



جامعة وهران 2
كلية العلوم الاجتماعية
أطروحة

للحصول على شهادة دكتوراه علوم في علم النفس وعلوم التربية
تخصص علم التربوي

الخصائص السيكومترية لاختبار تحصيلي في الرياضيات
وفق النظرية الحديثة في القياس التربوي والنفسي،
دراسة ميدانية على مستوى السنة الخامسة ابتدائي.

مقدمة ومناقشة علنا من طرف

الطالب : بغداد براهيم محمد

أمام لجنة المناقشة

اللقب والاسم	الرتبة	المؤسسة الأصلية	الصفة
أ. ياسين أمينة	أستاذة	جامعة وهران 2	رئيسا
أ. تليوين حبيب	أستاذ	جامعة وهران 2	مشرفا ومقررا
أ. مرياح فطيمة زهرة	أستاذة	جامعة وهران 2	مناقشا
أ. قيدوم أحمد	أستاذ	جامعة مستغانم	مناقشا
أ. بشلاغم محي	أستاذ	جامعة تلمسان	مناقشا
أ. بلعالية محمد	أستاذ	جامعة ابن خلدون تيارت	مناقشا

السنة: 2026/2027

بسم الله الرحمن الرحيم

كلمة شكر وتقدير

الحمد لله حمدا كثيرا يليق بجلال وجهه وعظيم سلطانه على ما وهبنا من صبر،
وهدى، وتوفيق لإنجاز هذا العمل. أتقدم بالشكر الجزيل إلى الأستاذ الدكتور
المشرف: أ.د. تيليوين حبيب ذلك نبع العطاء والرفقة والعطف، صاحب
الأستاذية الفذة وصاحب التواضع الذي أشرف على هذه الرسالة والذي قدم
لي النصح وأمدني بملاحظاته وصبره، فجزاه الله عن العلم وأهله خيرا.
وأيضا نتقدم بالشكر إلى جميع الأساتذة الذين تعاملوا معنا في تطبيق أدوات
الدراسة.

وأخيرا نتقدم بالشكر والعرفان إلى كل أعضاء لجنة المناقشة أقطاب المعرفة لقبولهم
مناقشة هذا العمل المتواضع.

الخصائص السيكومترية لاختبار تحصيلي في الرياضيات وفق النظرية الحديثة في القياس التربوي والنفسي: دراسة ميدانية على

" مستوى السنة الخامسة ابتدائي

: الملخص

هدفت هذه الدراسة إلى استخدام نظرية الاستجابة للفقرة في بناء اختبار محكي المرجع في مادة الرياضيات وفق نموذج "راش". ولتحقيق ذلك، تم بناء اختبار تحصيلي مكون من (55) فقرة ثنائية التصحيح، صُممت لقياس الفهم المفاهيمي، والمعرفة الإجرائية، وحل المشكلات لدى تلاميذ الصف الخامس الابتدائي. طبقت الأداة على عينة بلغت (559) تلميذاً، وحُللت البيانات باستخدام برنامج WINSTEPS (نسخة 3.72.3) للتحقق من صدق البناء، وأداء الفقرات، ودقة القياس.

وأظهرت النتائج أن معظم الفقرات تمتعت بمؤشرات ملاءمة جيدة (MNSQ) بين 0.5 و 1.5)، مما يؤكد استقرار هرمية الصعوبة وثبات الأداة. وبالرغم من أن تحليل المكونات الأساسية للبواقي كشف عن وجود بُعد ثانوي (بقيمة كامنة قدرها 5.0، تفسر 9% من التباين)، إلا أن بُعد "راش" الأساسي ظل هو المهيمن، مما يدعم أحادية بُعد المقياس. ولمعالجة التبعية المحلية، تم اعتماد نموذج "راش" متعدد الأبعاد الاستكشافي، حيث صُنِفَت الفقرات كأبعاد كامنة مستقلة مع الحفاظ على بُعد القدرة العامة. وبشكل عام، قدم التحليل أدلة قوية على الصدق الداخلي للأداة، مع إرجاع عدم الملاءمة البسيط في بعض الفقرات إلى "أخطاء عشوائية" في الاستجابات وليس إلى تعددية في البنية الجوهرية. وبناءً عليه، تُعد الأداة صالحة للاستخدام والتفسير في السياقات التربوية.

الكلمات المفتاحية: اختبار تحصيلي، نظرية الاستجابة للفقرة، نموذج راش، الرياضيات، الخصائص السيكومترية، الثبات، الصدق،

مطابقة الفقرات

« Caractéristiques psychométriques d'un test d'enseignement en mathématiques selon la théorie moderne de la mesure éducative et psychologique : une étude de terrain au niveau de la cinquième année de primaire » »

Résumé :

Cette étude a utilisé la Théorie de la Réponse aux Items (TRI) pour élaborer un test de mathématiques à référence critériée destiné aux élèves de CM2 (cinquième année primaire), en s'appuyant sur le modèle de Rasch. Un ensemble de 55 items dichotomiques, mesurant les compétences conceptuelles, procédurales et de résolution de problèmes, a été administré à 559 élèves. Les données ont été analysées à l'aide du logiciel WINSTEPS (v.3.72.3) afin d'évaluer la validité du concept, le fonctionnement des items et la précision de la mesure.

Les résultats indiquent que la plupart des items présentent un ajustement productif (MNSQ compris entre 0,5 et 1,5), confirmant ainsi une hiérarchie stable des items et une fidélité élevée. Bien que l'analyse en composantes principales (ACP) des résidus ait identifié une dimension secondaire (valeur propre de 5,0, expliquant 9 % de la variance), la dimension principale de Rasch est demeurée dominante, soutenant l'unidimensionnalité fondamentale de l'échelle. Pour tenir compte de la dépendance locale, une approche multidimensionnelle exploratoire a été employée, modélisant les grappes de résidus comme des dimensions latentes distinctes tout en préservant la structure de l'aptitude générale.

En conséquence, l'instrument présente de solides propriétés psychométriques, les légers défauts d'ajustement étant attribués à un "bruit" dans les réponses plutôt qu'à des failles structurelles. Le test est donc jugé valide et fiable pour une évaluation éducative approfondie.

Mots clés : Test , théorie de la réponse à l'item, Rasch model, mathématiques, propriétés psychométriques, fiabilité, validité, adéquation des items

« "Psychometric Characteristics of an Achievement Test in Mathematics According to the Modern Theory of Educational and Psychological Measurement: A Field Study at the Level of the Fifth Year of Primary School"»

Abstract :

This study utilized Item Response Theory (IRT) to develop a criterion-referenced mathematics test for fifth-grade students, anchored in the Rasch model. A pool of 55 dichotomous items, measuring conceptual, procedural, and problem-solving skills, was administered to 559 students. Data were analyzed using WINSTEPS (v.3.1.3) to assess construct validity, item functioning, and measurement precision. Results indicated that most items demonstrated productive fit (MNSQ 0.5–1.5), confirming stable item hierarchy and high reliability. While Principal Component Analysis (PCA) of residuals identified a secondary dimension (eigenvalue of 5.0, explaining 9% of the variance), the primary Rasch dimension remained dominant, supporting the scale's fundamental unidimensionality. To account for local dependence, an exploratory multidimensional approach was employed, modeling residual clusters as distinct latent dimensions while preserving the general ability structure. Ultimately, the instrument exhibited strong psychometric properties, with minor misfits attributed to response noise rather than structural flaws. The test is thus deemed a valid and reliable tool for substantive educational assessment.

Keywords

Key words : Achievement test, Item Response Theory, Rasch model, mathematics, psychometric properties, reliability, validity, Item fit

''

الصفحة	فهرس المحتويات
أ	الإهداء
ب	كلمة شكر وتقدير
ج	ملخص البحث باللغة العربية
	ملخص البحث باللغة الانجليزية
	ملخص البحث باللغة الفرنسية
د	قائمة المحتويات
هـ	قائمة الجداول
و	قائمة الأشكال
ز	قائمة الملاحق
1	مقدمة
3	الفصل الأول: موضوع البحث
4	خلفية الدراسة: - أهمية اختبار التحصيل في الرياضيات.
5	مصادر اشكالية الدراسة
7	- نظرة عامة على الخصائص السيكو مترية.
8	- الفجوات في الأبحاث الحالية حول اختبارات التحصيل في الرياضيات.
9	. التوجهات الحديثة في تصميم اختبارات التحصيل
7	إشكالية البحث
8	فرضيات البحث
10	التحديد والضبط الاجرائي للمفاهيم الأساسية للبحث
11	حدود الدراسة
13	أهداف الدراسة
12	أهمية الدراسة
13	مصطلحات الدراسة
	الفصل الثاني: الإطار النظري للدراسة

16	أولاً: اختبارات التحصيل في الرياضيات
19	تمهيد
19	- التطور التاريخي
20	. الاتجاهات الحالية.
23	ثانياً: نظرية الاختبار الكلاسيكية (CTT)
23	تمهيد
25	افتراضات النظرية الكلاسيكية (CTT)
25	- نماذج نظرية الاختبار الكلاسيكية. CTT
28	مزايا نظرية الاختبار الكلاسيكية
29	الخصائص السيكمومترية
	ثالثاً: نظرية الاستجابة للفقرة (IRT)
30	تمهيد
32	نشأة نظرية الاستجابة للفقرة ومتطلبات وشروط القياس الموضوعي
36	مفهوم نظرية الاستجابة للفقرة
35	مسلمات نظرية الاستجابة للفقرة:
37	افتراضات نظرية الاستجابة للفقرة
38	• أحادية البعد
38	• الاستقلال المحلي
39	• عامل السرعة
39	• تساوي مؤشرات التمييز (خاص بنموذج راش)
39	• الحد الأدنى من أثر التخمين
42	نماذج نظرية السمات الكامنة
43	- أنواع نماذج نظرية الاستجابة للفقرة
47	معايير اختيار النموذج المناسب للبيانات:

47	● التحقق من افتراضات النموذج
50	● التحقق من الخصائص المتوقعة للنماذج
55	مميزات نماذج نظرية الاستجابة للفقرة:
56	عيوب نظرية الاستجابة للمفردة
57	مفهوم نموذج راش
56	الأساس النظري لنموذج راش
60	الخصائص السيكو مترية لنموذج راش
60	● الإطار النظري للثبات في نماذج نظرية الاستجابة للمفردة
63	● المقارنة والاختلاف بين النظرية الكلاسيكية للاختبار ونظرية القياس لراش
63	دوافع ظهور نظرية القياس لراش
65	CTT و RMT خصائص المقارنة بين
66	تقدير الدرجة الحقيقية في النظرية الكلاسيكية
66	تقدير القدرة في نموذج راش
67	الخطأ المعياري في النظرية الكلاسيكية
67	الخطأ المعياري في نموذج راش
70	تقدير القدرة في نموذج راش
74	رابعاً: الدراسات السابقة
77	مزايا نظرية الاختبار الكلاسيكية
73	دراسات عربية
75	الدراسات الأجنبية
77	التعليق على الدراسات السابقة
	ملخص الفصل
82	الفصل الثالث: الدراسة الميدانية
82	تمهيد
88	تصميم اختبار التحصيل في الرياضيات

90	برنامج السنة الخامسة ابتدائي
91	. أداة الدراسة: خطوات بناء الاختبار التحصيلي
93	. الدراسة الاستطلاعية 1 و 2
94	تحليل نتائج الدراسة الاستطلاعية
95	. الدراسة الاساسية
96	وصف مجتمع الدراسة وطريقة أخذ العينات.
	الفصل الرابع : نتائج الدراسة ومناقشة النتائج
99	تمهيد
99	تحليل البيانات باستخدام نظرية الاختبار الكلاسيكية (CTT)
104	خصائص فقرات الاختبار: الصعوبة والتمييز
107	تقدير معامل الثبات في النظرية الكلاسيكية
108	خاصية الصدق:
111	عرض النتائج وتفسيرها في ضوء نموذج راش ونظرية الاستجابة للمفردة
113	تقدير معلمات الفقرات وفق نموذج راش
116	مصفوفة اتخاذ القرار اتجاه الفقرات (وفق نموذج راش)
117	مناقشة معاملات الصعوبة وفق النظريتين
120	معامل الثبات وفق نموذج راش
139	تفسير دالة المعلومات
125	تقدير القدرة
131	تحليل حدود النتائج
132	تحليل إحصاءات المطابقة
133	تحليل خصائص منحنيات الفقرات ICC
137	تحليل دالة معلومات الفقرات
146	خريطة توزيع الافراد والفقرات

150	تحليل دالة معلومات الاختبار
154	الخطأ المعياري للقياس
156	الدلالات السيكومترية للنتائج
157	بناء نموذج متعدد الأبعاد:
171	الصدق: صدق المفهوم ، الصدق الترتيبي
181	مقارنة موسعة بين CTT و Rasch و IRT
182	تحليل ثبات واستقرار تقديرات التلاميذ
183	استقرار ترتيب التلاميذ معامل سبيرمان
187	حساب استقرار الإحصائيات وفق النظرية الكلاسيكية:
189	استقرار إحصاءات وفق نموذج راش
200	مناقشة الفرضيات
202	الاختلافات والتشابهات في النتائج بين نظرية الكلاسيكية ونظرية الاستجابة للمفردة
203	الفصل الخامس: دلالة النتائج في سياق خصوصيات الوسط التعليمي التربوي النحلي
207	التوصيات والاقتراحات
220	قائمة المراجع
229	الملاحق

قائمة الجداول

الجدول	عناوين الجداول	الصفحة
1	جدول المقارنة بين النظرية الكلاسيكية ونظرية راش	66
2	مقارنة نظرية بين CTT و Rasch و IRT	67
3	مقارنة المعادلات الرياضية	67
4	مثال تحويل Raw Score إلى Logit	79
5	الكفاءات في نهاية السنة الخامسة	94
6	مواصفات الاختبار التحصيلي ¹	95
7	مواصفات الاختبار التحصيلي ²	95
8	اتفاق المحكمين على فقرات الاختبار	96
9	المؤسسات التربوية حسب الولايات	98
12	احصائيات الدرجة الكلية النظرية الكلاسيكية	101
13	الاحصاء الوصفي	103
14	معامل الصعوبة لأسهل الفقرات	103
15	معامل الصعوبة لأصعب الفقرات	103
16	معامل الثبات وفق النظرية الكلاسيكية للاختبار	106
17	التجزئة النصفية	106
18	معاملات الثبات وفق نظرية التعميم	107
19	احصائي المجموعتين	110
20	ت لعينتين مستقلتين	110
21	أسهل الفقرات في الاختبار	113
22	أصعب الفقرات	114
23	تحويل الدرجة الخام الى اللوجيت	114
24	معايير اتخاذ القرار لحذف الفقرات	117

118	ترتيب الفقرات	25
118	مقارنة معامل الصعوبة وفق النظريتين	26
119	معامل الثبات من خلال نموذج راش	27
120	تقدير الثبات الهامشي للاختبار	28
121	مقارنة بين الثبات الهامشي والشرطي	29
121	الثبات الشرطي لمستويات القدرة	30
130	الفرق بين نوعين من "قلة المعلومات	31
132	تصنيف الفقرات وفق جودة المطابقة للنموذج	32
133	تقديرات المطابقة الخارجية للفقرات الأكثر عدم المطابقة وفق نموذج راش	33
133	إحصائيات ملائمة الفقرات الأفضل في المطابقة	34
157	مقارنة مع نظرية الاختبار الكلاسيكية	35
158	التحليل العاملي للمكونات الرئيسية للبراهين	36
160	الخلاصة السيكومترية	37
161	تحليل المكونات الرئيسية للبراهين المعيارية	38
164	التحليل العاملي للمكونات الأساسية للبراهين المعيارية	
165	مواصفات نموذج متعدد الأبعاد	39
167	البنية النهائية لنموذج متعدد الأبعاد	40
170	المقارنة بين نموذج راش أحادي البعد والنموذج متعدد الأبعاد	41
168	اختيار الأبعاد	42
171	جدول صدق الترتيب	43
180	تقييم اجابات التلاميذ سوء التنبؤ	44
181	مقارنة موسعة بين CTT و Rasch و IRT	45
183	مقارنة المعادلات الرياضية	46
184	مدى دقة الاختبار	47
187	جدول مقارنة القدرات	48

189	مقارنة القدرات الشامل	49
189	تقسيم العينة	50
189	معامل ارتباط سيرمان	51
190	احصاء وصفي للفوجين	52
190	اختبار ت للمقارنة	53
194	مقارنة معامل الصعوبة ومعامل التمييز حسب كل فقرة بين المجموعتين	54
194	معامل الارتباط الصعوبة بين المجموعتين	55
194	الارتباط بين معامل التمييز مج 1 ومعامل التمييز مج 2	56
201	مقارنة الفقرات مع المنهاج المقرر	57
202	تحليل قدرات التلاميذ	58

قائمة الاشكال

الصفحة	التمثيل البياني	
43	المنحنى المميز للفقرة	1
44	المنحنى المميز لفقرة ثنائية الدرجة	2
51	التمثيل البياني للنموذج اللوغاريتمي أحادي المعلم	3
52	التمثيل البياني للنموذج اللوغاريتمي ثنائي المعلم	4
53	التمثيل البياني للنموذج اللوغاريتمي ثلاثي المعلم	5
65	احتمال الاجابة الصحيحة	6
103	التمثيل البياني: التوزيع الطبيعي للبيانات	7
108	خصائص الفقرات 7 و 8	8
124	دالة المعلومات للاختبار	11
124	الثبات الشرطي	12
125	مخطط ملائمة الافراد 4.1 WINSTEPS	13
134	منحنيات خصائص فقرات الاختبار	14
135	المجموعة 1: من الفقرة 1 الى الفقرة 11	15
135	المجموعة 2: من الفقرة 12 الى الفقرة 23	16
136	المجموعة 3: من الفقرة 24 الى الفقرة 35	17
136	المجموعة 4: من الفقرة 36 الى الفقرة 47	18
137	المجموعة 5: من الفقرة 48 الى الفقرة 55	19
139	دالة معلومات الفقرات	20
141	دالة معلومات الفقرات مع 1 و 2	21
142	دالة معلومات الفقرات مع 3	22
143	دالة معلومات الفقرات مع 4	23
145	دالة معلومات الفقرات مع 5	24

148	خريطة الافراد والفقرات	25
150	توزيع الفقرات والافراد	26
153	معلومات الاختبار وكثافة القدرة	27
193	العلاقة بين المجموعتين في درجات التلاميذ	28
193	العلاقة بين المجموعتين في فقرات الاختبار	29
193	المقارنة في معامل الصعوبة	30
194	معامل التمييز بين المجموعتين	31
194	لتوزيع بيانات الفوجين	32
195	للمقارنة في معامل الصعوبة للمجموعتين	33
196	للمقارنة لمعامل التمييز	35

قائمة الملاحق

الرقم	الملاحق	الصفحة
01	قائمة المحكمين	228
02	صدق الاتساق الداخلي	
03	معامل الصعوبة وفق النظرية الكلاسيكية	
04	مخرجات برنامج WINSTEPS 3.72.3 معامل الصعوبة	
05	خصائص الفقرات	
06	خريطة ورايت من برنامج JMetrik	
07	الاختبار التجريبي في الرياضيات	
08	معامل الثبات الشرطي	
09	عدم ملائمة التلاميذ	
10	خريطة ورايت وفق برنامج Winsteps	

مقدمة الدراسة:

الحمد لله رب العالمين والصلاة والسلام على أشرف المرسلين سيدنا محمد وعلى آله وصحبه أجمعين ومن تبع هداهم بإحسان إلى يوم الدين، وبعد:

تحتل قضية تطوير الاختبارات النفسية والتربوية مكانة مميزة في مجال علم النفس عامة، والقياس النفسي خاصة، ويرجع ذلك إلى كونها من أهم الوسائل المستخدمة في مجال التقويم، الأمر الذي يتطلب إعداد وبناء اختبارات نفسية موثوقة في كفاءتها تستخدم في مجالات مختلفة، وإذا ما انتقلنا إلى مجال قياس التحصيل الدراسي، سنجد أنّ هذا المجال يمثل حجر الزاوية في مجال القياس النفسي.

ونتيجة لتطور القياس النفسي والذي ينعكس بدوره على تطور الاختبارات والمقاييس كان الهدف الذي يسعى إليه علماء القياس النفسي هو تحقيق موضوعية القياس، لذا شهد هذا الميدان تطورات متزايدة تتعلق بأساليب تصميم وبناء وتحليل فقرات الاختبار لتحقيق هذا الهدف. (أبو حطب، 1992: 321)

يعالج التقييم التربوي قضايا مثل ثبات وصدق الاختبارات والامتحانات، وربط ومعادلة الاختبارات، وتقييم تحيز الاختبار وأداء الفقرات التفاضلية (التمييزي) يتناول التقييم التربوي مجموعة واسعة من القضايا من المستوى الجزئي (التدريس، والوسائل، المناهج) إلى المستوى الكلي (أبحاث فعالية المدارس ودراسات التحصيل واسعة النطاق، مثل TIMSS و PISA).

يبدو أن النظرية الإحصائية للتقييم والتقويم التربوي لها مقاربتان: مقارنة نظرية الاختبار الكلاسيكية (CTT) و نظرية الاستجابة للفقرة (IRT). على الرغم من أن جذور نظرية الاختبار الكلاسيكية تعود إلى بيرسون (1904، 1907)، وعمل Gulliksen الرائد "نظرية الاختبارات العقلية" (Gulliksen، 1950) يمكن اعتباره أول تأسيس شامل لمقاربة نظرية الاختبار الكلاسيكية. أصبح امتداد مهم عن هذه النظرية يسمح بمزيد من التفصيل لمصادر عدم الثبات في درجات الاختبارات يعرف بنظرية التعميم (Cardinet، 1997; Cardinet, Tourneur & Allal, 1976, Cronbach, Glaser, Nanda & Rajaratnam, 1972). وهناك مقارنة

آخري مرتبطة ارتباطا بنظرية الاختبار الكلاسيكية وهي النمذجة متعددة المستويات (Goldstein, 1986) والتي تستخدم كثيرا في أبحاث فعالية المدارس .

تم نشر أولى لصيغ نظرية الاستجابة للمفردة خلال وبعد الحرب العالمية الثانية (Lawley, 1943,)
(Rasch 1960)،
(Lord, 1952, 1953) لكن أكثر المساهمات تأثيرا قدمها راش (Lord 1980). وأخيرا، ما
بيرنباوم (Birnbaum 1968)، بوك (Bock 1972) ولورد (Lord 1980). وأخيرا، ما
بين CCT و IRT تطورات في نمذجة المتغيرات الكامنة مثل التحليل العاملي ونمذجة المعادلات
الخطية (انظر، على سبيل المثال، Jöreskog، 1970). غالبا ما تبدو هذه المقاربات المختلفة غير
مرتبطة. من ناحية أخرى، أشار عدة مؤلفين إلى وجود هذه الارتباطات. على سبيل المثال، يظهر
(Takane and de Leeuw 1987) أن التحليل العاملي وبعض نماذج نظرية الاستجابة للمفردة
متعددة المستويات IRT (multilevel IRT) متكافئة تماما (A.W. Glas , 2008).ⁱ
ويعز القياس النفسي والتربوي بين مدخلين رئيسيين في تصميم وبناء الاختبارات والمقاييس وتحليل
البيانات المستمدة منها، هما:

- المدخل التقليدي المتمثل في النظرية الكلاسيكية للاختبارات Classical Test Theory (CTT) وما ينطوي عليه من مفاهيم ومبادئ بعضها يتعلق بخصائص فقرات الاختبار (الصعوبة، التمييز)، والبعض الآخر بخصائص الاختبار ككل (الصدق، الثبات والمعايير) (علام 2005: 48).
- المدخل المعاصر المتمثل في نظرية السمات الكامنة Latent Trait Theory (LTT)، أو نظرية الاستجابة للمفردة (أو الفقرة) Item response theory (IRT)، كما يسميها بعض علماء القياس أمثال هامبلتون Hambleton ولورد Lord، لاهتمامهما بالربط بين استجابة الفرد لفقرة اختبارية وبين خصائص هذه الفقرة. (علام 2001: 204)

لقد كانت النظرية الكلاسيكية للاختبار لسنوات عديدة مضت، هي الأساس النظري والعملي للقياس في العلوم السلوكية، والتي استندت على نموذج الدرجة الحقيقية للفرد، والذي ينص على أن لكل فرد قدر ما من السلوك غير الملاحظ (الدرجة الحقيقية) والذي لا يمكن أن يُقاس بصورة مباشرة، وعليه

فإن عدد الاجابات الصحيحة الملاحظة هي التي تقيس نفس قدرة الفرد على أساس أن هذا القياس يتضمن قدرا من الخطأ أثناء تقدير الدرجة الحقيقية" (عبد الحافظ: 2:1999).

يمكن أن يسبب التقييم التقليدي (كما هو الحال في النظرية الكلاسيكية) العديد من المشاكل في التعليم، مثل التبعية المتبادلة:

- (1) تعتمد إحصائية المفحوص (الدرجة الملاحظة) على عينة الفقرات؛
- (2) تعتمد إحصاءات الفقرات (معاملي الصعوبة والتمييز) على عينة المفحوصين (Fan, 1998, pp. 357-381) ;

- (3) لا يتم تصور البنود بالترتيب من السهل إلى الصعب؛
- (4) لم يُختبر الترتيب النظري لمعامل صعوبة الفقرات باستخدام الدرجات "الحقيقية" لإنشاء مقياس خطي؛

- (5) لم تُعابر صعوبات الفقرات (من السهل إلى الصعب) ومقاييس الأفراد (من الأدنى إلى الأعلى) على نفس مقياس ذو مستوى فقوي.

لذا جاءت نظرية الاستجابة للمفردة للتغلب على هذه النقائص، من خلال نماذجها المختلفة. ويعتبر نموذج راش أشهرها في بناء الاختبارات.

يتمتع القياس من خلال نموذج راش (راش، 1980/1960) في بناء الاختبارات النفسية والتربوية بالقدرة على التغلب على بعض هذه المشكلات. من الممكن إنتاج فقرات عالية الجودة لا تضمن دقة أكبر في تقييم التحصيل الدراسي فحسب، بل توفر أيضًا طريقة بديلة لتحسين النظام التعليمي ككل. يمكن أن يؤدي بنك الفقرات إلى جانب قياس من خلال نموذج راش إلى تحسينات في التعلم المدرسي وفي تقارير التحصيل المدرسية

لذا من خلال هذه الدراسة نحاول تقديم هذه الاسهامات التي قدمتها نظرية الاستجابة للمفردة من خلال مقارنة الخصائص السيكومترية بين النظرية الكلاسيكية ونظرية الاستجابة للمفردة. تم تقسيم هذه الرسالة الى ستة فصول، حيث تم تقديم الموضوع واشكاله البحث في الفصل الأول، وفي الفصل

الثاني خصص للجانب النظري والذي قسم الى اربعة اقسام يضم القسم الاول خصص لمادة الرياضيات ومختلف المراحل التي مر بها تقييم تحصيل مادة الرياضيات. اما القسم الثاني خصص للنظرية الكلاسيكية واهم مبادئها وحدودها. والقسم الثالث تناول نظرية الاستجابة للمفردة واهم نماذجها وخصائصها وحدودها. والقسم الأخير خصص للدراسات السابقة والتي اثبتت الاختلاف بين النظريتين وبين التي توافق بينهما. الفصل الثالث تناول الدراسة الميدانية انطلاقا من بناء الاختبار التحصيلي في الرياضيات وفق الأطر النظرية لبناء الاختبارات التحصيلية، وتطبيق الاختبار على عينة الدراسة الاستطلاعية ثم التعديلات التي أوصى العمل بها من خلال نتائج الدراسة. ثم الدراسة الأساسية والفصل الرابع خصص لتقديم النتائج ومناقشتها والفصل الخامس خصص لعلاقة النتائج المتوصل عليها وربطها بالمجال التربوي المحلي ومقارنتها بالدراسات السابقة مع تقديم الاقتراحات للدراسات المستقبلية وتوجيهات من خلال تقديم آفاق جديدة في القياس التربوي والنفسي.

الفصل الأول : تقديم الموضوع

خلفية الدراسة

- أهمية اختبارات التحصيل في مادة الرياضيات.

- نظرة عامة عن الخصائص السيكو مترية.

تحديد المشكلة

- الثغرات في الأبحاث القائمة حول اختبارات الرياضيات.

- الاسئلة الفرعية

الفرضيات:

اهداف الدراسة

أهمية الدراسة

تمهيد

الرياضيات هي تخصص أساسي يلعب دورا مهما في مختلف جوانب الحياة اليومية والتعليم والمجالات المهنية. إنه بمثابة حجر الزاوية للتفكير المنطقي وحل المشكلات والتفكير التحليلي. وبالتالي، أصبح اختبار التحصيل في الرياضيات مكونا أساسيا في التقييم التربوي، والذي يهدف إلى تقييم فهم الطلاب وإتقان المفاهيم الرياضية. لا توفر هذه التقييمات رؤى حول أداء الطلاب الفردي فحسب، بل تثير أيضا الممارسات التعليمية وتطوير المناهج وقرارات السياسة التعليمية. تعتمد فعالية هذه الاختبارات على خصائصها السيكو مترية، والتي تضمن أن تكون الاختبارات التحصيلية ذات خصائص سيكو مترية مقبولة (ثبات، صدق....) وصالحة وعادلة.

خلفية الدراسة

أهمية اختبارات التحصيل في الرياضيات

تم تصميم اختبارات التحصيل في الرياضيات لقياس معرفة الطلاب ومهاراتهم في الرياضيات على مختلف المستويات التعليمية. تمتد أهمية هذه الاختبارات إلى ما هو أبعد من مجرد التقييم. إنها مفيدة في تحديد نقاط القوة والضعف لدى الطلاب، وبالتالي توجيه المدرسين الى تصميم ادوات التقييم لتلبية احتياجات التعلم المتنوعة. علاوة على ذلك، تلعب اختبارات الرياضيات الموحدة دورا حاسما في تقييم الأنظمة التعليمية، مما يؤثر على معايير المناهج الدراسية والإصلاحات التعليمية. يمكن أن تؤثر نتائج هذه التقييمات على المسارات الأكاديمية للطلاب والمشهد التعليمي الأوسع، مما يجعل من الضروري أن يتم بنائها وتقييمها بصرامة ودقة.

بالإضافة إلى دورها في التعليم، يتم استخدام اختبارات التحصيل في الرياضيات بشكل متزايد في البحث لتحليل الاتجاهات في الأداء التعليمي عبر تركيبة المجتمع والمناطق المختلفة. على هذا النحو، فإن الخصائص السيكو مترية لهذه الاختبارات أمر حيوي لضمان أن تكون الاستنتاجات المستخلصة من

بيانات الاختبار دقيقة وذات مغزى. وبدون خصائص القياس النفسي السليمة، قد تؤدي النتائج إلى تفسيرات مضللة وسياسات تعليمية غير فعالة.

الثغرات في الأبحاث الحالية حول اختبارات الرياضيات

تعد اختبارات التحصيل في الرياضيات أدوات أساسية في التقييم التربوي، حيث توفر رؤى حول تعلم الطلاب وتوجيه الممارسات التعليمية. ومع ذلك، على الرغم من أهميتها، لا تزال هناك فجوات ملحوظة في البحث حول هذه الاختبارات. يعد فهم هذه الفجوات أمراً بالغ الأهمية للنهوض بمجال القياس التربوي وتحسين فعالية تقييم الرياضيات. وهناك مجالات رئيسية تفتقر إلى البحث، بما في ذلك التقييم السيكو متري للاختبارات، وتأثير العوامل الثقافية والسياقية، واستخدام التكنولوجيا في التقييمات، ومواءمة عمليات التقييم مع معايير المناهج الدراسية.

التقييم السيكو متري لاختبارات الرياضيات

واحدة من الفجوات الرئيسية في الأبحاث الحالية هي التقييم السيكو متري لاختبارات التحصيل في الرياضيات. في حين أن الكثير من الأبحاث ركزت على تطوير هذه الاختبارات، إلا أن هناك نقصاً في الدراسات الشاملة التي تفحص الثبات والصدق من خلال أطر نظرية مختلفة وفي مجموعات سكانية وسياقات متنوعة.

- الاستخدام المحدود للنماذج السيكو مترية المتقدمة: لا تزال العديد من الدراسات تعتمد بشكل كبير على نظرية الاختبار الكلاسيكية (CTT)، والتي توفر فهماً أساسياً للثبات والصدق ولكنها لا تأخذ في الحسبان الأداء على مستوى الفقرات وتعقيدات قدرات الطلاب. تقدم نظرية الاستجابة للفقرة (IRT) نهجاً أكثر تعقيداً من حيث التطبيق، ومع ذلك لا يزال تطبيقها في اختبارات تحصيل الرياضيات غير مستكشف. هناك حاجة إلى البحث لمقارنة فعالية CTT و IRT في مختلف البيئات التعليمية (Hambleton & Jones ، 1993).

تأثير العوامل الثقافية والبيئية

تؤثر العوامل الثقافية والبيئية بشكل كبير على تعلم الرياضيات وتقييمها، ومع ذلك فإن البحث في هذا المجال محدود.

- التحيز الثقافي في فقرات الاختبار: أظهرت الأبحاث الحالية أن اختبارات الرياضيات يمكن أن تعكس التحيزات الثقافية التي تضر بمجموعات معينة من الطلاب. على سبيل المثال، قد لا يكون للفقرات ذات الصلة بسياق مجموعة ثقافية ما صلة مع مجموعة ثقافية أخرى، مما يؤثر على أداء التلاميذ (Steele & Aronson 1995). هناك حاجة إلى مزيد من الدراسات للتحقيق في كيفية تأثير العوامل الثقافية على أداء الاختبار وتطوير فقرات اختبار تتفادى هذا العامل.

- التأثيرات السياقية البيئية على الأداء: تأثير عوامل البيئة، مثل بيئة الفصل الدراسي وممارسات التدريس وتأثيرات الأقران، على درجات اختبار التحصيل في الرياضيات غير مفهوم جيداً. يجب أن يستكشف البحث كيفية تفاعل هذه العوامل مع أداء الاختبار وكيف يمكن دمجها في تصميم عمليات التقييم (Eccles & Roeser ، 2011).

كما ان هناك نقص في تمثيل المجتمع: ركزت العديد من الدراسات السيكو مترية على تلاميذ مناطق معينة في مدن كبيرة في حين هناك نقص فادح في المناطق الاخرى، مما أدى إلى تساؤلات حول قابلية تعميم النتائج. هناك حاجة للبحث الذي يتضمن خلفيات ثقافية واجتماعية واقتصادية متنوعة لضمان أن تكون الاختبارات صادقة وثابتة لجميع التلاميذ (Gonzalez et al. ، 2010).

استخدام التكنولوجيا في التقييمات

أدى التقدم السريع للتكنولوجيا إلى تغيير الممارسات التعليمية، ومع ذلك لم يتم بحث اندماجها في اختبار الرياضيات بشكل كاف.

- فعالية الاختبارات عبر الكمبيوتر: في حين أن التقييمات المستندة إلى الكمبيوتر تقدم مزايا مثل التغذية الراجعة الفورية (Huang et al. ، 2018).

دمج الأدوات التفاعلية: اكتسب استخدام الأدوات التفاعلية، مثل برامج الهندسة الديناميكية والمحاكاة عبر الإنترنت، في تقييم الرياضيات شهرة. ومع ذلك، هناك نقص في الدراسات التي تدرس كيفية تأثير هذه الأدوات على نتائج تعلم الطلاب وصلاحياتها كأدوات تقييم (Kershner et al. ، 2011).

مواءمة التقييمات مع معايير المناهج الدراسية

فجوة أخرى مهمة في البحث هي مواءمة اختبارات التحصيل في الرياضيات مع معايير المناهج الدراسية الحالية.

- اختلال المناهج الدراسية: لا تعكس العديد من الاختبارات الحالية بشكل كاف المحتوى والمهارات التي تم التأكيد عليها في مناهج الرياضيات المعاصرة، مما يؤدي إلى مخاوف بشأن أهميتها وفعاليتها. هناك حاجة إلى البحث لتحليل مواءمة التقييمات مع المعايير الوطنية والمعايير الدولية (Popham ، 2017).

الاشكالية

تظهر معظم الدراسات العربية عامة والجزائرية خاصة ان طريقة تصميم الاختبارات التحصيلية وتحليل الفقرات، وإطار التقييم تعتمد بشكل كبير على طريقة نظرية الاختبار الكلاسيكية (CTT)، التي تعرضت للانتقاد بسبب عيوبها، مثل أن الاعتماد على الاختبار الكلي بدلا من الفقرات، ويفترض علاقات خطية بين المتغيرات الكامنة والدرجات الملاحظة. لذا من المستحيل تقدير الدرجة الحقيقية مباشرة، أو دون تقديم افتراضات قوية، فبارامترات الفقرة مثل معامل التمييز ومعامل صعوبة لفقرات الاختبار تعتمد على العينة المستخدمة، كما ان تحديد مستوى التحصيل للتلاميذ يعتمد على درجة صعوبة الفقرات وتمييزها، وكذلك الخطأ المعياري للقياس و دالة ثبات وتباين درجات الاختبار، هو معمم لجميع الممتحنين. هذه القيود قد تشكل عدة تحديات عند استخدامها في الامتحانات لقياس القدرات العليا . على سبيل المثال، يفشل اختبار CTT في أخذ التوزيعات الملاحظة لدرجات الاختبارات التي تتأثر بالقيم المتطرفة حيث يحصل نسبة كبيرة من الممتحنين على درجات قريبة من الحد الأدنى أو الأعلى من المجال (Demars, 2017; Jabrayilov et al., 2016; Rusch et al., 2017). ونظرا للصعوبات في حل هذه المشكلات ضمن إطار نظرية القياس الكلاسيكية، تحول اهتمام الاخصائيين في القياس النفسي ومنظمات التقييم إلى نظرية حديثة تعرف بنظرية الاستجابة للفقرة (IRT). (Embretson & Reise, 2013).

كما يجب أن تتمتع أدوات التقييم التعليمي بالصدق والثبات لضمان استنتاجات دقيقة حول تحصيل التلاميذ. في التعليم الابتدائي، وخاصة الصف الخامس (مرحلة انتقالية حاسمة تربط العمليات الرياضية الملموسة بالتفكير المجرد)، وهذه الادوات بدورها تؤثر على جودة التقييم مباشرة واتخاذ القرارات التربوية وتخصيص الموارد (Hattie, 2009).

بينما تستخدم تقييم التحصيل في الرياضيات للصف الخامس على نطاق واسع، فإن التحقق السيكومتري غالبا ما يتوقف عند تقدير الثبات. هذا النطاق المحدود يفشل في التحقق من مبادئ القياس الأساسية، مثل البعد الواحد والاستقلال الموضوعي، والتي تعتبر ضرورية عند تقييم لتتبع التقدم أو تخطيط

للتدخل على مستوى القدرات العليا. فمن الضروري المقارنة المباشرة بين البيانات المستمدة من CTT وتلك المستمدة من نموذج راش لتحديد ما إذا كانت المقاربة التقليدية تحمي بشكل كاف جودة أدوات التقييم المستخدمة لهذه المرحلة التنموية الحاسمة.

- تقييم مهارات التفكير او القدرات العليا: غالبا ما تؤكد اختبارات الرياضيات الحالية على الحفظ والمهارات الإجرائية بدلا من المستويات العليا من التفكير وقدرات حل المشكلات، كما بينت العديد من الدراسات لتحليل هذه الاختبارات في مراحل التعليم المختلفة ومنها دراسة سليمان بن عبد الله بن صالح الدويش (2003) وغيرها وقد توصلت معظم نتائج هذه الدراسات إلى أن الاختبارات لا تزال تركز على المستويات العقلية الدنيا، وتهمل المستويات العقلية العليا كالتحليل والتركيب والتقويم، فاختبارات الرياضيات بحاجة إلى التحليل والتقويم المستمر لكي يتم الكشف عن جوانب القصور والنقص فيها ومدى اشتغالها على المستويات المعرفية.

يمكن تحديد مشكلة الدراسة في التساؤلات الرئيسة التالية:

التساؤل الرئيسي: ما هي الخصائص السيكو مترية للاختبار التحصيلي في الرياضيات وفق نظرية الاستجابة للفقرة (نموذج راش انموذجا)؟

ويتفرع من هذا التساؤل التساؤلات الفرعية الآتية:

1. ما خصائص درجات عينة الدراسة على الاختبار التحصيلي وفق النظرية الكلاسيكية؟
2. ما خصائص درجات عينة الدراسة على الاختبار التحصيلي وفق نموذج راش؟
3. ما قيم مؤشرات الصعوبة ل فقرات الاختبار التحصيلي وفق النظرية الكلاسيكية؟
4. ما قيم مؤشرات الصعوبة ل فقرات الاختبار التحصيلي وفق نموذج راش؟
5. ما قيم مؤشرات معامل التمييز ل فقرات الاختبار التحصيلي وفق النظرية الكلاسيكية؟
6. ما قيم مؤشرات معامل التمييز ل فقرات الاختبار التحصيلي وفق نموذج راش؟
7. ما قيم مؤشرات ثبات الاختبار التحصيلي وفق النظرية الكلاسيكية؟
8. ما قيم مؤشرات ثبات الاختبار التحصيلي وفق نموذج راش؟

9. ما قيم مؤشرات صدق الاختبار التحصيلي وفق النظرية الكلاسيكية؟

10. ما قيم مؤشرات صدق الاختبار التحصيلي وفق نموذج راش القدرة ؟

11. هل يختلف تقييم قدرة الرياضيات عبر النظريتين؟

فرضيات الدراسة:

الخصائص السيكو مترية للاختبار التحصيلي في الرياضيات وفقراته وفق القياس الكلاسيكي ووفق

نظرية الاستجابة للفقرة نموذج راش انموذجا تتميز بما يلي:

1. خصائص درجات عينة الدراسة على الاختبار التحصيلي تتوزع توزيعا طبيعيا وفق النظرية الكلاسيكية

وتزايد مستوى القدرات بتزايد الدرجات الخام. كما يمكن تطبيق الاحصاء البرامتري.

2. اما نموذج راش تتميز بانتشار مستوى القدرات توافقا مع مستوى صعوبة الفقرات في نموذج راش وفق

نموذج خطي للبيانات.

3. قيم مؤشرات الصعوبة لفقرات الاختبار التحصيلي لا تتميز بالاستقرار وفق النظرية الكلاسيكية حيث

تختلف حسب العينة ومستوى القدرات يتأثر بمستوى الفقرات.

4. قيم مؤشرات الصعوبة لفقرات الاختبار التحصيلي تتميز بالاستقرار وفق نموذج راش حيث لا تختلف

حسب العينة ومستوى القدرات يتأثر بمستوى الفقرات.

5. قيم مؤشرات التمييز لفقرات الاختبار التحصيلي لا تتميز بالاستقرار وفق النظرية الكلاسيكية حيث

تختلف حسب العينة ومستوى القدرات يتأثر بمستوى الفقرات.

6. قيم مؤشرات التمييز لفقرات الاختبار التحصيلي تتميز بالاستقرار وفق نظرية الاستجابة للمفردة حيث

لا تختلف حسب العينة ومستوى القدرات بتأثر بمستوى الفقرات.

7. تتميز قيم مؤشرات ثبات الاختبار التحصيلي بالارتفاع نتيجة عدد افراد العينة المرتفع في النظرية

الكلاسيكية.

