخدام الطرق الأخرى لفحص نتيجة معامل ألفا وإضفاء صلاحية أكثر عليه ويفترض نموذج راش أن تكون القدرة المقاسة أحادية البعد. وبسبب هذا الافتراض تكون العلاقة بين مفردات المقياس قوية وارتباطاتها متينة ويؤدي ذلك إلى ارتفاع مستوى الثبات بغض النظر عن طول المقياس أو قصره. ومن فوائد نموذج راش أيضا أنه تتوفر فيه خاصية الاستقلالية، أي عدم تباين Invariance تقديرات قدرة الأفراد بتباين عينة المفردات أو المقياس، وعدم تباين الخصائص السيكومترية للمفردات بتباين عينة الأفراد التي أجري عليها الاختبار (علام، ٢٠٠٦). وعلى الرغم من تباين صعوبة مفردات الاختبار أو المقياس، فإن نموذج راش يفترض كذلك أن جميع المفردات تتساوى في قدرتها على التمييز.

جدول (١-١)

تلخيص ثبات للأشخاص مدخل: ٧٥٢ شخصا، و٢٩ مفردة

SUMMARY OF 737 MEASURED (NON-EXTREME) PERSON

	TOTAL SCORE	COUNT	MEASURE	MODEL ERROR	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD
MEAN	158.6	29.0	1.42	.25	1.08	-.3	1.04	-.4
S.D.	27.1	.0	1.39	.11	.89	2.7	.80	2.6
MAX.	202.0	29.0	6.05	1.01	6.50	9.9	6.08	9.5
MIN.	36.0	28.0	-2.74	.14	.06	-6.7	.06	-6.8

REAL RMSE	.31	TRUE SD	1.35	SEPARATION 4.34	PERSON RELIABILITY .95
MODEL RMSE	.27	TRUE SD	1.36	SEPARATION 5.04	PERSON RELIABILITY .96
S.E. OF PERSON MEAN = .05					

جدول (٢)

تلخيص ثبات للمفردات مدخل: ٧٥٢ شخصا، و ٢٩ مفرداة

SUMMARY OF 29 MEASURED (NON-EXTREME) ITEM

	TOTAL SCORE	COUNT	MEASURE	MODEL ERROR	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD
MEAN	4134.6	751.9	.00	.04	.99	-.6	1.04	-.3
S.D.	184.1	.0	.33	.00	.30	4.4	.41	4.3
MAX.	4383.0	752.0	.75	.05	1.90	9.9	2.54	9.9
MIN.	3685.0	751.0	-.48	.04	.69	-5.7	.74	-4.7

REAL RMSE	.05	TRUE SD	.32	SEPARATION	7.13	ITEM	RELIABILITY	.98
MODEL RMSE	.04	TRUE SD	.32	SEPARATION	7.46	ITEM	RELIABILITY	.98
S.E. OF ITEM MEAN = .06								

فضلا عن ذلك، يعتمد معامل ألفا على بيانات خام في حساب الثبات وتقديره، وهي إجراءات تعد مضللة؛ لأن البيانات الخام غير خطية (Non-linear) والاعتماد على هذا النوع من البيانات يكون غير علمي لأنه يظهر عكس ما تتضمنه البيانات من تأثيرات مما يترتب عليه التوصل إلى معلومات غير دقيقة وغير واقعية.

أما الخطأ المعيار ي للقياس (Measurement Standard Error)، فإنه من الخصائص السيكومترية الأساسية للاختبارات أو المقاييس، حيث يعتمد عليه دائما في تحديد دقة الثبات، وتحليله، وتفسيره. وتقع الطرق التقليدية في المشكلة نفسها بسبب اعتبار الخطأ المعياري ثابتا عبر المفردات، واستخدامه للتحقق من الثبات لكل درجة ممكنة في المقياس أو الاختبار دون مراعاة للدرجات المتطرفة Extreme scores للقياس، علما بأن الدرجات المتطرفة أقل دقة من متوسط الدرجات (الذي يعتمد عليه نموذج راش في تقدير الثبات) مما يؤثر سلبا على دقة تحديد الثبات وتقديره.

وتحل هذه المشكلة عن طريق استخدام نموذج راش؛ لأن قدرة الأشخاص وصعوبة المفردات يكونان على المقياس الخطي (Linear Scale) ويقدم اختبار قدرة المختبرين وصعوبة المفردات على المقياس الخطي مصداقية على التقديرات (Estimation) وصلاحية لحساب المتوسطات (Means) والتباينات (Variances) بصورة دقيقة.

علاوة على ذلك، يوفر نموذج راش التقدير المباشر لنموذج تباين الخطأ لكل من قدرة الأشخاص وصعوبة المفردات. كما توفر هذه الأخطاء القياسية طريقة تحديد كمية الدقة للقياس وصعوبة المفردات كما يمكن استخدامها لوصف فترة الثقة (Confidence Interval) التي يقع فيها الشخص أو المفردة. وتعطي هذه الأخطاء القياسية المعلومات الدقيقة مقارنة إلى المتوسطات أو التباين المحصلة من البيانات الخام حسب تقديرات الطرق التقليدية لأنها (أي الطرق التقليدية) تقلل من التأثير السلبي للدرجات المتطرفة أو غير الحساسة.

فإذا أريدَ تقدير الثبات لمجموعة (group) عن طريق نموذج راش يمكن تربيع الخطأ القياسي لكل فرد من أفراد العينة وتلخيصها لإنتاج أفضل تقدير للاتساق الداخلي؛ لأن القيمة –إذا كانت البيانات التي بها مطابقة وتم استخدام تباين معدل الأخطاء الحقيقي للعينة (Actual Average Error Variance of the sample) بدلا من تباين الخطأ لمتوسط الأشخاص (Error Average) – أصبحت حينئذ خطية.

ويتم اختبار معامل الاختلاف أو دليل التباين (Person Separation Index) (أي ثبات الأفراد) والذي حل محل ثبات الانفصال للشخص بسبب اللاخطي (Non-linear) والتأثير السقفي Ceiling effect (وهو عبارة عن مستوى معين لم يعد التباين في المتغير المستقل فيه يقدر)، عن طريق نسبة التباين المعدل (Ratio of Adjusted Variance) إلى التباين المشاهد وأنه يمثل نسبة التباين (Proportion of Variance) التي وقعت بسبب الخطأ.

نتائج نموذج راش:

اعتمد الباحثان في تحليل هذا النموذج على برنامج الإحصاء (Win steps) الذي صممه مارك ليناكار Mark Linacre 1991-2004. يوفر هذا البرنامج كغيره من البرامج مثل Quest and Conquest نوعين من جودة التوفيق، وهما (١) جودة التوفيق الداخلية Infit Meansquare (٢) جودة التوفيق الخارجيةOutfit Meansquare، ولكل واحدة منهما امتيازاتها ومجالاتها. أما جودة التوفيق الداخلية فإنها مجموع الثقل للمعلومات في البيانات المجتمعة لدى المفحوصين، وأما جودة التوفيق الخارجية فإنها عبارة عن مجموع المربعات للبواقي المعيارية standardized residuals بطريقة تقليدية. وتعد كلتا الجودتين (الداخلية والخارجية) مربع المتوسطات مقسم بدرجة الحرية Degree of Freedom، وتتراوح هذه درجة الحرية بين ٠-١ (بوند وفوكس ٢٠٠١، سلفار، سميث و غرين، ٢٠٠١). وأكد غرين وزملاؤه أن مربع المتوسط الإحصائي الأقرب من الواحد يدل على التكرار والاتكاء والتقييد للمعلومات، كما أن مربع المتوسط الإحصائي الأكبر من الواحد يدل على تفاوت غير متوقع، وعلى التناقض الداخلي للمعلومات. وتكون جودة التوفيق الداخلية والخارجية بدرجة مقار بتها من الواحد؛ لأن الدرجة ١ تدل على دقة تلائم البيانات وا إصابتها للنموذج المدروس. والجدير بالذكر أن قلة الإنحراف بين المعلومات والنموذج هو المطلوب في نموذج راش. وكلما ابتعدت جودة التوفيق الداخلية والخارجية من هذا النطاق أدت ْ العناصر الأخرى غير المقاسة دور ا حيويا في قرار أفراد العينة. علاوة على جودتي التوفيق: الداخلية والخارجية فإن نموذج راش يوفر الثبات الكلي، ومجموع الخطأ المحتمل للتحليل. ويستخدم الجودة الداخلية والجودة الخارجية للأشخاص وعملية تعيير المفردات في اختبار صدق البناء للمقياس سواء من الناحية المحتوى أو التعميم.

وتختبر جودة المفردات أو الاختبار عن طريق جودة المفردة الإحصائية (Item infit statistics)؛ لمعرفة مدى اتفاق الإجابة مع كل بند من بنود المقياس أو فقراتها مرتبطة مع كل الأسئلة أم البنود في الاختبار أو المقياس. كما يدرس كذلك مدى اتساق إجابة كل شخص من الأشخاص على كل بند أو فقرة على حدة مع إجابته على البنود أو الفقرات الأخرى في المقياس. وهل نظام ترتيب إجابة على البند أو المفردة لكل شخص متحدا لطريقة إجابته على كل مفردات المقياس. ويهدف هذا الاختبار لدراسة الاتساق الداخلي بين مفردات المقياس و السمات الكامنة لها (انظر جدول رقم ١.٣ مع رسمه البياني جدول رقد ١.٤).

ويفترض نموذج راش أن احتمال إجابة شخص ما على إحدى مفردات الاختبار إجابة صحيحة هي دلالة مطردة لقدرته وصعوبة المفردة. كما يهتم بتدريج مستويات القدرة باختبار معين على نفس ميزان تعيير المفردات (Item calibration). ويفترض نموذج راش تساوي جميع مفردات الاختبار في التمييز بين مستويات القدرة المقاسة (علام، ٢٠٠٦).