8. تتميز قيم مؤشرات ثبات الاختبار التحصيلي في نموذج راش بتعددده حيث هناك معامل ثبات الصعوبة للفقرات ومعامل ثبات خاص بالأفراد وكل منها يحدد مستويات الصعوبة ومستويات القدرة، والثبات الهامشي والشرطي مع تحديد الخطأ المعياري الشرطي.
9. تتميز قيم مؤشرات صدق الاختبار التحصيلي في النظرية الكلاسيكية بإمكانية تحقيقها لتوفر طرق عديدة لتحقيق ذلك، بناء الاختبار التحصيلي وفق البرنامج المقرر لمنهاج الرياضيات للسنة الخامسة ابتدائي بإنشاء جدول المواصفات واجراء عملية التحكيم مما سيعزز بيئة صدق المحتوى وصدق المفهوم والاتساق الداخلي.
10. نموذج راش يتحقق الصدق بتحقق افتراضات النموذج والتي تدل على صدق بياناته وهي احادية البعد والاستقلال الموضوعي. وكذلك مطابقة بيانات الدراسة لنموذج راش.

أهداف الدراسة:

سعت هذه الدراسة إلى تحقيق الأهداف التالية:

1. التعرف على خصائص درجات عينة الدراسة على الاختبار التحصيلي.
2. التعرف على قيم مؤشرات الصعوبة لفقرات الاختبار التحصيلي وفق القياس الكلاسيكي
3. التعرف على قيم مؤشرات التمييز لفقرات الاختبار التحصيلي وفق القياس الكلاسيكي
4. التعرف على قيم مؤشرات ثبات الاختبار التحصيلي وفق القياس الكلاسيكي .
5. التعرف على قيم مؤشرات صدق الاختبار التحصيلي وفق القياس الكلاسيكي .
6. التعرف على درجة ملائمة نموذج راش لبيانات الاختبار التحصيلي.
7. تقدير صعوبة فقرات الاختبار التحصيلي وتدريبها وفق نموذج راش.
8. تقدير قدرات التلاميذ المقابلة لكل درجة كلية خام على الاختبار التحصيلي وفق نموذج راش .
9. التعرف على مدى ثبات الاختبار التحصيلي وثبات الافراد وفق نموذج راش.
10. التعرف على مدى صدق الاختبار التحصيلي وفق نموذج راش .

أهمية الدراسة :

1. إن استخدام نموذج راش في تدريج اختبار تحصيلي تبعاً لتقديرات المفردات في قياس القدرات يتيح إمكانية تطويرها عن طريق حذف بعض المفردات غير الملائمة لقياسها.
2. قد تؤكد نتائج الدراسة الحالية على مميزات نموذج راش بما يفترضه من موضوعية في القياس قد تتمثل في دقة أكبر في التنبؤ بالظاهرة موضوع الدراسة مما يؤيد استخدامه في مجالات القياس النفسي والتربوي ويدفع الباحثين المهتمين إلى إجراء المزيد من الدراسات في استخدام نموذج راش في فحص الخصائص السيكومترية لقياس القدرات
3. قد تعد هذه الدراسة إضافة إلى رصيد الدراسات الذي يحفل بتطبيق نموذج راش في مجالات القدرات العقلية والتحصيلية.
4. استخدام نموذج راش يوفر متطلبات الموضوعية في قياس المتغيرات موضع القياس، مما يعني توفر شرطي الصدق والثبات لتقديرات كل من صعوبات المفردات والسمات للأفراد أي تحقق صدق وثبات القياس.
5. يتيح الفرصة لتقدير بارامترات الفرد والمفردة بوحدة قياس معرفة هي وحدة اللوجيت مما يحقق دقة القياس وموضوعيته.
6. يتيح الفرصة لقياس التغير الحادث في التحصيل في مستوى زمني معين.

حدود الدراسة:

يتحدد البحث الحالي بموضوعه المتمثل في البحث السيكومتري المقارن بين نظرية القياس الكلاسيكية ونظرية الاستجابة للمفردة باستخدام نموذج راش أحادي البارامتر، في بناء وتحليل وتدريج الاختبار التحصيلي في الرياضيات.

كما يتحدد بالعينة التي شملها البحث والمكونة من 559 تلميذ في الدراسة الأساسية و78 تلميذ في الدراسة الاستطلاعية من تلاميذ السنة الخامسة ابتدائي، والموزعين على مدارس من ولاية معسكر وولاية وهران، خلال السنة الدراسية 2015/2016.

استخدام أحد نماذج نظرية الاستجابة للمفردة وهو نموذج راش الذي يعتمد على معلم واحد هو صعوبة

الفقرة في تصميم الاختيارات التحصيلية والنفسية.

اقتصرت هذه الدراسة على:

أولاً: الحدود المكانية: جميع تلامذة السنة الخامسة ابتدائي لدائرتي سيق، عقاز (معسكر) ووادي تليلات

(وهران).

ثانياً: الحدود الزمانية: تم تطبيق هذه الدراسة ميدانياً في نهاية الفصل الدراسي الثاني 2016.

ثالثاً: الحدود الموضوعية:

- استخدام أحد نماذج نظرية الاستجابة للفقرة وهو نموذج راش الذي يعتمد على معلم واحد هو صعوبة الفقرة.
- استخدام اختبار تحصيلي للسنة الخامسة ابتدائي أعد من طرف الطالب.
- حساب الخصائص السيكومترية للاختبار وفق النظرية الكلاسيكية ونموذج راش .

مصطلحات الدراسة

الاختبار : Test

عرف (الظاهر،ⁱⁱ1999) الاختبار بأنه "وسيلة منظمة تستهدف قياس كمية المعلومات التي يحفظها

الطالب ويتذكرها في حقل من حقول المعرفة، وقدرته على فهمها، أو تطبيقها، أو تحليلها والانتفاع بها

في مواقف الحياة المختلفة" (الظاهر، 1999، ص50).

في حين عرفه (البج، 2000) بأنه "موقف يطلب في أثناءه من المفحوص أن يظهر معارفه أو مهاراته

أو اتجاهاته وميوله أو جوانب منها، تتصل بموضوع معين أو عدد من الموضوعات". (البج، 2000،

ص160).

بينما عرفه (ابو جادو، 2000) بأنه " إجراء منظم لتحديد مقدار ما تعلمه الطلبة في موضوع ما في ضوء الأهداف المحددة. (ابو جادو، 2000، ص453).

اما (الزغول، 2001) فعرف الاختبار بأنه "عينة من الأسئلة أو المهام التعليمية المصاغة على نحو يمكن معه قياس مدى تحقق الأهداف المحددة مسبقا لمادة تعليمية معينة أو مهارة ما ". (الزغول، 2001، ص327).

من خلال ما تقدم من تعريفات فيمكن للطالب أن يعرف الاختبار إجرائيا في هذا البحث بأنه "مجموعة من الأسئلة التي أعدت لقياس مقدار ما تعلمه التلاميذ من مستويات المعرفة والتطبيق والاستدلال لمادة الرياضيات للمستوى السنة الخامسة ابتدائي.

3- الخصائص السيكومترية: Psychometric Properties

ويعرف الطالب الخصائص السيكومترية بأنها مؤشرات للاختبار المعتمد في هذا البحث وفقراته التي تكشف عن مدى دقتها في قياس ما أعدت لقياسه، من خلال الصدق والثبات للاختبار ومعاملات التمييز والصعوبة ومؤشرات الثبات لفقراته، وفيما يأتي تعريف لكل خاصية من الخصائص المعتمدة في هذا البحث وهي:

أ - ثبات الاختبار: Test Reliability

يعرف "ايل ، 1972 (Ebel) الثبات بأنه "الاتساق في مجموعة درجات فقرات المقياس التي يفترض أن تقيس ما يجب قياسه " (Ebel 1972 , p408).

ويعرف إحصائيا بأنه "نسبة التباين الحقيقي (المنسوب للسمة المقاسة أولا يرتبط بها) الى التباين الكلي"، وبلغه احصائية فان هذه النسبة هي مربع معامل الارتباط بين العلامة الحقيقية والعلامة الظاهرية. (الأنصاري، 2000، ص119).

ويعرف كل من "مارك ورونالد، 2004 "معامل ثبات الاختبار بأنه نسبة التباين الحقيقي الى التباين الكلي (Mark & Ronald , 2004. P. 129).

ويعرف الثبات إجرائيا في هذا البحث بأنه التجانس بين درجات فقرات الاختبار الذي يحسب بمعادلة الفا كرونباخ والتجزئة النصفية.

ب - صدق الاختبار: Test Validity

يذكر " عودة، 2004 " ان الصدق يعني ان الاختبار يقيس ما أعد لقياسه ، وبلغة الإحصاء، هو نسبة التباين الحقيقي المرتبط أو المنسوب للسمة المقاسة الى التباين الكلي (عودة، 2004 ص422). بينما نجد " ايل ، 1972 (Ebel) يعرف الصدق بأنه " دقة درجات الاختبار في قياس ما يجب ان يقيسه " (Ebel 1972 , p409) . فالصدق ان يقيس الاختبار ما ينتظر منه قياسه (ربيع ، 2008، ص 92).

ويعرف الصدق اجرائيا في هذا البحث بأنه الصدق الارتباط بين الفقرة والدرجة الكلية والمقارنة الطرفية، بالنسبة للنظرية الكلاسيكية وتحقيق فرضيات نموذج راش (احادي البعد والاستقلال الموضوعي)

معامل صعوبة الفقرة : Item Difficulty

تعد صعوبة الفقرات من اهم الخصائص السيكومترية للاختبار التحصيلي وتمثل نسبة الذين أجابوا عن الفقرة إجابة صحيحة، وينبغي أن يكون متوسط معاملات الصعوبة قريبا من 50% (يعطي أدق ما يمكن من عملية القياس (Hetzel , 1997, p.1-6) وبالتالي ينبغي حذف الفقرات السهلة جدا أو الصعبة جدا من الاختبار (Nitko ,1983, p.288).

ويعرف معامل صعوبة الفقرة إجرائيا "نسبة التلاميذ الذين أجابوا عن الفقرة إجابة صحيحة الى العدد الكلي لأفراد العينة. "

معامل تمييز الفقرة: يقيس مدى جودة تمييز الفقرة بين الدرجات العالية والمنخفضة في الاختبار العام.

والطريقة المتبعة هي الارتباط ثنائي التسلسل بين درجات الفقرات وإجمالي الدرجات..

ويفضل ان لا يقل معامل تمييز الفقرة عن 0.30 اذ تستبعد أي فقرة يكون معامل تمييزها اقل من

هذه

النسبة ، وتستبعد ايضا أي فقرة معامل تمييزها سالباً مهما كانت نسبة التمييز، وإذا كانت درجات

الفقرة (1،0) فيمكن ان نحسب معامل التمييز من خلال الفرق بين عدد الذين اجابوا عن الفقرة

إجابة صحيحة في المجموعة العليا وبين عدد الذين اجابوا إجابة صحيحة عنها في المجموعة الدنيا

مقسوما على عدد أفراد مجموعة واحدة (عوده، 2004)

يذكر (رلورد ونوفك ، 1986) بأن معامل تمييز الفقرة هو "قدرة الفقرة على التمييز بين ذوي

الدرجات العليا و ذوي الدرجات الدنيا من المفحوصين" (Lord&Novick,1986.p331),

والفقرات التي لا تميز يجب استبعادها إذ لا بد من اختيار الفقرات التي تميز بين المفحوصين

(Matlock ,1997, p.9).

ويعرف معامل تمييز الفقرة إجرائيا في هذا البحث بأنه "نسبة الفرق في عدد الذين أجابوا عن الفقرة

إجابة صحيحة من المجموعتين المتطرفتين في الدرجة الكلية للاختبار الى عدد طلبة مجموعة واحدة الذي

ينبغي ان يكون متساويا في المجموعتين.

. مؤشر ثبات الفقرة: Item Reliability

يذكر (عبدالرحمن، 1998) ان ثبات الفقرة يعني استقرارها في قياس ما اعدت لقياسه. (عبد الرحمن

، 1998، ص 264). ويذكر (كروكر و الجينا ، 1986) ان مؤشر ثبات الفقرة يعني "استقرارها الذي

يحسب من خلال ضرب الانحراف المعياري للفقرة في معامل ارتباطها بالدرجة الكلية" (Algina&,

(.Crocker1986,p.320

معامل صدق الفقرة: Item Validity

تذكر (انستازي، 1988) بأن صدق الفقرة يعني "قدرتها على قياس المفهوم المراد قياسه الذي يقيسه

الاختبار"، وتحسب من خلال ارتباط الفقرة بمحك داخلي أو خارجي، وان أفضل محك داخلي لصدق

الفقرة حينما لا يتوافر محك خارجي، هو درجة المفحوص الكلية على الاختبار ، (Anastasi ,

1988 , p.211) لذلك يعرف معامل صدق الفقرة إجرائيا في هذا البحث بأنه "معامل ارتباط الفقرة

بالدرجة الكلية للاختبار.

الفصل الثاني: الاطار النظري

اولا: اختبارات تحصيل الرياضيات

ثانيا : النظرية الكلاسيكية

ثالثا : نظرية الاستجابة للمفردة

رابعا : نموذج راش

اولا: اختبارات تحصيل الرياضيات

• نظرة عامة على اختبارات التحصيل في الرياضيات

• التطور التاريخي

• الاتجاهات الحالية .

تمهيد

الرياضيات هي تخصص أساسي يلعب دورا مهما في مختلف جوانب الحياة اليومية والتعليم والمجالات المهنية. إنه بمثابة حجر الزاوية للتفكير المنطقي وحل المشكلات والتفكير التحليلي. وبالتالي، أصبح اختبار التحصيل في الرياضيات مكونا أساسيا في التقييم التربوي، والذي يهدف إلى تقييم فهم الطلاب وإتقان المفاهيم الرياضية. لا توفر هذه التقييمات رؤى حول أداء الطلاب الفردي فحسب، بل تثير أيضا الممارسات التعليمية وتطوير المناهج وقرارات السياسة التعليمية. تعتمد فعالية هذه الاختبارات على خصائصها السيكومترية، والتي تضمن أن تكون التقييمات موثوقة وصالحة وعادلة.

التطور التاريخي لاختبارات التحصيل في الرياضيات

1. أوائل القرن العشرين: حركة التوحيد القياسي

- شهدت أوائل القرن العشرين بداية الاختبار الموحد في التعليم، بسبب الحاجة إلى قياس موضوعي لأداء الطلاب. وضع إدخال اختبار ستانفورد بينيه للذكاء في عام 1916 سابقة للتقييمات الموحدة، بما في ذلك الرياضيات (بينيه وسيمون، 1916).
- حدث تطوير أول اختبارات تحصيل الرياضيات في عشرينيات وثلاثينيات القرن العشرين. تم إنشاء اختبارات مثل اختبارات أيوا للمهارات الأساسية (ITBS) واختبارات التحصيل في كاليفورنيا (CAT) لتقييم إتقان الطلاب للمفاهيم والمهارات الرياضية (Harris ، 2007).

2. منتصف القرن العشرين: التركيز على التوافق مع المناهج الدراسية

- شهدت الخمسينيات والستينيات تحولا نحو مواءمة التقييمات مع معايير المناهج الدراسية. تأسس المجلس الوطني لمعلمي الرياضيات (NCTM) في عام 1920، وبحلول الثمانينيات، بدأ في التأثير على تطوير مناهج وتقييم الرياضيات (NCTM، 1989). وهذا مرتبط بالأطر النظرية للاختبارات التحصيلية حيث صُنفت أساليب القياس الكلاسيكي نوعين من القياس هما:

1. القياس جماعي المرجع : Norm-referenced measurement :

نشأ هذا النظام مرتبطاً بالفلسفة التربوية التي سادت في أوائل قرن العشرين، وهي تصنيف الأفراد بحسب مركزهم النسبي بين أقرانهم في القدرات المختلفة، وكان علماء النفس أكثر من علماء التربية فاعلية في نشأة هذا النوع من القياس، فقد كان بينيه "Binet" أول من صمم اختباراً لعزل الأطفال المتخلفين عقلياً عن العاديين (الشرقاوي وآخرون، 1996: 83)

وقد أدى ظهور ونجاح أول اختبار جمعي لتصنيف الضباط والجنود قبيل الحرب العالمية الأولى إلى تشجيع حركة القياس جماعي المرجع. (علام 2001: 21)

وهكذا تقوم المقاييس جماعية المرجع - وهي أكثر المقاييس شيوعاً وانتشاراً - على تقدير الفروق الفردية بين الأفراد والتمييز بينهم، وفي هذه المقاييس، لا تكون لدرجة الفرد معنى ما لم تقارن بمعيار يعتمد على مستوى جماعة الأقران التي ينتمي إليها هذا الفرد، ويتمثل هذا المستوى بمتوسط درجات هذه الجماعة، بينما تتمثل المقارنة بمدى انحراف درجة الفرد عن هذا المتوسط وبوساطة المعايير المحسوبة لدرجات المجموعة التي ينتمي إليها هذا الفرد (كاظم، 1988: 17)

2. القياس محكي المرجع: Criterion-reference measurement:

مع شيوع مفهوم التعلم من أجل الإتقان "Mastery" "Learning" لم يعد الهدف هو التركيز أساساً على الفروق الفردية، وظهرت المناذاة بالابتعاد عن شكل التوزيع ألعندالي - الذي تعتمد عليه المقاييس جماعية المرجع - حيث إن النشاط التربوي نشاط مقصود يبدل بهدف أن يتقن الطلبة ما تعلموه، وإذا ما تحقق الهدف من العملية التعليمية، فإن توزيع الأداء يختلف تماماً عن المنحنى الاعندالي، لذلك ظهر أسلوب جديد في القياس، عرف بالقياس محكي المرجع (كاظم 1988: 18)

ويعتبر جليسر Glaser 1963، أول من وجه الاهتمام إلى هذا الأسلوب في القياس، وكان أول من أطلق عليها مصطلح الاختبارات محكية المرجع، ثم جاء بعده كارفر Carver 1974 ليكون أول

من استخدم مصطلح اديو ميري (محكي المرجع) في مقابل سيكو ميري (جماعي المرجع) (منصور،

1987: 22)

والهدف من القياس محكي المرجع هو تحديد المستوى الذي يصل إليه أداء الفرد وتقدير مدى تقدمه

بالنسبة للأهداف الموضوعية للقياس، بغض النظر عن مكانته بين أقرانه (عبد السلام، 1992: 340)

- أكد إدخال حركة "الرياضيات الجديدة" في الستينيات على المفاهيم الرياضية المجردة، مما أدى إلى تطوير

الاختبارات التي قيمت مهارات التفكير العليا في الرياضيات. (Kilpatrick, 2003)

3. أواخر القرن العشرين: المساءلة والاختبار الموحد (المعياري)

- شهدت الثمانينيات والتسعينيات زيادة في تدابير المساءلة، استجابة لتقارير وطنية مثل "أمة في خطر"

(اللجنة الوطنية للتميز في التعليم، 1983). سلط هذا التقرير الضوء على الحاجة إلى تحسين المعايير

التعليمية، مما يؤدي إلى تنفيذ اختبارات التحصيل الموحدة في الرياضيات عبر الولايات. (Linn,

2000)

- أكد إدخال قانون عدم ترك أي طفل خلف الركب (NCLB) في عام 2001 على الاختبار الموحد

كوسيلة للمساءلة، حيث فرض اختبارات سنوية في الرياضيات للطلاب في الصفوف 3-8 (U.S.

Department of Education, 2002).

الاتجاهات الحالية في اختبارات التحصيل في الرياضيات

1. التركيز على مهارات التفكير المستويات العليا

- تؤكد الاتجاهات الحديثة على تقييم مهارات التفكير العليا، مثل حل المشكلات والتفكير النقدي، بدلا

من الحفظ عن ظهر قلب للحقائق الرياضية. يتماشى هذا التحول مع معايير الدولة الأساسية المشتركة

(the Common Core State Standards (CCSS)، والتي تعزز فهما وتطبيقا أعمق

للمفاهيم الرياضية (مركز جمعية الحكام الوطنية لأفضل الممارسات ومجلس كبار مسؤولي المدارس

الحكومية، 2010). (National Governors Association Center for Best

Practices & Council of Chief State School Officers, 2010).

2. دمج التكنولوجيا

- زاد استخدام التكنولوجيا في تقييم الرياضيات، حيث يتم الآن إجراء العديد من الاختبارات عبر الإنترنت.

يسمح الاختبار المستند إلى الكمبيوتر بالتقييم التكيفي التي تضبط صعوبة الأسئلة بناء على أداء الطالب

(Van der Linden & Pashley ، 2006).

- بالإضافة إلى ذلك، يتم دمج الفقرات المعدلة عن طريق التكنولوجيا، مثل المحاكاة التفاعلية وأدوات

الرسوم البيانية، في التقييم لقياس التفكير الرياضي وقدرات حل المشكلات لدى الطلاب بشكل أفضل

(Bennett et al. 2012).

3 . التركيز على التقييم التكويني

- هناك اتفاق متزايد بأهمية التقييم التكويني في تعليم الرياضيات. يوفر التقييم التكويني تغذية راجعة مستمرة

للطلاب والاساتذة ، مما يساعد على تحديد مجالات التحسين وإبلاغ الممارسات التعليمية (Black

& Wiliam ، 1998).

- تستخدم أدوات مثل مهام الأداء والمحفز (PortFolio) بشكل متزايد جنباً إلى جنب مع الاختبارات

الموحدة التقليدية لتوفير رؤية أكثر شمولاً لفهم الطلاب الرياضي (Hattie & Timperley ،

2007).

4. المساواة وإمكانية الولوج

- تركز الاتجاهات الحالية أيضاً على المساواة وإمكانية الولوج الى تقييم الرياضيات. وتبذل الجهود لضمان

أن تكون الاختبارات ذات صلة ثقافياً ومتاحة لجميع التلاميذ، بمن فيهم ذوي الاعاقة ومتعلمي اللغة

المختلفة. (Thompson et al.، 2006).

- يهدف استخدام مبادئ التصميم الشامل للتعلم (UDL) في تصميم الاختبار إلى إنشاء اختبارات تستوعب احتياجات التعلم المتنوعة وتعزز الفرص العادلة لجميع التلاميذ (Rose & Meyer ، 2002).

استنتاج

شهدت اختبارات التحصيل في الرياضيات تحولات كبيرة منذ إنشائها، مما يعكس اتجاهات تعليمية أوسع نطاقا واحتياجات مجتمعية. من الاختبارات الموحدة الأولى إلى الاختبارات الحديثة التي تؤكد على مهارات التفكير العليا والمساواة في اجتياز الاختبارات، يستمر مجال تقييم الرياضيات في التطور. مع تقدم المعايير والتقنيات التعليمية، مما ينتج عنه تقدم أيضا في الأساليب المستخدمة لتقييم التحصيل الرياضي للطلاب.

ثانيا : نظرية الاختبار الكلاسيكية (CTT)

تمهيد

افتراضات النظرية الكلاسيكية

نماذج نظرية الاختبار الكلاسيكية

مزايا نظرية الاختبار الكلاسيكية

الخصائص السيكمترية

تمهيد

نظرية الاختبار الكلاسيكية (CTT) هي أحد الأطر الأساسية في مجال القياس التربوي والقياس النفسي. يوفر نهجاً إحصائياً لفهم ثبات وصدق درجات الاختبار، والتي تعتبر ضرورية للتقييم الفعال لتعلم الطلاب. سنتعمق في هذا الفصل في مبادئ CTT ومزاياها وحدودها، وستسلط الضوء على الأبحاث ذات الصلة التي توضح تطبيقاتها وقيدوها.

تم تسمية نظرية القياس الأولى بنظرية الاختبار الكلاسيكية (CTT) لأنها تمت صياغتها من الافتراضات البسيطة التي وضعها منظرو الاختبار في بداية القياس وتسمى أيضاً نظرية الدرجات الحقيقية والخطأ، لأنه يعتقد أنها تضم كلا من الدرجات الحقيقية ونسبة الخطأ، وكذلك نظرية الثبات الكلاسيكية، لأن وظيفتها الرئيسية هي تقييم ثبات الدرجات الملاحظة في الاختبار، أي أنه يحسب قوة العلاقة بين الدرجة الملاحظة والدرجة الحقيقية. تستفيد النظرية الكلاسيكية من عدد من النماذج التي تعتبر مجموعة من الإجراءات المختلفة يحاول بها مطورو الاختبار تقديم حلول للمشاكل المعقدة من أجل قياس المتغيرات النفسية. لا يهتم علماء النفس بالنتيجة نفسها ولكن بالاستنتاجات التي يمكنهم استخلاصها والتفسيرات التي يمكنهم تقديمها على أساس السلوك المقاس للفرد.

دوافع ظهور النظرية الكلاسيكية للاختبار (CTT)

- ظهرت النظرية الكلاسيكية للاختبار في أوائل القرن العشرين نتيجة عدة عوامل أساسية، من أهمها:
1. التطور في مجال الاختبارات النفسية والتربوية خاصة اختبارات الذكاء التي استخدمت لتحديد ما إذا كان الأطفال الصغار قادرين على الاستفادة من التعليم المدرسي العادي.
 2. تطور استخدام معامل الارتباط في العلوم الإنسانية حيث كان يتم حسابه باستخدام الدرجات الكلية الناتجة عن جمع استجابات مفردات ثنائية التصحيح (0 أو 1).
 3. التطور في تحليل البيانات الإحصائية حيث أصبح من المقبول أن المشاهدات قد تحتوي على تباين عشوائي، ويُنظر إلى هذا التباين على أنه خطأ في القياس.

4. التطور النظري لمفهوم تباين الخطأ مما أدى إلى استخدام التوزيع الطبيعي وافترض أن الدرجة المشاهدة تتكون من مجموع درجتين:

$$X = T + E$$

حيث:

- X الدرجة المشاهدة
- T الدرجة الحقيقية
- E خطأ القياس

5. فكرة أن مفردات الاختبار تمثل تكرارات لقياس نفس الصفة أي أن مفردات الاختبار الثنائية التصحيح يمكن اعتبارها قياسات مكررة للمتغير نفسه ونتيجة لذلك أصبح جمع المتغيرات العشوائية (درجات المفردات) للحصول على درجة كلية للفرد أمراً مبرراً نظرياً وليس مجرد افتراض. ومع ذلك، فقد كان من المفهوم أيضاً أن الأفراد يختلفون في مستوى الكفاءة أو القدرة التي يقيسها الاختبار.

افتراضات النظرية الكلاسيكية:

تنظر النظرية الكلاسيكية إلى درجة الاختبار على أنها تحتوي على مكونين، الدرجة الحقيقية والخطأ العشوائي. تعتبر الدرجة الحقيقية هي متوسط القيم المتطابقة المأخوذة من القياسات المتكررة. يعتبر الخطأ غير مرتبط إما بالدرجة الحقيقية أو بالخطأ الذي يمكن أن يحدث في قياس آخر لنفس السمة. على الرغم من أن النظرية عبارة عن تبسيط مفرط ولا تمثل النتائج، إلا أنها تبرز علاقات مفيدة في تصميم الاختبار وبنائه وكذلك في تقييم درجات الاختبار. الافتراض الأساسي الأول للنظرية الكلاسيكية هو أن النتيجة التي تم الحصول عليها هي مجموع الدرجة الحقيقية بالإضافة إلى الخطأ. أي أن الدرجة الحقيقية ودرجة الخطأ مندمجة بشكل لا يمكن الفصل بينهما، ويمكن التعبير عن هذا المفهوم كمعادلة بسيطة:

$$X = T + E,$$

أين X هي النتيجة التي تم الحصول عليها (الدرجة الملاحظة)، T هي الدرجة الحقيقية، ويمثل E أخطاء القياس.

وتجدر الإشارة إلى أن الدرجة الحقيقية غير معروفة أبداً. ومع ذلك، فإنه تظل ضمن مجال معين، ويمكن الحصول على أفضل تقدير لها. خطأ القياس هو كل شيء ما عدا النتيجة الحقيقية. يمكن أن تنشأ أخطاء القياس من مصادر عديدة، مثل اختيار الفقرات، وتطبيق الاختبار، وتصحيح الاختبار، والأخطاء المنهجية في القياس. تسمى المصادر الثلاثة الأولى للخطأ بشكل مشترك خطأ القياس غير المنتظم، مما يعني أن تأثيرها غير متوقع وغير متسق. يحدث خطأ القياس المنتظم عندما يقيس الاختبار باستمرار شيئاً مختلفاً عن السمة التي تم تصميمه لقياسها. يقلل خطأ القياس من ثبات أو تكرار نتائج الاختبار. يكشف افتراض أن الدرجة التي تم الحصول عليها تتكون من الدرجة الحقيقية ودرجة الخطأ عن العديد من الافتراضات الإضافية. الافتراض المستمد من الدرجات الحقيقية هو أن خطأ القياس غير منتظم يؤثر على درجات الاختبار بشكل عشوائي. إن عشوائية خطأ القياس هي افتراض أساسي للنظرية الكلاسيكية نظراً لوجود أحداث عشوائية، فإن أخطاء القياس غير المنتظمة لديها بعض الاحتمالات أن تكون إيجابية أو سلبية، وبالتالي فهي تصل إلى متوسط الصفر عبر مجموعة كبيرة من الأفراد. ويترتب على ذلك أن متوسط خطأ القياس هو صفر. افتراض آخر لنظرية القياس الكلاسيكية هو أن أخطاء القياس لا ترتبط بالدرجات الحقيقية.

$$\sigma_x^2 = \sigma_T^2 + \sigma_E^2$$

حيث:

$$\sigma_x^2 \text{ تباين الدرجة الملاحظة:}$$

$$\sigma_T^2 \text{ تباين الدرجة الحقيقية:}$$

$$\sigma_E^2 \text{ تباين درجة الخطأ:}$$

الافتراض الأخير هو أن أخطاء القياس لا ترتبط بالأخطاء في الاختبارات الأخرى. يمكن تلخيص جميع

هذه الافتراضات على النحو التالي :

- أخطاء القياس عشوائية، وموزعة توزيعاً طبيعياً
- متوسط خطأ القياس هو صفر،
- الدرجات الحقيقية والخطأ غير مرتبطين،
- الأخطاء في الاختبارات المختلفة غير مرتبطة .
- يمكن تقدير ثبات الاختبار من خلال فحص اتساق الدرجات عبر التطبيقات المختلفة.
- من الافتراضات المذكورة أعلاه، يمكننا الوصول إلى بعض الاستنتاجات المهمة حول ثبات القياس .على سبيل المثال، عندما نطبق اختباراً لعدد كبير من الأشخاص، نجد تبايناً في الدرجات التي يمكن التعبير عنها كتباين σ_x^2 وفقاً للنظرية الكلاسيكية، فإن تباين الدرجات التي تم الحصول عليها له مصدران منفصلان، تباين الدرجة الحقيقية وتباين أخطاء القياس.
- الافتراض الأساسي الثاني للنظرية الكلاسيكية هو الاختبارات المتوازية .مصطلح الاختبارات المتوازية يأتي من حقيقة أنه يمكن النظر إلى كل فقرة على حدة على أنها اختبار .أي أن كل بند يتوافق مع قيمة المتغير الكامن .إن افتراض نموذج الاختبارات المتوازية مرض لأنه يؤدي إلى استنتاجات مفيدة حول علاقة كل بند بالمتغير الكامن التي تركز على ملاحظة علاقات الفقرات ببعضها البعض .وبالتالي، يضيف نموذج الاختبارات المتوازية افتراضين إلى النظرية الكلاسيكية:

 1. يعتبر المقدار الذي يؤثر فيه المتغير الكامن على كل بند هو نفسه لجميع البنود
 2. يعتبر كل بند يمتلك نفس مقدار الخطأ مثل أي بند آخر .هذا يعني أن العوامل تؤثر على جميع البنود بالتساوي .

- تشير هذه الافتراضات إلى أن ارتباط كل عنصر بالدرجة الحقيقية متطابق، وهذا الاستنتاج يؤدي إلى تحديد الثبات.
- باختصار، يفترض نموذج الاختبارات المتوازية:

 - الخطأ عشوائي،
 - أخطاء غير مرتبطة ببعضها البعض،

- الأخطاء غير المرتبطة بالدرجة الحقيقية،

- تأثير المتغير الكامن على جميع الفقرات بالتساوي،

- مقدار متساو من الخطأ لكل فقرة. وبالتالي، يسمح لنا النموذج بإجراء استنتاجات حول المتغير

الكامن الذي يركز على الارتباط بين الفقرات. ومع ذلك، فإن النموذج يحقق ذلك من خلال اقتراح

افتراضات صارمة.

النماذج البديلة للنموذج المتوازي:

بصرف النظر عن الافتراضات الصارمة المشار إليها أعلاه، يمكن إجراء استنتاجات مفيدة بافتراضات

أقل صرامة. يعتمد نموذج آخر على ما يسمى الاختبارات المكافئة لتاو (essentially tau-

equivalent tests) ويقدم افتراضا أكثر تسامحا أي أن كل بند يتطلب نفس مقدار تباين الخطأ

الذي تمتلكه البنود الأخرى. هذا يعني أنها متوازنة فيما يتعلق بالمتغير الكامن ولكنها لا تتأثر بشكل

أساسي بنفس الدرجة بالعوامل الدخيلة التي تسمى الخطأ. وبالتالي، نظرا لأن الخطأ يمكن أن يختلف،

يمكن أن تختلف أيضا متوسطات وتباين البنود. يسمح لنا هذا النموذج بالوصول إلى نفس الاستنتاج

الذي نحققه من خلال نموذج الاختبارات المتوازنة، ولكن مع افتراضات أكثر تساهلا.

ومع ذلك، يعتقد بعض المنظرين أنه حتى النموذج المكافئ للتاو (essentially tau-

equivalent model) بشكل أساسي صارم للغاية. لهذا السبب، طرحوا ما يسمى بالنموذج

المتجانس (congeneric model)، مع افتراضات أكثر تحمرا، حيث يفترض النموذج فقط أن

جميع البنود لها متغير كامن مشترك. لا تحتوي الفقرات على علاقات قوية بنفس القدر مع المتغير الكامن،

ولا يلزم أن تكون فروق الخطأ الخاصة بها متساوية. الشيء الوحيد الذي يجب أن نفترض هو أن كل

بند يعرض النتيجة الحقيقية إلى حد ما..

مزايا نظرية الاختبار الكلاسيكية

1. البساطة والحدس: CTT سهلة الفهم والتطبيق نسبيا. إن مفاهيم الدرجة الحقيقية ودرجة الخطأ بديهية، مما يجعلها في متناول التربويين والباحثين الذين قد لا يحصلون على تدريب إحصائي مكثف (Crocker & Algina ، 1986).
2. تستخدم على نطاق واسع ورأسخة: CTT لها تاريخ طويل في القياس التربوي ومقبولة على نطاق واسع في هذا المجال. يتم تطوير العديد من الاختبارات والمقاييس الموحدة بناء على مبادئ CTT، مما يوفر إطارا موحدا لتقييم فعاليات الاختبارات (Kline ، 2000).
3. سهولة الحساب: يمكن حساب معاملات الثبات، مثل ألفا كرونباخ، بسهولة باستخدام برامج إحصائية، مما يسمح للباحثين بتقييم ثبات اختباراتهم بسرعة (Tavakol & Dennick ، 2011).
4. مفيد للتحليل على مستوى المجموعة: CTT فعال في تحليل درجات الاختبار على مستوى المجموعة، مما يجعله مناسباً للبيئات التعليمية حيث ينصب التركيز على الأداء العام للفصل أو المدرسة بدلا من الدرجات الفردية (Harris ، 2007)..

الخصائص السيكمترية

الثبات

معامل ثبات الفقرة : Item Reliability

يعد ثبات الفقرة من الخصائص الأساسية للاختبارات معيارية المرجع، لأن الحصول على فقرات ذات ثبات عال يهيئ الفرصة للحصول على اختبار ثابت بدرجة جيدة (عبد الرحمن، 1998، ص264) وهناك أكثر من طريقة لحساب ثبات الفقرات. إذا كانت الاجابة عن الفقرة، اما صحيحة أو خاطئة ومن ثم حساب معامل ارتباط (فاي) الذي يمثل معامل ثبات الفقرة (احمد، 1981، ص280)، وهناك معادله أخرى تعطي مؤشرا لثبات الفقرة وتعتمد على ضرب الانحراف المعياري للفقرة بمعامل ارتباطها بالدرجة الكلية (Algina , 1986 , P.320 & Crocker).

ثالثاً : نظرية الاستجابة للمفردة :

تمهيد

نشأة نظرية الاستجابة للفقرة ومتطلبات وشروط القياس الموضوعي

مفهوم نظرية الاستجابة للفقرة

مسلّمات نظرية الاستجابة للفقرة

افتراضات نظرية الاستجابة للفقرة

نماذج نظرية السمات الكامنة

مميزات وعيوب النظرية

نموذج راش

تقدير القدرة والخطأ المعياري وفق النظريتين

تمهيد:

نظرًا للمشكلات التي ظهرت من تبني النظرية الكلاسيكية، تجلت الحاجة إلى تطوير مقاربة جديدة للقياس بما يحقق كل من دقته وموضوعيته، وقد أسفرت جهود العلماء في سبيل تحقيق ذلك عن نظرية الاستجابة للفقرة.

القياس الموضوعي:

يعد القياس الموضوعي من التطورات المعاصرة في القياس النفسي التربوي، وقد ارتبط هذا النظام بمدخل جديد يطلق عليه مدخل السمات الكامنة في القياس؛ بما يشتمل عليه من نظريات ونماذج سيكو مترية مستحدثة.

وقد انتشر هذا النظام في الأوساط التربوية كرد فعل للنقد الذي يوجه إلى فلسفة القياس بعامة والنظام مرجعي المعيار بخاصة، والأسس التي يستند إليها هذا النظام في بناء الاختبارات والمقاييس النفسية والعقلية والتربوية وتحليل فقراتها، إذ يعاب على الاختبارات والمقاييس التي يتم بناؤها وفق الطرق الكلاسيكية للقياس؛ أنها تعتمد على موازنة أداء المتعلم بأداء أقرانه، وفقًا للمجموع الكلي لدرجاتهم في الاختبار. وسواء أجريت هذه الموازنة على أساس الدرجات الخام أو الدرجات المعيارية بأنواعها المختلفة، فإن هذه الدرجات تخضع لخصائص عينة الأسئلة التي يشتمل عليها الاختبار، واختلاف هذه الخصائص باختلاف عينة الأفراد المختبرين وعينة الأسئلة يؤثر - بلا شك - في صدق إجراء هذه الموازنات، وبذلك يصعب تعميم نتائج هذه الاختبارات أو الاستفادة العملية منها. فإن تدرج صعوبة أسئلة الاختبارات يتباين بتباين قدرات الأفراد الذين يخضعون لهذه الأسئلة، وقياس قدرات أو سمات الأفراد يتباين بتباين صعوبة الأسئلة، كما أن ثبات درجات الاختبار يتأثر تأثرًا كبيرًا بتباين درجات الأفراد الذين يطبق عليهم الاختبار ومستوى هذه القدرات.

لكل هذه الأسباب وغيرها فإن من الصعب إجراء العمليات الحسابية للموازنة بين مجموعات الأفراد وتقدير التحسن الذي يطرأ على أنماط سلوكهم، لذلك اهتم علماء القياس النفسي والتربوي بمواجهة

هذه المشكلات والتوصل إلى نماذج جديدة تجعل القياس موضوعيًا (علام، 2001: 19:18)

ويقصد بموضوعية القياس ألا تعتمد درجة المتعلم في الاختبار على عينة الأفراد المختبرين التي يوازن على أساسها أو على عينة الفقرات التي يتكون منها الاختبار أي: لا تكون قدره أو سمة معينة محكومة بعينة الأفراد أو بعينة الفقرات، وهذا يعني أن تتحرر درجة الفرد من التقيد بأداة قياس معينة، وكذلك أن تتحرر من الانتساب إلى أداء مجموعة معينة من الأفراد.

فلكي يكون القياس موضوعيًا ينبغي ألا يتأثر باختلاف الأداة المستخدمة طالما أنها أداة قياس مناسبة، وألا يتأثر أيضًا بالعناصر التي استخدمت هذه الأداة في تقديرها، وأن تدرج هذه الأداة بوحدة قياس مطلقة ثابتة تتوافق مع تدرج مستويات المتغير المراد قياسه، وهذا ما نراه مألوفًا في مجال قياس الظواهر الفيزيائية. في القياس النفسي والتربوي فإن الأمر يختلف، فكثيرًا ما يجد المكلف بعملية القياس أن درجات الطالب تختلف باختلاف الاختبار الذي يقيس المحتوى ذاته، كما أن مركز الطالب بالنسبة لأقرانه يختلف باختلاف العينة وبذلك لا يعد هذا قياسًا موضوعيًا (كاظم، 1988: 2725).

وهناك عدة تساؤلات يمكن طرحها: كيف يمكن جعل القياس النفسي والعقلي والتربوي موضوعيًا؟ أي: كيف تتحرر أدوات القياس من خصائص الأفراد الذين نطبق عليهم هذه الأدوات؟ وكيف تتحرر قياس قدرات وممات الأفراد من خصائص عينة الأسئلة التي يشتمل عليها الاختبار الذي يطبق عليهم؟ (علام، 2001: 20)

ويشير عراقي (2004: 20) إلى أن علماء القياس من أمثال رايت (Wright 1974) ولورد (Lord 1978) وجروجر (Gruijter 1983) و هامبلتون (Hambleton 1984) يرون أن بناء نظام موضوعي لتحويل الملاحظات إلى قياسات يتطلب أن تتحرر تدرج أدوات القياس من خصائص الأشياء التي تقيسها هذه الأدوات، وأن يتحرر قياس الخصائص أو الأشياء من خصائص أدوات القياس التي استخدمت في هذا القياس، وهذا يتطلب أن يشترك كل من قياس الفرد وتعبير الفقرات في وحدة قياس ثابتة على المدى المتصل للمتغير، أي يكون تدرج هذا المتصل خطيًا ووحداته متساوية وبذلك يسمح هذا النظام بتعميم القياسات على أي أدوات قياس أخرى غير المستخدمة في القياس، كما يسمح بالمقارنة بين خصائص الأشياء المراد قياسها بأدوات قياس متشابهة (غير متطابقة)،

وأيضاً يسمح بتوسيع أو تضيق مدى المحتوى الذي تقيسه أدوات القياس بما يتناسب مع المواقف والأغراض المختلفة للقياس، بحيث تناسب متطلبات قياس جديدة كما هو الحال في القياس الفيزيائي. (عراقي، 2004: 31.30).

متطلبات وشروط القياس الموضوعي:

- صدق الفقرات الاختبارية في تعريف المتغير موضوع القياس تعريفاً إجرائياً.
- صدق تدرج الفقرات الاختبارية التي تقيس متغيراً معيناً؛ بحيث يمكن تمثيل هذا المتغير بواسطة خط مستقيم.
- صدق أنماط الاستجابات على الفقرات الاختبارية المتدرجة؛ بحيث يمكن تحديد موضع كل فرد على متصل المتغير موضوع القياس.
- التوافق بين تدرج الأفراد على الاختبار وخصائص هذه الفقرات، بحيث تؤدي إلى تقديرات لمستويات الأفراد لا تتقيد باختبار معين.
- ضرورة توافر قياسات خطية يمكن استخدامها لدراسة النمو، أو للمقارنة بين الأفراد والمجموعات (كاظم، 1988: 39).

ويرى الباحث أنه لكي تتحقق هذه المتطلبات في القياس النفسي والعقلي والتربوي ينبغي تطوير أدوات القياس من حيث طرق بنائها وإعدادها وأساليب تحليل وانتقاء فقراتها، وهذا يتطلب توفر قدر كبير من الإمكانيات المادية لتطوير هذه الأدوات.