جدول (٣)

مدخل: ٧٥٢ شخصا، و٢٩ مفردة

ITEM STATISTICS: ENTRY ORDER

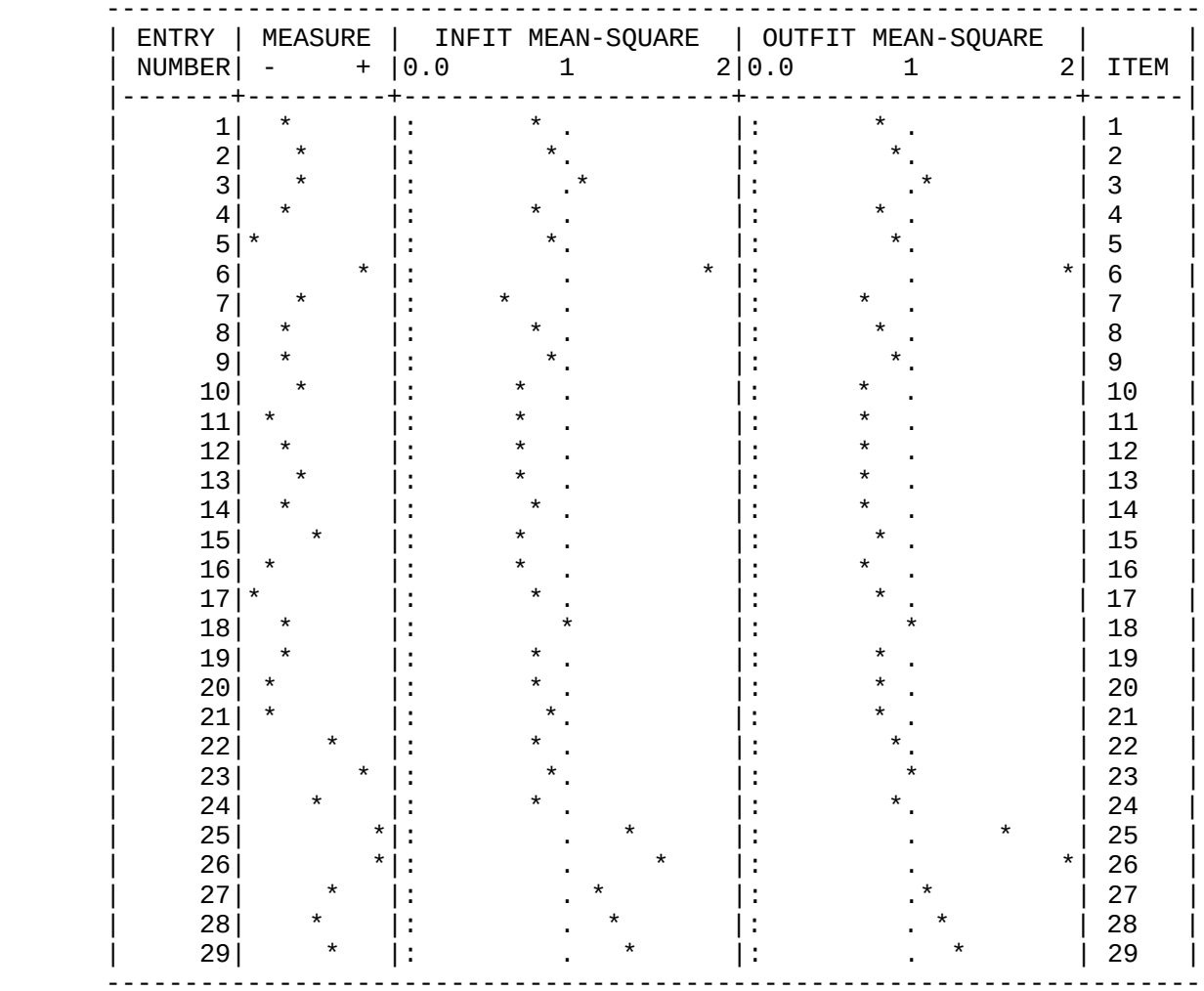
ENTRY NUMBER	TOTAL SCORE	COUNT	MEASURE	MODEL| MATCH| S.E. | EXP%	INFIT | MNSQ	INFIT | ZSTD	OUTFIT | MNSQ	OUTFIT | ZSTD	PT-MEASURE | CORR.	PT-MEASURE | EXP.	EXACT MATCH | OBS%	EXACT MATCH | EXP%	ITEM
1	4249	752	-.20	.04	.83	-3.0	.84	-2.8	.70	.66	52.2	49.0	1
2	4167	752	-.04	.04	.93	-1.1	.97	-.4	.68	.67	48.7	47.8	2
3	4164	752	-.03	.04	1.20	3.1	1.16	2.5	.63	.67	47.2	47.8	3
4	4221	752	-.14	.04	.89	-1.8	.88	-2.0	.69	.66	53.5	48.6	4
5	4383	752	-.48	.05	.91	-1.6	.92	-1.2	.67	.64	53.1	51.0	5
6	3831	752	.53	.04	1.90	9.9	2.54	9.9	.53	.71	40.3	43.4	6
7	4188	752	-.08	.04	.69	-5.7	.74	-4.7	.72	.67	56.4	48.1	7
8	4208	752	-.12	.04	.81	-3.3	.90	-1.7	.69	.66	57.0	48.5	8
9	4265	752	-.23	.05	.92	-1.3	.92	-1.3	.65	.66	53.7	49.2	9
10	4192	752	-.09	.04	.73	-4.9	.77	-4.1	.71	.67	58.1	48.1	10
11	4321	752	-.35	.05	.78	-3.9	.77	-4.0	.70	.65	57.8	50.0	11
12	4202	752	-.11	.04	.71	-5.4	.76	-4.2	.71	.66	55.9	48.3	12
13	4170	752	-.05	.04	.75	-4.5	.76	-4.2	.71	.67	55.9	47.8	13
14	4250	752	-.20	.04	.82	-3.2	.79	-3.6	.70	.66	58.3	49.0	14
15	4036	752	.20	.04	.74	-4.7	.87	-2.3	.72	.68	48.6	45.8	15
16	4301	752	-.31	.05	.77	-4.0	.74	-4.6	.70	.65	57.8	49.8	16
17	4362	752	-.44	.05	.83	-2.9	.80	-3.3	.68	.64	57.3	50.7	17
18	4217	752	-.14	.04	1.01	.2	1.01	.1	.66	.66	55.4	48.6	18
19	4245	751	-.20	.04	.82	-3.1	.82	-3.1	.70	.66	56.1	49.1	19
20	4283	752	-.27	.05	.85	-2.7	.84	-2.7	.69	.65	52.8	49.5	20
21	4338	752	-.38	.05	.98	-.3	.88	-1.9	.67	.64	55.6	50.4	21
22	3990	751	.27	.04	.88	-2.0	.93	-1.1	.71	.69	48.2	45.5	22
23	3802	752	.58	.04	.99	-.1	1.06	1.0	.71	.71	44.5	42.9	23
24	4052	752	.17	.04	.86	-2.5	.90	-1.7	.71	.68	52.5	46.2	24
25	3729	752	.68	.04	1.44	6.9	1.66	9.8	.63	.72	40.7	42.3	25
26	3685	752	.75	.04	1.70	9.9	2.10	9.9	.58	.72	36.1	41.7	26
27	3962	752	.32	.04	1.21	3.3	1.20	3.3	.69	.69	50.3	44.8	27

ENTRY NUMBER	TOTAL SCORE	COUNT	MEASURE	MODEL MATCH S.E. EXP%	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PT-MEASURE CORR.	PT-MEASURE EXP.	EXACT MATCH OBS%	ITEM
28	4072	752	.13	.04 46.5	1.34	5.2	1.22	3.5	.67	.68	49.8	28
29	4017	752	.23	.04 45.7	1.41	6.1	1.33	5.2	.66	.69	43.8	29
MEAN	4134.6	751.9	.00	.04 47.5	.99	-.6	1.04	-.3			51.6	
S.D.	184.1	.0	.33	.00 2.5	.30	4.4	.41	4.3			5.9	

جدول (٤)

الرسم البياني للمفردات

ITEM FIT GRAPH: ENTRY ORDER



من جهة آخرى، فقد اختبر الباحثان وظيفة فارقة للمفرداتDifferential Items Functioning طبقا للجنس. وقد افترض نموذج راش أن للمفردات قدرة التمييز المتساوية. لذا، فإن أية فروق في تحليل هذه

المفردات ناتجة عن الفروق الأساسية من ناحية التصورات أو التوجهات، أو الكفاءة أو الأداء، وليست ناتجة عن المفردات أو الأدوات أو الاستبانة. ولا يجوز أن تكون الفروق كامنة في المفردات أو المقياس؛ لأن ذلك يعد ضربا من ضروب الانحياز غير المقبول في التحليل الإحصائي، أو في منهج البحث العلمي السليم.

وبسبب ظروف هذه الدراسة الميدانية فإن الباحثين استخدما النطاق العريض - وهو الفرق بين أصغر وأكبر قيمة في التوزيع العشوائي وسَلطَ الضوء على أية مفردة وصل مؤشر جودتها إلى ١.٤٠ أو أكبر. يتضح من جدول رقم (2 &1) أن اختبار الثبات للأشخاص وصل إلى 96. مع التفريق 4.19، مما يدل على أعلى معدل ثبات، حيث أوصى علماء الإحصاء بأفضلية وصول ثبات الأشخاص والمفردات إلى 90. وما فوقها. ومن الجانب الثاني، فإن الثبات لمفردات (الاستبانة) وصل إلى أعلى درجة (98.) مما يدل على جودة مفردات هذه الدراسة (انظر جدول رقم١ وجدول (٢)). ويدل ثبات الأشخاص في نموذج راش إلى تقييم فاعلية التمييز بين الأشخاص المفحوصين في المتغيرات المقاسة. أما ثبات المفردات فإنه يدل على أن جميع المفردات التي تقيس العامل الداخلي الواحد تُمكن من إنتاج درجة (نقطة) الملاحظة لكل الأشخاص المفحوصين في الدراسة (بوند و فوكس، ٢٠٠١).