نشأة نظرية الاستجابة للفقرة

يعتقد Vander Linden أنه يجب علينا تكريم Binet في المقام الأول لإدخال فكرته في تدرج مقياس. لقد تجرأ على قياس شيء لم يكن موجوداً كمتغير يمكن ملاحظته أو واضح بشكل مباشر، ومع ذلك شعر بضرورة تعيين الفقرات والأفراد على مقياس واحد لذلك، وقدّر قيم فقرات المقياس باستخدام بيانات تجريبية تم جمعها في الاختبار القبلي، ثم قيم طلابه باستخدام قيم فقرات المقياس - تماماً نفس الممارسة المتبعة من قبل أي مستخدم لـ IRT اليوم. في الواقع، كانت ثقة Binet في تدرج نطاق

فقرات اختباره قوية للغاية لدرجة أنه تجنب الفكرة الساذجة المتمثلة في الحاجة إلى اختبار موحد لتطبيق نفس الفقرات لكل طالب من طلابه. بدلاً من ذلك، تم اختيار الفقرات بشكل تكيفي. تضمن بروتوكول اختباره للذكاء مجموعة من القواعد التي تحدد بدقة كيف كان اختيار الفقرات يتم لأعلى ولأسفل على طول المقياس، حيث أصبحت فكرة الاختبار التكيفي مقبولة بشكل عام كمييار للاختبار الفعال.

في أعمال بينيت، تم دفن عدة مفاهيم كان الإحصائي الحديث قد أعاد تبنيتها. على سبيل المثال، تضمنت طريقة القياس (الترجيح) الخاصة به ضرورة التقدير لتوزيعات المجتمع. ومع أفكاره الحدسية حول كفاءة (فعالية) الاختبار، حقق ما يعرف الآن باسم التصميم الأمثل في الإحصاء. لكن، بشكل ملحوظ، لم يعترف سبيرمان بأهمية عمل بينيت. بقي وفيًا لنموذجه الخطي الذي يضم الدرجة الحقيقية ودرجة الخطأ في القياس وأصبح اب التحليل العاملي. (وربما احتاج إلى قدر كبير من وقته للجدال مع كارل بيرسون حول تصحيح معامل الارتباط لبيرسون وعدم ثبات القياس).

كما يشير علام (2000) إن جذور نظرية الاستجابة للفقرة (IRT) ترجع إلى أعمال ريتشارد سون (Richardson 1936) ولاولي (Lawley 1943-1944) ، و توك (Tucker 1946) ويعد توك أول من استخدم مصطلح المنحنى المميز للفقرة Item Characteristic Curve والذي يعد أحد المفاهيم الأساسية والرئيسية في مجال نظرية الاستجابة للفقرة، ثم قدم لاولي Lawley عددًا من التطورات النظرية الأولية في مجال IRT واهتم بدراسة العلاقة بين معالم نماذج نظرية الاستجابة للفقرة ومعالم النماذج الكلاسيكية (التقليدية) ، وتوصل إلى بعض الإجراءات الخاصة بطرق تقدير المعالم ، ولكن هناك أوجه نقد وجهت لأعمال " لاولي " وكان أهمها أن نماذج " لاولي " كلاسيكية حيث اعتمدت على فرضين هما:

- تساوي الارتباطات الداخلية للفقرة.
- التخمين ليس عاملا في أداء الاختبار.

وقام ريتشاردسون (Richardson 1936) وتوكر (Tucker 1946) بتطوير العلاقة بين معالم

النماذج التقليدية، والمعالم المرتبطة بالمنحنيات المميزة للفقرة.

وفي عام 1950 قام لازارسفيلد «Lazarsfeld» بإجراء معظم أبحاثه في مجال قياس الاتجاهات،

ويعد لازارسفيلد أول من قدم مصطلح السمات الكامنة Latent Traits .

ويرجع الفضل في تقديم أسس نظرية الاستجابة للفقرة الاختبارية للمهتمين بالقياس النفسي والتربوي

إلى العالم لورد (Lord 1953) حينما أطلق عليها اسم النظرية الحديثة في الاختبارات.

وظلت أعمال "لورد" وغيره من العلماء حتى السبعينيات من القرن الماضي لا تهم إلا عدد محدود من

خبراء القياس في الولايات المتحدة الأمريكية، وبعض الدول المتطورة أمثال رايت "Wright"

وبيرنوم "Birnbaum" وشوبين "Choppin" والعالم الدنماركي جورج راش "G.Rasch". إلا

أن التقدم السريع الذي حدث في مجال البحوث التطبيقية لهذه النظرية كان في العام 1968 حينما

نشر كل من "لورد" و"نوفيك" كتابهما: (النظريات الإحصائية لدرجات الاختبارات العقلية. ثم جاء

"بوك" و"وود" (Book & Wood 1971) بعرض كثير من البحوث المتصلة بهذه النظرية وذلك

في الكتاب السنوي لعلم النفس "Annual Review of Psychology" وكذلك سايبوفاك

وبيكر (Subkoviak & Baker, 1977) في بحث بعنوان: نظرية الاختبارات، ثم

هامبلتون "و" سواميناثان "و" كوك Hambleton, Swaminathan & Cook, 1977.

بمقال: تطورات في نظرية السمات الكامنة ، وبعد ذلك جمع "هامبلتون" و"سواميناثان"

(Hambleton, Swaminathan, 1985) أهم تطبيقات النظرية في العشرين سنة السابقة

في كتابهما (النظرية الحديثة في القياس : الأسس والتطبيقات) ولاحقاً قام العديد من علماء القياس

النفسي والتربوي في البحث والكتابة في مجال المقارنة بين النظرية الكلاسيكية والنظرية الحديثة في القياس

أمثال "كروكر" و"ألجينا Crocker, Algina 1986 وكلاين Kline 2000، حينما وصفا

العلاقة بين الإحصائيات الكلاسيكية على مستوى (الفقرة، والمفحوص)، والإحصائيات الناتجة من

تطبيق النظرية الحديثة في القياس بأنها غير واضحة.

وفي نهاية القرن الماضي وبداية القرن الحديث أصدر كل من " أمبريستون " و " رايس (Embrestson & Reise,2000)" كتابهما: " نظرية الاستجابة للفقرة للأخصائيين النفسيين " الذي وصف فيه أسس هذه النظرية، ومقارنة بينها وبين النظرية الكلاسيكية، وكذلك بعض التطورات الحديثة على نماذج هذه النظرية وتطبيقاتها، وبرامج الحاسوب المستخدمة والملائمة لكل حالة ونموذج. (علام،2000: 682).

❖ مفهوم نظرية الاستجابة للفقرة

تعد نظرية الاستجابة للفقرة طريقة مشهورة لنمذجة البيانات، بمعنى أنها تحاول نمذجة العلاقة بين المتغير غير الملاحظ (قدرة الفرد) واحتمال استجابة الممتحن صوابًا على فقرة ما (المتغير الملاحظ) ويعتمد صدق طرق نظرية الاستجابة للفقرة في جزء كبير على البيانات المستمدة من أداء الفرد. وتقوم نظرية الاستجابة للفقرة على جمع متغيرين في علاقة رياضية وتمثل نموذج احتمالي؛ نظرًا لأن طبيعة البيانات تسمح بالتنبؤ باحتمال النجاح على أي فقرة من الفقرات المدرجة، كما توفر تقديرات لقدرات الأفراد بوحدة النموذج المستخدم (إسماعيل، 2007: 13).

وإن الفكرة الأساسية لنماذج نظرية الاستجابة للفقرة تتلخص في ربطها خصائص الفقرات بمعلم مقدار السمة التي يمتلكها الفرد واحتمال إجابته عن فقرة محددة بمستوى معين من الإجابة حسب نوع الفقرة، ويقوم هذا الرابط على تحديد موقع الفرد على مقياس السمة بصرف النظر عن مجموعة الفقرات التي يجيب عنها الفرد (عودة، 1992).

❖ مسلمات نظرية الاستجابة للفقرة:

تقوم نظرية الاستجابة للفقرة على مسلمة أساسية هي أن القيمة الاحتمالية لاستجابة فرد لفقرة اختبارية ما تكون دالة لكل من السمة أو القدرة التي يفترض أن الاختبار يقيسها لدى الفرد وخصائص الفقرة التي يحاول الإجابة عنها، أي أنها تفترض أن هناك دالة احتمالية تربط بين بارامترين أحدهما يتعلق بالفرد والآخر يتعلق بالفقرة التي يختبرها، وتهدف النظرية للتوصل إلى قيم تقديرية لكل من هذين البارامترين (Linden & Hambleton, 1997: 5).

ويذكر الوليلي(2002) أن نظرية الاستجابة للفقرة تقوم على بعض المسلمات التي يجب أن يدركها

كل مستخدم لهذه النظرية أو لأحد النماذج المتعلقة بها، ومن أهم هذه المسلمات ما يلي:

- يمكن التنبؤ بأداء أي مختبر في أي اختبار بواسطة مجموعة من العوامل يطلق عليها سمات أو قدرات

كامنة

- يمكن وصف العلاقة بين أداء المختبر على أي فقرة اختبارية ومجموعة السمات أو القدرات الكامنة التي

يفترض أنها تؤثر في أدائه على هذه الفقرة بدالة تزايدية (تزايد طردي) ويطلق عليها دالة خصائص

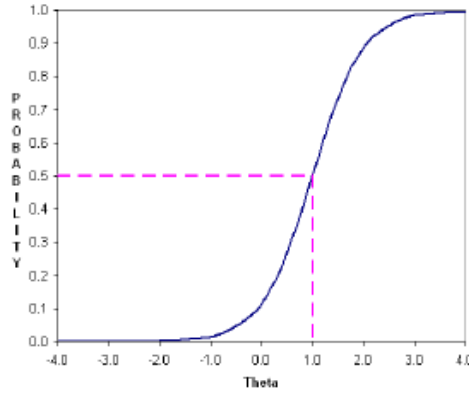
الفقرة؛ لأن هذه الدالة تحدد المختبرين الذين حققوا درجات مرتفعة في السمات التي لها توقعات

احتمالية عالية للإجابة الصحيحة للفقرة من المختبرين الذين حققوا درجات منخفضة على

السمات(الوليلي، 2002: 48)

ويمكن تمثيل هذه العلاقة بالشكل التالي:

الشكل 1: المنحنى المميز للفقرة



ويلاحظ من الشكل (1) العلاقة بين احتمال الإجابة الصحيحة $P(\theta)$ لفقرة ما ومقياس القدرة

(Ability) ويعرف بأنه المنحنى المميز للفقرة، وكل فقرة في الاختبار يكون لها منحنى مميز خاص

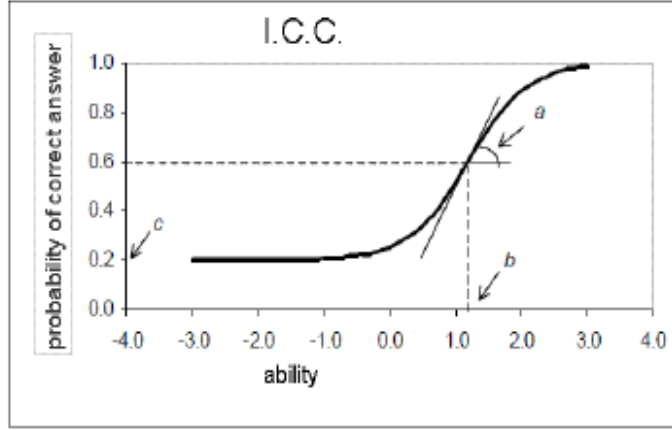
بها. واحتمال الإجابة الصحيحة تقترب من الصفر عند مستويات منخفضة من القدرة. ويزيد احتمال

الإجابة الصحيحة عند المستويات العالية للقدرة حتى تصل احتمال الإجابة الصحيحة إلى واحد صحيح

(Baker 2001)

ويأخذ المنحنى المميز للفقرة في حالة الفقرة ثنائية الدرجة الشكل التالي:

الشكل 2: المنحنى المميز لفقرة ثنائية الدرجة



حيث يرمز:

θ : قدرة الفرد

b_i : صعوبة الفقرة

$P_i(\theta)$ احتمالية أن يجيب المفحوص الذي قدرته θ على الفقرة i بطريقة صحيحة.

ويعد مفهوم المنحنى المميز للفقرة من المفاهيم المهمة في نماذج الاستجابة للفقرة التي تتناول سمة كامنة

متصلة يتم قياسها بفقرات اختبارية من النوع الثنائي أي: تتطلب إجابة صحيحة أو إجابة خطأ. ويمثل

هذا المنحنى بدالة رياضية تربط بين احتمال نجاح الفرد في إجابة الفقرة، وبين السمة أو القدرة التي

تقيسها مجموعة فقرات أو يقيسها اختبار معين . أي أنها دالة انحدار درجة الفقرة على السمة الكامنة

التي يقيسها الاختبار، وهذه الدالة غير خطية "Non-Linear" (علام، 2000: 690)

❖ افتراضات نظرية الاستجابة للفقرة:

تتميز نظرية الاستجابة للفقرة بعدة افتراضات التي ينبغي أن تتوفر في البيانات المتعلقة بالاختبارات، حيث يتم اختيار النموذج الملائم للبيانات وفقًا لمدى تحقق هذه الافتراضات في البيانات، حيث إن ذلك يؤدي إلى دقة وإحكام القياس وما يستتبعه من أحكام وقرارات وفيما يلي توضيح للافتراضات الأساسية التي تركز إليها نماذج نظرية الاستجابة للفقرة:

1. أحادية البعد Unidimensionality

يرى وبيبرغ Wiberg2004 أن أحادية البعد تشير إلى الافتراض بأن الاختبار يقيس قدرة كامنة واحدة، وهي تعني وفق ما يراه أورلاندو Orlando(w.d.) أن التغير بين الفقرات يمكن أن يفسر بواسطة بعد واحد يقف خلف ذلك التغير. وهذا يتفق مع أن المفحوص ذو القدرة العالية يفترض أن تكون احتمالية إجابته على الفقرة بشكل صحيح عالية أيضًا، والعكس صحيح (1980 Lord) و(كاظم، 1988ب)

ويثير دودين (2004) نقطة تطبيقية في غاية الأهمية، وهي أن العديد من الاختبارات عادة ما تحتوي على عدة مجالات، والسؤال كما يثيره دودين: كيف ندعي عندئذ في مثل هذا الوضع أن الاختبار أحادي البعد في حين أنه يقيس عدة مجالات؟ ويجب دودين نفسه قائلًا: "والحقيقة أنه لا تناقض في ذلك إذا ما أخذ في الاعتبار أن هناك عاملاً مشتركاً أساسياً بين هذه المجالات؛ وهي قياسها للسمة المراد قياسها.

ثم يضيف دودين ما اقتبسه من سميث (Smith) 1996 حيث قال: "من المهم التفريق بين أحادية البعد النظرية والعملية. فمن الناحية النظرية فإن مسائل الحساب تتكون من أربع مكونات أو أبعاد مثلاً هي: الجمع والطرح والضرب والقسمة، والتحليل العاملي لاختبار الحساب يمكن أن يظهر هذه البناءات أو الأبعاد الأربعة. ولكننا من الناحية العملية نفترض أن هذه البناءات الأربعة تقيس فعلياً موضوعاً واحداً وهو القدرة الحسابية. وفي الواقع فإن من الصعب إيجاد اختبار حساب لا يجمع بينها. (دودين، 2004: 115).

2. الاستقلال المحلي: Local Independence

يرى الدوسري أن خاصية الاستقلال المحلي تعني أن الإجابة على فقرة ما مستقلة إحصائيًا عن الإجابة عن أي فقرة أخرى في الاختبار الدوسري (2001: 50).

والمفهوم الأدق كما عرفه هامبلتون وآخرون (Hambleton et al.) بأنه: "عند تثبيت القدرة المسؤولة عن أداء المفحوصين على الاختبار ، فإن استجابات المفحوصين على أي زوج من الفقرات تكون مستقلة إحصائيًا (Hambleton et al 199:10) ويرى أولاندو (Orlando(w.d.) بأن هذا الافتراض ، فنيًا يعتبر ، امتدادًا للافتراض الأعم وهو أحادية البعد ، وهو يعني أنه إذا تم إزالة أثر العامل أو البعد أو السمة الكامنة خلف المقياس ، فلن يوجد أي تباير منتظم إضافي بين الفقرات .وبذلك فإن الاستقلال المحلي يمكن أن يظهر بين المجموعات الجزئية من الفقرات التي تسير في اتساق واحد (Orlando , w.d.) .

3. عامل السرعة في الإجابة Speededness :

هناك افتراض أساسي عام لجميع نماذج نظرية الاستجابة للفقرة المستخدمة، وهو أن الاختبار الذي يسعى النموذج لمطابقة بياناته لم يتم تطبيقه تحت ظرف السرعة، بمعنى أن المفحوصين الذين أخفقوا في الإجابة على فقرات الاختبار لم يكن ذلك بسبب إخفاقهم في السرعة الكافية لإنجاز الاختبار ، وإنما يعود ذلك إلى محدودية قدراتهم (Hambleton & Swaminathan, 1985) (علام، 1986)

ويشير هامبلتون وسواميناثان Hambleton and Swaminathan 1985 إلى أنه من النادر الإشارة إلى هذا الافتراض؛ لأنه قد يكون متضمنًا في افتراض أحادية البعد، فعندما تؤثر السرعة في الأداء على الاختبار، فإن هذا يعني أن هناك سمتين على الأقل تؤثران في الأداء هما : سرعة الأداء والسمة الأخرى التي يسعى الاختبار لقياسها.

الافتراضات السابقة التي تمت مناقشتها هي افتراضات عامة لجميع نماذج نظرية الاستجابة للفقرة.

4. تساوي مؤشرات التمييز Equal Discrimination Indices

يرى الدوسري والمرشد(2004) أنه يرتبط بمشكلات التحقق من فرضيات النظرية مسألة اختيار النموذج المناسب لتحليل البيانات الاختبارية، بمعنى هل يتم استخدام النموذج ذي الثلاثة معالم أو المعلمتين أو المعلمة الواحدة؟ (الدوسري والمرشد، 2004).

الخطوة الأولى لتحديد أنواع المعالم التي يمكن تضمينها في النموذج تعتمد على التحليل وفق النظرية الكلاسيكية فعلى سبيل المثال: معلمة التمييز (الميل Slope) في نظرية الاستجابة للفقرة والتي يرمز لها بالرمز (a_i) تناظر ارتباط الفقرة بالدرجة الكلية في النظرية الكلاسيكية (Reeve, w.d.).

ويقصد بتساوي مؤشرات التمييز، أن تكون جميع المنحنيات المميزة ل فقرات الاختبار متوازية (متساوية الميل) وتختلف عن بعضها البعض في نقط التقائها بالمحور الأفقي الذي يمثل متصل القدرة أو السمة الكامنة.

5. الحد الأدنى من أثر التخمين Minimal Guessing :

- أما معلمة التخمين (c) فيمكن فحص مدى مناسبة تضمينها في النموذج المقترح، وبذلك يمكن استخدام النموذج 3 PLM، حيث إن النموذج ذا المعالم الثلاثة 3 PLM له معلمة خاصة بالتخمين، وهي عبارة عن المقدار الذي يقترب فيه المنحنى من نقطة الصفر.
- وهذا يعني أن معلمة التخمين يجب أن تكون جزءًا من النموذج ، خاصة إذا كانت الفقرات هي الاختيار من متعدد فإن إمكانية التخمين تزداد بشكل عام (Wiberg,2004).
- إلا أن سبب عدم انتشار مبادئ وأسس هذه النظرية انتشارًا واسعًا في بداية ظهورها يرجع لما يلي:
- تمثل هذه النظرية فرعًا معقدًا من فروع نظرية الاختبارات، ويحتاج دارسها ومطبقها إلى قدر كبير من الفهم في الرياضيات العالية، والموضوعات الإحصائية، وعلى وجه الخصوص الإحصاء متعدد المتغيرات وإحصاء بيز.
 - معظم بحوث هذه النظرية وجهت لمنظري القياس، وليس لمن سيقومون بتطبيق هذه النظرية في بناء الاختبارات، وتحليل وتفسير نتائجها وغيرها من التطبيقات النفسية والتربوية.

- صعوبة تحليل البيانات عن طريق برامج الحاسوب الفعالة والمناسبة لتقدير معالم نماذج السمات الكامنة التي تتضمنها هذه النظرية، مثل النموذج الأحادي والثنائي والثلاثي المعلم، إضافة إلى النموذج المتعدد.
- صعوبة تحقق بعض فروض نماذج النظرية المتعلقة بالبيانات الفعلية المستمدة من الاختبارات العقلية والتحصيلية.

إلا أنه، وحديثاً، اتجه علماء القياس للتغلب على معظم هذه الأسباب التي أدت لعدم انتشار هذه النظرية انتشاراً واسعاً، حيث أجريت الدراسات لتحويل النماذج من الجانب النظري إلى الجانب العملي التطبيقي، وتبسيطها لتمكين المهتمين من التطبيق دون عناء، كما أعدت البرامج الحاسوبية لتقدير معالم النموذج مثل برامج:

Logist, Bical, Quest,Parascale 3.2, Big Steps, Win Steps, Dicot, Multilog, Microscale, Dicot, Rumm 2010, Bilog Mg3...
JMetrik, JASP, Bical : Logist, وهذه مدفوعة، وعناك برامج مجانية اثبتت فعاليتها مثل JMetrik, JASP, Bical : Logist .R.

كما اتجهت كثير من المؤسسات والهيئات الأمريكية المسؤولة عن إعداد الاختبارات ونشرها، لاستخدام نماذج النظرية في تصميم وبناء مختلف أنواع الاختبارات والمقاييس العقلية والتحصيلية، وموازن التقدير، ومقاييس الاتجاهات، كما واستعانت بها إدارات التقويم بمختلف الولايات، في إعداد الاختبارات في المواد الدراسية المختلفة، فيما حذت كل من كندا، وإنجلترا وأستراليا، ونيوزلندا، حذو الولايات المتحدة الأمريكية في هذا الأمر (علام، 2000: 684).

❖ نماذج نظرية السمات الكامنة:

لقد أدت الجهود المتواصلة للمهتمين بنماذج الاستجابة للفقرة إلى التوصل إلى مجموعة من، النماذج السيكو مترية الجديدة تعرف بنماذج الاستجابة للفقرة Item Response Models وتتلخص الفكرة الأساسية لهذه النماذج في أنها تحاول اشتقاق قيم تقديرية للسمة أو السمات التي تتضمنها

الاستجابة لمجموعة من فقرات الاختبار، بمعنى أن هذه النماذج تقدم تفسيراً لاستجابة الفرد لفقرة الاختبار التي تنطوي على السمة التي تقيسها هذه الفقرة، والسمة المقاسة قد تكون قدرة معينة أو خاصية من خصائص الفرد، بحيث تكون هناك علاقة منتظمة بين مستويات السمة المقاسة لدى أفراد مختلفين وبين احتمالات الاستجابة الصحيحة لفقرات الاختبار (علام، 2000: 686)

إن الهدف الأساسي من أي نموذج ينتج من أي نظرية اختبارية بصفة عامة هو استخدامه (أي النموذج) في التدرّج (Scaling)، أي إعطاء قيم للأفراد على التدرّج (Scale values) ولن يتحقق ذلك إلا إذا حدد النموذج العلاقة بين أداء الفرد، والقدرة التي تكمن خلف هذا الأداء، وعلى هذا فإن نموذج السمة الكامنة يحدد العلاقة بين البيانات (أداء الفرد على الاختبار) وقيم التدرّج أو الدرجات على تدرّج السمة الكامنة، وكلما كان هناك ملائمة (Fit) بين النموذج المستخدم ومجموعة البيانات أدى ذلك إلى الحصول على تقديرات دقيقة للسمة المقاسة، حيث يجب وضع كل من الأفراد والفقرات على مقياس تدرّج السمة، ويتم ذلك من خلال عمليات التقدير طالما توجد علاقة تقارب ممكنة بين الاحتمالات المتوقعة للممتحنين، والاحتمالات الواقعية (الفعلية) لأدائهم في كل مستوى من مستويات السمة المقاسة، وتتميز هذه النماذج بأنها دوال احتمالية Probabilistic وليست دوال حتمية Deterministic وهذا يعني أن العلاقة بين أداء الفرد والسمة التي يتم قياسها تسير وفقاً لنظرية الاحتمالات (علام، 1982: 40).

إن استخدام الاحتمالات أقرب إلى طبيعة العلم والظواهر السلوكية بصفة خاصة ولا يتنافى ذلك مع دقة القياس، فتلك النماذج الرياضية هي وسائل قياس دقيقة يمكن الاعتماد عليها بدرجة عالية من الثقة في المجال النفسي والعقلي والتربوي، وتختلف النماذج فيما بينها في الشكل الرياضي للدالة المميزة للفقرة، كما تختلف في عدد البارامترات الواجب تحديدها في كل نموذج للوصول للعلاقة السابقة، فكل نموذج يحتوي على واحدًا أو أكثر من البارامترات التي تصف الفقرة وواحد أو أكثر من البارامترات التي تصف الفرد. (علام 2000: 698).

❖ انواع نماذج نظرية الاستجابة للفقرة:

توجد نماذج متعددة في إطار نظرية الاستجابة للفقرة، وذلك لاختلاف الافتراضات المتعلقة بالبيانات الاختبارية. إذ ربما يفترض أن الأداء في الاختبار ينطوي على سمة أحادية البعد، أو على سمات متعددة. أو ربما يفترض أن الفقرات تتباين في قدرتها على التمييز بين المستويات المختلفة للقدرة، أو أن الإجابات الصحيحة على فقرات الاختبار من متعدد تتأثر بالتخمين العشوائي أو لا تتأثر كما في الأسئلة المفتوحة، أو عما إذا كانت الدرجات ثنائية أم متعددة الأقسام. لذلك يمكن تقسيم هذه النماذج إلى:

✓ نماذج سكونية Static models

ويهتم هذا النوع من النماذج بالقياس في مدة زمنية واحدة، وكذلك بتحديد العمليات التي ينطوي عليها الأداء في الاختبارات السيكولوجية والتربوية. وتختلف هذه النماذج في صيغها الرياضية التي تربط أداء الفرد في الاختبار بدرجات قدرته، كما تختلف في عدد الخصائص أو البارامترات المستخدمة في وصف كل فقرة من فقرات الاختبار. (علام، 2005: 67).

ومن أهم النماذج السكونية ثلاثة نماذج أساسية شائعة الاستخدام وهي:

1. النموذج اللوغاريتمي أحادي المعلم

ويسمى بنموذج "راش" Rasch Model "نسبة إلى عالم الرياضيات" جورج را ش "بجامعة كوبنهاجن" (1960) وهو من أبسط نماذج نظرية الاستجابة للفقرة ويفترض أن استجابة كل فرد تكون إما 1 أو صفر على كل فقرة من فقرات الاختبار (المرجع السابق ، 2005: 69).

كما يفترض النموذج تساوي معاملات التمييز (a_i) وانعدام التخمين (c_i) ويقوم بتقدير صعوبة الفقرات (b_i) فقط ، وتمثل الصيغة الرياضية لهذا النموذج فيما يلي:

$$P(X_{ij} = 1 | \theta_i, \delta_j) = \frac{\exp(\theta_i - \delta_j)}{1 + \exp(\theta_i - \delta_j)}$$

i: رقم الفقرة 1، 2، 3.....i

$P_i(\theta)$: احتمالية أن يجيب المفحوص ذو القدرة (θ) على الفقرة i اجابة صحيحة.

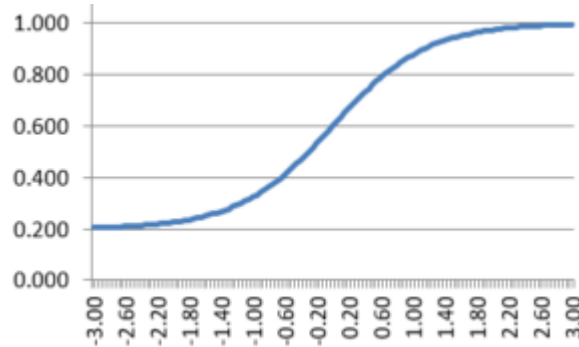
b_i : صعوبة الفقرة

e : الأساس اللوغاريتمي ويساوي 2.718 تقريبا

D : معامل القياس أو التدرج Scaling Factor وهو مقدار ثابت يساوي 1.7 ، وبحول هذا

المعامل المنحنى اللوغاريتمي تطابق تقريبي مع المنحنى ألعندالي ((Fan 1998: 369 .

ويمكن تمثيل الدالة التي تعبر عن نموذج "راش" بالشكل التالي:



شكل(3):التمثيل البياني للنموذج اللوغاريتمي أحادي المعلم(نموذج راش).

ويتضح من الشكل (3) أن المنحنيين متوازيان ويختلفان فقط في إزاحة الموقع أي في الصعوبة، حيث

إن الفقرة (2) تبدو أكثر صعوبة من الفقرة (1) وتطبيق الصيغة الرياضية لنموذج "راش" يتطلب تساوي

تمييز جميع فقرات الاختبار، أي أن تكون جميع المنحنيات المميزة

لفقرات الاختبار متوازية (متساوية الميل) وتختلف عن بعضها البعض في نقط التقائها بالمحور الأفقي

الذي يمثل متصل القدرة أو السمة الكامنة، وهذا يعني أن الفقرات تختلف فقط في صعوبتها، وهو

البارامتر الوحيد في نموذج "راش" (علام2000: 70)

وكما يرى أورلاندو Orlando فإن نموذج راش وإن كانت له إيجابيات بسبب بساطته وسهولة

حساباته، إلا أن النماذج الأخرى والتي تقدر ميل Slope مختلفة أي أن لها معالم تمييز مختلفة لكل

فقرة غالبًا ما تعكس البيانات بشكل أدق.(Orlando , w.d.) .

2. النموذج اللوغاريتمي ثنائي المعلم: Two –Parameter Logistic Model

ويسمى بنموذج "لورد" Lord Model وفيه يتم إضافة معلم جديد إلى نموذج راش وهو معلم

التمييز (a_i) لكل فقرة، وهذا يؤدي إلى تقاطع المنحنيات المميزة لفقرات الاختبار الذي يصمم وفق هذا

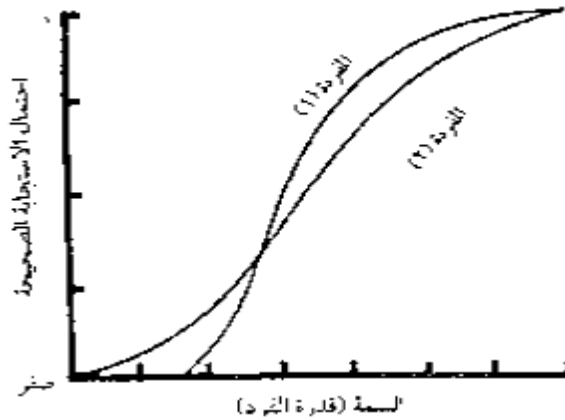
النموذج، أي أن هذا النموذج يفترض انعدام التخمين (c_i) ويقوم بتقدير صعوبة الفقرات (b_i) ومعاملات التمييز (a_i)، وتتمثل الصيغة الرياضية لهذا النموذج فيما يلي

$$P(\theta) = \frac{1}{1 + \exp[-D a (\theta - b)]}$$

حيث:

a_i هو تمييز الفقرة .

ويمكن تمثيل الدالة التي تعبر عن نموذج "لورد" بالشكل التالي:



شكل (4): التمثيل البياني للنموذج اللوغاريتمي ثنائي المعلم (نموذج لورد)

ويلاحظ من الشكل (4) الذي يوضح التمثيل البياني لنموذج "لورد" أن المنحنيات المميزة للفقرات المختلفة تتقاطع. وقد أضاف "لورد" معلم التمييز لكل فقرة إلى نموذج "راش" لأنه من الصعب إيجاد مجموعة من الفقرات تميز بدرجة واحدة بين مستويات القدرة أو السمة التي يقيسها الاختبار، وهو الافتراض الذي استند إليه نموذج "راش" (علام:2001: 188)

ويفتقر هذا النموذج إلى بعض الخصائص الإحصائية التي يتميز بها نموذج "راش"، لذلك فإن عملياته الحسابية أكثر صعوبة (71: 2007 Kang & Cohon)،.

3. النموذج اللوغاريتمي ثلاثي المعلم [Three –Parameter Logistic Model]

ويسمى بنموذج " بير نبوم Birnbaum Model " حيث أضاف " بيرنبوم " معلماً ثالثاً أطلق عليه معلم الخط التقاربي الأدنى، أو معلم التخمين Guessing Parameter فيما يتعلق بالفقرات الاختبارية التي تتطلب الاختيار من بدائل متعددة أو الصواب والخطأ ويرمز لهذا المعلم بالرمز (c_i) وتمثل الصيغة الرياضية لهذا النموذج فيما يلي:

(Salvia&Ysseldyke,1995,721)

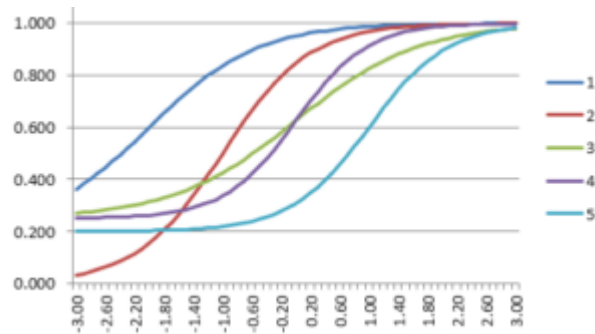
$$P(\theta) = c + \frac{1-c}{1 + \exp[-Da(\theta - b_i)]}$$

حيث:

c_i هو معلمة تأثير التخمين.

ويمكن تمثيل الدالة التي تعبر عن نموذج " بيرنبوم " بالشكل التالي:

شكل (5): التمثيل البياني للنموذج اللوغاريتمي ثلاثي المعلم (نموذج بيرنبوم)



ويلاحظ من الشكل (5) أن معلم التخمين يمثل بالخط الذي يلتقي بالطرف الأسفل للمنحنى المميز لكل فقرة في اللانهاية ، وهو يسمى بالخط التقاربي . ولذلك تتقاطع المنحنيات المميزة لفقرات الاختبار ، أي تختلف عن بعضها البعض في كل من الجزء الذي تلتقي فيه بالمحور الأفقي الذي يمثل متصل

السمة الكامنة وهو بارامتر الصعوبة، والميل يمثل بارامتر التمييز، والخط التقاربي الأدنى هو الخط الذي يلتقي بالطرف الأسفل للمنحنى في اللانهاية ، ويمثل بارامتر التخمين(علام،2005: 74.73)

✓ نماذج ديناميكية Dynamic Models

تتم هذه النماذج بقياس التغير الحادث في السمات الإنسانية عبر الزمن . فالبعض يرى أن هذا التغير يعد عملية تدريجية ، بينما يرى البعض الآخر أنه عملية تحول من حالة إلى أخرى ، وفي الحالة الأولى يكون المتغير الذي ينطوي عليه التغير متصلا ، بينما في الحالة الثانية يكون المتغير منفصلا ، غير أن الحالة الأولى هي التي نالت الاهتمام في النماذج الديناميكية(علام،2005: 84)

معايير اختيار النموذج المناسب للبيانات:

من أجل فحص وتقييم النماذج المختلفة وصولا للنموذج الأمثل لمطابقة البيانات، اقترح هامبلتون وسواميناثان(Hambleton & Swaminathan 1985) وهامبلتون وآخرون (1991) Hambleton et al. ثلاث معايير لتقييم نماذج نظرية الاستجابة للفقرة هي :

أولا – التحقق من افتراضات النموذج.

ثانياً – التحقق من الخصائص المتوقعة للنموذج.

ثالثاً – التحقق من قدرة النموذج على التنبؤ بالدرجات الفعلية.

أولاً: التحقق من افتراضات نموذج الاستجابة للفقرة

أحادية البعد: Unidimensionality

هناك أساليب عديدة يمكن استخدامها لفحص افتراض أحادية البعد لبيانات الاختبار .أحد هذه الأساليب هو فحص معامل الثبات فمعامل الثبات ألفا كرونباخ هو مؤشر للاتساق الداخلي الذي يشير إلى أحادية البعد (الحري،2006، Wiberg 2004) ومع ذلك فهناك أساليب أكثر تحديداً، تساعد في فحص تحقق افتراض أحادية البعد في البيانات المشاهدة من أهمها التحليل العاملي.

ويتم فحص خاصية أحادية البعد في التحليل العاملي من خلال تفحص قيم الجذور الكامنة Eigen values فإذا أعطى التحليل العاملي عاملاً واحداً مميزاً والعديد من العوامل الصغيرة، بحيث كانت قيمة الجذر الكامن لذلك العامل تشكل نسبة واضحة ومرتفعة من التباين الكلي للدرجات، فإن ذلك يعني أن هناك عاملاً واحداً مهيمناً في الاختبار مما يشير إلى تحقق افتراض أحادية البعد.

ويشير دودين (2004) نقلاً عن ريكاس Reckase أن التحليل العاملي الذي يفسر العامل الأول فيه 20 % من التباين الكلي للدرجات على الأقل يعتبر أحادي البعد. (دودين، 2004)

لكن هامبلتون "Hambleton" أثار ملاحظة تطبيقية في غاية الأهمية، وهي أن الحصول على عامل يفسر نسبة أقل من 10 % من التباين الكلي للدرجات أمر ليس مستغرباً، ولذا فإنه طالما كان ذلك العامل أكبر بشكل واضح ومميز من بقية العوامل، فإن ذلك يعني صحة افتراض أحادية البعد لذلك الاختبار (Wiberg 2004)

كما يمكن استكشاف ما سبق من خلال فحص الرسوم Scree Plot والتي هي عبارة عن تمثيل لقيم الجذور الكامنة لجميع العوامل الممكن استخلاصها من الاختبار. فإذا اتضح من فحص تلك الرسوم أن التغير في قيمتي الجذور الكامنة بين العامل الأول والعامل الذي يليه لافت للنظر وسريع، مقارنة بالتغير التدريجي والبسيط بين العوامل التي تأتي بعد ذلك، فإن هذا يشير إلى أن هناك عامل واحد مسيطر ومهيمن على درجات الاختبار، أي أن افتراض أحادية البعد متحقق.

والطريقة الأخرى للكشف عن أحادية البعد باستخدام التحليل العاملي هي تفحص معاملات تشبع الفقرات على العامل الأول. فإذا كانت جميع الفقرات متشعبة على العامل الأول فهذا يعني تحقق افتراض أحادية البعد.

ويرى أورلاندو "Orlando" بأن وجود فقرة واحدة أو أكثر غير متشعبتين على العامل الأول لا يعني عدم توفر خاصية أحادية البعد، ومع ذلك يجب فحص نتائج نموذج نظرية استجابة الفقرة من أي شذوذ ربما يظهر بسبب انتهاك هذا الافتراض، مثل أن يظهر بأن هناك فقرة واحدة أو أكثر لهما معالم تمييز منخفضة جداً. (Orlando, w.d.).

1. الاستقلال المحلي: Local Independence

يعتبر هذا الافتراض امتداداً للافتراض الأعم وهو أحادية البعد، وهو يعني أنه إذا تم إزالة أثر العامل أو البعد أو السمة الكامنة خلف المقياس، فلن يوجد أي تغاير منتظم إضافي بين الفقرات. وبذلك فإن الاستقلال المحلي يمكن أن يظهر بين المجموعات الجزئية من الفقرات التي تسير في اتساق واحد، بحيث يكون بمقدور الباحث أن يختبر إحصائياً أحد الافتراضين نيابة عن الآخر. ويمكن التحقق من هذا الافتراض أمبريقياً عن طريق تثبيت القدرة عند مستويات مختلفة وحساب معامل (F) لكل فئة من مستويات القدرة، حيث يفترض أن معاملات الثبات (F) تقترب عند مستوى ثابت من القدرة من الصفر. (Nandakumar & Ackerman, w.d.).

2. عامل السرعة في الإجابة: Speededness

يرى كل من هامبلتون وسواميناثان (Hambleton & Swaminathan, 1985) وعلام (1986) بأنه يمكن التحقق ما إذا كان الاختبار هو اختبار قوة أو سرعة من خلال حصر عدد المفحوصين الذين لم يتمكنوا من إكمال الاختبار موضع التطبيق

3. تساوي مؤشرات التمييز: Equal Discrimination Indices

يمكن التحقق من تساوي مؤشرات التمييز من خلال فحص ذلك النوع من الارتباط بين الفقرة والدرجة الكلية للاختبار والمعروف بمعامل الارتباط ثنائي التسلسل الحقيقي Point-biserial correlation فإذا كانت قيمة الانحراف المعياري لمعاملات الارتباط الثنائي التسلسل الحقيقي للفقرات صغيراً وتقترب من الصفر، فإن معنى ذلك أن مؤشرات التمييز متساوية في جميع الفقرات. (Wiberg 2004) فإذا تحقق مثل الوضع السابق، فإن هذا يعني أن النموذج سوف يكون أحادي المعلمة أي نموذج راش. (Rasch Model) (كاظم، 1988: 42).

أما إذا اتضح من فحص قيم معاملات الارتباط الثنائي التسلسل الحقيقي بأنها ليست متساوية فإن من الأفضل استخدام النموذج ثنائي المعلم 2PLM بدلا من النموذج أحادي المعلم (1991 Hambleton et al.)

4. الحد الأدنى من أثر التخمين Minimal Guessing:

يمكن التحقق ما إذا كان هناك أثر للتخمين (c) أم لا ، وذلك من خلال مؤشرات معاملات الصعوبة المحسوبة وفق النظرية الكلاسيكية (التقليدية) فإذا اتضح أن هناك فقرات صعبة ، فإن المفحوصين غالباً ما سيلجئون إلى التخمين ، وبذلك يمكن استخدام النموذج 3PLM، حيث أن النموذج ذا المعلم الثلاثة له معلمة خاصة بالتخمين ، والتي هي عبارة عن المقدار الذي يقترب فيه المنحنى من نقطة الصفر (Wiberg, 2004).

أما إذا كانت كل الفقرات سهلة نسبياً ، أو إذا كان الاختبار يتكون من فقرات تكون الاستجابة عليها بإجابات مفتوحة وقصيرة فإن من غير المحتمل أن يكون النموذج المختار هو نموذج المعلم الثلاثة. (Hambleton et al. 1991).

أما الطريقة الأخرى للتحقق من الحد الأدنى للتخمين فإنها تعتمد على فحص أداء المفحوصين ذوي القدرات المنخفضة، لأن هؤلاء المفحوصين هم الذين يلجئون لهذا السلوك عندما يواجهون الفقرات الصعبة (Leeson & Fletcher, w.d.).

ويتم ذلك التحليل من خلال اختبار 10% من المفحوصين الأقل قدرة على الدرجة الكلية للاختبار، ومن ثم دراسة أدائهم على الفقرات الأكثر صعوبة، ومقارنة نسب أولئك المفحوصين الأقل قدرة، الذين أجابوا إجابة صحيحة على تلك الفقرات الصعبة بالقيم النظرية، إذا

حدث أن خمن المفحوصين بطريقة عشوائية على تلك الفقرات، مع ملاحظة أن تلك القيم النظرية للتخمين تعتمد على عدد فئات الاستجابة لكل فقرة (Hambleton & 1985)

Swaminathan,

ويرى ويرغ بأن نسبة إجابة المفحوصين الأقل قدرة على الفقرات الصعبة إذا كانت قريبة من القيمة النظرية أو أكبر منها، فهذا يعني أن معلمة التخمين يجب أن تكون جزءًا من النموذج (Wiberg 2004).