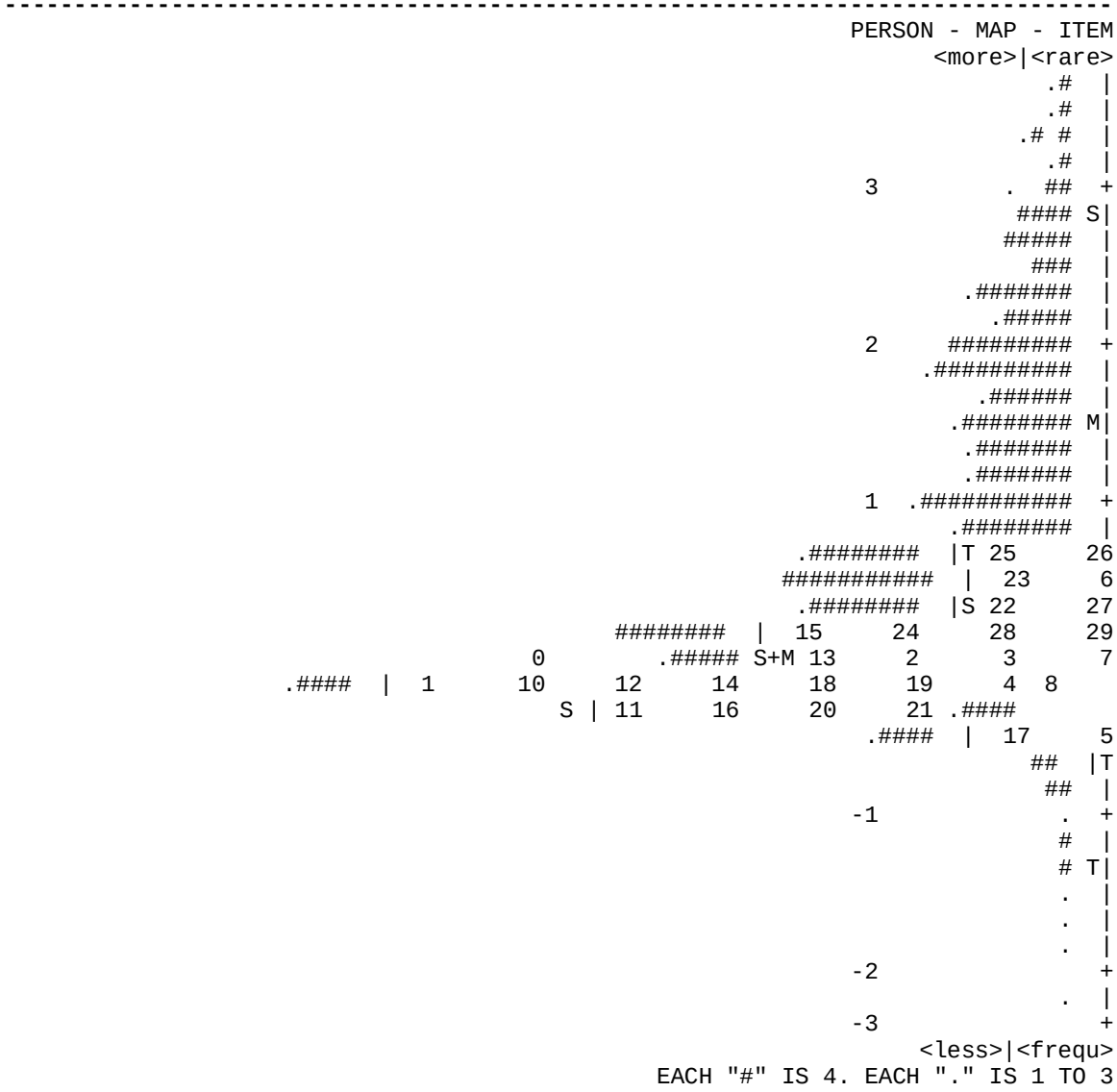
أما فيما يتعلق بجودة المقياس وصلاحيته للقياس النفسي لطلبة برنامج تدريب المعلمين فقد أظهرت نتيجة تطبيل نموذج راش أن جميع المفردات داخل النطاق المخصص (أي 67.-١.٤٠). وهذا يدل على جودة المفردات وقيمة ثباتها وصلاحيتها للقياس النفسي. وعلى الرغم من جودة مفردات هذه الدراسة إلا أن النتائج كشفت أن هناك بعض الأشخاص خارج النطاق المخصص. ويدل هذا على عدم جدية بعض المفحوصين، وقلة اهتمامهم في الإجابة عن المفردات أو الاستبانة (انظر جدول (٣) & ٤).

كما أشار الشكل الهندسي أو خريطة التغير (الشكل الأول) بصورة جلية على عدم وجود تناسب بين توزيع تقدير مفردات التقويم من الجانب (اليسار) وتدريج المفردات من الجانب الآخر (اليمين). إن صعوبة البنود أو المفردات تقيس نفس ما تقيسه قدرات الأفراد وتعبر عنه على نفس المقياس، وتعرف كل هذه الصعوبات وتلك القدرات بنفس وحدة القياس ونفس نقطة الصفر. فإن معلم قدرة الفرد يقيس ما يقيسه معلم صعوبة البند، ويعبر عنه على نفس المقياس. وتهدف خريطة المتغير إلى تحديد موضع كل من قدرات الأفراد وصعوبة البنود على هذا التدريج والتوزيع التكراري لهما. وتجدر الإشارة إلى أن كيفية تقدير الثبات من خلال راش تختلف نوعا ما من ناحية الممارسة العملية عن تقديرات الطرق الإحصائية التقليدية. ويشير هذا إلى مصداقية المفردات (الاستبانة) وثباتها، ومناسبتها واستيعابها وصلاحيتها في القياس النفسي.