ثانياً: التحقق من الخصائص المتوقعة للنماذج:

إن هذا المعيار للخصائص المتوقعة للنموذج هو معيار عام، فأيا كان نوع النموذج المستخدم، فإن هناك خاصيتان مهمتان لنماذج نظرية الاستجابة للفقرة يمكن فحصهما هما: خاصية الثباتية لمعلم القدرة، Invariance of ability parameter وخاصية الثباتية لمعلم الفقرات Invariance of Item parameter

تعني خاصية الثباتية لمعلم القدرة، أن تقديرات القدرة للمفحوصين يفترض ألا تعتمد على خصائص الفقرات أي على صعوبة أو سهولة الفقرات (Hambleton & al 1991).
بينما تعني خاصية الثباتية لمعلم الفقرات أن تقديرات معلم الفقرات يفترض ألا تعتمد على طبيعة وخصائص المفحوصين، سواء كانت قدراتهم مرتفعة أم منخفضة، أو مهما كان تصنيفهم داخل العينة الكبيرة للدراسة. (Wiberg.2004)

ويتم التحقق من الخاصيتين السابقتين على النحو التالي:

1. التحقق من ثباتية معلم القدرة:

يمكن التحقق من خاصية الثباتية لمعلم القدرة من خلال تقسيم فقرات الاختبار إلى فقرات سهلة وأخرى صعبة بناء على معاملات صعوبة الفقرة المستخرجة من النظرية الكلاسيكية (التقليدية)، ومن ثم يتم تقدير قدرات جميع المفحوصين في مجموعتي الفقرات السهلة والفقرات الصعبة، ثم تفحص العلاقة بين تقديرات قدرات المفحوصين في المجموعتين المختلفتين في خصائص الفقرات (معاملات الصعوبة).
فإن التمثيل البياني لرسم الانتشار بين تقديرات معلم القدرة المفحوصين بناء على مجموعتي الفقرات السهلة والصعبة يمكن أن يعكس مدى تقارب أو اختلاف التقديرات. فإذا لم يتم الحصول على قيمة

جيدة لمعامل الارتباط بين التقديرات، فإن ذلك يشير إلى أن خاصية ثباتية معالم القدرة لن تكون متوفرة عند نمذجة البيانات بذلك النموذج (Hambleton et al., 1991) (Wiberg, 2004).

2. التحقق من ثباتية معالم الفقرات:

ويتم ذلك من خلال تقسيم المفحوصين إلى مجموعتين أو فئتين؛ إحداها مرتفعة القدرة، والأخرى منخفضة القدرة، وذلك بناءً على درجاتهم الكلية المكتسبة على الاختبار (Hambleton 1985) (& Swaminathan,

وفي حالة إذا كان الاختبار محكي المرجع، فإن تقسيم المفحوصين إلى مجموعتين مرتفعة القدرة ومنخفضة القدرة يمكن أن يعمل بناءً على درجة القطع (Wiberg, 2004).

بعد ذلك يتم عمل تقدير لمعالم الفقرات (الصعوبة والتمييز والتخمين) لكل مجموعة من المجموعتين السابقتين، ثم تمثل تقديرات معالم الفقرات للمجموعتين على رسم الانتشار لملاحظة ما إذا كانت العلاقة قوية أم ضعيفة بين التقديرات إذا كانت العلاقة قوية فإن هذا يعني توفر خاصية الثباتية لمعالم الفقرات، بمعنى أن تقديرات معالم فقرات الاختبار لم تعتمد على خصائص عينة المفحوصين وتحدد الإشارة إلى أن الفحوصات السابقة لثباتية معالم القدرة وثباتية معالم الفقرات يتم إجراؤها في هذه الخطوة على النماذج الثلاثة المقترحة لتوفيق البيانات، بحيث يتمكن فاحص الاختبار من مقارنة درجة توفر خصائص الثباتية في النماذج الثلاثة (Leeson & Flecher, w.d.) (Wiberg 2004)

ثالثاً : التحقق من قدرة النموذج على التنبؤ بالدرجات الفعلية للاختبار:

إن جميع تطبيقات نظرية الاستجابة للفقرة تفترض ببساطة أن النموذج المستخدم هو نموذج صحيح، وفائدة وجدوى أي نموذج من نماذج نظرية الاستجابة للفقرة تعتمد على المدى الذي يعكس فيه النموذج البيانات الفعلية بدقة (Orlando, w.d; Sinharay, 2005)

وهناك عدة طرق للتحقق من قدرة النموذج على التنبؤ بالنتائج الفعلية للاختبار من ذلك ما يأتي:

1. مقارنة التوزيعات المشاهدة بالتوزيعات المتوقعة:

ويمكن ذلك من خلال استخدام الرسوم البيانية لتوزيعات الدرجات المشاهدة والدرجات المتوقعة. فعلى قدر ما تكون تلك التوزيعات متقاربة أو متطابقة يكون ذلك النموذج الأفضل في مطابقة البيانات مقارنة بالنماذج الأخرى.

2. استخدام اختبار " X^2 " لاختبار جودة المطابقة:

يستخدم هذا الاختبار لفحص مدى مطابقة كل فقرة من فقرات الاختبار للبيانات المشاهدة للاختبار، وهو يقارن نسبة الإجابات الصحيحة على الفقرة (i) للمجموعة الجزئية التي قدرتها (h) مع النسبة المتوقعة للإجابات الصحيحة وفقاً للنموذج المستخدم.

فإذا كانت قيمة " X^2 " المحسوبة أكبر من القيمة الحرجة، ترفض الفرضية الصفرية، وبذلك نتوصل إلى أن منحى خصائص الفقرة (ICC) لا يلائم الفقرة موضع الاختبار، (Wiberg, 2004).

3. فحص البواقي:

يعتبر فحص البواقي (البواقي المعيارية) أفضل طريقة لفحص مطابقة البيانات، والمعادلة المستخدمة لذلك كما يلي:

$$\text{البواقي} = \text{القيم الملاحظة} - \text{القيم المتوقعة}$$

$$R_{ij} = P_{ij} - E(P_{ij})$$

وتمثل (P) احتمالية الإجابة بشكل صحيح على كل فقرة، في مقابل القدرات، وبالتحديد في مقابل نوعين من القدرات إحداها القدرات المشاهدة أو الحقيقية أو ما تسمى أيضاً التقليدية، والأخرى القدرات المتوقعة أو المتنبأ بها من خلال النموذج. كما يمكن أيضاً تمثيل البواقي المعيارية حسب فئات القدرات لكل فقرة، حيث يظهر ما إذا كانت هناك نمطية معينة للبواقي المعيارية في أحد النماذج، كما تتم ملاحظة أي من النماذج يعطي توزيعاً أكثر انتشاراً وعشوائية للبواقي المعيارية (1990،

(Hambleton et al,

4. التوزيع الطبيعي للقدرات:

يفترض أن تتوزع القدرة بشكل طبيعي وبوسط قيمته (صفر) وانحراف معياري قيمته (واحد) أي توزيع طبيعي معياري. ولذا فإن تقديرات القدرات للبيانات المشاهدة يتم تمثيلها بـمدرج تكراري، وتقارن بالتوزيع الطبيعي المعياري بحيث ينظر هل تحقق ذلك التوقع أم لا؟ (Wiberg, 2000).

5. مقارنة إحصاءات عدم المطابقة مع معاملات الارتباط الثنائي التسلسل الحقيقي للفقرات: وهي إيجاد علاقة بين إحصاءات عدم المطابقة للفقرات والارتباطات الثنائية التسلسل الحقيقية-Point biserial correlations. إحصاء عدم التطابق misfit. يمكن أن تحدد بإيجاد معدل القيم المطلقة للبواقي المعيارية التي حصلنا عليها بعد مطابقة النموذج موضع الفحص بيانات الفقرة. ويكون الوضع المفضل هنا عندما لا يكون هناك نمط محدد لانتشار البواقي المعيارية المطلقة مع الارتباطات ثنائية التسلسل الحقيقية، وعندما تكون قيمة البواقي المعيارية المطلقة أقل ما يمكن. (Hambleton et al 1991).

❖ مميزات نماذج نظرية الاستجابة للفقرة:

- يمكن تلخيص أهم مميزات نماذج نظرية الاستجابة للفقرة فيما يلي:
- إن استخدام نماذج نظرية الاستجابة للفقرة يجعل بارامترات الفقرة مستقلة عن عينة الأفراد المستخدمة في تقدير هذه البارامترات، التي تمثل المجتمع المستهدف من تصميم الاختبار، كما أن تقديرات قدرات الأفراد مستقلة عن مجموعة فقرات الاختبار المستخدمة في الحصول على هذه التقديرات، المنتقاة من مجموعة كبيرة من الفقرات التي تم تدريبها باستخدام أحد نماذج الاستجابة للفقرة. (1988: 281). (Lord & Stocking 1989: 148).
- إمكانية الحصول على مؤشر إحصائي (إحصاءات صدق المطابقة) يوضح مدى دقة تقدير قدرات الأفراد وبارامترات الفقرات. (Hambleton & Swaminathan 1985: 11).
- تحسين دقة وثبات النتائج من خلال تحديد وحذف الأفراد والفقرات غير المطابقتين للنموذج المستخدم، إضافة إلى حذف التأثيرات المتطرفة لقياسات كل من القدرة والفقرة.
- سهولة التمثيل البياني لبارامترات الفقرات وقدرات الأفراد.

- استخدام نماذج نظرية الاستجابة للفقرة يسهل دراسة كل من تحيز الأفراد، وال فقرات، وصدق الاستجابات، والفروق الفردية، واستراتيجيات الاستجابة، وكذلك سهولة إعداد بنوك الأسئلة، والاختبارات الآلية، واستخدام الحاسب الآلي في تكرار العملية الاختبارية.
- يمكن تحديد الأفراد الذين يلجؤون إلى التخمين، كما يمكن تكميم مدى استجاباتهم الشاذة، كما يمكن ترتيب هؤلاء الأفراد عن طريق جودة المطابقة لنموذج احتمالي معين.
- تقليل عدد الفقرات اللازمة للقياس الدقيق، وبالتالي توفير الزمن اللازم للإجابة على فقرات الاختبار، وإعادة التحليل عند الصور البديلة للاختبار. (Henning, 1984: 132).

❖ عيوب نظرية الاستجابة للمفردة:

- يمكن تلخيص أهم عيوب نماذج نظرية الاستجابة للفقرة فيما يلي:
- تمثل هذه النظرية فرعاً معقداً من فروع نظرية الاختبارات، واستيعاب البحوث السيكو مترية التي تجري في هذا المجال ونشرها في الدوريات العلمية المتخصصة يحتاج إلى قدر كبير من الفهم الرياضي العالي، والإحصاء متعدد المتغيرات.
- معظم خبراء القياس الذين تناولوا هذا الموضوع وجهوا دراساتهم وبحوثهم لمنظري القياس، وليس للذين سيقومون بتطبيقه في بناء الاختبارات وتحليل وتفسير نتائجها، وفي غير ذلك من التطبيقات السيكلوجية والتربوية.
- صعوبة تحقق بعض فروض النماذج المتعلقة بهذه النظرية في البيانات الفعلية المستمدة من الاختبارات العقلية والتحصيلية. (علام, 2000: 684)

❖ مفهوم نموذج راش:

ويطلق على هذا النموذج نموذج راش اللوغاريتمي الاحتمالي البسيط، Rasch Simple Logistic Model وأحياناً يطلق عليه نموذج البارامتر الحر لتحليل الفقرات Sample Free.

Item Analysis

وقد ارتبط هذا النموذج باسم عالم الرياضيات الدانمركي "جورج راش" George Rasch الذي نادى بأهمية بناء نظام قياسي موضوعي في العلوم السلوكية، وكان يهدف إلى تحقيق مفهوم الموضوعية، بمعنى أن درجة الفرد في الاختبار لا يجب أن تكون دالة لعينة الأفراد التي استخدمت في التدرج الأصلي للفقرات التي يشتمل عليها الاختبار، كما أنه يجب أن يحصل الفرد على الدرجة نفسها في كل من اختبارين يقيسان السمة أو القدرة نفسها مهما اختلفت صعوبة فقرات كل منهما، وهذا يعني أن تكون القياسات مستقلة عن الفقرات المستخدمة في القياس، وكذلك أيضاً أن تكون أدوات القياس مستقلة عن خصائص عينة تقنينها. (علام، 1986: 118)

❖ ويعد نموذج راش أحادي المعلم من أكثر النماذج شيوعاً في تصميم وبناء الاختبارات والمقاييس النفسية والعقلية والتربوية، ويهتم بتحديد موقع الفقرة الاختبارية على ميزان صعوبة جميع الفقرات التي تشكل الاختبار (بارامتر الصعوبة)، كما يهتم بتدرج مستويات قدرة الفرد باختبار معين على نفس ميزان تعيير الفقرات. (علام، 2000: 693).

❖ الأساس النظري لنموذج راش:

يستخدم نموذج راش "في تحليل البيانات المستمدة من فقرات الاختبارات التي تعتمد في إجابتها على طريقة (صح أو خطأ)، أي الاختبارات التي تكون درجة الإجابة عن كل فقرة فيها (واحد) في حالة الإجابة الصواب، و(صفر) في حالة الإجابة الخطأ. (Henning. 1989: 92).

وعندما يستجيب الفرد لفقرة في اختبار ما، فإنه يحدث تفاعل بين قدرة هذا الفرد وصعوبة

هذه الفقرة، ومن خلال هذا التفاعل يتضح أن نموذج راش "يعتمد على أساسين هما:

- احتمال أن يجيب الفرد إجابة صحيحة على فقرة سهلة هو أكبر من احتمال أن يجيب إجابة صحيحة على فقرة صعبة.

- يزداد احتمال الإجابة الصحيحة بزيادة مستوى القدرة لدى الفرد. (عبد الحافظ، 1999: 65).
- فعندما يشرع الفرد في الاستجابة لفقرة معينة (i) فإن كلا من مستوى القدرة (θ) لدى هذا الفرد وصعوبة الفقرة (b_i) تعبران عن وضع هذا الفرد على متصل المتغير المقاس ، وتتحكمان في احتمال حدوث الاستجابة المناسبة عن الفقرات المتدرجة على متصل هذا المتغير (أمينة كاظم، 1988: 43). (Wright & Stone, 1979: 12). وتكون هذه الاحتمالية محكومة بالفرق بين معلم قدرة الفرد وصعوبة الفقرة ($\theta - b_i$) (اسماعيل، 2007: 35).

وعلى هذا يجب بذل الوقت والجهد لتنظيم الموقف الاختباري بحيث نقلل من تأثير أية عوامل أخرى —عدا عاملي قدرة الفرد وصعوبة الفقرة — على استجابات الأفراد على فقرات الاختبار، وتستخدم استجابات الفرد لتلك الفقرات المتدرجة في الصعوبة لتقدير مستوى القدرة لدى هذا الفرد على هذا المتغير.

ولما كانت (θ)، (b_i) تمثلان وضعين على متصل متغير واحد يشتركان فيه ، لذا فإن ($\theta - b_i$) هو الصيغة الأكثر ملاءمة للعلاقة بينهما، ومنطقيًا أنه كلما زادت قدرة الفرد (θ) عن صعوبة الفقرة (b_i) فإن احتمال حدوث الاستجابة المناسبة هو الأرجح والعكس ، ولكن قد يحدث أحياناً أن ينجح الفرد في الإجابة على فقرة تفوق صعوبتها مستوى القدرة لديه أو أن يحدث العكس وعلى هذا فليس من المناسب القطع بعلاقة تحديدية فاصلة بناءً على أن الفرق ($\theta - b_i$) يحكم قيمة الاستجابة X_{vi} بل من الأوفق أن تكون الطريقة التي يؤثر فيها الفرق ($\theta - b_i$) في الاستجابة احتمالية. (كاظم أ. 1988: 44).

وفي نموذج راش البسيط يكون هناك فئتان فقط للإجابة على الفقرة (نعم/لا) ويصور الشكل (6) ثلاث حالات (أ، ب ، ج)، توضح منطقية تأثير الفرق ($\theta - b_i$) على احتمالية الاستجابة الصواب. فإذا كانت (θ) أكبر من (b_i) أي أن مستوى قدرة الفرد أكبر من مستوى صعوبة الفقرة، فإن الفرق (θ)

(bi -) يكون أكبر من الصفر، وعندئذ يكون احتمال حدوث الاستجابة الصواب (نعم) أكبر من

(0.50) (العلاقة أ). أما إذا كان مستوى قدرة الفرد (θ) أقل من مستوى صعوبة الفقرة (bi)، فإن

الفرق ($\theta - bi$) يكون أقل من الصفر، وعندئذ يكون احتمال حدوث الاستجابة الصواب (نعم)

أقل من (0.50) (العلاقة ب). أما في حالة تساوي قدرة الفرد (θ) مع صعوبة الفقرة (bi)، فإن الفرق

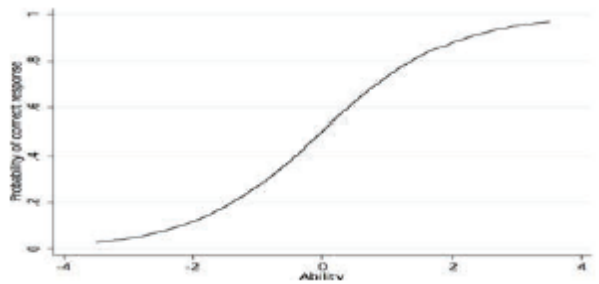
($\theta - bi$) يكون مساوياً للصفر وعندئذ يكون احتمال حدوث الاستجابة الصواب (نعم) يساوي

0,50 (العلاقة ج).

تأثير الفرق بين مستوى قدرة الفرد ومستوى صعوبة الفقرة في احتمال حدوث الاستجابة الصواب

ويلخص المنحنى الموضح بالشكل (7) تلك العلاقة المنطقية بين الفرق ($\theta - bi$) واحتمالات

الاستجابة الصواب. 84



• عند تفاعل المستويات المختلفة من قدرات أو سمات (الأفراد) مع فقرة معينة عندئذ يكون هذا المنحنى

وصفاً للفقرة، ويكون المتغير قدرة الفرد (θ) الذي يسمى بالمنحنى المميز للفقرة Item

Characteristic Curve (I C C)

• عندما تختبر الفقرات المختلفة الصعوبة قدرة أو سمة فرد معين، عندئذ يعد المنحنى وصفاً للفرد، ويكون

المتغير صعوبة الفقرة (bi) الذي يسمى بالمنحنى المميز للفرد Characteristic Curve

Person (p. c. c.) (كاظم، 1988أ: 46)

كما انبثق عن هذه النظرية طرق رياضية لتقدير قدرة الأفراد الكامنة التي تقف وراء إجاباتهم على مجموعة من الفقرات، منها:

طريقة الأرجحية العظمى (ML) (Maximum likelihood): وتعتبر من أكثر طرق التقدير شهرة حيث يتم إيجاد تقدير المعالم، من خلال إجراءات تعظيم الاحتمالية للمعلمة المراد تقديرها عندما يكون لدينا معلومات عن العينة.

2. طريقة القيمة العظمى للتوزيع البعدي (MAP) (Maximum A Posteriori): وتستخدم هذه الطريقة في العادة عندما لا نستطيع تطبيق طريقة الأرجحية العظمى، وتتميز هذه الطريقة بالإضافة لاستخدامها معلومات عن العينة بأنها تستخدم معلومات أولية متوفرة من خبرات سابقة.

3. طريقة توقع التوزيع البعدي (EAP) (Expected A Posteriori): وتستخدم هذه الطريقة التوزيع الطبيعي المعياري، حيث أن قيمة القدرة المستخرجة تمثل الوسط الحسابي للتوزيع الطبيعي. ونظرا لأهمية دقة تقدير القدرة للمفحوصين والتي تعتمد عليها دقة القرارات التنبؤية المختلفة، فقد تمحورت الدراسة الحالية على الطرق الثلاثة لتقدير القدرة في ضوء الخطأ المعياري للتقدير باستخدام النموذج المناسب، وذلك للكشف عن أكثر هذه الطرق دقة في تقدير قدرات الأفراد.

كما تم التحقق من افتراضات النموذج المستخدم وفقا لنظرية الاستجابة للفقرة، وهي: أحادية البعد (Unidimensionality) باستخدام مؤشرين: مؤشر تانكا Tanaka's Index ومؤشر الجذر التربيعي للوسط الحسابي لمربعات البواقي (Root Mean Square of Residuals, RMSR).

وأما افتراض الاستقلال الموضعي (Local Independence) فقد تم التحقق منه باستخدام مؤشر

(ZQ3) باستخدام برنامج (LDID (Local Dependence Indices for Dichotomous Items)

الخصائص السيكمترية لنموذج راش:

الثبات والصدق في مجال قياس راش، توجد تقنيات عديدة يمكن من خلالها تقييم صدق الأداة وتقييم الثبات. عند إجراء تحليل راش، يتيح جدول من برنامج Winsteps حساب ما يشار إليه ثبات الفقرات وثبات الأشخاص. تتراوح تلك القيمة من أدنى حد 0 إلى أعلى حد 1.00، والقيمة الأعلى أفضل. هذا الإحصاء مشابه لذلك الذي ينتج عن "ألفا كرونباخ"، لكن الحسابات تستند إلى فهم أن البيانات الخام ليست خطية. وبالتالي في تحليل راش يمكن التحدث عن ثبات الأشخاص لاختبار ما وثبات الفقرات. القيم المستهدفة التي تم الاستشهاد بها في الأدبيات لثبات الأشخاص وثبات الفقرات هي 0.8 و 0.9 (Malec et al., 2007)، غالباً ما يتم تقديم في تحليل راش عن معامل فصل الفقرات ومعامل فصل الأشخاص. القيم المستهدفة التي تم تقديمها في الأدبيات هي معامل فصل الأشخاص 2.0 و معامل فصل الفقرات 4.0 (Malec et al., 2007) كما نجد في الأدبيات

التقدير النظري للثبات الهامشي للاختبار Marginal Reliability Estimation

1. الإطار النظري للثبات في نماذج نظرية الاستجابة للمفردة

يمثل الثبات أحد المؤشرات الأساسية في تقييم جودة أدوات القياس التربوي والنفسي. ففي إطار نظرية الاستجابة للمفردة (IRT) يتم تعريف الثبات بطريقة تختلف عن التعريف المستخدم في النظرية الكلاسيكية للاختبار (CTT). إذ لا يعتمد الثبات في نماذج IRT على الاتساق الداخلي بين المفردات فقط، بل يرتبط بدرجة دقة تقدير القدرة الكامنة للأفراد عبر متصل السمة المقاسة.

يشير Embretson & Reise إلى أن الثبات في نماذج IRT يعبر عن نسبة التباين الحقيقي في القدرة الكامنة إلى التباين الكلي في التقديرات، وهو ما يطلق عليه غالباً الثبات الهامشي (Marginal Reliability) لأنه يحسب على مستوى التوزيع الكلي لقدرات الأفراد وليس عند مستوى قدرة

معين. (Embretson & Reise, 2000)

كما أوضح Bond & Fox أن معامل الثبات في نموذج راش يعتمد بصورة مباشرة على العلاقة بين تباين تقديرات القدرة ومتوسط مربع الخطأ المعياري للتقدير، مما يجعله مؤشراً دقيقاً لجودة القياس في الاختبارات التعليمية. (Bond & Fox, 2015)

2. الصيغة الرياضية للثبات الهامشي

يحسب الثبات الهامشي في نماذج IRT وفق المعادلة التالية:

$$Rel = 1 - \frac{SE^2}{Var(\theta)}$$

حيث:

- Rel : معامل الثبات الهامشي
- $Var(\theta)$: تباين تقديرات القدرة للأفراد
- SE : الخطأ المعياري لتقدير القدرة
- SE^2 : متوسط مربعات أخطاء القياس

وتشير هذه المعادلة إلى أن الثبات يزداد كلما انخفض الخطأ المعياري لتقدير القدرة مقارنة بتباين القدرة بين الأفراد.

.العلاقة بين الثبات ودالة معلومات الاختبار

في إطار نماذج IRT، يرتبط الثبات مباشرة بدالة معلومات الاختبار (Test Information Function) حيث:

$$SE(\theta) = \frac{1}{\sqrt{I(\theta)}}$$

حيث:

• $I(\theta)$: معلومات الاختبار عند مستوى القدرة θ

كلما زادت معلومات الاختبار انخفض الخطأ المعياري وارتفع الثبات الهامشي.

المقارنة والاختلاف بين النظرية الكلاسيكية للاختبار ونظرية القياس لراش

الدوافع والخلفية النظرية للنظرية الكلاسيكية للاختبار (CTT) ونظرية القياس لراش (RMT)

سنقوم بتلخيص مجموعة من أوجه التشابه والاختلاف بين النظرية الكلاسيكية للاختبار (Classical

Test Theory: CTT) ونظرية القياس لراش (Rasch Measurement Theory: RMT).

ونظراً لاختلاف الدوافع النظرية التي قامت عليها كل من النظريتين، قد يبدو للوهلة الأولى أن النظريتين

غير متوافقتين أو متعارضتين. ومع ذلك، فعلى الرغم من وجود اختلافات جوهرية بينهما، فإن الطريقة

التي تنعكس بها كل نظرية في افتراضاتها وصياغتها الرياضية كنماذج قياس تجعل من الممكن اعتبار نظرية

راش امتداداً أو تطويراً للنظرية الكلاسيكية للاختبار.

سنبدأ بعرض موجز للدوافع المختلفة التي أدت إلى ظهور كل من النظريتين. ونقترح أن اختلاف الدوافع

لا يمنع من أن تكون نظرية القياس لراش تطويراً للنظرية الكلاسيكية. إلى أن محاولة تطوير النظرية

الكلاسيكية للاختبار مباشرة لا تؤدي إلى نفس النوع من التطوير الذي تقدمه نظرية راش.

فعلى وجه الخصوص، لا يؤدي ذلك إلى اعتبار الدرجة الكلية للفرد على مجموعة من المفردات الإحصاء

الأساسي في القياس. وبدلاً من ذلك، تظهر تلك التطويرات غالباً على شكل إضافة معاملات جديدة

بشكل وصفي (ad hoc) بهدف تفسير مجموعات بيانات مختلفة بشكل أفضل.

دوافع ظهور نظرية القياس لراش (RMT)

كما ذكرنا سابقاً، فإن الدافع الأساسي لنظرية راش يتمثل في أن:

المقارنات بين الأفراد والمفردات يجب أن تكون ثابتة (Invariant) بغض النظر عن العينة المستخدمة.

أي أن:

- مقارنة الأفراد يجب ألا تعتمد على المفردات المختارة.
- مقارنة المفردات يجب ألا تعتمد على العينة المختيرة.

ويتم التعبير عن خصائص الأفراد والمفردات باستخدام أعداد حقيقية على متصل قياس واحد.

الخلفية التاريخية لنموذج راش: يمكن تتبع تطور نظرية راش من خلال:

- مقدمة كتابه الشهير:

Probabilistic Models for Some Intelligence and Attainment Tests (1960)

وكذلك من خلال أعمال:

- Andersen & Olsen (2001)
- Andrich (2005)

بداية تطبيق نموذج راش

كان جورج راش يعمل مستشاراً في المعهد الدنماركي للبحوث التربوية، حيث طُلب منه تصميم دراسة

لقياس فعالية برنامج لتحسين القراءة لدى الأطفال الذين يعانون صعوبات في القراءة.

بدلاً من الاعتماد على النظرية الكلاسيكية للاختبار، تعامل راش مع المشكلة بالطريقة التي كان يستخدمها في البحوث الطبية والبيولوجية.

اثناء الدراسة، طُلب من الأطفال قراءة نصوص مختلفة بصوت عالٍ تم تسجيل عدد الأخطاء في قراءة الكلمات

مشكلة القياس في هذه الدراسة: لكي يتم قياس تطور قدرة القراءة لدى الأطفال، يجب:

1. ألا تكون النصوص صعبة جداً بحيث يفشل الأطفال في قراءتها.

2. وألا تكون سهلة جداً بحيث لا يرتكبون أخطاء.

ومع تحسن قدرة القراءة بمرور الوقت، يجب تقديم نصوص أكثر صعوبة.

لذلك كان من الضروري:

وضع جميع النصوص على نفس مقياس الصعوبة. تصميم الربط بين النصوص.

كانت الكلمات داخل النص الواحد تختلف في الصعوبة، ولكن النصوص كانت مصممة بحيث تكون:

- متجانسة نسبياً داخلياً
- مختلفة في مستوى الصعوبة الكلي

ولربط النصوص على نفس المقياس تم استخدام تصميم الربط (Linking Design)

صياغة نموذج راش

قام راش بوصف الاستجابة باستخدام معاملين:

- قدرة الشخص على القراءة

- صعوبة النص

ولأن عدد الأخطاء كان صغيراً نسبياً، فقد لاحظ أن توزيع بواسون (Poisson) قد يكون مناسباً

لنمذجة البيانات. لكن استخدامه كان مميزاً، حيث استخدمه لوصف: عدد الأخطاء لشخص معين

في نص معين وليس توزيع الأخطاء على مستوى المجتمع. وبالتالي ركز على الفرد وليس على التوزيع

الطبيعي للأفراد كما في CTT.

خصائص المقارنة بين RMT و CTT

بدلاً من عرض الاختلافات بشكل منفصل، سيتم مناقشة الخصائص المشتركة لكل نظرية. ويبين الجدول

التالي أهم المقارنات.

جدول 1 : المقارنة بين النظرية الكلاسيكية ونظرية راش

الخاصية	نظرية راش (RMT)	النظرية الكلاسيكية (CTT)
الدافع النظري	ضرورة ثبات المقارنات بين الأشخاص والمفردات	المفردات تمثل تكراراً لقياس قدرة الشخص
الافتراضات	أحادية البعد - استقلال الاستجابات - لا يفترض توزيع معين للقدرة	أحادية البعد - استقلال الاستجابات - توزيع طبيعي للقدرة
موقع المفردة	معرف بمعامل صعوبة δ	خارج النظرية - مؤشر الصعوبة يعتمد على العينة
النموذج الرياضي	$u_{ni} = (\beta_n - \delta_i) + \varepsilon_{ni}$	$x_{ni} = \tau_n + \varepsilon_{ni}$
احتمال الإجابة الصحيحة	$P_{ni} = \frac{e^{\beta_n - \delta_i}}{1 + e^{\beta_n - \delta_i}}$	غير معرف صراحة
الخاصية الأساسية	الدرجة الكلية إحصاء كاف لتقدير القدرة	الدرجة الكلية تعبر عن قدرة الفرد
تقدير القدرة	علاقة غير خطية	علاقة خطية مع الدرجة الكلية
الخطأ المعياري	متغير حسب القدرة والصعوبة	ثابت لكل الأفراد
الثبات	مؤشر فصل الأفراد Person Separation	معامل ألفا
البيانات المفقودة	يمكن التعامل معها مباشرة	غالباً يتم تعويضها
التمييز	الانحراف عن المتوسط مؤشر على سوء المطابقة	يفضل أن يكون عالياً

تقدير الدرجة الحقيقية في النظرية الكلاسيكية

يتم تقدير الدرجة الحقيقية وفق المعادلة:

$$\widehat{T}_n = \bar{Y} + r_{yy}(Y_n - \bar{Y})$$

حيث:

- Y_n الدرجة الكلية للفرد
- $\bar{Y}$ متوسط الدرجات
- r_{yy} معامل الثبات

وتكون العلاقة بين الدرجة الحقيقية والدرجة الكلية علاقة خطية.

تباين الدرجات الحقيقية

$$V[\widehat{T}_n] = r_{yy}^2 V[Y]$$

أي أن الانحراف المعياري للدرجات الحقيقية يساوي:
 $SD[\widehat{T}_n] = r_{yy} SD[Y]$

تقدير القدرة في نموذج راش

في نموذج راش يتم تقدير قدرة الشخص وفق:

$$r_n = \sum_{i=1}^I \frac{e^{\beta_n - \delta_i}}{1 + e^{\beta_n - \delta_i}}$$

ويتم تحويل الدرجة الكلية إلى مقياس لوجيت (Logit) باستخدام تحويل غير خطي يأخذ شكل منحنى S (اللوجستي)

الخطأ المعياري في النظرية الكلاسيكية

الخطأ المعياري ثابت ويحسب من:

$$SE = s_y \sqrt{1 - r_{yy}}$$

الخطأ المعياري في نموذج راش

$$\sigma_{\hat{\beta}} = \frac{1}{\sqrt{\sum P_{ni}(1 - P_{ni})}}$$

ويزداد الخطأ المعياري عندما تكون القدرة قريبة من القيم المتطرفة.

مقارنة موسعة بين CTT و Rasch و IRT

تُعد كل من النظرية الكلاسيكية للاختبار (CTT) ونظرية القياس لراش (Rasch Measurement Theory) ونظرية الاستجابة للمفردة (Item Response Theory: IRT) من الأطر الأساسية في القياس النفسي والتربوي. وعلى الرغم من أن هذه النظريات تهدف جميعها إلى تقدير السمات الكامنة لدى الأفراد، فإنها تختلف في افتراضاتها الرياضية وطرق تقدير المعلومات وتفسير النتائج.

جدول (2) مقارنة نظرية بين CTT و Rasch و IRT

البعد المقارن	النظرية الكلاسيكية CTT	نظرية IRT العامة	نموذج Rasch
الهدف الأساسي	تقدير الدرجة الحقيقية للفرد	نمذجة احتمال الاستجابة	بناء مقياس قياس موضوعي
طبيعة النموذج	نموذج خطي	نماذج لوجستية متعددة	نموذج لوجستي بسيط
مستوى التحليل	الاختبار ككل	المفردة والفرد	المفردة والفرد
المعلومات الأساسية	الدرجة الحقيقية والخطأ	الصعوبة (θ) القدرة ، (a) ، التمييز (b) (c) التخمين	صعوبة (θ) القدرة (δ) المفردة
العلاقة بين المفردة والفرد	غير صريحة	صريحة	صريحة
استقلال المعلومات	غير متحقق	متحقق جزئياً	متحقق (خاصية الثبات)
تأثير العينة	قوي	متوسط	ضعيف
تقدير القدرة	يعتمد على الدرجة الكلية	يعتمد على النموذج الاحتمالي	يعتمد على النموذج الاحتمالي
مقياس القدرة	مقياس رتبي	مقياس فاصل	(logit) مقياس فاصل
التعامل مع البيانات المفقودة	صعب	سهل	سهل
تقدير الخطأ المعياري	ثابت	متغير	متغير
استخدام الدرجة الكلية	أساسي	ليس بالضرورة	إحصاء كاف

جدول (3) مقارنة المعادلات الرياضية

النظرية	المعادلة الأساسية
CTT	$X = T + E$
Rasch	$P_{ni} = \frac{e^{(\theta_n - \delta_i)}}{1 + e^{(\theta_n - \delta_i)}}$

$P_{ni} = \frac{1}{1 + e^{-a_i(\theta_n - b_i)}}$	2PL IRT
$P_{ni} = c_i + (1 - c_i) \frac{1}{1 + e^{-a_i(\theta_n - b_i)}}$	3PL IRT

حيث:

- θ قدرة الفرد
- δ صعوبة المفردة
- a معامل التمييز
- b الصعوبة
- c التخمين

ثانياً: العلاقة الرياضية بين Logit و Raw Score

1 مفهوم الدرجة الخام

الدرجة الخام للفرد في الاختبار هي:

$$r = \sum_{i=1}^I x_i$$

حيث:

- x_i استجابة المفردة (0 أو 1)
- I : عدد المفردات

وبالتالي فإن:

$$0 \leq r \leq I$$

تقدير القدرة في نموذج Rasch

في نموذج Rasch يتم تعريف احتمال الإجابة الصحيحة كما يلي:

$$P_{ni} = \frac{e^{(\theta_n - \delta_i)}}{1 + e^{(\theta_n - \delta_i)}}$$

وبالتالي فإن التوقع الرياضي للدرجة الكلية يساوي:

$$E(R_n) = \sum_{i=1}^I P_{ni}$$

3 تحويل الدرجة الخام إلى لوجيت

يتم تحويل الدرجة الخام إلى مقياس اللوجيت (**Logit**) باستخدام التحويل اللوجستي.

عند افتراض أن صعوبة المفردات متساوية تقريباً يمكن تقريب القدرة بالمعادلة:

$$\theta \approx \ln \left(\frac{r}{I - r} \right)$$

وهذا التحويل يسمى Log-odds transformation

- نسبة الإجابة الصحيحة (p-value)

- صعوبة الفقرة باللوجيت

$$\delta = \ln \left(\frac{1 - p}{p} \right)$$

4 تفسير اللوجيت

اللوجيت يمثل اللوغاريتم الطبيعي لنسبة النجاح إلى الفشل:

$$\text{logit} = \ln\left(\frac{P}{1-P}\right)$$

حيث:

$$P = \frac{r}{I}$$

مثال تطبيقي

اختبار مكون من 20 فقرة:

$$I = 20$$

إذا حصل الطالب على:

$$r = 15$$

فإن:

$$P = 15/20 = 0.75$$

إذن:

$$\text{logit} = \ln\left(\frac{0.75}{0.25}\right)$$

$$\text{logit} = \ln(3)$$

$$\text{logit} = 1.10$$

أي أن قدرة الطالب تساوي تقريباً: 1.10 لوجيت

ثالثاً: الخصائص الإحصائية للتحويل Logit

يتميز تحويل الدرجة الخام إلى لوجيت بعدة خصائص:

1 العلاقة غير الخطية

العلاقة بين Raw Score و Logit غير خطية وتأخذ شكل منحنى S.

2 التمدد في الأطراف عند القيم المتطرفة:

- الدرجات العالية تتمدد
- الدرجات المنخفضة تتمدد
- مما يزيد دقة القياس.

3 مقياس فاصل حقيقي حيث اللوجيت يمثل مقياساً فاصلاً. (Interval Scale) (فئوي)

بينما الدرجة الخام تمثل مقياساً ترتيبياً فقط.

جدول (4) مثال تحويل Raw Score إلى Logit

الدرجة الخام	الاحتمالية	لوجيت
2	0.10	-2.20
5	0.25	-1.10
10	0.50	0
15	0.75	1.10
18	0.90	2.20

رابعاً: الأهمية السيكمترية للتحويل إلى Logit

يسمح تحويل الدرجات إلى مقياس لوجيت بما يلي:

- مقارنة الأفراد والمفردات على مقياس واحد
- تحقيق القياس الموضوعي

- تقدير الخطأ المعياري لكل قدرة
 - بناء خرائط Wright Map
 - حساب دالة معلومات الاختبار TIF
- وهذا ما سنراه في الجانب التطبيقي

رابعاً : الدراسات السابقة

- مزايا النظرية الكلاسيكية
- مزايا نموذج راش
- المقارنة بين الدراسات
- التعليق على الدراسات السابقة

تمهيد

تعد نظرية الاختبار الكلاسيكية (CTT) ونظرية الاستجابة للفقرة (IRT) إطارين مهيمنين في القياس التربوي والقياس النفسي. بينما يهدف كلاهما إلى تقييم القدرات والسمات الفردية من خلال الاختبارات، إلا أنهما يعملان على مبادئ ومنهجيات مختلفة. ستحدد هذه المقارنة مزايا وحدود كل نظرية، مدعومة بالاقتباسات والمقالات والأطروحات ذات الصلة.

مزايا نظرية الاختبار الكلاسيكية

1. ومما يؤكد اتجاه العلماء نحو استخدام مبادئ نظرية الاستجابة للمفردة في بناء وتطوير المقاييس المختلفة ما قام به (Waugh&Addison,1998,95-112) عندما هدفا إلى تحليل الخصائص السيكو مترية باستخدام نموذج راش اللوغاريتمي وذلك لاستفتاء مداخل الدراسة المكون من خمسة مقاييس فرعية لقياس المدخل العميق والسطحي والاستراتيجي ونقص التعليمات والثقة بالنفس أكاديميًا، وتكونت عينة التحليل من 346 طالبًا وطالبة من طلاب الفرقة الأولى في إدارة الأعمال في جامعة Perth بأستراليا بواقع (170 طالبة، 176 طالبًا ، 212 تحت سن 23 سنة، 134 فوق سن 23 سنة)، وقد أسفرت النتائج عن مناسبة الخصائص السيكو مترية لمفردات المقاييس الفرعية الثلاثة (المدخل العميق والسطحي والاستراتيجي) لافتراضات نموذج راش، ولكن لم تتناسب الخصائص السيكو مترية لمفردات المقاييس الفرعيين (نقص التعليمات، والثقة بالنفس أكاديميًا) مع افتراضات النموذج.

2. وكذلك ما قامت به (Wiberger,2004,2) عندما هدفت إلى تحديد أي نماذج الاستجابة للمفردة (أحادية أو ثنائية أو ثلاثية البارامتر) أكثر مناسبة عند استخدام اختبار الترخيص للقيادة السويدية ومقارنة النموذج المختار عندما استخدمنا كل من النظرية الكلاسيكية. وما قام به (Ainol&Noor, 2006,1) عندما استخدمنا كل من النظرية الكلاسيكية ونموذج راش (Rasch Model) (وهو أحد نماذج الاستجابة للمفردة) في تقييم كفاءة مفردات اختبار مكون من (35) مفردة لقياس القراءة في اللغة الإنجليزية، والذي يستخدم لإجراءات قبول الطلاب في الجامعة الدولية الإسلامية بماليزيا، وقد أثبت الباحثان أنه لا يمكن الاعتماد على النظرية الكلاسيكية فقط في تحليل مفردات هذا الاختبار، ولكن لابد من الاستفادة أيضًا من نموذج راش في تحديد كفاءة هذه المفردات معللين ذلك بأن تحليل المفردات باستخدام النظرية الكلاسيكية لا يمدنا بمعلومات ضرورية عن أداء الطلاب ذوي المستويات المختلفة في القدرة على كل مفردة في الاختبار.

3. وقام (Thompson&Weiss, 2003,3) باستخدام نظرية الاستجابة للمفردة في تدريب مفردات

اختبار (The Multistate Bar Examination (MBE)) والمكون من 200 مفردة لقياس ستة مجالات

هي: القوانين الدستورية وقوانين العقود والجرائم والشواهد وقوانين العقارات والاعتداءات، وذلك باستخدام ثلاث طرق

للتعبير أو التدريب وهي طريقة الاحتمالية القصوى (Maximum Likelihood) وطريقة تعظيم الاقتران البعدي

(Maximum A Posteriori) و طريقة التوقع التعدي (Expected A Posteriori) وقد اتضح

وجود فروق طفيفة غير مؤثرة بين تقدير الافراد باستخدام الطرق الثلاثة المختلفة للتدريب.

4. إلا أن بعض الدراسات الأجنبية قد اختلفت في نتائجها عن الدراسات السابقة، فبالرغم من الفروق النظرية بين نظرية

الاستجابة للمفردة والنظرية الكلاسيكية إلا أن (Fan,1998,357) قد وجد نقصاً في المعرفة القائمة على أساليب

تجريبية عن مدى الاختلافات بين النظريتين في تحليل المفردات وتقدير قدرات الأفراد، فقام بدراسة أثبت فيها أن

الإحصاءات الخاصة بالمفردات والأفراد المشتقة من النظريتين متشابهة تماماً، وكانت درجة التجانس (عدم التباين

(Invariance) في إحصاءات المفردات بين العينات المختلفة متماثلة في الحالتين.