وقد أثبت الخبراء أن استخدام جودة نموذج راش دليل على وفاء البيانات بالصدق البنائي, وخاصة بعد التغيير حيت تحول عملية تغيير البيانات الخام إلى المقياس الخطي الهرمي وبالتالي تُعالج مشكلة التمثيل للصدق البنائي. علاوة على ذلك، فإن نموذج راش افترض استقلالية القياس، وهذا يعني أنه لا يعتمد تقدير صعوبة البند أو المفردة على تقديرات البنود أو المفردات الأخرى المكونة للاختبار أو المقياس ولا على تقديرات قدرة الأفراد الذين يجيبون عنها.كما لا تعتمد قدرة الفرد على تقديرات أي مجموعة من الأفراد الذين يؤدون الاختبار أو يجيبون عن المقياس أو على تقديرات صعوبة البنود التي يؤدونها (أمينة محمد كاظم، ١٩٨٨) انظر جدول رقم ٥.

جدول (٥)

خريطة التغير لتقويم برنامج تدريب المعلمينمدخل: ٧٥٢ شخصا، و٢٩ مفردة,



وحسب جدول السابق (جدول رقم ٥) فليست هناك مفردات تحدد موقع الطلبة في اللوغارتيمات الإيجابية من حيث تصوراتهم أو قدراتهم تجاه المقياس أو الاختبار، لذا، لا بد من صياغة المفردات أو الفقرات الجديدة التي تغطي جوانب النقص في المقياس.

التعميم:

تعد إمكانية تعميم الاختبار أو المقياس دليلا على تمتعه بالصدق البنائي أو التكويني. ويعرف صدق البناء بأنه عبارة عن مدى قياس الاختبار أو المقياس لتكوين معين أو سمة معينة. أو بعبارة أخرى، هو الحكم بإمكانية عزو الاستنتاجات المستخلصة من درجات الشخص على المتغير الآخر المسمى التكوين أو البناء. فمثلا، الذكاء تكوين أو بناء يمكن أو يدرك من خلاله أداء الطالب في المدرسة. والتكوين أو البناء عبارة عن السمات الكامنة التي لا يمكن ملاحظتها مباشر ة ولكن يمكن دراستها واختبارها عن طريق المتغيرات الصريحة (Manifest) أو المتغيرات المشاهدة (Observed variables). بمعنى آخر يمكن الاستدلال بها من خلال السلوك الظاهر, فمختبر صدق البناء أو التكوين

للاختبار أو المقياس لابد أن يصيغ الفرضيات حول الأداء المتوقع للمتفوقين والمتدنيين على الاختبار أو المقياس المعين. وقد تؤدي هذه الفرضيات إلى النظرية المبدائية لطبيعة البناء والتكوين الذي صُمم الاختبار أو المقياس لأجله، فكلما أدى الاختبار أو المقياس حسب المتوقع (أي قادر على التمييز بين مستويات الطلبة المتنوعة) عُد هذا دليلا عمليا على ثبات النظرية وصدقها.وإذا جاءت النتيجة عكسية لمستويات الطلبة المختلفة فعلى الباحث إعادة النظر في كيفية صياغة الاختبار أو المقياس، لأن ذلك يعني أساسا أن المقياس أو الاختبار لم يمتص أو يستوعب النظرية بصورة دقيقة. وقد يرجع السبب في ذلك إلى الإجراءات الإحصائية أو الطريقة التي تم من خلالها تنفيذ الإجراءات. ويدل صدق البناء أو التكوين للاختبار أو المقياس على أنه يتمتع بسمة (أحادي البعد) (Unidimensional).

لذلك، يعد نموذج راش المتعلق بإمكانية التعميم من المتطورات المعاصرة المهمة في حقل القياس وقد أسهم في إبراز التكامل بين مفهومي الثبات والصدق.

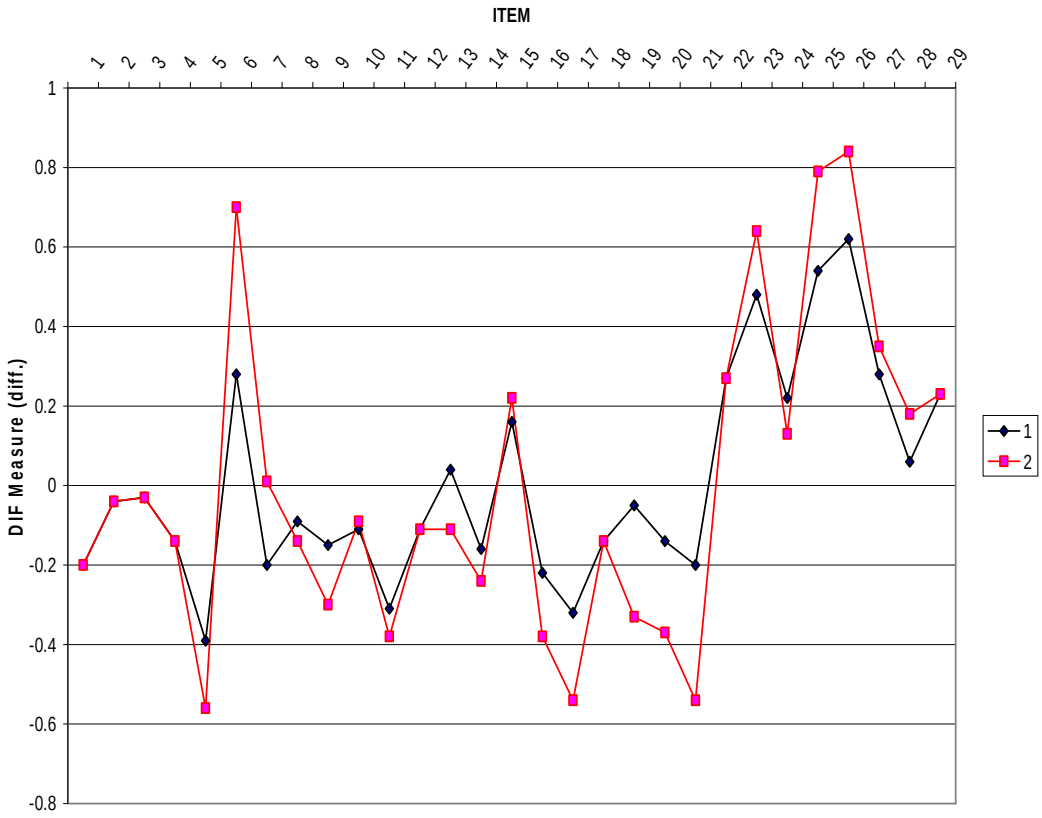
وقد اختبر الباحثان كذلك الوظيفة الفارقة للمفردات Differential Items Functioning طبقا للجنس، حيث أظهرت نتائج التحليل وجود فروق دالة إحصائيا بين الذكور والإناث في سبع مفردات. وتحدث الوظيفة الفارقة عند ما تتجاوب مجموعة مختلفة داخل المجموع الكلي بصورة مغايرة تجاه المفردة، أو المفردات المعينة على الرغم من تساوي سمات المفردات وتعادلها (انظر جدول رقم ٦وشكل رقم ١.

جدول (٦)

تحليل القلق اللغوي مدخل: ٧٥٢ شخصا، و ٢٩ مفردة

PERSON CLASSES	SUMMARY DIF CHI-SQUARE	D.F.	PROB.	ITEM Number	Name
2	.0000	1	1.0000	1	1
2	.0000	1	1.0000	2	2
2	.0000	1	1.0000	3	3
2	.0000	1	1.0000	4	4
2	3.2444	1	.0717	5	5
2	26.8481	1	.0000	6	6
2	5.2760	1	.0216	7	7
2	.3654	1	.5455	8	8
2	2.5557	1	.1099	9	9
2	.1420	1	.7063	10	10
2	.6871	1	.4071	11	11
2	.0000	1	1.0000	12	12
2	2.8723	1	.0901	13	13
2	.6690	1	.4134	14	14
2	.4328	1	.5106	15	15
2	2.8270	1	.0927	16	16
2	5.4755	1	.0193	17	17
2	.0000	1	1.0000	18	18
2	9.1994	1	.0024	19	19
2	6.2947	1	.0121	20	20
2	12.8216	1	.0003	21	21
2	.0000	1	1.0000	22	22
2	4.2700	1	.0388	23	23
2	1.0525	1	.3049	24	24
2	10.1508	1	.0014	25	25
2	8.4018	1	.0037	26	26

PERSON CLASSES	SUMMARY DIF CHI-SQUARE	D.F.	PROB.	ITEM Number	Name
2	.6405	1	.4235	27	27
2	1.9129	1	.1666	28	28
2	.0000	1	1.0000	29	29



شكل ١. الوظيفة الفارقة للمفردات

الخاتـــمة:

تبنى الباحثان في مناقشة ثبات المقياس وصدقه نموذج راش الإحصائي. وهو النموذج البارامتر الواحد. حيث تم تناول النموذج ومقارنته مع بعض الطرق الإحصائية التقليدية الأخرى وبيان الفرق بينمها من ناحية القدرات والإمكانيات. ويتميز نموذج راش في أنه يختبر قوة المفحوصين وصعوبة المفردات في آن واحد الأمر الذي قلما تجده في الطرق الإحصائية التقليدية. ولقد أثبت الباحثان من خلال المناقشة وجود إمكانية هائلة لهذه الطريقة الإحصائية في دراسة ثبات المقاييس وصدقها، من حيث إنها لا تتأثر بطول المقياس وعدد فقراته، بل بالعكس فإن المقياس القصير من حيث عدد فقراته قد يتسم بالثبات العالي مقارنة بالمقياس الطويل. ومن جهة أخرى، ناقش الباحثان بصورة مستفيضة كيفية تأثر معامل ألفا بطول المقياس. وهي نقطة غابت عن كثير من الباحثين أو بالأحرى عدم درايتهم بها، حيث نجد أن جل الدراسات والأبحاث تستخدم معامل ألفا لاختبار الثبات غير واضعين في الاعتبار حجم مفردات المقياس أو الاختبار.