5. ومثال ذلك أيضاً ما قامت به (Stage,2003,1) عندما أجرت دراسة هدفت فيها إلى تطبيق مبادئ نظرية

الاستجابة للمفردة في تحليل مفردات اختبار الاستعداد المدرسي في السويد والذي تم بناؤه في ضوء النظرية الكلاسيكية

في القياس، ومعرفة أثر ذلك في تحسين كفاءة الاختبار، وقد تكونت الدراسة من ثلاثة أجزاء، يهدف الجزء الأول منها

إلى معرفة إمكانية تطبيق نماذج الاستجابة للمفردة على الاختبارات الفرعية الثلاثة المكونة للاختبار منفصلة، أما الجزء

الثاني فقد هدف إلى المقارنة بين تحليل المفردات باستخدام النظرية الكلاسيكية وباستخدام نظرية الاستجابة للمفردة

وذلك عند حساب معاملات الصعوبة ومعاملات التمييز، أما الجزء الثالث فقد سعى إلى معرفة إمكانية تطبيق نظرية

الاستجابة للمفردة على الاختبار ككل، وقد كانت النتائج على غير المتوقع إذ إن الباحثة أوضحت أن البيانات التي تم

الحصول عليها من النموذج مشكوك في صحتها خاصة للاختبار ككل، كما لم تحدث أي فائدة بالانتقال من النظرية

الكلاسيكية إلى نظرية الاستجابة للمفردة.

6. وأجرى ادواردز والكوك (Edwards & Alcock, 2010) دراسة هدفت إلى تحليل نتائج اختبار في الرياضيات

في بريطانيا وفق أنموذج راش، ومن أجل تحقيق هدف الدراسة قام الباحثان ببناء اختبار مؤلف من (11) فقرة، وقد تم

توزيعه على عينة الدراسة المؤلفة من (164) طالبا وطالبة من طلبة السنة الأولى في المرحلة الجامعية، وقد طلب من أفراد العينة ترتيب فقرات الاختبار حسب مستوى صعوبتها، وبعد ذلك تم تحليل البيانات، التي تم الحصول عليها، باستخدام نموذج راش، وأشارت النتائج إلى عدم مطابقة فقرات الاختبار لستة أشخاص فقط من عينة الدراسة، بعد إعادة ترتيب فقرات الاختبار حسب نموذج راش .

7. أما في البيئة العربية فقد سعت (أمينة محمد كاظم، 119، 1988) إلى التعرف على شروط القياس الموضوعي للسلوك، ومدى تحقق هذه الشروط على اختبار تحصيلي في مادة علم النفس باستخدام نموذج راش، وقد تكون هذا الاختبار من 100 بند تغطي ثمانية موضوعات رئيسة من المقرر في ثلاثة مستويات معرفية وقد تم تحليل المفردات باستخدام برنامج Bical ، وبناء على هذا التحليل تم حذف خمسة بنود غير ملائمة ليصبح تحليل مفردات الاختبار عدد مفردات الاختبار 95 مفردة، كما تم التحقق من أن بنود الاختبار تقيس متغيراً أحادي البعد، وأن القياس تحرر من خصائص العينة المطبق عليهم الاختبار، ومن خصائص البنود التي يجب عليها الأفراد.

8. كما قام (عماد عبد المسيح يوسف، 472، 1991) باستخدام نموذج راش اللوغاريتمي الاحتمالي في التأكد من مطابقته للبيانات المستمدة من تطبيق مفردات اختبار الأشكال المتضمنة (الصورة الجمعية) المعد لقياس الأسلوب المعرفي ثنائي القطب (الاعتماد- الاستقلال عن المجال)، والذي سبق أن عرّبه كل من (أنور محمد الشرقاوي وسليمان الحضري الشبخ، 1988)، والذي تم بناؤه من قبل وفقاً لمبادئ النظرية الكلاسيكية، وقد ثبت أنه لا يتغير تدريج صعوبة مفردات الاختبار بتغير مستوى قدرات أفراد العينة المستخدمة في الحصول على هذا التدرج.

9. وقد سعى (عبد الرحمن سليمان الطيري، 457، 1996) إلى تطبيق نموذج راش اللوغاريتمي الاحتمالي على البيانات المستمدة من تطبيق اختبار الذكاء الإعدادي والذي أعده السيد محمد خيري ، وذلك على عينة قوامها 147 (طالباً في السنة الثانية المتوسطة (الإعدادية) في مدينة الرياض، وقد توصل الباحث إلى أن هذا الاختبار (تتوفر فيه الخصائص السيكمترية وموضوعية القياس من حيث الاستقلال بين صعوبة المفردات وقدرات الأفراد باستخدام نموذج راش، بعد حذف أحد مفردات الاختبار المكون في الأصل من (50) مفردة؛ وذلك لفشل جميع الأفراد في الإجابة عنها، وتم التأكد من مطابقة البيانات لافتراضات نموذج راش من خلال عينة قدرها (13) مفردة من مفردات الاختبار متباينة في مستوى صعوبتها، و(28) مفحوصاً من المفحوصين مختلفين في مستوى قدراتهم.

10. وفي دراسة دعنا Da'ana, 2005 التي هدفت لبناء اختبار مكيف هرمي في الرياضيات للصف الثامن الأساسي وفقاً لنموذج راش، وتم بناء اختبارين من نوع الاختيار من متعدد وبأربعة بدائل لكل فقرة تقيس فقراتها المستويات الخمسة الأولى من تصنيف بلوم في المجال المعرفي، الأول يتعلق بمحتوى الفصل الدراسي الأول عدد فقراته (35)، والثاني يتعلق بمحتوى الفصل الدراسي الثاني عدد فقراته (44)، وطبق الاختباران على عينة التحليل الإحصائي المؤلفة من 376 مفحوصاً اختيروا بطريقة عشوائية مرحلية ذات مرحلتين، وأظهرت النتائج حذف (7 فقرات) من الاختبار الأول وحذف 8 فقرات من الاختبار الثاني وذلك لأن قيم مربع كاي فيها دالة إحصائية عند مستوى دلالة (0.05) وأشارت النتائج إلى أن الاختبارين يقيسان عاملاً واحداً (نسبة التباين المفسر لم تقل عن 20 %)، وقد تم ترتيب الفقرات تصاعدياً وفق صعوبتها.

11. وهدفت دراسة العيثاوي (Ithawi, 2009) إلى استعمال نموذج راش أحادي المعلمة في بناء اختبار تحصيلي في مادة علم نفس الفروق الفردية، ولتحقيق أهداف الدراسة شرع الباحث في بناء الاختبار مستنداً إلى المفردات المقررة من قبل الهيئة القطاعية لوزارة التعليم العالي. أعد الباحث (173) فقرة من نوع الاختيار من متعدد تغطي مستويات الأهداف المعرفية الثلاث الأولى بحسب تصنيف بلوم وهي: المعرفة والفهم والتطبيق، واشتملت عينة البحث على (336) طالباً وطالبة من طلبة كلية التربية للجامعات العراقية، ثم أدخلت البيانات لبرنامج الحاسب الآلي راسكال Rascal (). ومن خلال معالجة البيانات تضمنت الصورة النهائية للاختبار 108 فقرة مطابقة لافتراضات النموذج ومتطلباته فقد تم حذف 28 فقرة لا تحقق أحادية البعد من خلال التحليل ألعاملي لفقرات الاختبار بطريقة المكونات الأساسية، فيما حذفت 29 فقرة لدلالة قيمة مربع كاي، واقتربت قيمة معامل التميز من 1 عدد صحيح حيث كانت القيمة ضمن المدى المقبول 0.629 كما تم استبعاد 8 فقرات لا تحقق استقلالية القياس. دُرّجت الفقرات 108 بحسب صعوبة الفقرات وقدرات الأفراد للحصول على الاختبار بصيغته النهائية والذي بلغت قيمة معامل ثباته 0.952 ومعامل تمييزه 0.713.

12. وهدفت دراسة الخياط Alkhaia, 2012 إلى التحقق من فاعلية النموذج اللوغاريتمي ذي المعلمة الواحدة نموذج راش ومطابقة البيانات للنموذج لاختبار TIMSS الذي يقيس المهارات الرياضية لمستوى الصف الثامن الأساسي، ولتحقيق هدف الدراسة تم تطبيق فقرات الاختبار على عينة من طلبة الصف الثامن الأساسي بلغ عددها 599 طالباً

وطالبة؛ وأشارت نتائج الدراسة إلى تمتع الاختبار المقدم بالصدق والثبات، حيث بلغ معامل الثبات للاختبار 0.9 وأما معامل الصدق بدلالة المحك بين درجات الطلبة على الاختبار ودرجاتهم في مادة الرياضيات فقد بلغت قيمته 0.87 وأشارت النتائج إلى أن الاختبار يصنف بأنه متوسط الصعوبة، حيث تراوح مستوى صعوبة الفقرات بين (0.305 - 0.782) لوجيت. ومن ثم أوجد الباحث قيمة مربع كاي بحيث دلت قيمة مربع كاي لعدم مطابقة 10 فقرات لنموذج راش. وأشارت النتائج أيضا إلى مناسبة 25 فقرة من فقرات الاختبار والبالغ عددها 35 فقرة للبيئة المحلية.

13. أما علي (Ali, 2012) فهدفت دراسته إلى تعرف فاعلية بناء اختبار محكي المرجع لمقرر القياس والتقويم في التربية وفق نموذج راش؛ بحيث تكون الاختبار من 99 فقرة اختبارية من نوع الاختيار من متعدد وبأربعة بدائل، وتكونت عينة البحث من 426 (طالبًا وطالبة من طلبة كلية التربية في قسم معلم الصف من جامعة دمشق، الذين درسوا المقرر في العام الدراسي . 2010 / 2009 وتم تقسيم العينة إلى قسمين، وتطبيق الاختبار عليهما في وقت واحد. وتم استخدام برنامج (WINSTEPS) في تحليل البيانات الناتجة عن تطبيق الاختبار محكي المرجع وفق نموذج راش، وقد نتج عن هذا البرنامج حذف عشر فقرات بسبب وقوعها خارج حدود المطابقة، التي حددت بين 0.7 - 1.3، وبالتالي أصبح العدد النهائي لفقرات الاختبار 89 فقرة، جميعها تقع ضمن حدود المطابقة الداخلية والخارجية. وبينت النتائج أن معامل الثبات للفقرات بلغ 0.89، ومعامل ثبات الأفراد بلغ 0.94 وكلاهما مرتفع. كما بينت النتائج أن تدريج الاختبار يحقق افتراضات نظرية الاستجابة للفقرة.

14. نزاي الزهراء ، نظرية الاستجابة للمفردة في الاختبارات التحصيلية لمادة الرياضيات: دراسة ميدانية على مستوى السنة الثانية ابتدائي بولاية مستغانم. (حذف 104 تلميذا و 6 فقرات) 2020/2019

15. لعزالي صليحة. التعرف: على أساليب التعلم والتفكير عند الطلبة والتحقق من موضوعية النتائج باستخدام النظرية الحديثة في القياس: -دراسة ميدانية على عينة من الطلبة ببعض جامعات الوسط- الدكتوراه في علم النفس جامعة بليدة 2017/2016

16. وهدفت دراسة أديدوين و موكوبي (Adedoyin & Mokobi, 2013) إلى تحليل فقرات الاختيار من متعدد في اختبار بوسطن في الرياضيات الورقة الأولى للعام 2010 . وتكون الاختبار من 40 فقرة تم بناءها من منهاج

الرياضيات للسنة الثالثة. تكون مجتمع الدراسة من جميع الطلبة الذين تقدموا للاختبار في العام 2010 وعددهم

36940 طالبًا وطالبة، وبلغ حجم العينة 10000 طالبًا وطالبة تم اختيارهم عشوائيًا باستخدام برمجية (SPSS)

وتم تحليل استجابات المفحوصين باستخدام أحد نماذج نظرية الاستجابة للفقرة وهو النموذج ثلاثي المعلمة، حيث تم حساب معالم الفقرات وهي: صعوبة الفقرات، وتمييزها، وتحمينها. كما تم رسم منحنيات معالم الفقرات المطابقة للنموذج ثلاثي المعلمة وعددها 23 فقرة من أصل 40 فقرة، وأظهرت نتائج الدراسة أن 12 من الـ 23 فقرة صُنفت كفقرات ضعيفة و 10 فقرات صُنفت كفقرات جيدة ولكن يمكن تنقيحها أو تحسينها، وفقرة واحدة تعتبر فقرة جيدة يمكن اعتمادها دون أي تعديل. وتمت التوصية بأن يتم تحسين جودة فقرات الاختبار لتكون مطابقة لنماذج نظرية الاستجابة للفقرة (IRT)

17 ط د بوعمامة مريم 1، أ د بحري نبيل 2 بناء اختبار تحصيلي في مادة الرياضيات للسنة الأولى من التعليم المتوسط باستعمال نموذج راش أحادي المعلمة هدفت الدراسة الحالية إلى بناء اختبار تحصيلي في الرياضيات المدرسة للسنة الأولى من التعليم المتوسط باستخدام النموذج اللوجستي أحادي البعد. لتحقيق هذا الهدف تم إعداد صيغة أولية للاختبار مكونة من 32 فقرة ثنائية التدرج روعي في بناءها الميادين الثلاثة في تعليم الرياضيات وهي: الأنشطة العددية الأنشطة الهندسية وتنظيم المعطيات والدوال طبق الاختبار على عينة من 255 تلميذ على مستوى متوسطات مديرية التربية لولاية ميلة ومعسكر. تمكن أغلبية التلاميذ من الإجابة عن جميع الفقرات. وقد أظهرت النتائج تحقق افتراضات نظرية الاستجابة للفقرة، بعد حذف 03 فقرات، وحذف 04 أفراد نظرا لعدم مطابقتهم لنموذج راش. كما بينت نتائج التحليل فيما يتعلق بمطابقة الفقرات الثنائية التدرج للنموذج أحادي المعلمة أن أغلبية الفقرات كانت مطابقة لنموذج راش.

18. وهدفت دراسة دبوس (Dabous, 2016) إلى استخدام نظرية الاستجابة للفقرة في بناء تجمع فقرات اختبار محكي المرجع في الرياضيات وفق النموذج اللوجستي ثنائي المعلم. ولتحقيق هذا الهدف تم إعداد تجمع من الفقرات ثنائية التدرج مكون من 50 فقرة، وتجمع من الفقرات متعددة التدرج مكون من 10 فقرات. وروعي في بناء هذه الفقرات أن تقيس تحصيل الطالب في الرياضيات في المعرفة المفاهيمية والمعرفة الإجرائية وحل المشكلات. وقد تكونت عينة الدراسة من 502 طالبًا وطالبة أجابوا عن جميع الفقرات. وقد أظهرت النتائج تحقق افتراضات نظرية الاستجابة للفقرة. كما بينت نتائج التحليل فيما

يتعلق بمطابقة الفقرات الثنائية التدرج للنموذج الثنائي المعلم أن جميع الفقرات كانت مطابقة للنموذج ثنائي المعلم، وأن الفقرات المتعددة التدرج كانت مطابقة لنماذج الاستجابة المتدرجة.

19 وهدفت دراسة محمود وصباح (Mahmoud, L. & Sabah, S. 2016) إلى بناء اختبار تحصيلي في الهندسة لطلبة الصف الخامس باستخدام نموذج راش، وتشخيص فهم الطلبة للمفاهيم الهندسية. تم بناء اختبار تحصيلي مكون من 30 فقرة من نوع الاختبار من متعدد. طبق الاختبار على 216 طالباً وطالبة. وأظهرت النتائج أن الاختبار ثابت، حيث بلغ معامل الثبات للفقرات 0.98 ومعامل الثبات للأفراد 0.78 وأن فقرات الاختبار غطت مدى الصعوبة بشكل جيد، فتراوحت قيمة الصعوبة للفقرات ما بين - 3.29 + 2.69 (لوجيت، وظهر أن معظم الفقرات تقع ضمن إحصائيات الملاءمة لنموذج راش. وأظهرت النتائج أيضاً وجود أخطاء متنوعة في المفاهيم الهندسية منها: خصائص متوازي الأضلاع، مفهوم المحيط، مفهوم الشبكات، والربط بين الأشكال الهندسية، وفي المقابل هناك مجموعة من المفاهيم الهندسية التي أجاب عليها الطلبة بنسبة أكبر من (88%) منها: مفهوم المضلع، مفهوم المضلع المنتظم، ومفهوم الزاوية وقياسها باستخدام المنقلة.

وهذه مقارنة بين بعض الدراسات الاجنبية

الجدول رقم: 5 مقارنة بين دراسات سابقة

الدراسة	أداة القياس / السياق	N	المنهجية	النتيجة الرئيسية
Janssen, Meier & Trace (2014)	اختبار تحديد المستوى، جامعة كولومبية (75 مفردة)	—	CTT (IF/ID) + راش (logit)	يحدد نموذج راش كمياً الفجوة بين صعوبات المفردات وبين الأفراد على مقياس فتري؛ بينما تظل CTT مناسبة للفحص الاستكشافي (رابط)
Planinic et al. (2019)	بحث في أبحاث التعليم الفيزيائي (PER)	على سبيل المثال RCI: 70 (قبل) / 63 (بعد)	راش مقابل "نسبة النجاح"	صعوبات المفردات مستقرة ومستقلة عن العينة، مما يسمح بمتابعة التقدم (رابط)
Hwang (2002)	اختبار مكون من 15 مفردة، عينة من 600	600	مقارنة CTT/IRT تحليلياً وإمبيريقياً	القابلية الجزئية للمقارنة بين تقديرات IRT (ERIC و CTT) (ED466779 رابط)
Séville et al. (2010)	النتائج المصرح بها من المرضى (QdV)	محاكاة	قوة CTT مقابل IRT	اختيار المنهجية يؤثر على حجم العينة المطلوب بناءً على الأثر المراد كشفه
Jabrayilov & Emons (2016)	تقييم التغير الفردي	—	مقارنة CTT مقابل IRT	تبالغ نظرية IRT في تقدير الثبات الإمبريقي وتتطلب عينات كبيرة
Aggen, Neale & Kendler (2005)	معايير DSM للاكتئاب الجسيم	—	1PL (مشابه لـ Rasch مقابل PL2)	نموذج 2 PL يطابق البيانات بشكل أفضل، لكن 1 PL أكثر إيجازاً؛ مقارنة المنحدرات

يسمح نموذج راش IRT / بإنشاء بنوك المفردات والاختبارات التكيفية، ولكن بتكلفة متطلبات عينة أكبر	مراجعة CTT مقابل IRT	—	تطوير مقاييس النتائج المصرح بها من المرضى (PROM)	Cappelleri, Lundy & Hays (2014)
---	----------------------	---	--	---------------------------------

التعليق على الدراسات من الجدول:

حدود النظرية الكلاسيكية للاختبارات (CTT)

1. مؤشرات تعتمد على العينة (غير قابلة للتعميم).
2. الدرجة الخام على مقياس غير خطي (محدودة بين 0-100%؛ الزيادة بـ 1% لا تمثل كسباً ثابتاً في القدرة).
3. خطأ معياري للقياس (SEM) فريد يفترض دقة ثابتة.
4. غير مناسبة للتصاميم المعقدة (قبل/بعد، الصور المتكافئة، المحكمون المتعددون).

(Planinic et al., 2019; Janssen et al., 2014).

حدود نظرية IRT / نموذج راش

1. تتطلب أحادية البعد (بعد واحد لكل مقياس) والاستقلال المحلي.
2. تتطلب عينات أكبر (خاصة مع نموذجي PL2 و PL3).
3. قد ينتج عن نموذج PL3 تقديرات غير مستقرة (صعوبة تقدير بارامتر التخمين c_i).
4. خطر التقييد المفرط: قد يؤدي اشتراط أحادية البعد الصارمة إلى إضعاف صدق المحتوى.
5. تعقيد التفسير، والاعتماد على افتراضات قوية.

الجدول رقم 6 : دراسات مقارنة بين النظريتين (Tate, 2003; Janssen et al., 2014).

Study	Instrument/Construct	Population	Model	Contribution
Rachmatullah et al. (2020)	CS Concept Inventory	457 Middle grades students	CTT & IRT	Validated 24 items for pre/post-testing.
Noroozi & Karami (2022)	Language test items	60 MA students/grads	Rasch	Linked cognitive load to difficulty estimates.
Youssef (2022)	Academic proficiency	477 University students	Rasch (ARSM)	Built a 48-item scale with validity.
DuPaul et al. (2020)	ADHD symptoms	2,079 children (parent) / 1,070 (teacher)	Rasch	Identified DIF across demographics.

Jin et al. (2024)	GenAI Literacy	355 Higher ed students	CTT & IRT	Developed robust 2PL instrument.
Bhatti (2023)	Innovation Skills	Undergraduate STEM students	IRT	Measured growth in affective domain.

التعليق على الدراسات

دراسات مثل تلك التي أجراها Rachmatullah et al. (2020) Jin et al. (2024) تدمج بشكل صريح كلا الإطارين لتعظيم دقة خاصية الصدق. وعلى العكس، تطبق دراسات مثل DuPaul et al (2020) Bhatti (2023) إطار راش أو IRT حصرياً لمعالجة تحديات القياس المحددة، مثل وظيفة العناصر التفاضلية (DIF) أو الدراسات الطولية. ركز E. C. Aybek 2023 بشكل خاص على الجسر المنهجي بين الأطر، حيث وجد أن الصعوبات المشتقة من CTT يمكن تحويلها إلى معلمات IRT ذات ارتباط عال في عينات تتجاوز 250 مشاركاً.

كما في الأدبيات الحديثة، وجد مجموعة حديثة من الدراسات، منها دراسات منشورة في 2023–2025، إضافة إلى بعض الدراسات الأقدم التي تستحق الإبقاء عليها لأنها شديدة الصلة بموضوع المقارنة.

الجدول رقم 6 : أهم الدراسات

الدراسة	السنة	المجال/العينة	المقارنة	أهم ما يفيد أطروحتك
Celen	2023*	تلاميذ الصفوف 6–8	CTT مقابل IRT	أظهرت تقارباً في النتائج مع اختلاف طريقة الوصول إليها
Ukofia & Essen	2023	5,192 استجابة في رياضيات UTME	CTT مقابل IRT/2PL	وجود أوجه تشابه واختلاف في صعوبة وتمييز المفردات
Aybek	2023	محاكاة/بيانات CAT	صعوبة CTT مقابل Rasch	إمكانية وجود علاقة قوية بين مؤشرات الصعوبة في النظريتين
Mutluer & Çakan	2023	بيانات PISA 2012 للرياضيات	CTT مقابل IRT في معادلة الاختبارات	مقارنة منهجية بين طرق التكافؤ في الإطارين
Santoso et al.	2023	تقييم مفاهيم الضوء	CTT مقابل IRT	تقارب عام بين استنتاجات النظريتين مع وجود مفردات تحتاج مراجعة
Ciptari et al.	2024	اختبار مفاهيم رياضية SPLDV	CTT مقابل Rasch	أظهرت النظريتان نتائج متقاربة في بعض الخصائص ومختلفة في أخرى

Priyani & Sugiharto	2024/25	اختبار أحياء	CTT مقابل Rasch	التكامل بين تحليل الصعوبة والتمييز والملاءمة والثبات
Polimi/HEAd	2025	اختبار فيزياء لطلبة ثانوي	CTT مقابل Rasch	توافق جيد بين الطريقتين، مع فائدة الجمع بينهما
Holland & Hoskens	2025	منهجية نظرية	CTT من منظور IRT	تقدم تفسيراً حديثاً للعلاقة النظرية بين IRT و CTT
Kartikasari & Sugiharto	2026	اختبار أحياء	CTT مقابل Rasch	مثال حديث جداً على استخدام النظريتين معاً في تحليل المفردات

تشير الأدلة التجريبية الحديثة إلى وجود قدر من التقارب بين مؤشرات النظرية الكلاسيكية ونظرية الاستجابة للمفردة في الحكم على عدد من خصائص الاختبار والمفردات، مع احتفاظ كل إطار بخصائص تحليلية مختلفة. وتظهر قيمة IRT بصورة خاصة عندما يكون الهدف هو تقدير خصائص المفردات وقدرات الأفراد على مقياس كامن مشترك، وفحص دقة القياس عبر مستويات القدرة المختلفة، في حين تظل CTT إطاراً عملياً وفعالاً للتحليل الأولي للمفردات والثبات والاتساق الداخلي. ولذلك فإن استخدام الإطارين بصورة تكاملية يمكن أن يوفر حجة أكثر شمولاً لجودة الاختبار من الاعتماد على أحدهما بصورة منفردة.

دقة القياس وتطور المقاييس

من حيث هذا الجانب فإن توفر نماذج IRT/Rasch مرونة فائقة في تطوير الاختبارات. نظراً لأن معايير البنود والأفراد توضع على نفس المقياس الكامن، يسمح IRT بحساب الأخطاء القياسية عند مستويات قدرة مختلفة، مما يوفر مستويات مختلفة من دقة القياس

. وهذا مفيد بشكل خاص في اختبار التكيف المحوسب (CAT)، حيث يتم اختيار العناصر لتناسب مع مستوى القدرة المحدد للمستجيب .

أما CTT يعتمد غالباً على مقاييس الاتساق الداخلية مثل ألفا كرونباخ، وغالباً ما ينظر إليه كطريقة أكثر "تقليدية" تقدم تقديراً عاماً للثبات بدلاً من معلومات محددة بعنصر.

الدقة والتطبيق العملي

يتم تطبيق النموذجين CTT و IRT بشكل متزايد لضمان أن تكون التقييمات ليست فقط سليمة سيكومترياً وأيضاً تنبؤاً بالأداء الواقعي .

تشير الأدلة الحديثة إلى أن:

- الدقة: تسمح الطرق المعتمدة على IRT بتقدير معايير صعوبة العنصر التي يمكن استخدامها في الاختبار التكيفي المحوسب (CAT)، مع تحسين عدد الفقرات المطبقة مع الحفاظ على دقة القياس

- التفسير العملي: من خلال ربط تقديرات الصعوبة بمؤشرات التشبع المعرفية — مثل زمن الاستجابة والصعوبة
- الملاحظة — يصبح الباحثون قادرين بشكل أفضل على تفسير المعنى السيكلوجي وراء معايير الفقرات
- درجة التعقيد: إن دمج الذكاء الاصطناعي والبيانات big data في القياسات النفسية يدفع الانتقال نحو التقييمات الشخصية في الوقت الحقيقي، رغم أن ذلك يتطلب اهتماما مستمرا بالمعايير الأخلاقية وخصوصية البيانات
- تعزز هذه المنهجيات القدرة على التمييز بين الأفراد ذوي القدرات العالية وأولئك الذين قد يتأثر أداؤهم بعوامل خارجية، مثل العشوائية التي تنطوي عليها الامتحانات الرسمية.
- . من خلال دمج هذه الأطر، يمكن للممارسين تصميم مقاييس أكثر عدالة وموضوعية، متجاوزين البيانات البسيطة التي يتم الادلاء بها ذاتيا إلى التحقق من صدق أكثر قوة قائم على الأداء

الفصل الثالث

الدراسة الميدانية

تمهيد

تصميم الاختبار التحصيلي

الدراسات الاستطلاعية

الدراسة الأساسية

تصميم اختبار التحصيل في الرياضيات

تمهيد

يهدف اختبار التحصيل الرياضي الذي تم تطويره لهذه الدراسة إلى تقييم القدرات الرياضية للتلاميذ في مختلف المجالات، بما في ذلك الاعداد والهندسة وتحليل البيانات. تم تصميم هذا الاختبار ليتماشى مع الأهداف التعليمية كما هو موضح في تصنيف TIMSS لتقييم الرياضيات.

نظرة عامة على الإطار النظري

Trends in International Mathematics (TIMSS)

and Science Study

تصنف صنف TIMSS محتوى الرياضيات والمجالات المعرفية لتوفير تقييم شامل لأداء التلاميذ. يتكون الإطار من مكونين رئيسيين:

- مجالات المحتوى: تمثل هذه الموضوعات الرياضية التي تم تقييمها.
- المجالات المعرفية: تمثل هذه أنواع التفكير المطلوبة لحل المشكلات في الرياضيات.

مجالات المحتوى

يتم تنظيم تقييم الرياضيات في TIMSS في أربعة مجالات محتوى رئيسية: الاعداد والحساب، الهندسة، القياس، التناسبية وتنظيم المعلومات.

1. الاعداد والحساب: يتضمن هذا المجال موضوعات مثل الأعداد الصحيحة والكسور العشرية والعلاقات بينها.

2. القياس: يشمل هذا المجال مختلف وحدات القياس والتحكم في العلاقة بينها.

3. الهندسة: يغطي هذا المجال المفاهيم المتعلقة بالأشكال والفضاء وخصائص الأشكال الهندسية.

4. التناسبية وتنظيم المعلومات: يشمل هذا المجال النسب وتمثيل البيانات وقراءتها وحل مشكلات.

المجالات المعرفية

تم تنظيم المجالات المعرفية في TIMSS لتعكس مستويات مختلفة من مهارات التفكير وحل المشكلات. لتشمل:

1. المعرفة: يركز هذا المستوى على تذكر الحقائق والمفاهيم الأساسية. من المتوقع أن يتعرف التلاميذ على المصطلحات والصيغ والإجراءات الرياضية ويتذكرونها.
2. التطبيق: في هذا المستوى، يستخدم التلاميذ معرفتهم الرياضية لحل المشكلات الروتينية. ويشمل ذلك تطبيق الصيغ والإجراءات على حالات محددة.
3. الاستدلال: يتطلب هذا المستوى من التلاميذ الانخراط في تفكير عالي المستوى. يقوم التلاميذ بتحليل وتفسير المعلومات الرياضية، وإجراء روابط بين المفاهيم، وتبرير تفكيرهم.

المواءمة مع الأهداف التعليمية

- يتوافق تصنيف TIMSS مع الأهداف والصناعات التعليمية المختلفة، بما في ذلك:
- صنافه بلوم: يمكن تعيين المستويات المعرفية ل TIMSS وفقا لتصنيف بلوم المنقح، حيث تتوافق كلمة "المعرفة" مع التفكير الأدنى (التذكر والفهم)، و "التطبيق" و "التفكير" يتماشى مع التفكير من الدرجة العليا (التطبيق والتحليل والتقييم والإبداع) (Anderson & Krathwohl ، 2001).
 - التقييم من أجل التعلم: يؤكد إطار صنافه TIMSS على أهمية تقييم ليس فقط المعرفة ولكن أيضا مهارات التطبيق والتفكير اللازمة لحل المشكلات في سياقات العالم الحقيقي.
- ولا ننسى ان نقدم تصنيفات اخرى تطبق في ميدان التقييم للأهداف التعليمية ومنها:

صنافه بلوم

تصنف أهداف التعلم إلى مستويات معرفية، مما يوفر إطارا منظما لتقييم نتائج تعلم الطلاب.

بالإضافة إلى تصنيف بلوم هناك تصنيف حديث لمارزانو.

يتميز تصنيف مارزانو الجديد: بأهمية العمليات ما وراء المعرفة والتنظيم الذاتي في التعلم، ويشجع الطلاب على مراقبة فهمهم وتطبيق استراتيجيات لتحسين مهاراتهم الرياضية (Marzano, 2007).

- عمق معرفة لويب (DOK) (Webb): يصنف إطار عمل DOK الخاص ب Webb المهام وفقا لتعقيد التفكير المطلوب، مما يضمن أن الاختبار التحصيلي يتضمن فقرات تغطي مجموعة من المستويات المعرفية (Webb ، 1997).

برنامج السنة الخامسة ابتدائي

تسمح الرياضيات باكتساب أدوات مفهوماتية وإجرائية مناسبة تمكن التلميذ من القيام بدوره بثقة وفاعلية، في محيط اجتماعي تتزايد متطلباته أكثر فأكثر وفي عالم يتحول باستمرار. إن الرياضيات حاضرة في المحيط الاجتماعي والاقتصادي والإعلامي والثقافي للإنسان أكثر من أي وقت مضى.

تساهم الرياضيات، مع المواد التعليمية الأخرى في تحقيق ملمح التلميذ، وتدريسها يرمي إلى تمكينه من اكتساب كفاءات قابلة للتحويل إلى مختلف المجالات (المدرسية، الحياة اليومية ...)، كما تسهم، بقدر كبير، في تطوير الكفاءات الخاصة بحل المشكلات والتواصل (التبليغ) وينتظر من تعلم الرياضيات تحقيق غرضين اثنين أحدهما ذو طابع تكويني ثقافي والآخر نفعي.

فبرنامج السنة الخامسة ابتدائي يندرج ضمن شبكة المفاهيم لسنوات التعليم الابتدائي الخمس، وكما هو الشأن في السنوات السابقة، تبنى المعارف كأدوات فعالة لحل المشكلات. ويعتبر نشاط حل المشكلات النشاط المفضل لتنمية سلوك البحث عند التلاميذ.

حل مشكلات في ميدان الأعداد والحساب تتعلق بـ:

- تعيين الأعداد (الطبيعية والعشرية والكسور).

- مقارنة الأعداد وترتيبها.

- الحساب على الأعداد بكل أنواعه.

- التعرف على علاقات حسابية بين أعداد متداولة.

3.2. حل مشكلات في ميدان التناسبية وتنظيم المعلومات والمتعلقة بـ:

- التناسبية وخاصيتي الخطية.

- تنظيم معلومات في جداول.

- قراءة جداول وبيانات بسيطة وتفسيرها.

3.3. حل مشكلات في ميدان الفضاء والهندسة تتعلق بـ:

- استعمال خريطة او مخطط لتقدير مسافات.

- التعرف على أشكال مستوية ووصفها وتسميتها ونقلها وإنشائها.

- تمييز علاقات وخصائص بعض الأشكال المستوية (التوازي والتعامد والتناظر).

- التعرف على المجسمات وملاحظتها ووصفها وتسميتها وصنعها وإنجاز مثيلات لها.

- تكبير وتصغير أشكال هندسية.

4.3. حل مشكلات في مجال القياس تتعلق بـ:

- استعمال أدوات ملائمة لقياس مقادير فيزيائية وهندسية.

- اختيار وحدات ملائمة لقياس مقادير فيزيائية وهندسية.

- التمييز بين مساحة ومحيط شكل هندسي.

- مقارنة مفهوم الحجم .

- تصنيف زوايا وترتيبها.

الجدول رقم 7: الكفاءات في نهاية السنة الخامسة

الأعداد والحساب	التناسبية وتنظيم المعلومات	الفضاء والهندسة	القياس
حلّ مشكلات متعلقة بقراءة وكتابة الأعداد الطبيعية والعشرية ومقارنتها وترتيبها والحساب عليها.	حلّ مشكلات متعلقة بالتناسبية والنسبة المئوية والسرعة والمقياس والتحويلات من وحدة قياس إلى أخرى وتنظيم معطيات عددية على شكل قوائم أو جداول.	حلّ مشكلات متعلقة ب: - التعليم في الفضاء. - العلاقات والخواص الهندسية. - وصف ونقل وإنشاء أشكال هندسية. - تكبير وتصغير أشكال هندسية مستوية.	حلّ مشكلات متعلقة ب: - وحدات القياس الاصطلاحية. - المساحة والمحيط. - إدخال مفهوم الحجم. - مقارنة زوايا وتصنيفها.

خطوات بناء جدول المواصفات للاختبار التحصيلي:

بدأ الطالب بتحديد عدد الأهداف التعليمية في منهاج الرياضيات للسنة الخامسة ابتدائي، حيث وجد

عدها 136 هدفا وعدد الدروس كذلك. وعليه بني جدول المواصفات على أساس هذه المعطيات.

1- يتم بناء الجدول ، ويتم تحديد الوزن النسبي لكل موضوع بقسمة عدد حصص كل موضوع على إجمالي

عدد الحصص ثم نضرب الناتج في مائة مع التقريب.

2- يتم تحديد الوزن النسبي للأهداف في كل مستوى بقسمة عدد أهداف المستوى على العدد الكلي

للأهداف وضرب الناتج في مائة.

3- احتساب عدد الأسئلة في كل مستوى لكل موضوع من المعادلة التالية:

عدد اسئلة الموضوع = العدد الكلي للأسئلة × الوزن النسبي لأهمية الموضوع × الوزن النسبي لأهداف

الموضوع

على أن يتم التقريب لأعداد صحيحة بشكل متوازن.

ومع اعتبار عدد الأسئلة في الاختبار 60 تمرينا يصبح جدول المواصفات على النحو التالي:

جدول 8 مواصفات الاختبار التحصيلي

عدد الفقرات	المجموع 100%	الاستدلال 22%	التطبيق 47%	المعرفة 30%	مستويات الاهداف
31	41%	9 % = 6	19 % = 6	13 % = 17	المجال الاعداد 40%
17	33%	7 % = 2	16 % = 4	10 % = 8	المجال القياس 34%
8	17%	3 % = 1	6 % = 2	4 % = 4	المجال الهندسة 12%
4	13%	3 % = 1	6 % = 1	4 % = 2	المجال التناسبية 13%
60	100%	22 % = 11	47 % = 13	31 % = 31	المجموع 100%

جدول (9) جدول المواصفات كاملا بعد التعديلات المقترحة من المحكمين

عدد الفقرات	فقرات الخور	المخاور
29 فقرة	من 1 إلى 22 44+35+34+ 51+50+46+45+	الأعداد والحساب
7 فقرات	23+27+28+29+30+31+ 33	الفضاء والهندسة
15 فقرة	24+25+26+32+36+37+38+3 9+ 40+41+42+43+47+ 48+49	القياس
4 فقرات	52+ 53+54+55+	التناسبية وتنظيم المعلومات

بعد اعداد الاختبار في صورته النهائية تم تقديمه الى مجموعة من المحكمين وعددهم 11، يتوزعون الى مفتشين

ومدراء واساتذة في التعليم الابتدائي من ذوي الخبرة تفوق 15 سنة من العمل.

التأكد من صدق محتوى الاختبار: تمّ عرض الصورة الأولى للاختبار، وجدول المواصفات، والاهداف التعليمية،

على (11) محكما ممن هم من ذوي الخبرة والاختصاص، خمسة من مفتشي الرياضيات في وزارة التربية والتعليم،

(2) مديرين من المدارس الابتدائية اربعة مدرّسين ممن يدّرّسون في المرحلة الابتدائية (الملحق 1)، حيث طلب

منهم الإجابة على اسئلة الاختبار للحكم على مدى ملائمة جدول المواصفات وفقرات الاختبار.

وفي ضوء تقديرات المحكّمين ومناقشة مقترحاتهم تمّ إجراء تعديلات بسيطة على فقرات الاختبار، حيث تمّ

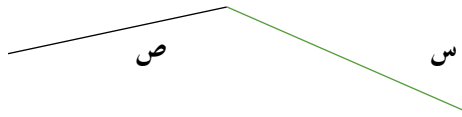
حذف 4 فقرات ، ليبقى الاختبار بصورته النهائية مكوّنًا من (55) فقرة.

الجدول رقم 10..: اتفاق المحكمين على فقرات الاختبار

عدد الفقرات المتفق عليها	عدد الفقرات المعدلة	عدد الفقرات المحذوفة
52	3	4

الفقرات المحذوفة:

1. ارسم قطعة على الخط المستقيم د بحيث يساوي طولها القطعتين س، ص



د

2. قطعة ارض مثلثة الشكل طول قاعدتها 80 مترا وارتفاعها 20 مترا اراد

صاحبها بناء منزل مستطيل الشكل طوله 20م وعرضه 15 مترا. ساعده في

امكانية تحقيق ذلك.

3. عندك ورقة نقدية من فئة 200 دج، وعندك قائمة من المأكولات حاول ان

تختار المجموعات التي يمكنك شراءها

المواد	الثلث	المواد	الثلث
مشروب غازي	25د	شطيرة بيتزا	50د
جبّ	59د	شكولاته	30د
برتقال	35د	كباب	70د

$$\text{●} \text{ ●} \text{ ●} = 27$$

4. اليك المساواة التالية ،

$$\text{■} \text{ ▲} = 12$$

$$\text{●} \text{ ▲} = 13$$

اوجد قيم الاشكال الثلاث

التجريب الأولي للاختبار:

بعد أن أصبح الاختبار بصورته الأولى جاهزا للتطبيق، تم تطبيق الاختبار على العينة الاستطلاعية عبر فترتين زمنيتين، وذلك بعد الزيارة الأولى لمدارس العينة المقصودة، والتي كان الهدف منها تعريف مد رسي ومدرسات مادة الرياضيات للصف الخامس الابتدائي في المدارس المعنية بالاختبار والاطلاع عليه، وآلية تطبيقه، وما يلزمها من إجراءات ومعلومات يستلزمها الاختبار ويُطلَّب تسجيلها من قبل المعلمين والمعلمات خلال عملية تطبيق الاختبار كوضوح تعليمات الاختبار، والتأكد من وضوح الصياغة اللغوية لل فقرات، وتحري الفقرات التي يتم التساؤل حولها للكشف عما إذا كان هنالك غموض فيها أم لا.

وبناءً على عملية تطبيق الاختبار على العينة التجريبية الاولى في فترتين حيث تم تحديد الزمن المناسب للإجابة على فقرات الاختبار، حيث وجد أن معدل (95 %) من التلاميذ تقريباً قد توقفوا عن الإجابة على فقرات الاختبار بعد مرور (90) دقيقة على بداية الاختبار، وعليه تم اعتماد زمن الاختبار (90) دقيقة؛ كما تبين وضوح فقرات، وتم تقسيم تطبيق الاختبار الى حصتين بسبب طول الفترة الزمنية، لكل فترة 45 دقيقة.

سيتم تصحيح الاختبار باستخدام نظام تسجيل ثنائي (1 نقطة للإجابة الصحيحة، 0 نقطة للإجابة غير الصحيحة). سيتم حساب الدرجة الإجمالية لتحديد مستويات التحصيل الإجمالية للتلاميذ.

الدراسة الاستطلاعية

طبق الاختبار على عينة الدراسة الاستطلاعية والتي كان عددها 75 تلميذ موزعين على ثلاث اقسام من دائرة سيق. تم تحليل النتائج للتأكد من الخصائص السيكو مترية (الثبات والصدق) وقد كان معامل الفا كرونباخ 0.82 ومعامل الصدق عن طريق المقارنة الطرفية $f(10.339), sig 0.000$ وبالتالي هناك فروق لصالح المجموعة العليا التي لها متوسط قدره 37.54 ومتوسط المجموعة الدنيا 18.65.

وبالتالي يمكن تطبيق الاداة في الدراسة الاساسية.

الدراسة الاساسية

نوع تصميم البحث

سيتم استخدام تصميم تجريبي لهذه الدراسة، مما يسمح بإجراء مقارنات بين الاطارين باستخدام كل من طرق IRT و CTT. سيتمكن هذا التصميم الطالب من تقييم فعالية كل نظرية في البيئات التعليمية، مع مراعاة المتغيرات المحتملة أيضا والتي تؤثر على النتائج.

3.1 مجتمع الدراسة

ستشمل الفئة المستهدفة لهذه الدراسة تلاميذ السنة الخامسة ابتدائي من مختلف المؤسسات التعليمية في ولاية معسكر، 374 مؤسسة وولاية وهران 402 مؤسسة للسنة الدراسية 2017/2016 وسيعزز هذا التنوع إمكانية تعميم النتائج.

3.2 طريقة أخذ العينات

سيتم استخدام طريقة أخذ العينات العشوائية لضمان التمثيل عبر المجموعات الديموغرافية المختلفة التوزيع الجغرافي. ستسمح هذه الطريقة بجمع عينة تعكس مجتمع الدراسي، وبالتالي تعزيز صحة النتائج (Creswell ، 2014).

سيتم تحديد حجم العينة بناء على تحليل القدرة لضمان القوة الإحصائية الكافية للكشف عن الاختلافات بين المجموعات. يوصى بحجم عينة لا يقل عن 300 مشارك لتحقيق تقديرات موثوقة

ونتائج قابلة للتعميم (Cohen, 1988). وعليه تم اختيار 600 تلميذ من دوائر ولاية معسكر وبعض دوائر ولاية وهران.

الجدول رقم: 11 المؤسسات التربوية حسب الولايتين

عدد التلاميذ	عدد الاقسام	المؤسسات التربوية حسب الولايتين	
71	2	ولاية معسكر	شيقري الحبيب
63	2	دائرة سيق	حسيبة بن بوعلي
68	2		بعلاوي محمد
35	1		مراد جلول
56	2	دائرة عقاز	جبور لخضر
26	1		مقني ميلود
31	1		مهني حلفاية سي عابد
76	2	ولاية وهران	سبيت لخضر
74	2		مالك بن نبي
69	2		مهور ادريس
579	17 مؤسسة	المجموع	10 مدارس

تطبيق الاختبار على عينة الدراسة: طبق الاختبار على عينة الدراسة والتي كان عددها (577) تلميذا، وقد تم التأكد من الإجابة على جميع الأسئلة قبل استلام الاختبار من التلاميذ، وصححت فقراته تمهيدا لتحليل البيانات وفق النظرية الكلاسيكية ونظرية الاستجابة للمفردة، من خلال البرامج الاحصائية

.SPSS20, WinSteps 3.1, JMetrik , FACTOR

استعمال الحاسوب في إدخال العلامات: تمّ تفريغ استجابات جميع أفراد العينة والبالغ عددهم (559)

تلميذا على فقرات الاختبار تمهيداً لاستكمال عملية تحليل البيانات . وفق النظريتين بدلا من 579،

حيث لم نتوصل على اوراق اجاباتهم لأسباب عدم اكتمال الاجابة على الاختبار.

الفصل الرابع

نتائج الدراسة

تحليل البيانات باستخدام نظرية الاختبار الكلاسيكية (CTT)

1. جمع البيانات: بعد إجراء الاختبار على عينة ممثلة.

الإحصاء الوصفي: حساب المتوسطات على مستوى الفقرة، والانحرافات المعيارية، وتوزيعات

الاستجابة؛ وحساب إجمالي درجات الاختبار لكل تلميذ.

2. تحليل الفقرات:

حساب معامل صعوبة الفقرات من خلال نسبة التلاميذ الذين يجيبون بشكل صحيح على الفقرة. والبنود

القريبة من 0 أو 1 تشير إلى صعوبة أو سهولة كبيرة بالترتيب.

حساب معامل تمييز الفقرات عبر ارتباطات كل بند والمجموع (من خلال معامل الارتباط المصحح، مما

يشير إلى مدى جودة التمييز بين أصحاب الدرجات العالية والمنخفضة.

3. تقدير الثبات: استخدم معاملات مثل ألفا كرو نباخ لتقييم ثبات الاتساق الداخلي للاختبار.

4. بينة الصدق: إجراء التحليل العاملي لتقييم خاصية أحادية البعد وفحص الارتباطات مع المحركات

خارجية لصدق المفهوم.

5. التفسير والتنقيح: قد يتم تعديل أو إزالة فقرات التي لها تمييز ضعيف أو صعوبة شديدة؛ تستخدم

الدرجات الإجمالية كتقديرات للقدرة (Devellis, 2006; Wilson et al., 2006).

الإحصاء الوصفي

- عدد التلاميذ 559
- عدد الفقرات 55
- متوسط الدرجة الكلية

$$\bar{X} = 27.92$$

- الانحراف المعياري:

$$SD = 10.22$$

وهذا يعني أن مستوى صعوبة الاختبار متوسط تقريباً لأن المتوسط قريب من منتصف الاختبار:

$$55/2 = 27.5$$

وهذا مؤشر جيد في القياس التربوي.

الجدول 10 إحصائيات الدرجة الكلية النظرية الكلاسيكية

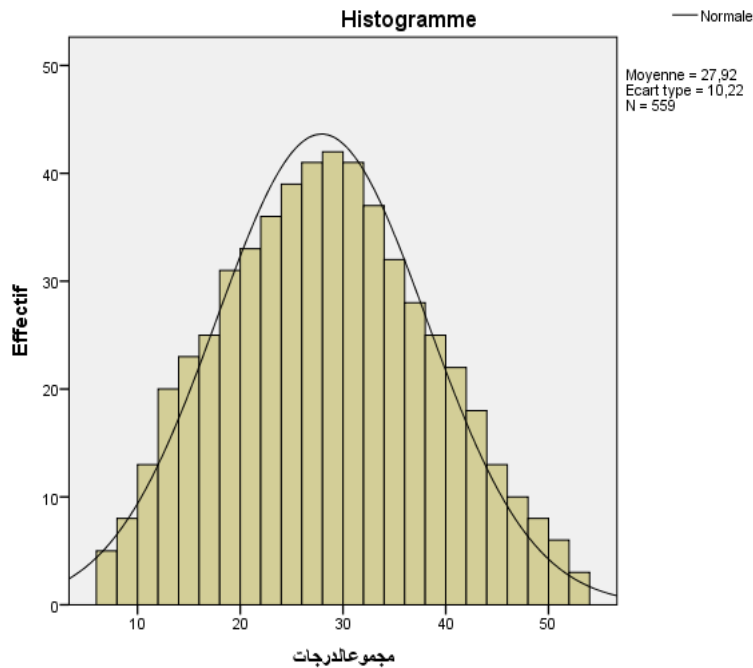
559	المجموع
27.1073	المتوسط
10.076	الانحراف المعياري
7	%25
20	%50
34	%75
52	أقصى درجة

تلخص الإحصائيات الوصفية الخصائص الرئيسية لمجموعة البيانات. تشمل المؤشرات التالية:

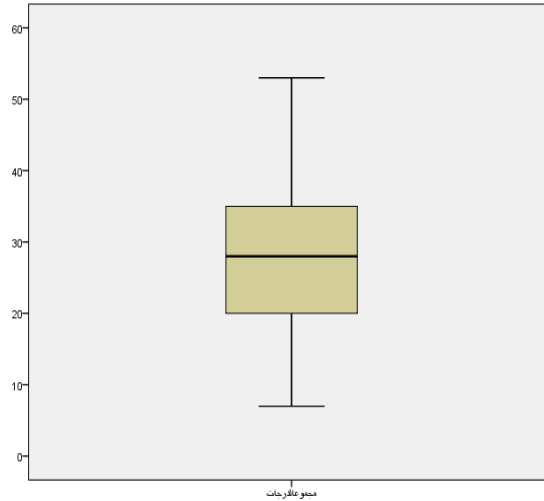
التوزيع الطبيعي: نلاحظ من الشكل 6 أن التوزيع للبيانات طبيعي. وبمتوسط قدره 27.92 وانحراف

معيارى قدره 10.22

التمثيل البياني 7: التوزيع الطبيعي للبيانات



وكذلك غياب الدرجات المتطرفة من خلال التمثيل التالي 8 :



- المتوسط: متوسط الدرجة والذي يوفر اتجاهها مركزيا للبيانات. وقيمتها 27.92 وهو متقارب من الوسيط 28.00 وهذا يدل التوزيع طبيعي.

- الانحراف المعياري: مقياس للتباين يشير إلى مدى انتشار الدرجات حول المتوسط. وهو 10.22

كما نلاحظ قيمتي التفلطح والالتواء 0.118 و -0.573

الجدول 11 الاحصاء الوصفي

Descriptives			Statistique	Erreur standard
مجموع الدرجات	Moyenne		27,92	,432
	Intervalle de confiance à 95% pour la moyenne	Borne inférieure	27,07	
		Borne supérieure	28,77	
	Moyenne tronquée à 5%		27,82	
	Médiane		28,00	
	Variance		104,440	
	Ecart-type		10,220	
	Minimum		7	
	Maximum		53	
	Intervalle		46	
	Intervalle interquartile		15	
	Asymétrie		,118	,103
	Aplatissement		-,573	,206

معامل الصعوبة وفق النظرية الكلاسيكية : p-value Item Difficulty

معامل صعوبة الفقرة: نسبة المستجيبين الذين يجيبون على الفقرة بشكل صحيح. تشير النسبة الأقل

إلى بند أكثر صعوبة. وفق الصيغة التالية :

$$P=Nc/Nt$$

التالية حيث تشير

P : معامل الصعوبة

Nc : عدد الاجابات الصحيحة

Nt : عدد الاجابات الكلية.

والذي يمتد من 0.97 الى 0.11 وبالتالي يمكن تحديد الفقرات السهلة والفقرات ذات معامل صعوبة

أكبر. فاننا نلاحظ من خلال الجدول (الملحق رقم 4) معامل صعوبة الفقرة تتنوع قيم p للفقرة بشكل

واسع، من الفقرات السهلة جدا مثل البند 11 بقيمة 0.957 p، البند 19 بقيمة 0.971 p إلى

الفقرات الصعبة جدا مثل البند 39 بقيمة 0.134 p، البند 41، البند 42، والبند 43 بقيمة p 0.111. هذا يشير إلى توزيع جيد في صعوبات الفقرات، وهو أمر مرغوب فيه عموما في الاختبار. وعليه يمكن حذف الفقرات التي يزيد معامل صعوبتها عن 0.9. وهي (19، 11، 20، 21، 22) والتي تقل عن 0.20 (39، 41، 42، 43)، إضافة إلى الفقرات التي لا تميز بين التلاميذ وهي (الفقرات 11، 12، 39) والتي يقل معامل تمييزها عن 0.30. وبالتالي يصبح عدد الفقرات 45 وهذا وفق النظرية الكلاسيكية.

الجدول رقم 12 معامل الصعوبة لأسهل الفقرات

الفقرات	معامل الصعوبة	معامل التمييز	الانحراف المعياري
19	0,9714	0,6647	0,1669
11	0,9571	0,0136	0,2029
22	0,9571	0,7113	0,2029
20	0,941	0,7777	0,2359
21	0,9374	0,7661	0,2425
5	0,898	0,562	0,3029
9	0,8623	0,3769	0,3449
55	0,8336	0,4173	0,3727
6	0,8247	0,4642	0,3806

الجدول رقم 13 معامل الصعوبة لأصعب الفقرات

الفقرات	معامل الصعوبة	معامل التمييز	الانحراف المعياري
52	0,2254	0,5633	0,4182
53	0,2254	0,5633	0,4182
48	0,2236	0,7939	0,417
49	0,2165	0,8	0,4122
46	0,2039	0,5101	0,4033
39	0,1342	0,2544	0,3411
41	0,1109	0,7612	0,3143
42	0,1109	0,7612	0,3143
43	0,1109	0,7612	0,3143

معامل التمييز وفق النظرية الكلاسيكية: معامل الارتباط المصحح بين الفقرة والدرجة

الكلية Point-Biserial Correlation :

- معامل تمييز الفقرة: يقيس مدى جودة تمييز الفقرة بين الدرجات العالية والمنخفضة في الاختبار

العام. والطريقة المتبعة هي الارتباط ثنائي التسلسل بين درجات الفقرات وإجمالي الدرجات. (الملحق).

معامل الارتباط -ثنائي التسلسلي: هذه هي الطريقة الأكثر استخداماً لحساب التمييز بين الفقرات

للعناصر ذات الدرجات الثنائية (مثل الصواب/الخطأ). يقيس العلاقة بين درجة التلميذ في فقرة واحدة

ومجموع نتيجته في الاختبار بأكمله (باستثناء ذلك العنصر).

الحساب: صيغة الارتباط -ثنائي التسلسلي

$$R_{pb} = \frac{(Mp - Mq)}{St \cdot \sqrt{p \cdot q}} \quad \text{صيغة الارتباط -ثنائي التسلسلي}$$

Mp: المتوسط: متوسط الدرجة الإجمالية للمقدمين للاختبار الذين أجابوا على السؤال بشكل صحيح.

Mq: متوسط الدرجة الإجمالية للمقدمين للاختبار الذين أجابوا على السؤال بشكل خاطئ.

ST: انحراف معياري لإجمالي درجات الاختبار.

p: نسبة المتقدمين للاختبار الذين أجابوا على السؤال بشكل صحيح (صعوبة المادة).

q: نسبة المتقدمين للاختبار الذين أجابوا على السؤال بشكل خاطئ $q = 1 - p$

نلاحظ ان ثلاث فقرات لم يكن لها معامل تمييز مقبول أي اصغر من 0.30 وهي 11، 12، 39.

وبالتالي يمكن حذفها. بينما باقي الفقرات ارتبطت ارتباطاً دالاً احصائياً بالدرجة الكلية.

ويفضل ان لا يقل معامل تمييز الفقرة عن 0.30 اذ تستبعد أي فقرة يكون معامل تمييزها اقل من

هذه النسبة، وتستبعد ايضاً أي فقرة معامل تمييزها سالباً مهما كانت نسبة التمييز، وكذلك إذا كان

المعامل للدرجات (1,0)

تختلف أيضاً ارتباطات التي تشير إلى معامل التمييز للفقرة بشكل كبير. بعض الفقرات تظهر تمييزاً جيداً

مثل الفقرة 24، البند 25، البند 26 جميعها حول 0.61-0.60، البند 33 عند 0.646 ، مما

يعني أنها تميز فعلياً بين التلاميذ ذوي القدرات العالية والمنخفضة. ومع ذلك، بعض الفقرات لديها تمييز منخفض مثل البند 11 عند 0.030، البند 12 عند 0.157، البند 39 عند 0.189. البند 11 الذي يحمل قيمة $p = 0.957$ والتمييز 0.030 يمثل مشكلة خاصة، مما يشير إلى أنه سهل جداً ولا يميز جيداً بين التلاميذ.

كما يمكن أن نحسب معامل التمييز من خلال الفرق بين عدد الذين أجابوا عن الفقرة إجابة صحيحة في المجموعة العليا. وبين عدد الذين أجابوا إجابة صحيحة عنها في المجموعة الدنيا مقسوماً على عدد أفراد مجموعة واحدة (عوده، 2004)

تقدير الثبات باستخدام المؤشرات الإحصائية

تقدير معامل الثبات في النظرية الكلاسيكية

يشير الثبات إلى اتساق أداة القياس. تشمل الطرق الشائعة لتقييم الثبات في النظرية الكلاسيكية CTT ما يلي:

كخطوة أولية لتقييم الاتساق الداخلي للاختبار، تم حساب معامل ألفا كرونباخ، والذي يمثل أحد أشهر مؤشرات الثبات في النظرية الكلاسيكية.

جدول 16 معامل الثبات وفق النظرية الكلاسيكية للاختبار

القيمة	معامل الثبات
0.918	Cronbach's Alpha
55	عدد المفردات
559	حجم العينة

تشير هذه القيمة إلى مستوى اتساق داخلي مرتفع جداً، حيث تتجاوز القيمة الحد المقبول في

الدراسات التربوية والذي غالباً ما يقدر بـ 0.70.

• معامل ألفا 0.918

• الخطأ المعياري للقياس (SEM): تعتمد هذه القيمة على الانحراف المعياري . ب 10.4 التفسير :

تشير القيمة 0.918 إلى استقرار ممتاز. وهذا يعني أن 91.8% من تباين الدرجات الملحوظة يعود

إلى القدرة "الحقيقية" للطلاب، و 8.2% فقط يعود إلى خطأ القياس العشوائي.

- التجزئة النصفية (الجدول 15): تتضمن هذه الطريقة تقسيم الاختبار إلى نصفين وتحديد العلاقة

الارتباطية بين النصفين مع التصحيح بمعادلة سبيرمان. حيث وجد 0.823

Statistiques de fiabilité			
Alpha de Cronbach	Partie 1	Valeur	.836
		Nombre d'éléments	28 ^a
	Partie 2	Valeur	.883
		Nombre d'éléments	27 ^b
	Nombre total d'éléments		55
	Corrélation entre les sous-échelles		.710
Coefficient de Spearman-Brown	Longueur égale		.830
	Longueur inégale		.830
Coefficient de Guttman split-half		.823	

في حين في الاختبارات محكية المرجع هناك أربع معاملات ارتباط تم حسابها عن طريق برنامج JMetrik ونلاحظ

أن كل المؤشرات أكبر من 0.9 وهذا يعطي الثقة في النتائج من عملية التصميم إلى تطبيق الاختبار وتطبيقه وعملية

التصحيح.

الجدول رقم 18 : معاملات الثبات وفق نظرية التعميم

Method	Estimate	95% Conf. Int.	SEM
Guttman's L2	0,9242	(0,9149, 0,9329)	2,8142
Coefficient Alph	0,9178	(0,9078, 0,9272)	2,9300
Feldt-Gilmer	0,9206	(0,9109, 0,9297)	2,8805
Feldt-Brennan	0,9203	(0,9106, 0,9294)	2,8854
Raju's Beta	0,9178	(0,9078, 0,9272)	2,9300

تشير هذه القيمة إلى أن مفردات الاختبار تقيس بقدر كبير من الاتساق البنية الكامنة نفسها، وهو ما يتوافق مع المعايير المتعارف عليها في القياس التربوي، حيث يعد معامل الثبات مقبولاً إذا تجاوز 0.70 وفق (Embretson & Reise, 2000).

كما تدل هذه النتيجة على أن الاختبار يتمتع بدرجة عالية من التجانس الداخلي، الأمر الذي يعزز من ثبات النتائج التي يتم الحصول عليها من خلال تطبيقه.

خاصية الصدق :

لدعم افتراضات الصدق حول الفقرات والاختبارات يجب توفر بيانات الأدلة النوعية والكمية من مراجعات الخبراء، والتصنيف المعرفي، وتحليل الفقرات الإحصائي ضرورية بشكل مشترك ، كما يوصى بإجراءات قائمة التحقق لضمان الجودة لتوثيق مصدر البند، والمراجعات، والتعديلات، ونتائج الاختبار الميداني، والقرارات المستخدمة لتضمنين أو استبعاد الفقرات في النماذج التطبيقية : كما يجب تحديد طرق تحليل الفقرات (المؤشرات الكلاسيكية : الصعوبة ، التمييز ، وظائف المشتت، وأنماط الاستجابة) لتحديد الفقرات غير النمطية للمحتوى أو المراجعة السيكو مترية .

اضافة الى ما سبق هناك الحساسية التعليمية يجب فحص حساسية التعليم وملاءمتها للمجال بشكل تجريبي حتى تظهر الفقرات العلاقات المتوقعة مع أهداف التعليم والتعلم، مما يعزز بيئة الصدق المتعلقة بالمحتوى.

كما سبق وان تم اجراء عملية التحكيم لفقرات الاختبار في الدراسة الاستطلاعية مما يحقق صدق المحتوى، اضافة الى حساب الصدق عن طريق الاتساق الداخلي في الدراسة الاساسية من خلال معامل الارتباط الفقرة بالدرجة الكلية. صدق الاتساق الداخلي

تم حساب معامل الارتباط بين كل فقرة والدرجة الكلية فكانت كل الفقرات مرتبطة بالدرجة الكلية ما عدا الفقرة 11 (0.026) بينما كان مجال الارتباط يمتد من الفقرة 12 ب 0.163 والفقرة 24 ب 0.62 وكلها دالة احصائيا عند مستوى 0.01. مما يدعم مصداقية الدرجات.(الملحق رقم:).

كما سيتم حساب خاصية الصدق عن طريق المقارنة الطرفية. وذلك بتقسيم العينة الى مجموعات ثلاث وتم أخذ المجموعة العليا (27%) بعد ترتيب التلاميذ من اعلى درجة الى ادناها وتم اخذ المجموعة الثانية (27%) والتي تضم أدنى الدرجات. وتم المقارنة بين متوسطي المجموعتين العليا والدنيا وفق احصائي t. يتضح من الجدول رقم (17.) والذي يبين وجود فروق بين متوسط درجات التلاميذ الذين تحصلوا على درجات مرتفعة في الاختبار ومتوسط درجات التلاميذ الذين تحصلوا على درجات منخفضة في الاختبار التحصيلي المطبق على عينة الدراسة الاستطلاعية في الدراسة الحالية، وذلك لصالح مرتفعي الدرجات (27%)، ويظهر ذلك من خلال ارتفاع متوسطات التلاميذ الذين ينتمون إلى المجموعة الأعلى في كل من الدرجة الكلية حيث تراوح متوسط التلاميذ (40.814) بينما المتوسط في المجموعة الدنيا (15.34)، نلاحظ أن المتوسطات مختلفة بين المجموعتين، لكن هل لهذا الاختلاف دلالة معنوية؟ لمعرفة إذا ما كان الاختلاف معنوي لا بد من اللجوء إلى اختبار "ت" لعينتين مستقلتين، وقبل حساب اختبار "ت"، لا بد من اختبار تجانس العينتين، وهو ما يوضحه الجدول ()، فقد كانت قيم اختبار "ت" كلها دالة عند (0.00)، كانت العينتين متجانسين بين المجموعتين (الأعلى والأدنى) أي تم اختيار اختبار "ت" لعينتين مستقلتين متجانستين حيث كانت قيمها وهي دالة عند (0.000) كما ذكر سابقا. أنها كلها اقل من مستوى الدلالة 0.05 ، فإننا نختار اختبار "ت" لعينتين غير متجانستين ، وتراوحت قيمة "ت" (0.04 : sig f=4.102) وكانت كلها دالة عند (0.04)، بمعنى الاختلاف بين المتوسطات له دلالة معنوية فعلا، أي توجد فروق ذات دلالة إحصائية لصالح مرتفعي الدرجات في الاختبار، إذا الاختبار يميز بين مرتفعي ومنخفضي الدرجات، وهو ما يدل على صدق المقياس وبالتالي يمكن تطبيقه على عينة الدراسة الأساسية .

جدول 17 احصائي المجموعتين

	VAR00003	N	Moyenne	Ecart-type	Erreur standard moyenne
VAR00002	دنيا عليا	151	15.3444	3.94976	.32143
		151	40.8146	4.74469	.38612

جدول 18 ت لعينتين مستقلتين

	Test de Levene sur l'égalité des variances		Test-t pour égalité des moyennes						
	F	Sig.	t	ddl	Sig.	Différence moyenne	Différence écart-type	Intervalle de confiance 95%	
								Inférieur e	Supérieur e
VAR00002	4.344	.038	-50.697	300	.000	-25.4702	.50240	-26.45	-24.48
			-50.697	290.44	.000	-25.4702	.50240	-26.45	-24.48

عرض النتائج وتفسيرها في ضوء نموذج راش ونظرية الاستجابة للمفردة

تمهيد

يهدف هذا الفصل إلى عرض النتائج التي تم التوصل إليها من خلال تحليل البيانات الخاصة باختبار الرياضيات المطبق على عينة الدراسة باستخدام كل من النظرية الكلاسيكية للاختبار CTT ونظرية الاستجابة للمفردة IRT ، مع التركيز بصورة أساسية على نتائج تطبيق نموذج راش Rasch Model 1PL .

وقد بلغ حجم العينة 559 تلميذاً، بينما اشتمل الاختبار على 55 فقرة من نوع الاختيار الثنائي 0،1. وقد تم تحليل البيانات باستخدام برامج القياس السيكو متري المتخصصة: Winsteps و JMetrik .

ويتضمن هذا الفصل عرضاً مفصلاً للنتائج المتعلقة بما يلي:

1. تقدير معلمات الفقرات وفق نموذج راش
2. تحليل إحصاءات المطابقة Fit Statistics للفقرات والافراد.
3. تحليل منحنيات خصائص الفقرات ICC
4. تحليل دوال معلومات الفقرات IIF
5. تحليل خريطة رايت المخصوص-الفقرة Person-Item Map / Wright Map
6. تحليل دالة معلومات الاختبار Test Information Function
7. تحليل الخطأ المعياري الشرطي للقياس Conditional SEM
8. فرضيات نموذج راش.

9. مناقشة الدلالات السيكومترية للنتائج

قبل البدء في عملية التحليل لا بد من التأكد من قابلية البيانات للتحليل وفق نموذج راش.

UMEAN=.0000 USCALE=1.0000

Item RAW SCORE-TO-MEASURE CORRELATION = .84

30745 DATA POINTS. LOG-LIKELIHOOD CHI-SQUARE: 27658.61 with 30132 d.f. p=1.0000

Global Root-Mean-Square Residual (excluding extreme scores): .3841

في برنامج : WINSTEPS هناك مصطلحين يعتمد عليهما التحليل الاحصائي:

الدرجة الخام = تقييم الشخص بدرجة خام

القدرة = تقدير راش للقدرة (θ)

يجب أن تكون هذان الاحصائيان مرتبطين بشكل إيجابي قوي.

وفق: (Linacre (2023)

المتوقع: من 0.80+ إلى 0.99+

وفق المعطيات الأولية وجد معامل الارتباط بين الاحصائيين 0.74 وهو $< 0.70+$ والتي تعتبر

التوافق جيد

تشير هذه القياسات قياسات للأفراد مركزها عند 0 لوجيت يتم تعيين الانحراف المعياري إلى 1 والتفسير هذا توحيد

خاص بنموذج راش يسهل: مقارنة الافراد والفقرات من جهة والاستقرار العددي للتقدير من جهة اخرى.

درجة البيانات 30745 المعنى إجمالي عدد الاجابات المقبولة = (عدد الأشخاص × عدد الفقرات)

فحجم العينة كاف جدا دقة إحصائية جيدة.

LOG-LIKELIHOOD CHI-SQUARE: 27658.61 with 30132 d.f. p = 1.0000

مربع كاي الاحتمالي اللوغاريتمي 27658.61: مع درجة حرية 30132 وقيمة $p=1.00$

تشير χ^2 ان الكفاية العامة لنموذج راش للبيانات والتفسير $p = 1.0000$ ملائمة مثالية أو ملائمة زائدة، مع الانتباه ان مع العينات الكبيرة جداً، يصبح هذا الاحصائي غير مفيد توصي ادبيات WINSTEPS بتفضيل الاحصائيات: المطابقة الداخلية والخارجية والجذر التربيعي لمربع البواقي infit, outfit, RMSR .
الجذر التربيعي لمربع البواقي هو الفرق المتوسط بين: الاجابات الملاحظة والاجابات المتوقعة من النموذج والقيم المرجعية > 0.40 ملائمة جيدة وبين $0.40 - 0.50$ مقبول اما < 0.50 فهناك مشكلة . هنا: 0.3841 ملائمة عامة جيدة.

1. تقدير معلمات الفقرات وفق نموذج راش

تحليل صعوبة الفقرات Item Difficulty

تم حساب:

- نسبة الإجابة الصحيحة p-value
- صعوبة الفقرة باللوغيت

$$\delta = \ln\left(\frac{1-p}{p}\right)$$

الجدول 21 أسهل الفقرات في الاختبار

الفقرة	نسبة النجاح	الصعوبة Logit
الفقرة 19	0.97	-3.52
الفقرة 11	0.96	-3.10
الفقرة 22	0.96	-3.10
الفقرة 20	0.94	-2.77
الفقرة 21	0.94	-2.71

هذه الفقرات سهلة جداً وذلك بسبب:

• ضعف التمييز

• انخفاض معلومات الاختبار عند القدرات العالية.

الجدول رقم 22 : أصعب الفقرات

الفقرة	نسبة النجاح	الصعوبة Logit
الفقرة 46	0.20	1.36
الفقرة 39	0.13	1.86
الفقرة 42	0.11	2.08
الفقرة 41	0.11	2.08
الفقرة 43	0.11	2.08

Item يعتمد نموذج راش على تقدير معلمة واحدة لكل فقرة وهي معلمة الصعوبة

Difficulty Parameter

ويتم التعبير عن احتمال الإجابة الصحيحة عن الفقرة وفق المعادلة:

$$P_{ni} = \frac{e^{(\theta_n - b_i)}}{1 + e^{(\theta_n - b_i)}}$$

حيث:

P_{ni} = احتمال أن يجيب الشخص اجابة صحيحة عن الفقرة i

θ_n = قدرة الشخص

b_i = صعوبة الفقرة

وقد تم تقدير معلمات الصعوبة باستخدام مقياس اللوجيت Logit Scale ، وهو المقياس القياسي

المستخدم في نماذج نظرية الاستجابة للمفردة.

ولو اخذنا على سبيل المثال الدرجات الخام التالية:

الجدول 23: تحويل الدرجة الخام الى اللوجيت

الدرجة الخام	القدرة بوحدة اللوجيت
5	-2.16
10	-1.46
20	-0.47
28	0.00
35	0.47
45	1.46
50	2.16

نلاحظ أن:

- العلاقة غير خطية
- الدرجات المنطرفة تتحول إلى قيم لوجيت كبيرة، كالدرجات 5 و 50.

3تفسير سيكو متري: تشير النتائج الأولية إلى ما يلي:

1وجود فقرات سهلة جداً

مثل: الفقرات 19، 11، 22.

هذه الفقرات قد تكون:

- منجزة في القسم مسبقاً
- أو لا تتطلب مهارة عالية.

2 وجود فقرات صعبة جداً

مثل الفقرات: 41، 42، 43.

قد تشير إلى:

- سوء صياغة
- محتوى غير مدروس
- أو مهارة أعلى من مستوى التلاميذ.

الاستنتاج الأولي

الاختبار يظهر توزيع الصعوبة بشكل جيد نسبياً لكن يحتاج مراجعة لبعض الفقرات المتطرفة.

مصفوفة اتخاذ القرار اتجاه الفقرات (وفق نموذج راش)

القاعدة الثلاثية: معال الارتباط × ملاءمة الفقرات × عدم مطابقة الافراد

(Triangulation : PTMEA CORR × Item Fit × Person Misfit)

مصفوفة اتخاذ القرارات حول الفقرات (استناداً على نموذج راش). مع الأخذ بعين الاعتبار ثلاث عوامل:

معامل الارتباط، ملاءمة الفقرات، التلاميذ غير ملائمين.

القاعدة المنهجية لا يجب الحكم على أي فقرة بناء على مؤشر واحد. تستند القرارات إلى تقارب عدة

تشخيصات في نموذج راش.

الجدول 24: معايير اتخاذ القرار لحذف الفقرات

المؤشر	التفسير	القرار
معامل الارتباط $0.20 \leq$	العلاقة إيجابية بين الفقرة والقدرة	فقرة وظيفية
معامل الارتباط (0.1- 0.19)	تمييز محدود	فحص التفاعل بين التلاميذ
معامل الارتباط سالب	اجابات متناقضة	يجب فحص الحساسية
ملاءمة الفقرة الداخلي والخارجي (0.5 الى 1.5)	ملاءمة مقبولة	الحفاظ على الفقرات
ملاءمة الفقرة الداخلي والخارجي $(1.5 \leq)$	تغاير مفرط	إمكانية المراجعة
عدم وجود عدم التلاؤم لأشخاص	وجود تشويه	يجب إعادة النظر
يوجد عدم تلاؤم للأفراد	احتمال مشكلة حول الفقرة	مراجعة او حذف

بالاستناد على ما مضى (الجدول 4.1 الخاص بمخطط ملاءمة الافراد، والجدول 10 الخاص بمعاملات الارتباط) فإن التوافق العام جيد.

الجدول 25 ترتيب الفقرات

مجموعات الفقرات	التشخيص	القرار
اغلب الفقرات	معاملات الارتباط موجبة + ملاءمة جيدة	الحفاظ عليها
مجموعة فقرات	معاملات الارتباط ضعيفة او سالبة + وجود عدم الملاءمة	الاحتفاظ بها مع المراقبة
ليس هناك فقرات	معاملات الارتباط سالبة + عدم المطابقة بدون وجود عدم ملاءمة الافراد	حذف غير مبرر

القرار النهائي: لا ينبغي استبعاد أي فقرة فقط بناء على نتيجة معامل الارتباط PTMEA سلبي بمفرده.

مناقشة معاملات الصعوبة وفق النظريتين

من خلال ما سبق نلاحظ ان هناك علاقة ارتباطية قوية جداً وعكسية سالبة بين البيانات وفق النظريتين:

1. تحديد البيانات

- البيانات في الجدول 12) تمثل صعوبة الفقرة في النظرية الكلاسيكية CTT وتحديدًا "نسبة الإجابة

الصحيحة. Proportion Correct / P-value "

- امثلة: الفقرة item11 قيمتها 0.957، وفي ملف Winsteps حصلت على 535 درجة من

559، أي $0.957 = 559/535$ وبالتالي تصبح الصعوبة (-3.62)

كذلك الفقرة item41 قيمتها 0.11، وفي الملف حصلت على 62 درجة من 559، أي

$$0.11 = 559/62 \text{ وفي نموذج راش تصبح الصعوبة } (2.77)$$

○ كلما كانت القيمة أعلى في هذه القائمة، كانت الفقرة أسهل بالنسبة للنظرية الكلاسيكية والعكس بالنسبة لنموذج راش.

• بيانات نموذج راش (الجدول رقم 10) يمثل العمود MEASURE صعوبة الفقرة في نموذج راش Rasch Difficulty بوحدة اللوجيت.

○ كلما كانت القيمة أعلى موجبة، كانت الفقرة أصعب.

○ كلما كانت القيمة أدنى سالبة، كانت الفقرة أسهل.

طبيعة العلاقة

العلاقة بين "نسبة الإجابة الصحيحة" و "صعوبة راش" هي علاقة لوغاريتمية عكسية.

• علاقة عكسية: Negative Correlation

○ عندما تكون نسبة الإجابة الصحيحة عالية جداً فقرة سهلة، تكون صعوبة راش منخفضة جداً قيم سالبة كبيرة.

○ عندما تكون نسبة الإجابة الصحيحة منخفضة جداً فقرة صعبة، تكون صعوبة راش مرتفعة جداً قيم موجبة كبيرة.

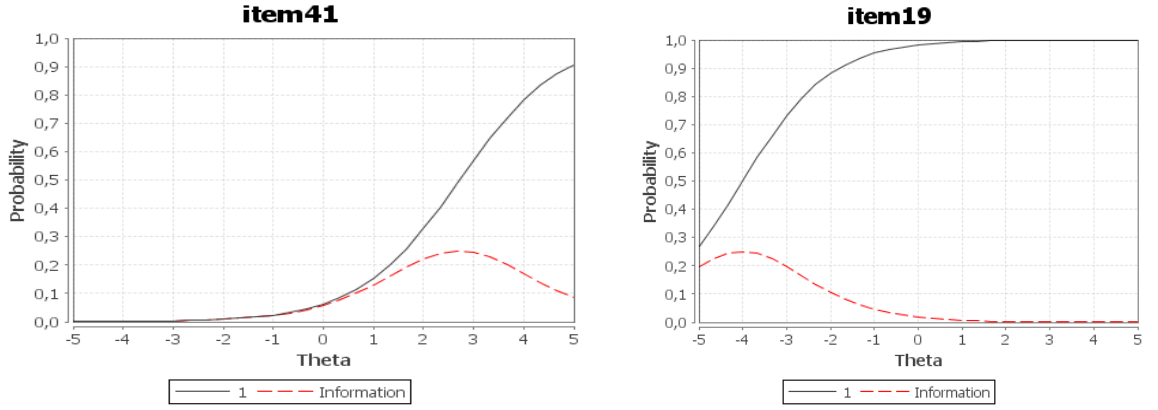
3. أمثلة توضيحية من البيانات يمكننا ملاحظة هذه العلاقة بوضوح شديد عند مقارنة الفقرات المتطرفة

الأسهل والأصعب: الجدول 24 مقارنة معامل الصعوبة وفق النظريتين

الفقرات	صعوبة نموذج راش MEASURE	صعوبة النظرية الكلاسيكية القائمة
19	-4.07 سهولة شديدة	0.97 أسهل فقرة تقريباً
11	-3.62 سهولة جدا	0.95 سهلة جداً
41	+2.77 صعوبة عالية	0.11 صعبة جداً
42	+2.77 صعوبة عالية	0.11 صعبة جداً

2	0.20 قريبة من الصفر	0.49 متوسطة
---	---------------------	-------------

التمثيل البياني: 9 و 10 خصائص الفقرات 19 (سهل فقرة) و 41 (صعب فقرة)



الاستنتاج الرياضي

رياضياً، يحول نموذج راش نسبة النجاح إلى صعوبة عبر المعادلة اللوغاريتمية :

$$\ln \frac{p}{1-p} = \text{معامل الصعوبة}$$

هذا يعني أن الارتباط بين الاحصائيين للصعوبة سيكون قريباً جداً من (1-) أي أن هناك تطابقاً كبيراً

بين ترتيب الفقرات من حيث السهولة والصعوبة في كلتا النظريتين، مع اختلاف وحدة القياس، الوحدة

في راش "لوجيت" تسمح بقياس دقيق أكثر من النسبة المئوية.

معامل الثبات وفق نموذج راش

ولحساب معدل ثبات الاختبار محكي المرجع عند تدريبه بنموذج راش تم حساب معامل الثبات Item

(G) "المتعلق بالأفراد، والثبات المتعلق بمفردات الاختبار من خلال معامل الفصل، لكل منهما. وقد

بلغ معامل الثبات للأفراد (0.91)، ومعامل الفصل المتعلق بالأفراد (Separation Index)

(3.16)، وهذا يشير الى ثلاث مستويات من القدرات في الاختبار. ومعامل الثبات للمفردات

(0.99) في حين وصل معامل الفصل المتعلق بالمفردات (13.32) معنى هذا ان هناك 13 مستوى

من صعوبة الفقرات.

جدول رقم: 25 معامل الثبات من خلال نموذج راش

PERSON 558 INPUT 558 MEASURED INFIT OUTFIT									
	TOTAL		COUNT	MEASURE	REALSE	IMNSQ	ZSTD	OMNSQ	ZSTD
MEAN	27.9	55.0	.14	.38	1.01	.0	.97	-.1	
S.D.	10.2	.0	1.26	.06	.24	1.3	.48	1.0	
REAL RMSE .38 TRUE SD 1.20 SEPARATION 3.16 PERSON RELIABILITY .0.91									

ITEM 55 INPUT 55 MEASURED INFIT OUTFIT									
TOTAL	COUNT	MEASURE	REALSE	IMNSQ	ZSTD	OMNSQ	ZSTD		
MEAN	282.9	55	8.0	.00	.12	.99	.0	.97	-.2
S.D.	139.7	.0	1.69	.03	.15	2.8	.40	3.0	
REAL RMSE .13 TRUE SD 1.68 SEPARATION 13.32 ITEM RELIABILITY .0.99									

اضافة الى معامل الثبات لكرونباخ ، فان أدبيات نظرية الاستجابة للمفردة ونموذج راش هناك ثبات

خاص بالفقرات للاختبار ومعامل ثبات خاص بالأفراد، اضافة الى تحديد مستويات الصعوبة لفقرات

الاختبار ومستويات القدرة للتلاميذ.

اضافة الى ذلك يوفر نموذج راش معاملين آخرين للثبات: الثبات الهامشي والثبات الشرطي.

الثبات الهامشي في نموذج راش

إلى جانب معامل ألفا كرونباخ، تم تقدير الثبات وفق إطار نظرية الاستجابة للمفردة باستخدام معامل

الثبات الهامشي (Marginal Reliability) ، والذي يعتمد على العلاقة بين تباين تقديرات

القدرة ومتوسط مربعات أخطاء القياس.

جدول 28. تقدير الثبات الهامشي للاختبار

المؤشر	القيمة
تباين القدرة	0.73
متوسط مربع الخطأ المعياري	0.097
الثبات الهامشي	0.87

تشير قيمة الثبات الهامشي إلى أن الاختبار يتمتع بدرجة مرتفعة من الدقة في تقدير القدرة الكامنة للأفراد. ويعني ذلك أن الجزء الأكبر من التباين في تقديرات القدرة يعود إلى فروق حقيقية بين الأفراد وليس إلى أخطاء القياس.

وقد أشار Bond & Fox إلى أن معاملات الثبات التي تتجاوز 0.80 تعد مؤشراً قوياً على جودة القياس في نماذج راش. (Bond & Fox, 2015)

2. الثبات الشرطي (Conditional Reliability)

هذا هو المفهوم الأكثر دقة وحداثة تعريفه: هو معامل ثبات "متغير" يتغير بتغير مستوى قدرة الطالب .

- يشير أين يكون الاختبار قوياً وأين يكون ضعيفاً. فمثلاً، قد يكون الثبات عند التلاميذ المتوسطين

0.90 (ممتاز)، بينما عند التلاميذ المتفوقين جداً ينخفض إلى 0.50 (ضعيف)

يفيد في "تحديد حدود النتائج" و "بيئة سيرورات الإجابة"؛ لأنه يثبت أن دقة القياس مشروطة بمستوى

القدرة. الجدول رقم 29: مقارنة بين الثبات الهامشي والشرطي

وجه المقارنة	(Conditional) الثبات الشرطي	(Marginal) الثبات الهامشي
كمية المعلومات	قيم متعددة تختلف باختلاف مستوى القدرة	قيمة ثابتة واحدة لكل التلاميذ
دقة المعلومات	"ما مدى دقة القياس عند مستوى قدرة معين؟"	"هل الاختبار بشكل عام جيد؟"
العلاقة بمتغيرات	كل مستوى قدرة يقابله ثبات	هو المتوسط الحسابي للقيم الموجودة

على الرغم من أن الثبات الهامشي للاختبار مقبول (0.87)، إلا أن تحليل الثبات الشرطي كشف عن

تباين في دقة القياس؛ حيث تزداد درجة الثبات في نطاق القدرات المتوسطة وتنخفض عند الأطراف،

مما يستوجب الحذر عند تفسير درجات التلاميذ ذوي الأداء المتميز جداً أو المنخفض جداً. "

هذا النوع من التحليل يرفع من القيمة العلمية ويظهر فهماً عميقاً لخصائص القياس وليس مجرد سرد

للأرقام

وصف "دالة المعلومات (Information Function) المرتبطة بهذا الثبات:

الجدول 30 الثبات الشرطي لمستويات القدرة

القدرة (Theta)	الثبات الشرطي	القدرة (Theta)	الثبات الشرطي	القدرة (Theta)	الثبات الشرطي	القدرة (Theta)	الثبات الشرطي
2.58	0.825	0.05	0.898	-2.47	0.815	-5.00	0.484
2.68	0.816	0.15	0.899	-2.37	0.822	-4.90	0.505
2.78	0.807	0.25	0.899	-2.27	0.828	-4.80	0.525
2.88	0.797	0.35	0.899	-2.17	0.834	-4.70	0.544
2.98	0.786	0.45	0.899	-2.07	0.839	-4.60	0.563
3.08	0.774	0.56	0.899	-1.97	0.844	-4.49	0.582
3.18	0.762	0.66	0.899	-1.87	0.849	-4.39	0.599
3.28	0.749	0.76	0.898	-1.77	0.853	-4.29	0.616
3.38	0.735	0.86	0.897	-1.67	0.858	-4.19	0.633
3.48	0.720	0.96	0.895	-1.57	0.862	-4.09	0.648
3.59	0.704	1.06	0.894	-1.46	0.866	-3.99	0.663
3.69	0.687	1.16	0.892	-1.36	0.869	-3.89	0.678
3.79	0.669	1.26	0.890	-1.26	0.873	-3.79	0.691
3.89	0.651	1.36	0.887	-1.16	0.876	-3.69	0.704
3.99	0.632	1.46	0.885	-1.06	0.879	-3.59	0.717
4.09	0.612	1.57	0.881	-0.96	0.881	-3.48	0.728
4.19	0.591	1.67	0.878	-0.86	0.884	-3.38	0.739

4.29	0.570	1.77	0.874	-0.76	0.886	-3.28	0.750
4.39	0.548	1.87	0.870	-0.66	0.888	-3.18	0.759
4.49	0.526	1.97	0.865	-0.56	0.890	-3.08	0.769
4.60	0.504	2.07	0.860	-0.45	0.892	-2.98	0.778
4.70	0.481	2.17	0.854	-0.35	0.894	-2.88	0.786
4.80	0.458	2.27	0.848	-0.25	0.895	-2.78	0.794
4.90	0.435	2.37	0.841	-0.15	0.896	-2.68	0.802
5.00	0.412	2.47	0.833	-0.05	0.897	-2.58	0.809

هذه البيانات تؤكد أن أعلى ثبات للاختبار يقع عندما تكون قدرة التلاميذ قريبة من (0.35) لوجيت

(بثبات قدره 0.899، وهو مؤشر ممتاز على جودة القياس في منطقة الوسط .