علاو ة على ذلك، فقد أثبت من خلال التحليل إمكانية استخدام نموذج راش في التأكد من صدق المقياس سواء عن طريق التعميم أو عدم التغير (Invariance)، أو الصدق البنائي. ولقد ذ َوه الباحثان إلى عدم الغرور بنتيجة كرونباخ ألفا، وليس في هذا القول تقليل لقيمة ألفا، بل فيه إشارة إلى تأثره بالعوامل الخارجية غير المقاسة. ولذلك لا بد من

استخدام الطرق الأخرى خاصة نموذج راش للتأكد من صلاحية المقياس المتبنى لتفادى تلك العوامل الخارجية. كما ينوه الباحثان على عدم الاعتماد على الصدق الظاهري فقط، أو ما يسمى بصدق المحكمين فهذا النوع من الصدق يعتمد على الشكل الظاهري للمقياس وأن الفقرات ظاهريا تقيس ما أعدت لقياسه لكن جوهريا لا نعلم ما إذا كانت تلك الفقرات جميعها فقرات جيدة، ولها نسبة مساهمة عالية مع الدرجة الكلية للمقياس أم أنها فقط عبارة عن فقرات ظاهرية جوفاء تستخدم لتقدير صدق المقياس المستخدم، في حين أن هناك أنواع صدق متنوعة لابد من استخدامها للتحقق من أن المقياس الذي يستخدم في الدراسة يقيس فعلا ما أُعد لقياسه.

ولم ينو الباحثان في هذا التحليل دراسة نموذج راش حيث لهها دراسات خاصة في هذا المجال وإنما يريدان إثبات أن استخدام نموذج راش في اختبار جودة المقياس أكثر مصداقية وأدق علما من معامل ألفا لأسباب سبق أن ذكرناها.

المراجع العربية:

كاظم، أمينة محمد (١٩٨٨). **استخدام نموذج راش في بناء اختبار تحصيلي في علم النفس وتحقيق التفسير الموضوعي للنتائج**. الكويت: مطبوعات جامعة الكويت.

علام، صلاح الدين محمود (٢٠٠٦). **القياس والتقويم التربوي والنفسي**. دار الفكر العربي، القاهرة.

علام، صلاح الدين محمود (٢٠٠٧). **نماذج الاستجابة للمفردة الاختبارية**. دار الفكر العربي، القاهرة.

وطفة، علي (د.ت) التفاعل **التربوي بين الطلاب وأعضاء الهيئة التدريسية في الجامعة، موازنة بين آراء طلاب جامعتي الكويت ودمشق**، مجلة اتحاد الجامعات العربية،٩٥، ١-١٥٢

المراجع الإنكليزية:

Bond, G & Fox, M. (200١). *Applying the Rasch model; fundamental measurement in human sciences*. London: Lawrence Erlbaum associates publishers.

Bond, G & Fox, M. (200٧). *Applying the Rasch model; fundamental measurement in human sciences*. London: Lawrence Erlbaum associates publishers.

Cohen, R. J. & Swerdlik, M. E. (2005). *Psychological testing and assesement*. McGraw-Hill. London.

DeVellis, R. J (1991). *Scale development: theory and application. Applied social research methods series.*

Ghiselli, E. E. Campbell, J. P. & Zedeck, S. (1981*). Measurement theory for the behavioral science*. San Francisco: Freeman.

Hair, F., Anderson, E., Tatham, L., & Black, C. (1998). *Multivariate data Analysis*. New Jersey: Prentice-Hall International, INC.

Harwell, R. & Gatti, G. (2001). Rescaling ordinal to interval data in educational research. *Review of Educational Research, 72, 105-131*.

Harwell, R. & Gatti, G. (2001). *Rescaling ordinal to interval data in educational research*.

Review of Educational Research, 72, 105-131. Institute for Objective Measurement. Measurement for social science and education: a history of social sciencemeasurement :http://www.rasch.org/memo62.htm

In Smith, E. V & Smith, R. M. *introduction to Rasch measurement*. Jam press, Minnesota.

In Smith, E. V & Smith, R. M. *Rasch Measurement: advanced and specialized applications*. Jam press, Minnesota.

Kaplan, R. M. & Saccuzzo, D. P. (2005*). Psychological testing*. Thomson, Canada.

Linacre, J. M (2004). *A user's guide to Winsteps Ministep Rasch model computer programs, Chicago.*

Linacre, J. M. (1996). True-score reliability or Rasch statistical validity*? Rasch measurementtransactions, 9, 455-696.*

Linacre, J. M. (1999). Relating Cronbach and Rasch reliabilities. *Rasch measuremen transactions, 13, 696 – 713.*

Schumaker, R. E. (2003). *Reliability in Rasch measurement: avoiding the rubber ruler*. Paper presented at annual meeting of the Amedican educational research association, Illinois, Chicago

Sliver, B., Smith, V. & Greene, A. (2001). A study strategies self-efficacy instrument for uswith community college students. *Educational and Psychological Measuremnt 61(5), 849-865.*

Smith Jr. E. V. (2004). Evidence *of the reliability of measures and validity of measure interpretation: A Rasch measurement perspective.*

Wolfe, E. W. & Smith, E. V. (2007). *Instrument development tools and activities*

for measure validation using Rash models: part II- Validation activities.

Wright, B.D. & Linacre, J. M. (1994). *Reasonable mean-square fit value*. Rasch**.**

Wright, B.D. & Linacre, J. M. (1994). *Reasonable mean-square fit value. Rasch measurement transaction,* 8(3), 370 -384.