كما ان الثبات الشرطي مرتبط بالقدرة باختلاف موقعها على متصل القدرة فانه مرتبط بدالة المعلومات

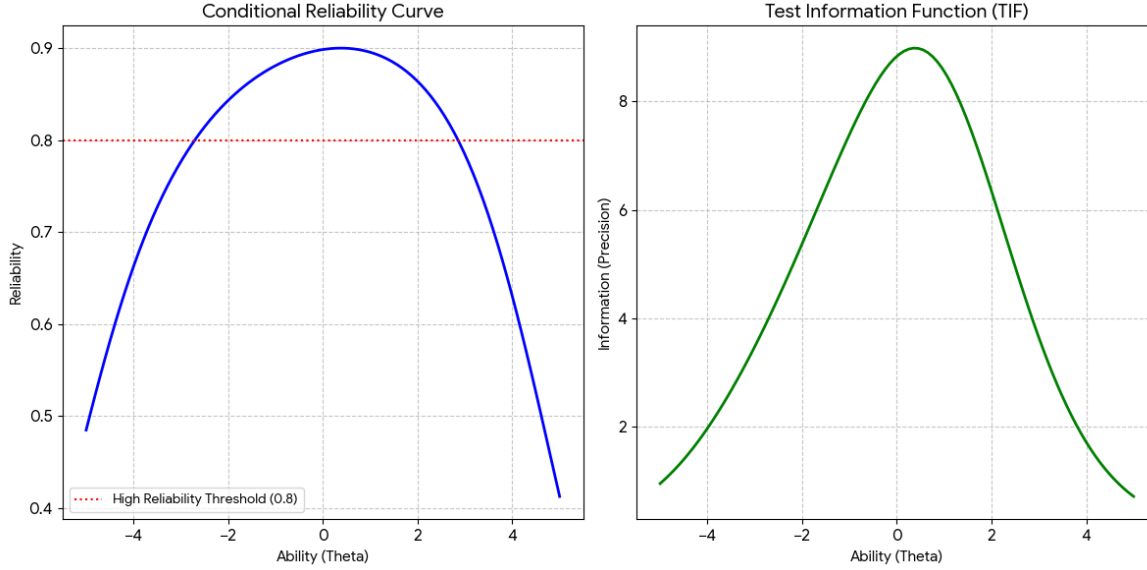
الاختبار

فمن خلال الشكل:

التقييم: الوضعية الأمثل - الفقرات موزعة بالتساوي عبر متصل القدرة

التمثيل البياني 9: دالة المعلومات للاختبار

التمثيل البياني 10: للثبات الشرطي



المنحنيات البيانية (Reliability & Information Curves).

يظهر الرسم أعلاه منحنين متكاملين:

- المنحنى الأزرق (الجهة اليسرى) الثبات الشرطي: يوضح كيف يتغير ثبات الاختبار بناءً على قدرة التلميذ نلاحظ أنه يتجاوز عتبة الـ 0.80 (الخط الأحمر المنقط) في النطاق ما بين قدرة -2.5 و +2.5 لوجيت تقريباً .
- المنحنى الجهة اليسرى دالة معلومات الاختبار: (TIF - يعبر عن مقدار "المعلومات" أو "الدقة" التي يوفرها الاختبار عند كل مستوى . كلما ارتفع المنحنى، قل الخطأ المعياري للقياس (SEM).

2. تفسير دالة المعلومات (Test Information Function).

بناءً على البيانات المستخرجة والتحليل البياني، يمكننا استنتاج :

- ذروة القياس (Peak Precision): يحقق الاختبار أقصى قدر من المعلومات (بقيمة 8.988) عند مستوى قدرة 0.35 لوجيت . هذا يعني أن الاختبار مصمم بدقة عالية جداً لتمييز الطلاب ذوي القدرة المتوسطة والمائلة قليلاً نحو الارتفاع .

- نطاق الفعالية: يوفر الاختبار معلومات غنية وموثوقة (ثبات $0.85 <$) في المدى الواقع بين -1.5 و 2.2 لوجيت. هذا النطاق يغطي غالبية التلاميذ في عينة الـ 559 مشاركاً.
- انخفاض المعلومات عند الأطراف: تلاحظ هبوطاً حاداً في المنحنى عند القدرات المتطرفة (أقل من -4 أو أكثر من +4).
- التفسير السيكمومتري: هذا الهبوط يعكس "فجوة في المعلومات"؛ فالاختبار يفتقر لفقرات سهلة جداً لقياس الضعاف بدقة، أو فقرات صعبة جداً لتحدي المتميزين. وهذا يدعم تماماً ما اشرنا اليه حول "تحديد حدود النتائج".

الاستنتاج:

"أظهرت نتائج دالة معلومات الاختبار (TIF) المستندة إلى نموذج راش أن الاختبار يوفر أعلى قدر من الدقة الإحصائية عند مستوى قدرة 0.35 لوجيت، حيث بلغ معامل الثبات الشرطي في هذه المنطقة ذروته بنسبة 0.90. (ويشير شكل المنحنى الجرسى المتماثل إلى أن الاختبار يتمتع بكفاءة قياس عالية في تشخيص المستويات المتوسطة، بينما تنخفض قدرته المعلوماتية عند المستويات المتطرفة، مما يعزز بيئة الصدق المتعلقة بملاءمة الاختبار للفترة المستهدفة وحدود تعميم نتائجه على الحالات الشاذة من القدرات".

بهذا التحليل، نكون قد تم الربط بين الأرقام الصماء وبين التفسير النظري لصدق الاختبار.

2تقدير القدرةPerson Ability

تم تحويل الدرجة الخام إلى لوجيت باستخدام:

$$\theta = \ln\left(\frac{r}{I - r}\right)$$

مع تصحيح:

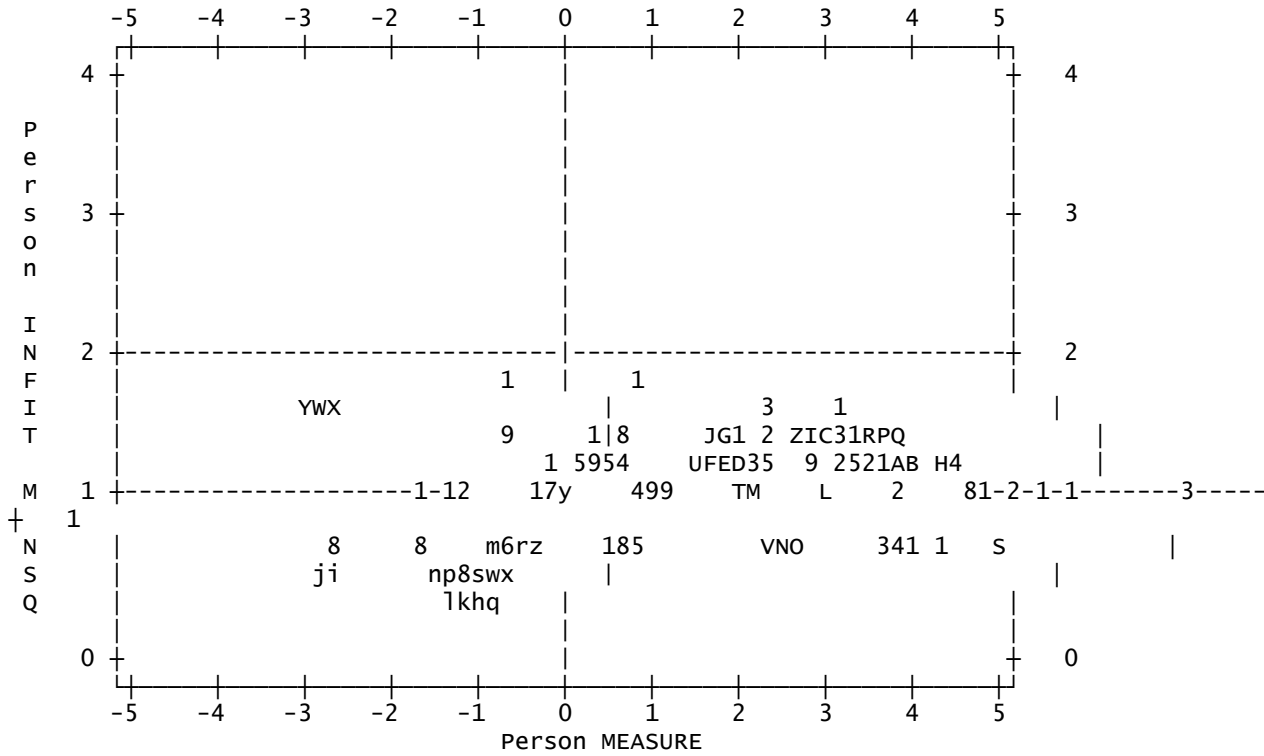
$$\theta = \ln\left(\frac{r + 0.5}{I - r + 0.5}\right)$$

حيث:

• الدرجة الخام

• $I = 55$

TABLE 4.1 C:\winsteps\doct25.txt ZOU935WS.TXT Feb 6 15:13 2026
INPUT: 559 Person 55 Item REPORTED: 559 Person 55 Item 2 CATS WINSTEPS 3.72.3



Item	1	2	11	1	1	212	13	1122231153342	32	1	3			
			T			S		M	S		T			
%TILE	0			10		20	30	40	50	60	80	90	99	

مخطط ملائمة الافراد (WINSTEPS)

هو مخطط يشير الى مربع متوسط التطابق الداخلي للافراد . Person INFIT Mean-Square

Plot.

الهدف منه:

• اظهار ملائمة الافراد لنموذج راش

• اكتشاف التلاميذ ذوي الاجابات الشاذة (تخمين، عدم انتباه، استراتيجيات غير متسقة)

على عكس خريطة Wright، هنا لا ننظر إلى الفقرات، بل إلى صدق الاجابات الفردية .

المحور العمودي: مربع متوسطات الملاءمة (MNSQ)

- من المتوقع ≈ 1.0

- المجال المقبول:

✓ $1.3 - 0.7$ (صارم)

✓ $1.5 - 0.5$ (شامل، مقبول في التحصيل الدراسي).

المحور الأفقي: مقاييس الافراد (logits)

- مستوى قدرة التلاميذ

- من منخفض جدا (-5) إلى مرتفع جدا ($+5$)

كل علامة في الشكل تعبر عن تلميذ أو أكثر.

المساحة الوسطى تمثل توزيع الافراد طبيعي حيث نلاحظ ان هناك تركيز عال من التلاميذ:

- مطابقة داخلية $\approx 0.8 - 1$.

- القدرات بين -2 و $+2$ لوجيت

التفسير

- الغالبية العظمى من التلاميذ يجيبون بإجابات متسقة

- نموذج راش يحظى بتطابق كبير على مستوى التلاميذ

هذا مؤشر ممتاز على الصدق الداخلي.

التلاميذ غير متلائمين مع النموذج حيث ($INFIT > 1.5$)

نلاحظ ان هناك عدة رموز فوق 1.5 ، وخاصة:

- من 0 إلى $+2$ لوجيت

- بعض الحالات القريبة من $INFIT \approx 2$

وهذا يشير الى ان التلاميذ أظهروا

- اجابات غير متوقعة
 - أخطاء في الفقرات السهلة
 - اجابات صحيحة على الفقرات الصعبة .
 - والاسباب الممكنة
 - الإجابات العشوائية
 - سوء فهم التعليمات
 - الغش أو استراتيجيات غير منتظمة
- لكن عددهم يبقى صغيرا مقارنة ب $N = 559$.
- الأشخاص ذوي الافراط في الملاءمة ($INFIT < 0.5$)
- Personnes en sur-ajustement ($INFIT < 0.5$)
- هناك بعض النقاط منخفضة جدا:
- $INFIT \approx 0.3-0.5$
 - غالبا ما يكون لدى التلاميذ ذوي المستوى المتوسط
 - ويفسر هذا بـ:
 - إجابات متوقعة جدا
 - إمكانية الحفظ
 - فقرات متشابهة جدا ومتجانسة جدا
- وهذا لا يسبب مشكلة كبيرة لانها لا يمكن ان تحدّد خاصية الصدق، لكن يجب الاشارة الى ان هؤلاء التلاميذ يقدمون معلومات قليلة.
- سبق وان اشرنا الى ان التلاميذ في المنطقة الوسطى من التوزيع يقدمون معلومات أكثر من التلاميذ المتطرفين لكن اشرنا قبل هذه العبارة هؤلاء التلاميذ يقدمون معلومات قليلة.

في الواقع، هناك فرق جوهري بين "كمية المعلومات الإحصائية" التي يقدمها التلميذ بناءً على موقعه في المنحنى، وبين "جودة ونوع المعلومات" التي تعكسها مؤشرات المطابقة (Infit/Outfit).

لماذا التلاميذ في "الوسط" يقدمون معلومات أكثر (نظرياً)؟

من الناحية الإحصائية، التلميذ الذي يقابل مستوى صعوبة الفقرة (أي أن قدرته تساوي تقريباً صعوبة السؤال) هو الذي يقدم أكبر قدر من المعلومات (Information Function). وسبب ذلك لأن احتمالية إجابته هي 50/50. هذه الحالة تعطينا أدق قياس لقدرته.

بالمقابل: التلاميذ "المتطرفون" (أذكىء جداً لدرجة حل كل شيء، أو ضعاف جداً لدرجة الفشل في كل شيء) لا يقدمون معلومات مفيدة لأن إجاباتهم معروفة مسبقاً.

وعند اشارتنا أن أصحاب الـ INFIT المنخفض (0.3 - 0.5) يقدمون معلومات قليلة، هنا ننتقل من "موقع التلميذ" إلى "نمط استجابته". عندما ينخفض مؤشر Infit عن 1.0 وصولاً إلى 0.3، نحن نتحدث عن حالة تسمى Guttman-like pattern أو الاستجابة "المثالية زيادة عن اللزوم".

هؤلاء التلاميذ يقدمون معلومات قليلة لأن:

التوقع الشاذ: إجاباتهم متوقعة بنسبة 100%. التلميذ يجيب صح على كل ما هو أسهل من مستواه، ويخطئ في كل ما هو أصعب، دون أي "مخاطرة" أو تفاعل حقيقي مع الأسئلة التي في مستوى قدرته. **التكرار (Redundancy):** إذا كانت الفقرات متشابهة جداً، فبمجرد حل السؤال الأول، تصبح الإجابة على البقية مجرد تحصيل حاصل (تكرار لنفس المعلومة).

الخلاصة الإحصائية: إحصائياً، لكي يصبح للقياس دلالة، نحتاج إلى "تنوع" أو "تباين". الاستجابات المتطابقة تماماً مع النموذج (التي تعطي Infit منخفض) تجعل الخطأ المعياري للقياس يبدو أصغر مما هو عليه في الحقيقة، لكنها لا "تختبر" قدرة التلميذ فعلياً هي فقط تؤكد ما نعرفه مسبقاً.

الفرق بين نوعين من "قلة المعلومات"

الجدول 31 مثال: لنقارن بين هذين النوعين من التلاميذ الذين لا يعطوننا معلومات:

نوع التلميذ	سبب قلة المعلومات	حالة Infit
التلميذ المتطرف (ضعيف جداً) أو العكس	لأن الاختبار صعب جداً عليه، فلا نعرف حدود قدرته الفعلية. أو العكس	قد يكون Infit طبيعي (1.0) لكن "المعلومة" كقيمة رياضية منخفضة.
التلميذ المتوسط (Infit < 0.5)	لأن إجاباته "آلية" أو "محفوطة"، لا يوجد فيها التفاعل الطبيعي الذي يتضمن احتمالية الخطأ.	ال Infit منخفض جداً، مما يعني أن استجابته لا تزيد من دقة القياس بل تكرر فقط.

الشيء المهم كذلك الذي يجب الإشارة إليه أن التلاميذ المتطرفون (الضعفاء جداً أو الأقوياء جداً)

لديهم INFIT أكثر تشتتاً وهذا أمر طبيعي في نموذج راش:

• قليل من الفقرات المتطرفة تعطي معلومات.

• تباين البواقي أعلى.

وبالتالي ليست مشكلة بيانات ، وإنما مشكلة افراد وبالتالي نتساءل هل يجب حذف هؤلاء التلاميذ؟

والاجابة على السؤال حسب (Linacre, 2012) يجب النظر الى المؤشرات التالية حيث يمكن حذف

الافراد فقط إذا:

1. $INFIT > 2.0$ و

2. $ZSTD > +3$ و

3. نسبة التلاميذ $< 5\%$

ونحن لدينا:

1. بعض الحالات المعزولة

2. لا يوجد بناء نسقي

3. $> 2-3\%$ من التلاميذ

"بناءً على توصيات (Linacre 2012)، تم الإبقاء على جميع أفراد العينة رغم وجود بعض القيم

الخارجة (Misfit). ويعود ذلك إلى أن نسبة الأفراد الذين حققوا قيم (Infit > 2.0) و (ZSTD

(3) > لم تتجاوز 3% من إجمالي العينة (559 طالباً)، وهي نسبة ضئيلة لا تؤثر على استقرار بارامترات الفقرات. كما أن عدم المطابقة لم يظهر أي بناء نسقي منتظم (هذه النقطة الأهم. لو وجدت مثلاً أن كل التلاميذ الضعفاء لديهم عدم مطابقة، فهذا يعني وجود مشكلة في "بناء الاختبار" نفسه. لكن بما أن عدم المطابقة عشوائي وغير منتظم، فهذا يثبت أن المشكلة في "سلوك التلميذ" (تعب، سرحان، تخمين) وليست في "جودة الفقرات."، مما يشير إلى أن هذه الاستجابات هي حالات معزولة ناتجة عن عوامل ظرفية لا تمس جوهر البناء السيكومتري للاختبار".

وبالتالي لا ينبغي حذف الافراد.

فما تأثير ذلك على تحليل الفقرات؟

ملاحظات مهمة جدا ينبغي الاشارة اليها:

- الأفراد غير الملائمين لم يقللوا من توافق الفقرات.
- متوسط الملاءمة لا يزال مركزياً
- الاستنتاجات حول الفقرات لا تزال ذات مصداقية.

وبالتالي يمكن إبقاء العينة كاملة.

تحليل حدود النتائج (تحديد الثغرات)

في أسفل المخطط، تظهر النسبة المئوية (TILE%) وتوزيع الفقرات (Items):

- تلاحظ وجود فقرات في أقصى اليمين وأقصى اليسار، لكن الكثافة العظمى في المنتصف .
- التفسير هذا يدعم نقطة "تحديد حدود النتائج". انخفاض كثافة الفقرات عند القدرات العالية جداً (أقصى اليمين) يعني أن دقة القياس تنخفض عند التلاميذ الموهوبين جداً لأن الاختبار لا يتحداهم بفقرات كافية في مستواهم .
- كما هو الحال في مستويات القدرة العالية، فإن انخفاض كثافة الفقرات في مستويات القدرة المنخفضة (أقصى اليسار) يؤدي إلى حدوث تأثير ادن (Floor Effect) هذا النقص في الفقرات السهلة

جداً يحد من قدرة الاختبار على التمييز بدقة بين ذوي المستويات المتدنية، مما يرفع من قيمة الخطأ المعياري للقياس لديهم، ويجعل تقديرات قدرتهم أقل استقراراً وثباتاً.

- باختصار: الاختبار الجيد في نموذج راش هو الذي يغطي "كامل المدى (Range) الخاص بالقدرات الموجودة في العينة، وأي فراغ في أي من الطرفين (أقصى اليمين أو أقصى اليسار) يضعف من الخصائص السيكمترية للاختبار في تلك المنطقة تحديداً.

الخلاصة:

بناءً على جدول 4.1، تظهر البيانات سلوك اجابات متسق لغالبية العينة (559 تلميذا)، حيث يتمركز معظمهم حول خط الملاءمة المثالي (1.0) وجود بعض الانحرافات في الأعلى هو أمر طبيعي في الدراسات التربوية، لكنه يتطلب فحص هؤلاء الأفراد للتأكد من عدم وجود "تخمين" أو "سرعة في الإجابة". هذا الاتساق هو ما يدعم في النهاية الاستنتاج حول خاصية الصدق بأن الدرجات لها معنى علمي وموثوق .

تحليل إحصاءات المطابقة Fit Statistics

تعد إحصاءات المطابقة من أهم المؤشرات المستخدمة لتقييم مدى توافق البيانات التجريبية مع افتراضات نموذج راش.

وتشمل هذه الإحصاءات أساساً:

• **Infit Mean Square**

• **Outfit Mean Square**

ويعد المجال المقبول عادة لهذه القيم كما تشير الأدبيات السيكمترية:

$$0.5 \leq MNSQ \leq 1.5$$

حيث تشير القيم المرتفعة إلى مشكلات في البيانات، بينما تشير القيم المنخفضة جداً إلى تكرار زائد

في الاستجابات Redundancy

وقد أظهرت نتائج التحليل أن معظم الفقرات تقع ضمن الحدود المقبولة للمطابقة، مما يشير إلى ملاءمة

جيدة لنموذج راش.

جدول 32 تصنيف الفقرات وفق جودة المطابقة للنموذج

النسبة	عدد الفقرات	نوع الفقرة
87%	47	فقرات مطابقة للنموذج
9%	5	فقرات بحاجة مراجعة
4%	2	فقرات غير مطابقة

جدول رقم (33): تقديرات المطابقة الخارجية للفقرات الأكثر عدم المطابقة وفق نموذج راش

الفقرات	معامل الصعوبة	احصائي المطابقة الداخلي Infit_MNSQ	Outfit_MNSQ احصائي المطابقة الخارجي	معامل الارتباط
11	3.62-	1.18	2.23	0.02
12	1.24-	1.28	2.02	0.18
6	1.83-	0.93	1.68	0.37
13	0.45-	1.28	1.61	0.25
17	0.36	1.17	1.60	0.31
10	1.03-	1.28	1.50	0.24

الجدول 34 إحصائيات ملاءمة الفقرات الأفضل في المطابقة

الفقرات	معامل الصعوبة	احصائي المطابقة الداخلي Infit_MNSQ	Outfit_MNSQ احصائي المطابقة الخارجي	معامل الارتباط PTMEA CORR
48	1.74	0.80	0.59	0.58
41	2.77	0.80	0.51	0.49

0.49	0.51	0.80	2.77	42
0.49	0.51	0.80	2.77	43
0.58	0.58	0.79	1.80	49
0.63	0.65	0.78	0.80	33

الجدول لبقية الفقرات في الملاحق (الجدول 10 من مخرجات البرنامج)

إحصائيات الملاءمة

تم اعتبار قيم INFIT و OUTFIT بين 0.50 و 1.50 منتجة للقياس. كان من المتوقع أن تكون قيم معاملات الارتباط إيجابية.

تحليل منحنيات خصائص الفقرات ICC

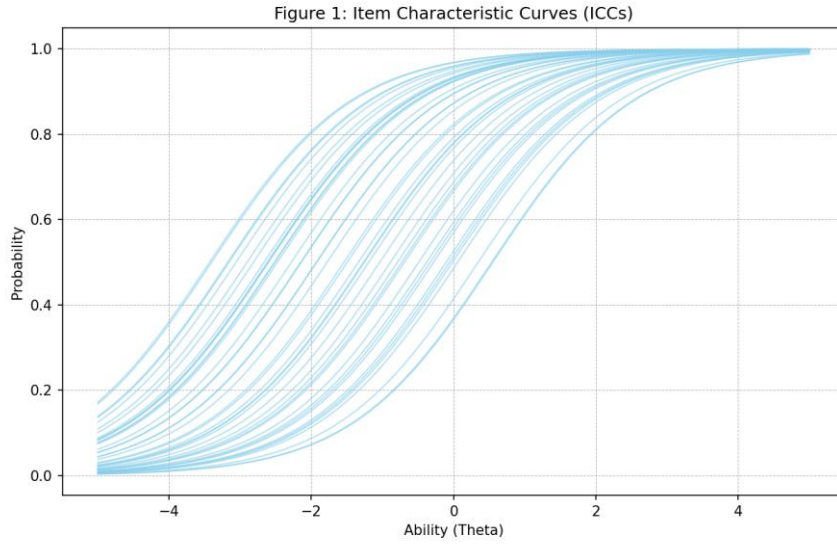
ويُعد تحليل منحنيات خصائص الفقرات Item Characteristic Curves – ICCs و دوال معلومات الفقرات Item Information Functions – IIFs من أهم الإجراءات التحليلية في إطار نظرية الاستجابة للمفردة، حيث يتيح هذا التحليل فهماً دقيقاً لكيفية تفاعل الفقرات مع مستويات القدرة المختلفة، إضافة إلى تحديد درجة إسهام كل فقرة في دقة القياس.

تم رسم منحنيات خصائص الفقرات لجميع فقرات الاختبار، حيث توضح هذه المنحنيات العلاقة بين:

- القدرة الكامنة للمفحوص

- احتمال الإجابة الصحيحة عن الفقرة.

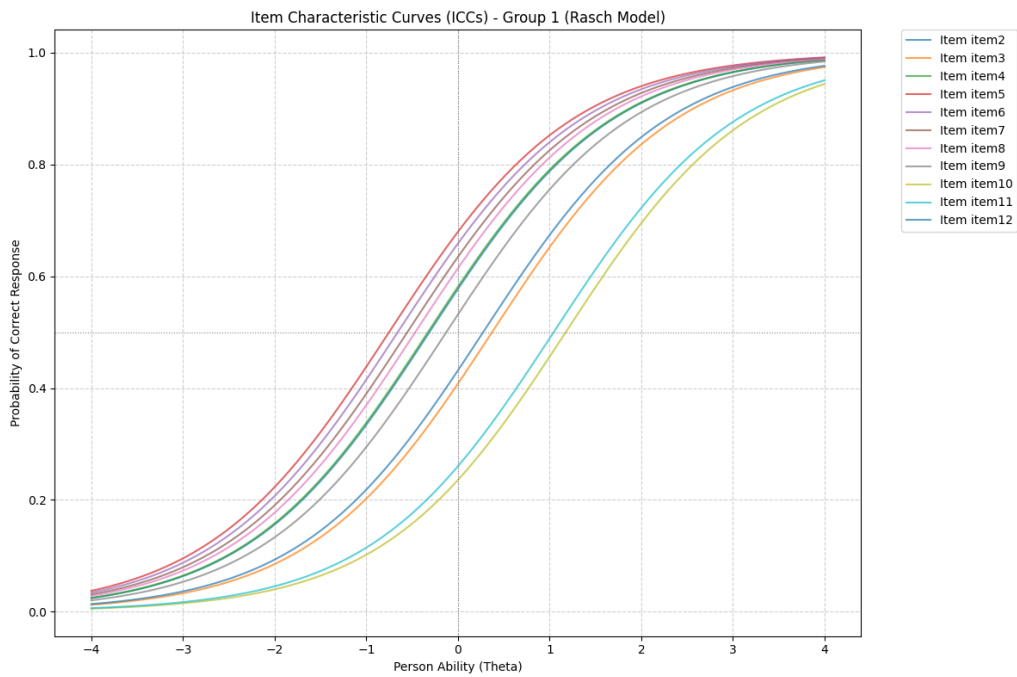
التمثيل البياني 14 لفقرات الاختبار



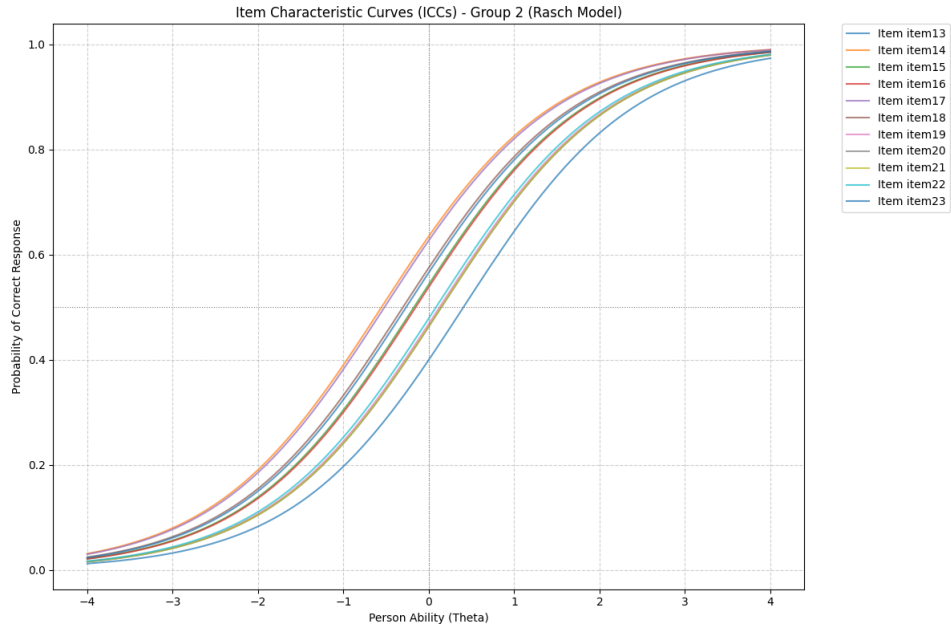
يتكون الاختبار من 55 فقرة، وقد تم تنظيم هذه الفقرات في خمس مجموعات لغرض تسهيل الفحص البياني للمنحنيات السيكو مترية. وقد تم تقسيم الفقرات على النحو الآتي:

خمس مجموعات تتكون كل منها من 11 فقرة.. ويهدف هذا التقسيم إلى تحسين وضوح الرسوم البيانية، حيث إن عرض جميع الفقرات في رسم بياني واحد يؤدي غالباً إلى تداخل المنحنيات وصعوبة تفسيرها، بينما يسمح عرضها في مجموعات صغيرة بإجراء تحليل بياني أكثر دقة لخصائص كل فقرة.

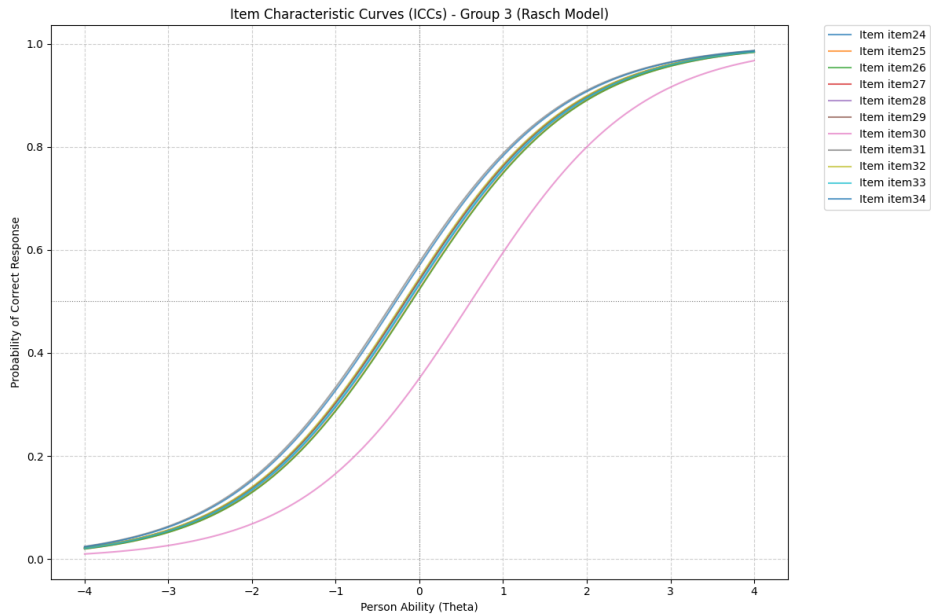
التمثيل البياني 15 المجموعة الاولى: من الفقرة 1 الى الفقرة 11



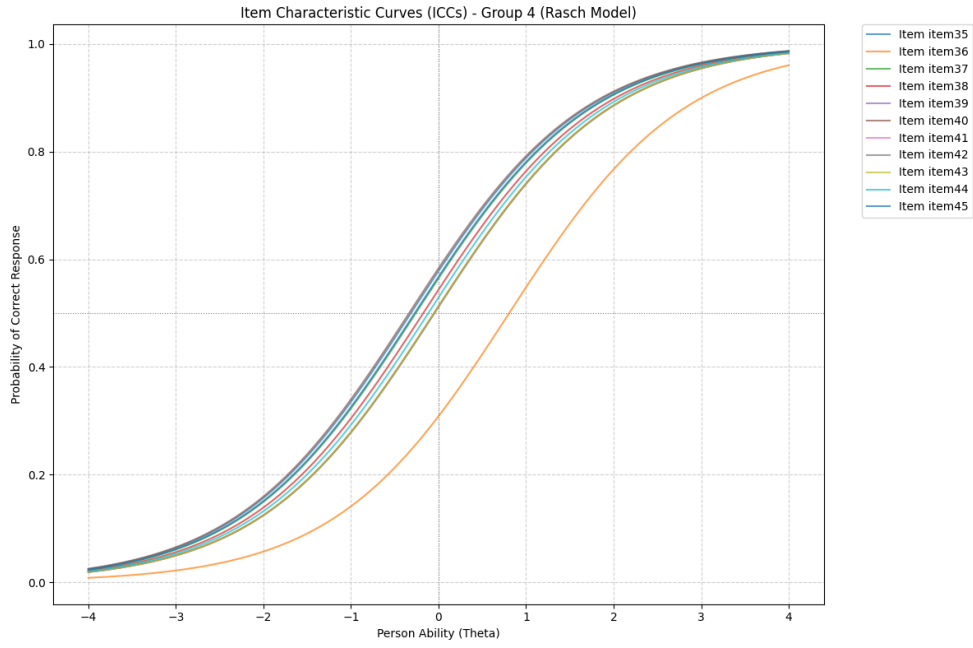
التمثيل البياني 16 المجموعة الثانية من الفقرة 12 الى الفقرة 23



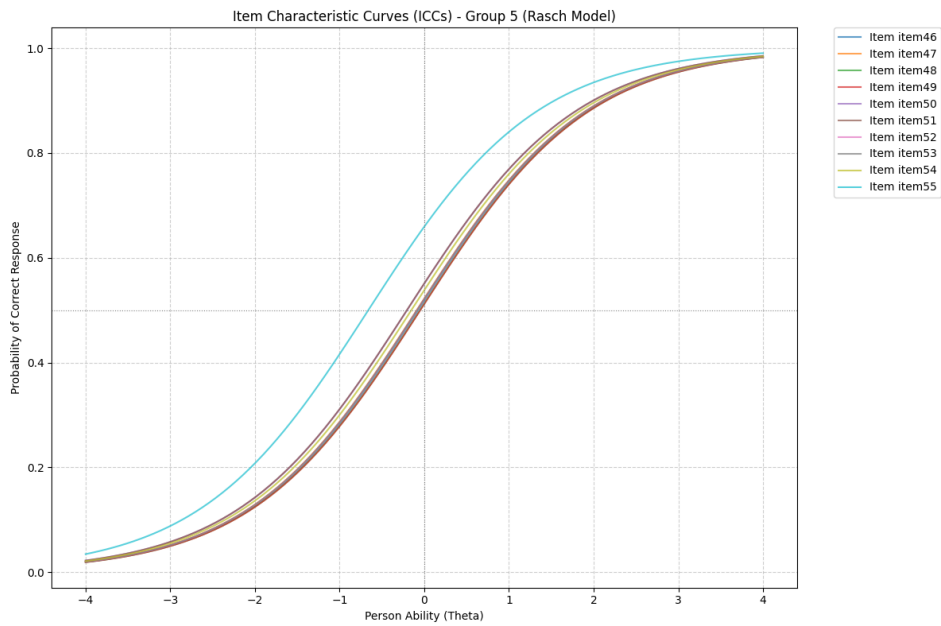
التمثيل البياني 17 المجموعة الثالثة من الفقرة 23 الى الفقرة 34



التمثيل البياني 18 المجموعة الرابعة الفقرة من 35 الى 46



التمثيل البياني 19 المجموعة الخامسة من الفقرة 46 الى 55



وقد أظهرت هذه المنحنيات أن الفقرات تتوزع عبر مدى واسع من مستويات الصعوبة، مع التمرکز

قليلا في الوسط.

فالفقرات ذات الصعوبة المنخفضة ظهرت منحنياتها في الجزء الأيسر من متصل القدرة، بينما ظهرت الفقرات الأكثر صعوبة في الجزء الأيمن من المتصل.

وقد سمح هذا التحليل بالتأكد من أن الاختبار يغطي مدى مناسباً من مستويات القدرة لدى التلاميذ. وهو ما يعد مؤشراً إيجابياً من حيث تغطية الاختبار لمجال السمة المقاسة، وهو ما يحقق بينة صدق المحتوى.

5.5 تحليل دوال معلومات الفقرات IIF

تشير دالة معلومات الفقرة إلى مقدار المعلومات التي تقدمها الفقرة حول قدرة المفحوص عند مستوى معين من القدرة.

وتبلغ دالة المعلومات أقصاها عندما يكون احتمال الإجابة الصحيحة قريباً من 0.50.

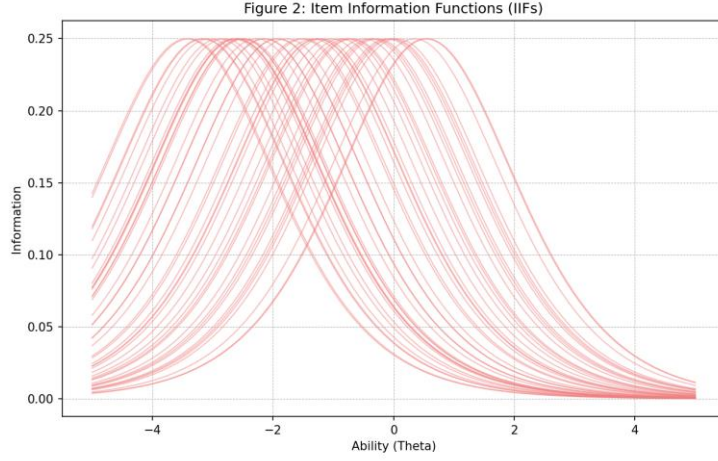
أظهرت الرسوم البيانية لدوال معلومات الفقرات مجموعة من الخصائص المهمة المتعلقة بدقة القياس في الاختبار.

قدم الاختبار أعلى دقة قياس عند مستويات القدرة القريبة من متوسط العينة، مع انخفاض المعلومات عند مستويات القدرات القصوى.

تحليل سيكو متري لدوال معلومات الفقرات IIFs

الشكل رقم : دوال معلومات الفقرات. IIFs الموضوع : تصور دقة القياس لفقرات الاختبار الفردية عبر طيف القدرة.

الخط البياني 20 : دالة معلومات الفقرات



النظرة العامة

يعرض هذا الرسم البياني دوال معلومات الفقرات IIFs لمجموعة من فقرات الاختبار. في نظرية استجابة الفقرة IRT مصطلح "المعلومات" هو مصطلح تقني يشير إلى الدقة . Precision كلما كان منحني المعلومات أعلى، كلما كانت الفقرة تقيس قدرة الطالب بدقة أكبر عند ذلك المستوى المحدد.

2. تفسير المحاور

- المحور الأفقي: القدرة - ثيتا
 - يمثل السمة الكامنة القدرة للمفحوص، وعادة ما يكون على مقياس معياري حيث يمثل الصفر متوسط القدرة.
 - المدى: من -4 قدرة منخفضة جداً إلى +4 قدرة عالية جداً .
- المحور العمودي المعلومات :
 - يمثل كمية المعلومات السيكومترية التي توفرها الفقرة.

- المدى: من 0.00 إلى 0.25.
- ملاحظة إحصائية: مقلوب المعلومات هو الخطأ المعياري للقياس SEM. المعلومات العالية تعني خطأ منخفضاً وبالتالي ثقة أكبر في الدرجة .

3. شكل المنحنيات الشكل "الجريسي"

يمثل كل منحني أحمر فقرة اختبار واحدة. ستلاحظ أن جميع المنحنيات تتبع نمطاً على شكل جرس مقعرة من الأعلى. يوضح هذا قاعدة أساسية في القياس السيكومتري.

- القمة: توفر الفقرة أكبر قدر من المعلومات أعلى دقة عندما تتطابق قدرة الطالب مع صعوبة الفقرة.
- إذا كانت الفقرة صعبة، ستكون قممتها جهة اليمين القدرة العالية .
- إذا كانت الفقرة سهلة، ستكون قممتها جهة اليسار القدرة المنخفضة .
- الانخفاض: تتناقض المعلومات كلما ابتعدنا عن القمة.
- تحليل تفصيلي للرسم البياني

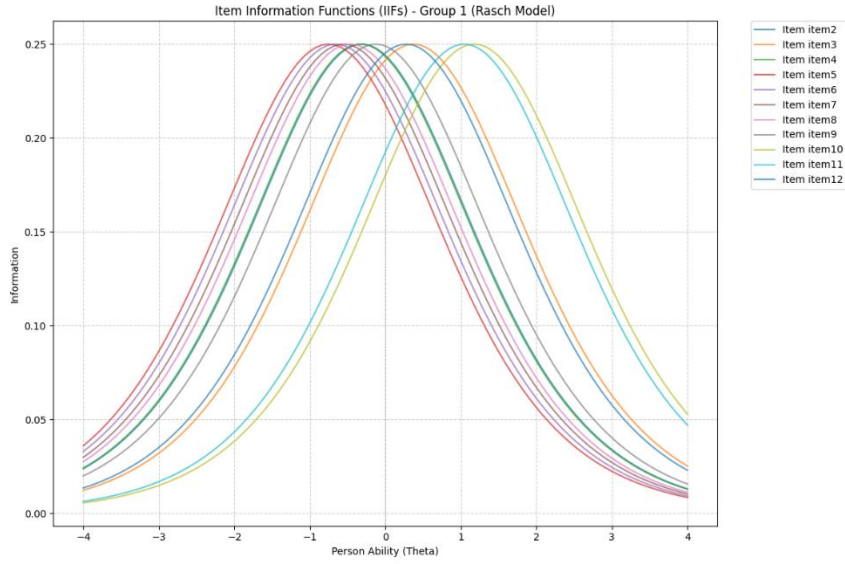
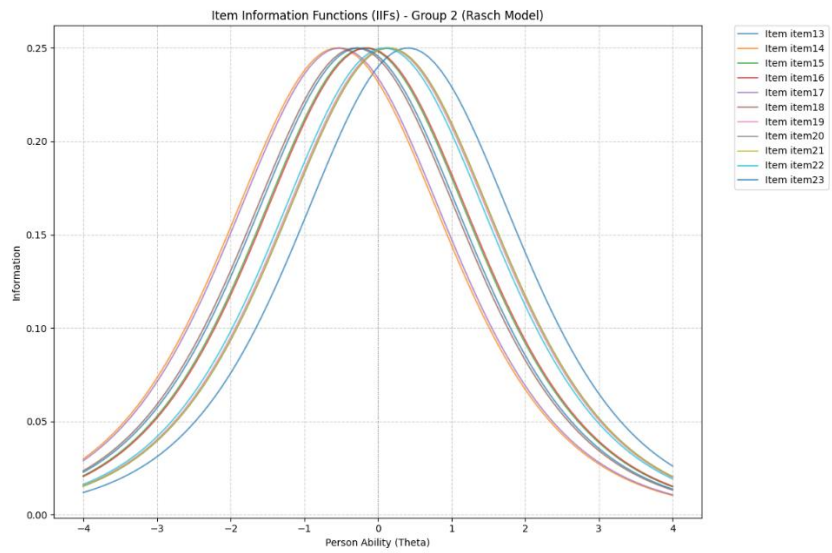
أ. ارتفاع أقصى معلومات 0.25~ تظهر قمم معظم المنحنيات أنها تصل إلى حوالي 0.25 على المحور العمودي.

- الدلالة السيكمترية: في نموذج راش - Rasch Model نموذج اللوجستي ذو المعامل الواحد ، الحد الأقصى للمعلومات للفقرة ثنائية الاستجابة صح/خطأ نظرياً ثابت عند 0.25.
- الاستنتاج: يؤكد هذا الرسم البياني أن التحليل يعتمد على نموذج راش. لو كان هذا نموذجاً ذو معاملين 2PL أو ثلاثة 3PL ، لتباينت ارتفاعات القمم بعض الفقرات كانت ستكون أكثر حدة وبالتالي أكثر تمييزاً من غيرها. هنا، جميع الفقرات متساوية في التمييز؛ فهي تختلف فقط في الصعوبة.

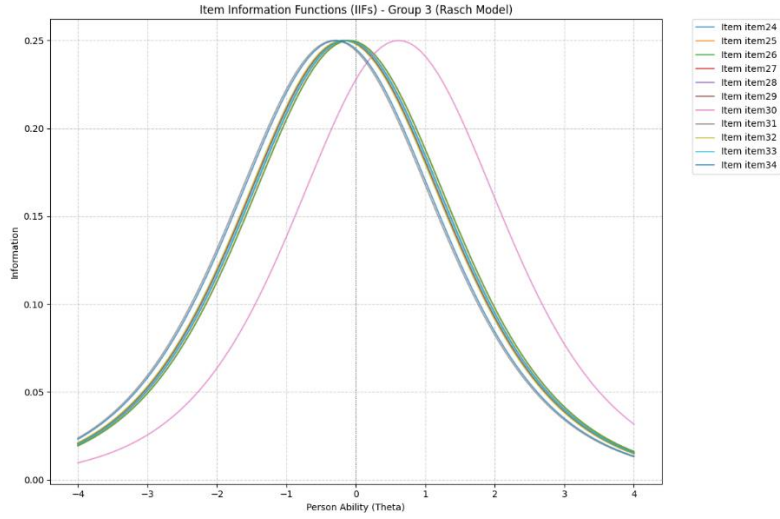
ب. التوزيع الأفقي انتشار صعوبة الفقرات المنحنيات منتشرة عبر المحور الأفقي.

- الجانب الأيسر ثيتا سالبة: الفقرات التي تبلغ ذروتها هنا هي سهلة .وهي مفيدة للتمييز بين الطلاب منخفضي القدرة .
 - ثيتا في الوسط :العديد من المنحنيات تبلغ ذروتها قرب المركز. هذه الفقرات موجهة لمستوى القدرة المتوسطة.
 - الجانب الأيمن ثيتا موجبة الفقرات التي تبلغ ذروتها هنا هي صعبة .وهي ضرورية لقياس الطلاب ذوي القدرات العالية بدقة.
 - من الناحية المثالية، يجب أن يحتوي الاختبار على فقرات تبلغ ذروتها في جميع مستويات القدرة لضمان أن دالة معلومات الاختبار TIF مجموع كل هذه المنحنيات تكون عالية عبر النطاق بأكمله.
 - يشير الرسم البياني إلى تغطية جيدة، ولا سيما حول مستوى القدرة المتوسطة ثيتا $\theta = 0$ ، حيث تتقاطع العديد من المنحنيات.
- ويمكن ان نستخلص أن من خلال التشخيص :يعمل الاختبار بشكل متسق وفقاً لافتراضات نموذج راش.
- نقاط القوة :يوجد انتشار جيد لصعوبات الفقرات. هناك فقرات "سهلة" لقياس أدنى المستويات وفقرات "صعبة" لقياس أعلى المستويات.
- لتحسين الاختبار، يجب ان نجمع هذه المنحنيات لإنشاء دالة معلومات الاختبار TIF. إذا كانت هناك "فجوات" في المحور الأفقي لا توجد قمم فيها مثلاً بين ثيتا 1.5 و 2.5 ، فإن الاختبار سيكون له دقة أقل للطلاب في هذا النطاق المحدد من القدرة. لكن بالنظر إلى هذا الجزء، التغطية تبدو كثيفة نسبياً، مما يشير إلى اختبار قوي.
- كما تم تقسيم الفقرات الى خمس مجموعات مثل التقسيم السابق.

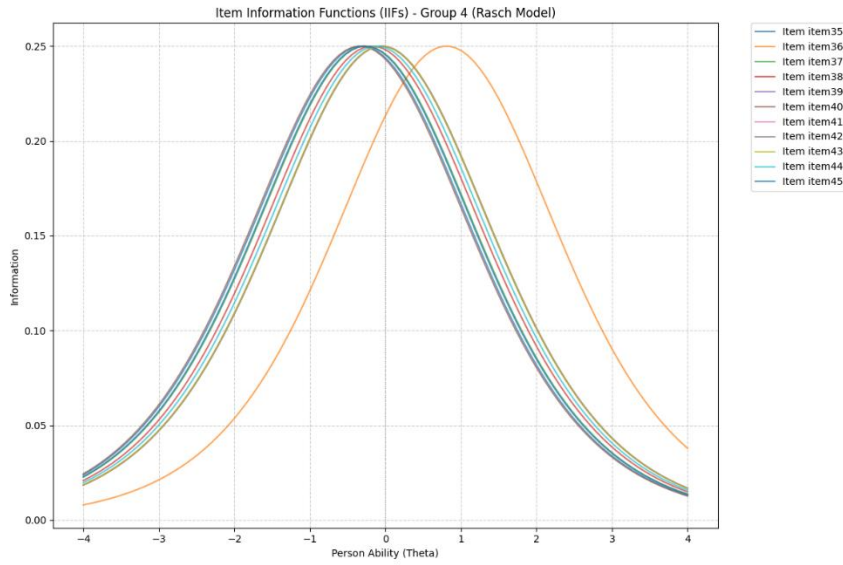
التمثيل البياني 19 للمجموعة 1 والملج 2



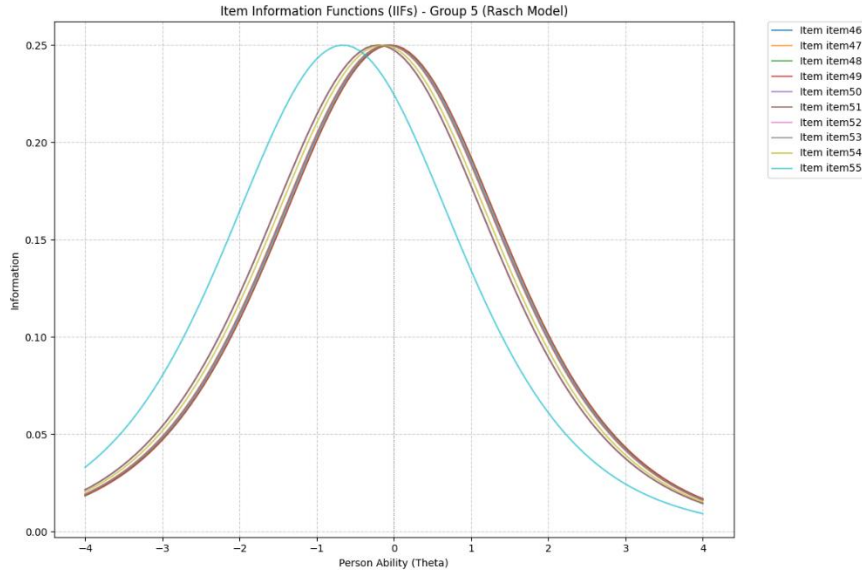
التمثيل البياني 23 للملج 3



التمثيل البياني 24 للمج 4



التمثيل البياني 25 للمجموعة 5



التمثيل البياني الأول مقارنته بالرسوم البيانية التفصيلية اللاحقة للمجموعات الخمس ووجد أنها متسقة

تماماً ولا تتناقض، حيث التحقق من الشكل والارتفاع Item Information Shape

في الشكل الأول : نرى منحنيات ذات شكل "جرسي" ومحد أقصى للارتفاع يصل تقريباً إلى 0.25،

في باقي الأشكال: جميع المنحنيات في المجموعات تتخذ نفس الشكل الجرسى وتبلغ ذروتها عند القيمة

0.25.

الاستنتاج: هذا يؤكد أن البيانات تتبع نموذج راش للبيانات الثنائية، يكون الحد الأقصى لمعلومات

الفقرة دائماً 0.25 عند نقطة تعادل القدرة مع صعوبة الفقرة. هذا تناسق رياضي دقيق.

2. التحقق من توزيع الصعوبة Difficulty Distribution

الشكل الأول يقدم نظرة عامة، بينما تقسم الأشكال التالية هذه الفقرات إلى مجموعات للربط بين

التحليل الإحصائي السابق والرسوم البيانية:

المجموعة 1 : الفقرات من 1 إلى 11 تشمل الفقرات

في التحليل الإحصائي السابق، وجدنا أن الفقرة 11 هي الأسهل بمقياس صعوبة (-3.62) و الفقرة 2 (0.20).

في الرسم البياني للمجموعة 1: يجب أن نرى منحنيات منزاحة ليسار نحو القدرات المنخفضة. وهذا ما يظهر في الوصف الذروة تكون على الجانب الأيسر. هذا يتطابق مع كونها فقرات "سهلة جداً".
المجموعة 2 من الفقرة 13 إلى 23:

في التحليل الإحصائي، كانت معظم هذه الفقرات متوسطة الصعوبة الفقرة 16 (-1.06)
في الرسم البياني للمجموعة 2: الوصف يوضح أن المنحنيات تتمركز حول النقطة 0 وسط المحور . هذا دليل على أن صعوبة هذه الفقرات تقع في المتوسط.

المجموعة 3 الفقرات 24 إلى 34

في التحليل الإحصائي، كانت الفقرة 32 هي الأصعب (0.65) والفقرة 30 (-0.98)
في الرسم البياني للمجموعة 3: نرى منحنيات منزاحة لليمين نحو القدرات العالية. وهذا يتطابق تماماً مع كونها فقرات "صعبة".

المجموعة 4 تغطي الفقرات من 35 إلى 45. وأصعب الفقرات 42، 43، 44، 45 (-2.77)

○ ملاحظة: في البيانات الأصلية، هناك فجوة بين الفقرة 40 الفقرات 41 و 42 غير موجودة.

○ الرسم البياني يعكس هذا الواقع بدقة يظهر منحنيات للفقرات الموجودة فقط.

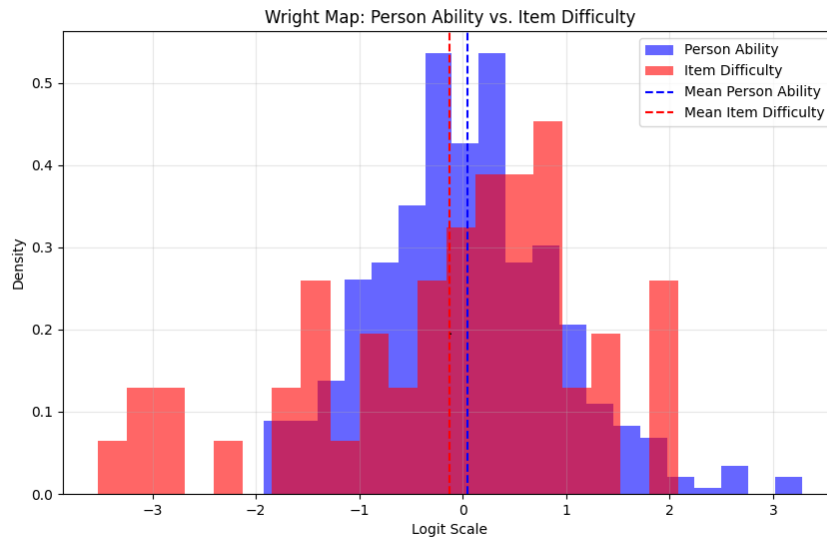
● المجموعة 5: تغطي الفقرات من 46 إلى 55 وهي نهاية البيانات .

الفقرة الأصعب رقم 46 (1.89) والفقرة الأسهل 55 (-1.91)

نلاحظ ان الفقرات تغطي كامل الاختبار بشكل متسلسل ومنطقي.

من خلال البيانات السابقة نلاحظ ان اغلب الفقرات تقع على يسار المتوسط وذلك لسهولة الفقرات
بينما تقع فقرات قليلة صعبة على يمين القدرة المتوسطة.

التمثيل البياني خريطة الافراد والفقرات Wright person-items map.



خريطة توزيع الافراد والفقرات:Wright Map

- قدرات التلاميذ
- صعوبة الفقرات

لتحديد مدى انتشار صعوبة الفقرات على متصل القدرة.

تم إنشاء **خريطة الشخص-الفقرة** التي تعد من أهم أدوات تحليل نموذج راش من مخرجات برنامج

Winsteps لتقديم توزيع التلاميذ والفقرات على مقياس موحد بوحدة اللوجيت.

نلاحظ من الشكل السابق ان متوسط الاختبار قريب من متوسط العينة وهذا يشير الى ملائمة الاختبار

للعينة وللمجتمع الدراسة، حيث نلاحظ متوسط الاختبار اقل بقليل من متوسط القدرة وهذا معناه ان

الاختبار كان سهلاً.

كما نلاحظ في المخطط التالي توزيع الفقرات والافراد على متصل القدرة بوحدة اللوجيت.

التمثيل البياني: 28: خريطة رايت

```

2 INPUT: 559 Person 55 Item REPORTED: 559 Person 55 Item 2 CATS WINSTEPS 3.72.3
3 -----
4
5 Person - MAP - Item
6 <more>|<rare>
7 5 +
8 |
9 |
10 |
11 |
12 . |
13 4 +
14 |
15 |
16 |
17 .# |T
18 . |
19 3 +
20 . | I0041 I0042 I0043
21 .# T|
22 .# | I0039
23 .# |
24 .# |
25 2 .# +
26 .#### | I0046 I0049
27 ## |S I0048 I0052 I0053
28 .### |
29 .##### S| I0036 I0047
30 .#### | I0015 I0025 I0026 I0044
31 1 .##### + I0016 I0024 I0037
32 .#### | I0033 I0051 I0054
33 .##### | I0027 I0028 I0029 I0032 I0038
34 ##### | I0023
35 .##### | I0017
36 .##### M| I0002 I0004 I0031
37 0 .##### +M I0018 I0050
38 .#### | I0035 I0045
39 .##### | I0034 I0040
40 .#### | I0013
41 .##### | I0008
42 .### |
43 -1 .#### + I0003 I0010 I0030
44 .##### S| I0012
45 .# |
46 ##### | I0007 I0014
47 .## |S I0001
48 .## | I0006 I0055
49 -2 +
50 .### | I0009
51 T|
52 | I0005
53 ## |
54 .# |
55 -3 +
56 | I0021
57 |T I0020
58 |
59 | I0011 I0022
60 |
61 -4 + I0019
62 <less>|<frequ>
63 EACH "#" IS 4. EACH "." IS 1 TO 3

```

وتعرض هذه الخريطة توزيع:

- قدرات الأفراد
 - صعوبات الفقرات
 - على نفس مقياس اللوجيت.
 - وقد أظهرت الخريطة أن:
 - معظم التلاميذ يتركزون حول القدرة 0 لوجيت
 - تتوزع الفقرات حول هذا المستوى بصورة مناسبة.
 - توزيع الفقرات يغطي توزيع الأشخاص رغم أن الفجوات الطفيفة لوحظت عند مستويات القدرات الأعلى.
 - معظم الفقرات أسهل من قدرة الأفراد وبالتالي الاختبار سهل Wright & Masters, 1982 .
- خريطة رايت Wright Map "أو خريطة الشخص-الفقرة Person-Item Map ، وهي من أهم الأدوات في تحليل راش . Rasch Analysis تُظهر هذه الخريطة توزيع الأفراد الطلاب والفقرات الأسئلة على نفس المقياس لوجيت .

التحليل المفصل للخريطة:

1. هيكل الخريطة والمحاور

- المحور العمودي مقياس اللوجيت يمثل مقياس القدرة/الصعوبة المشترك.
 - الأعلى مثلاً 5 :يمثل قدرة عالية جداً طلاب أذكاء جداً وفقرات صعبة جداً.
 - الأدنى مثلاً -3 :يمثل قدرة منخفضة طلاب أقل قدرة وفقرات سهلة جداً.

• الجانب الأيسر – Persons الأشخاص:

○ كل حرف X أو نقطة يمثل تلميذاً أو مجموعة من التلاميذ.

○ توزيع الـ X يوضح "توزيع القدرة" في العينة.

• الجانب الأيمن – Items الفقرات :

○ كل فقرة ممثلة برقم مثل الفقرة 01 والفقرة 020.

○ موقع الفقرة يحدد صعوبتها.

التحليل التفصيلي للتوزيع

أ. تحليل جانب الأفراد :

• الموقع: يتركز معظم التلاميذ في المنطقة العليا من الخريطة فوق الصفر، وبالتحديد بين 1 و 3 لوجيت

تقريباً .

• القدرة: يشير هذا إلى أن عينة التلاميذ تمتلك قدرات عالية High Ability معظم التلاميذ

يتمتعون بقدرات تفوق المتوسط.

ب. تحليل جانب الفقرات الأسئلة:

• الموقع: تتمركز معظم الفقرات في المنطقة الوسطى والسفلى حول الصفر وما دونه .

• الصعوبة: تشير إلى أن الفقرات سهلة إلى متوسطة الصعوبة.

3. تقييم التطابق Targeting

هنا تكمن المشكلة الرئيسية في هذا الاختبار:

• الفجوة الواضحة: يوجد عدم تطابق Mismatch واضح بين مستوى الطلاب ومستوى الفقرات.

- الطلاب يقعون في الأعلى قدرات عالية .
 - الفقرات تقع في الأسفل أو الوسط سهلة/متوسطة .
 - التفسير: الاختبار سهل جداً بالنسبة لهؤلاء التلاميذ.
 - التلاميذ الموجودون في القمة عند 3 أو 4 لوجيت لا توجد فقرات بنفس مستوى صعوبتهم لقياس قدراتهم بدقة.
 - هذا يؤكد ما أشرنا إليه في تحليل دالة المعلومات السابق TIF الذي كانت ذروته منحازة لليسار نحو -
 - 2 . هنا نرى الصورة بوضوح: الفقرات "سهلة" ولا تصل لارتفاع قدرات التلاميذ.
- تحليل الفقرات النوعي
- الفقرات السهلة أسفل الخريطة فقرات مثل 20، 21 إذا كانت في الأسفل هي سهلة جداً. هي مفيدة للطلاب الضعفاء الذين هم قليلون في هذه العينة، لكنها لا تفيد في تمييز الطلاب الأذكياء.
 - الفقرات الصعبة أعلى الخريطة نلاحظ ندرة الفقرات في المنطقة العليا فوق 2 لوجيت. هذا يعني أن الاختبار يفتقر إلى "سقف عالٍ High Ceiling" قياس التفوق.

دالة معلومات الاختبار Test Information Function

توليد دالة معلومات الاختبار TIF

دالة معلومات الاختبار TIF هي مجموع دوال معلومات الفقرات IIFs لجميع العناصر في الاختبار. يشير إلى مقدار المعلومات التي يوفرها الاختبار بالكامل حول قدرة الشخص في نقاط مختلفة على متصل القدرة. قيمة TIF أعلى عند مستوى قدرة معين تعني أن الاختبار أكثر دقة في قياس القدرة للأفراد في ذلك المستوى. في سياق نموذج اللوجستيات ذات المعامل الواحد Rasch ، الصيغة الخاصة ب IIF هي:

الفقرات عالية المعلومات.

لوحظ أن بعض الفقرات تمتلك قمماً مرتفعة وضيقة في دالة المعلومات، وهو ما يدل على أنها تقدم قدراً كبيراً من المعلومات حول المفحوصين ضمن نطاق محدد من القدرة.

وتعد هذه الفقرات ذات أهمية كبيرة لأنها تسهم بشكل فعال في تحسين دقة القياس عند مستويات القدرة القريبة من صعوبة الفقرة.

الفقرات منخفضة المعلومات.

في المقابل، ظهرت بعض الفقرات بقيم معلومات منخفضة نسبياً، حيث بدت قمم دوال المعلومات الخاصة بها منخفضة أو واسعة الانتشار.

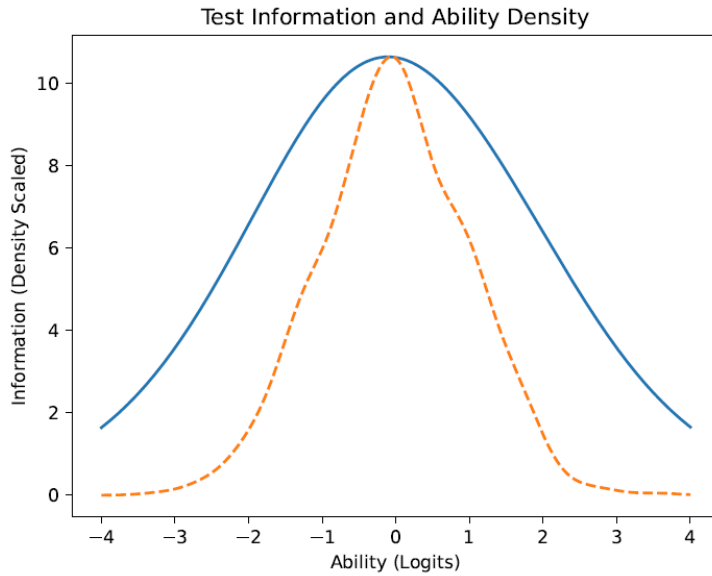
ويشير ذلك إلى أن هذه الفقرات تقدم مقداراً محدوداً من المعلومات، أو أن معلوماتها موزعة عبر نطاق واسع من القدرة.

وقد يتطلب الأمر في هذه الحالة مراجعة هذه الفقرات من حيث محتواها أو صياغتها.

توزيع المعلومات عبر متصل القدرة

عند تحليل مجموع دوال المعلومات الخاصة بجميع الفقرات، تبين أن الاختبار يقدم معلومات جيدة عبر نطاق متوسط من القدرة، وهو ما يعني أن الاختبار أكثر دقة في قياس المستويات المتوسطة من القدرة أما عند المستويات العالية أو المنخفضة من القدرة، فقد تنخفض كمية المعلومات المتاحة نسبياً كما في الشكل:

.التمثيل البياني 28 معلومات الاختبار وكثافة القدرة



تحليل سيكو متري: معلومات الاختبار وكثافة القدرة

شكل بعنوان "Test Information and Ability Density" : معلومات الاختبار

وكثافة القدرة لتقييم دقة الاختبار بالنسبة لتوزيع عينة المفحوصين.

1. النظرة العامة

يقدم هذا الرسم البياني طبقتين من المعلومات السيكو مترية اساسية:

1. دالة معلومات الاختبار: TIF مدى دقة قياس الاختبار عند مستويات القدرة المختلفة.

2. كثافة القدرة توزيع العينة أين يقع التلاميذ فعلياً في العينة على مقياس القدرة.

الهدف الأساسي من هذا التمثيل البياني هو تقييم "الاستهداف Targeting" هل يوفر الاختبار

دقة كافية في المناطق التي يتركز فيها التلاميذ فعلياً؟

✓ تحليل المنحنى الأزرق (المتصل) دالة معلومات الاختبار TIF

- التعريف: يمثل الخط الأزرق المتصل المعلومات التراكمية لجميع الفقرات الفردية مجموع منحنيات "معلومات الفقرة" من الرسم البياني السابق .

- الشكل: يتبع المنحنى الشكل الجرسى الكلاسيكي، ويصل إلى ذروته بالقرب من المركز تقريباً عند 0 لوجيت .

- التفسير:

- دقة القمة: الاختبار أكثر دقة أعلى معلومات للتلاميذ ذوي القدرة المتوسطة.
- الثبات: بما أن القمة مرتفعة نسبياً تبدو ذا حجم جيد بصرياً، فإن الاختبار يقدم ثباتاً عالياً لتلاميذ القدرة المتوسطة.
- الانخفاض: تتناقض الدقة بالنسبة للتلاميذ ذوي القدرة المنخفضة جداً الذيل الأيسر والقدرة العالية جداً الذيل الأيمن . هذا يعني أن الاختبار يقدم درجات أقل دقة للأداء المتطرف.

✓ تحليل المنحنى البرتقالي المقطع كثافة القدرة

- التعريف: يمثل هذا المنحنى توزيع قدرات التلاميذ المقدرة الشكل: يشبه التوزيع الطبيعي المنحنى الجرسى
- الموقع: تتطابق قمة توزيع العينة تقريباً مع مركز مقياس القدرة ثباتاً لوجيت . يشير هذا إلى أن العينة تتكون في معظمها من طلاب ذوي قدرة متوسطة، مع وجود عدد أقل من الأداء المنخفض أو المرتفع ..

التقييم النهائي: تأثير "الاستهداف"

- الميزة الأكثر أهمية في هذا المخطط هي التوافق أو التداخل بين المنحنيين.
- استهداف ممتاز: تتطابق قمة المنحنى الأزرق الدقة تقريباً مع قمة المنحنى البرتقالي توزيع العينة .
- الدلالة السيكو مترية: يشير هذا إلى تصميم اختبار مثالي . يكون الاختبار أكثر دقة تماماً حيث يوجد غالبية التلاميذ.

○ إذا كانت قمة المنحنى الأزرق بعيدة جهة اليمين اختبار صعب وقمة البرتقالي بعيدة جهة اليسار عينة ذات قدرة منخفضة، فسيشير ذلك إلى عدم تطابق حيث يكون الاختبار صعباً للغاية، وتكون أخطاء القياس عالية بالنسبة لغالبية التلاميذ.

○ هنا، التطابق جيد. الاختبار معايير بدقة لهذه العينة المحددة.

الخطأ المعياري للقياس SEM

في نظرية استجابة الفقرة IRT ترتبط كمية المعلومات والخطأ المعياري بعلاقة عكسية :

- بما أن معلومات الاختبار (اللون الأزرق) هي الأعلى في المركز، فإن الخطأ المعياري هو الأدنى في المركز.
- هذا يعني أنه يمكننا الوثوق بدرجات الطلاب المتوسطين أكثر من غيرهم.
- بالنسبة للطلاب عند الأطراف مثلاً -2 لوجيت أو $+2$ لوجيت، تنخفض المعلومات، وبالتالي يزداد الخطأ المعياري. وبالتالي نحن أقل يقيناً بشأن الدرجة الدقيقة لتلميذ ممتاز جداً أو تلميذ يواجه صعوبة كبيرة بناءً على هذا الاختبار المحدد.

الخلاصة

- حالة الاستهداف: مثالية فقرات الاختبار متطابقة جيداً مع مستوى قدرة العينة.
- الثبات: يعمل الاختبار بشكل أفضل للتلاميذ "المتوسط"، وهو أمر مناسب نظراً لأن العينة تتوزع توزيعاً طبيعياً.
- الحدود: هناك انخفاض طفيف في المعلومات للتلاميذ ذوي القدرة العالية الذيل الأيمن . إذا كان التمييز بين الطلاب "المتفوقين" و "المتفوقين جداً" أولوية لهذا الامتحان، فقد يحتاج مطور الاختبار إلى إضافة بعض الفقرات الصعبة لتحويل منحنى TIF قليلاً نحو اليمين لتغطية تلك المنطقة بشكل أفضل. ومع ذلك، كتقييم عام، يعتبر هذا اختباراً فعالاً جداً.

الدلالات السيكومترية للنتائج

تشير النتائج المتحصل عليها من تحليل ICC و IIF إلى مجموعة من الدلالات السيكومترية المهمة.

أولاً، يبدو أن الاختبار يحتوي على تنوع مناسب في مستويات صعوبة الفقرات، وهو ما يسمح بقياس مستويات مختلفة من القدرة لدى التلاميذ.

ثانياً، تشير بعض المؤشرات البيانية إلى احتمال وجود تباين محدود في معاملات التمييز بين الفقرات، وهو ما قد يشير إلى أن افتراض التمييز المتساوي في نموذج راش قد لا يتحقق بصورة كاملة.

ثالثاً، أظهرت دوال المعلومات أن بعض الفقرات تقدم معلومات عالية في نطاقات قدرة محددة، في حين أن فقرات أخرى تقدم معلومات أقل نسبياً.

دلالات النتائج لتطوير الاختبار

تعد هذه النتائج ذات أهمية كبيرة في سياق تحسين جودة الاختبار وتطويره.

فقد تشير الفقرات التي تظهر معلومات منخفضة أو خصائص غير متوقعة إلى الحاجة إلى:

- مراجعة صياغة الفقرة
- تعديل مستوى صعوبتها
- أو استبدالها بفقرات أكثر ملاءمة.

كما يمكن الاستفادة من تحليل المعلومات في بناء اختبارات يتم اختيار الفقرات التي تقدم أكبر قدر من المعلومات عند مستوى قدرة معين.

مشكلة "تحديد السقف" : Ceiling Effect "الاختبار يعاني من قصور في قياس التلاميذ

المتفوقين. الغالبية العظمى من التلاميذ سيحصلون على درجات عالية جداً، ولن نستطيع التفريق بين

التلميذ "المتميز" والتلميذ "العبقري" لأن الأسئلة ليست صعبة بما يكفي.

1. توصية: يحتاج الاختبار إلى إضافة فقرات أصعب لرفع مستوى التحدي وتمكين قياس التلاميذ ذوي

القدرات العالية الموجودين في هذه العينة.

2. جودة القياس: دقة القياس ستكون جيدة للتلاميذ المتوسطين والضعفاء، لكنها ستكون ضعيفة خطأ

معياري كبير للتلاميذ المتفوقين الموجودين أعلى الخريطة.

باختصار: الخريطة توضح أن الاختبار كان "سهلاً" بالنسبة لهذه العينة المحددة من التلاميذ.

الجدول 35 مقارنة مع نظرية الاختبار الكلاسيكية CTT

الاحصائيات	نموذج راش	النظرية الكلاسيكية
معامل صعوبة الفقرات	مقياس لوجيت متحرر من العينات	قيم p تعتمد على العينة
ملاءمة الفقرات للنموذج	المطابقة الداخلية/الخارجية	معامل الارتباط المصحح
معامل الثبات	معامل الفصل ومعامل ثبات النموذج	معامل كرونباخ الفا
الأبعاد	التحليل العاملي للمكونات الأساسية للبواقي	التحليل العاملي الاستكشافي والتوكيدي
درجات التلاميذ	θ المجال بوحدة اللوجيت	درجات خام رتبية

الفرق الرئيسي

النظرية الكلاسيكية لا يمكن أن تكتشف:

- الاستقلال الموضعي للفقرات
- تأثيرات الاختبارات الفرعية
- عدم الملاءمة بسبب تعدد الأبعاد

المخرجات المتوقعة للنظرية الكلاسيكية:

- معامل الفا كرونباخ مرتفع
- كل الفقرات تظهر جيدة مقبولة
- المشاكل الاساسية ومصادرها تبقى مخفية.

الاستنتاج النهائي

بينما تشير المؤشرات الكلاسيكية إلى اتساق داخلي قوي، كشف تحليل راش عن عدم توافق مهم على مستوى الفقرة وبنى أبعاد ثانوية. تسلط هذه النتائج الضوء على القيمة التشخيصية المضافة لنموذج راش في تطوير الاختبار والتحقق من خاصية الصدق.

بناء نموذج متعدد الأبعاد:

تم بناء نموذج متعدد الأبعاد بعد ما أظهرت النتائج لنموذج راش أحادي البعد:

الجدول 36 : التحليل العاملي للمكونات الرئيسية للبواري

Table of STANDARDIZED RESIDUAL variance (in Eigenvalue units)

	-- Empirical --	Modeled
Total raw variance in observations =	93.2 100.0%	100.0%
Raw variance explained by measures =	38.2 41.0%	41.1%
Raw variance explained by persons =	15.5 16.6%	16.7%
Raw Variance explained by items =	22.7 24.3%	24.4%
Raw unexplained variance (total) =	55.0 59.0%	100.0% 58.9%
Unexplned variance in 1st contrast =	5.0 5.4%	9.1%
Unexplned variance in 2nd contrast =	3.7 4.0%	6.7%
Unexplned variance in 3rd contrast =	3.4 3.6%	6.1%
Unexplned variance in 4th contrast =	3.3 3.5%	5.9%
Unexplned variance in 5th contrast =	3.1 3.3%	5.6%

هذا الجدول هو مخرجات تحليل المكونات الأساسية للبواقي (PCAR – Principal Components Analysis of Residuals)، وهو جزء مهم في نموذج "راش" (Rasch Model) "للقياس النفسي". يُستخدم هذا التحليل تحديداً للتحقق من أحادية البعد (Unidimensionality)، أي التأكد من أن المقياس يقيس قدرة واحدة فقط

. تحليل التباين الكلي

- يمثل إجمالي التشتت أو التباين في البيانات. الرقم (93.2) هو مجموع "الوحدات" (Eigenvalues) "التي يتكون منها المقياس: Total raw variance (93.2 units) .
- يشير إلى أن النموذج (الذي يجمع بين قدرات الأشخاص وصعوبة الفقرات) استطاع تفسير 41% من التباين الكلي: Raw variance explained by measures (41.0%) .
- ملاحظة: في مقاييس علم النفس، تعتبر النسبة التي تتجاوز 40% مؤشراً جيداً جداً على قوة البناء ونقاء القياس.

التباين المفسر بواسطة الأشخاص والفقرات

- التباين الناتج عن الاختلاف في مستويات قدرة المستجيبين: Explained by persons (16.6%) .
- التباين الناتج عن اختلاف مستويات صعوبة الفقرات (الأسئلة): Explained by items (24.3%) .

التباين غير المفسر (Unexplained Variance / Residuals)

- هذا هو الجزء الأهم للحكم على جودة المقياس. يمثل التباين الذي لم يستطع نموذج "راش" تفسيره (59.0%). نقوم بتحليله لمعرفة ما إذا كان هناك "بعد ثانٍ" خفي في عملية القياس.
- هذا هو التباين المرتبط بأكبر تجمع للفقرات التي قد تقيس شيئاً خارج السمة الأساسية .

Unexplained variance in 1st contrast (5.4%):

- **قوة التباين غير المفسر:** القيمة الذاتية (Eigenvalue) للتباين الأول تبلغ 5.0 وحدات، وهي تمثل 5.4% من التباين الكلي. أكاديمياً، إذا زادت هذه القيمة عن 2.0، فإنها تعتبر إشارة قوية على وجود بُعد آخر يجب فحصه .

القيمة: (5.0 units) في المجال السيكو متري، إذا كانت قيمة التباين في التباين الأول (1st contrast) أقل من 3.0، فإننا نجزم بأحادية البعد. هنا القيمة هي 5.0، مما يشير إلى احتمالية وجود "بعد فرعي (Sub-dimension)" بسيط، لكنه ليس بالضرورة مهدداً للمقياس ما لم تكن النسبة المئوية عالية جداً.

- **Contrasts (2nd to 5th) :** نلاحظ أن القيم تتناقص تدريجياً (من 4.0% إلى 3.3%)، مما يعني أن الحالات أو الأبعاد الثانوية الأخرى ضعيفة جداً ولا تشكل خطراً على صدق المقياس

التوصية البحثية: يُنصح بمراجعة المحتوى العلمي للفقرات (27، 28، 29) ومقارنتها بالفقرات (24، 25، 26) لمعرفة ما إذا كانت تقيس مهارة أو سمة مختلفة عن بقية فقرات الاختبار.

1 الجدول 37 الخلاصة السيكومترية

المؤشر	الحالة	التفسير السيكومتري
القوة التفسيرية	41.0%	ممتازة؛ النموذج يفسر قدرًا كافيًا من البيانات.
أحادية البعد	قيد التحقق	القيمة في التباين الأول (5.0) أعلى من المحك المثالي (3.0)، مما يستوجب فحص الفقرات المرتبطة بهذا التباين للتأكد من عدم وجود "قدرة ثانوية" تؤثر على دقة النتائج.
التطابق (Empirical vs Modeled)	متطابق جداً	هناك تقارب كبير بين القيم التجريبية (Empirical) والقيم المتوقعة (Modeled)، مما يدل على أن البيانات تتوافق مع افتراضات النموذج الرياضي.

ملاحظة:

الجدول 38 تحليل المكونات الرئيسية للبقايا المعيارية

PRINCIPAL COMPONENTS ANALYSIS OF STANDARDIZED RESIDUALS

Name	F1	F2	F3	F4	F5	H2	U2
item1	0,25	-0,06	-0,08	0,14	0,08	0,10	0,90
item2	0,10	0,01	0,04	0,12	0,22	0,08	0,92
item3	0,27	0,01	0,22	-0,18	0,03	0,16	0,84
item4	0,41	-0,15	0,15	-0,11	-0,06	0,23	0,77
item5	0,12	-0,04	-0,04	0,04	-0,02	0,02	0,98
item6	0,02	-0,07	-0,34	0,27	0,01	0,19	0,81
item7	-0,06	-0,28	0,20	-0,24	0,16	0,21	0,79
item8	0,34	-0,22	0,08	-0,01	0,21	0,22	0,78
item9	0,04	-0,20	-0,14	-0,11	0,44	0,27	0,73
item10	0,09	-0,25	-0,30	0,05	0,29	0,25	0,75
item11	0,15	-0,19	0,08	-0,05	0,32	0,17	0,83
item12	0,06	-0,32	-0,26	-0,19	0,18	0,24	0,76
item13	-0,05	-0,27	-0,16	-0,07	0,25	0,16	0,84
item14	0,07	-0,05	-0,05	0,33	0,26	0,19	0,81
item15	0,11	0,11	-0,01	0,20	0,22	0,11	0,89
item16	-0,10	0,41	0,11	-0,24	-0,11	0,26	0,74
item17	-0,14	0,18	0,17	-0,25	-0,16	0,17	0,83
item18	-0,24	0,33	-0,18	-0,06	-0,17	0,23	0,77
item19	-0,19	0,20	-0,16	0,29	0,07	0,19	0,81
item20	-0,17	0,36	-0,21	0,45	0,25	0,47	0,53
item21	-0,17	0,48	-0,24	0,61	0,30	0,78	0,22
item22	-0,06	0,38	-0,16	0,50	0,30	0,52	0,48
item23	-0,31	0,18	0,12	0,18	0,05	0,18	0,82
item24	0,57	0,35	0,45	0,02	0,14	0,66	0,34
item25	0,62	0,37	0,42	0,01	0,17	0,73	0,27
item26	0,62	0,37	0,42	0,01	0,17	0,73	0,27
item27	-0,83	0,23	-0,04	-0,21	0,06	0,79	0,21
item28	-0,83	0,23	-0,04	-0,21	0,06	0,79	0,21
item29	-0,83	0,23	-0,04	-0,21	0,06	0,79	0,21
item30	-0,26	-0,32	-0,07	-0,20	0,13	0,23	0,77
item31	-0,15	-0,46	-0,07	0,05	0,37	0,37	0,63
item32	-0,26	-0,28	-0,04	0,05	0,25	0,21	0,79
item33	0,05	0,02	0,21	0,14	-0,12	0,08	0,92
item34	0,08	-0,24	-0,11	-0,17	-0,19	0,14	0,86
item35	0,18	-0,37	-0,01	-0,17	-0,48	0,43	0,57

item36	-0,14	0,08	0,12	0,04	-0,30	0,14	0,86
item37	0,09	0,07	0,05	-0,12	0,05	0,03	0,97
item38	0,05	0,28	0,15	0,09	-0,38	0,26	0,74
item39	0,04	-0,07	-0,04	0,20	0,16	0,07	0,93
item40	0,29	-0,09	-0,06	-0,04	-0,12	0,11	0,89
item41	-0,35	-0,24	0,71	0,31	-0,14	0,79	0,21
item42	-0,35	-0,24	0,71	0,31	-0,14	0,79	0,21
item43	-0,35	-0,24	0,71	0,31	-0,14	0,79	0,21
item44	0,22	-0,22	-0,15	0,29	-0,13	0,22	0,78
item45	-0,07	-0,18	-0,00	-0,26	-0,34	0,22	0,78
item46	-0,08	0,12	-0,21	0,01	-0,30	0,16	0,84
item47	0,14	0,38	-0,21	-0,37	-0,03	0,35	0,65
item48	0,22	0,37	-0,11	-0,48	0,08	0,44	0,56
item49	0,21	0,41	-0,08	-0,50	0,09	0,48	0,52
item50	0,06	0,12	-0,23	0,11	-0,30	0,17	0,83
item51	0,17	-0,38	-0,22	0,04	0,02	0,23	0,77
item52	0,27	-0,04	-0,38	0,39	-0,53	0,65	0,35
item53	0,27	-0,04	-0,38	0,39	-0,53	0,65	0,35
item54	0,01	0,37	0,04	0,10	-0,38	0,29	0,71
item55	-0,17	-0,27	-0,17	0,02	-0,33	0,24	0,76

هذا الجدول يمثل نتائج تحليل المكونات الأساسية للبواقي المعيارية (Principal Components)

(Analysis of Standardized Residuals)، وهو تحليل يُستخدم عادةً في إطار نموذج راش

(Rasch Model) للتحقق من شرط "أحادية البعد" (Unidimensionality).

1. الهدف من هذا التحليل

يهدف التحليل إلى التأكد مما إذا كان الاختبار يقيس سمة واحدة فقط (مثل الذكاء أو المهارة اللغوية).

إذا وُجدت أنماط قوية في "البواقي" (المعلومات التي لم يشرحها نموذج راش)، فهذا يعني وجود أبعاد ثانوية

(أبعاد أخرى تؤثر على الإجابات غير السمة الأساسية).

2. معاني الأعمدة الرئيسية

• F1, F2, F3, F4, F5 (العوامل/المكونات): هذه هي التباينات التي لم يستطع نموذج راش

تفسيرها. أهم عمود هو F1 (المكون الأول للبواقي).

○ إذا كانت هناك قيم عالية (أكبر من 0.40 أو أقل من -0.40) في العمود F1، فهذا يشير إلى وجود "بعد آخر" محتمل.

• H2 (Communalities - الاشتراكات): توضح مقدار التباين في الفقرة الذي تشرحه العوامل الخمسة المستخرجة. كلما اقتربت من 1، زادت قوة ارتباط الفقرة بالعوامل المكتشفة.

• U2 (Uniqueness - التفرد): هو التباين الخاص بالفقرة والذي لا يرتبط بالعوامل الأخرى. هو

$$\text{عكس } H2 \text{ حيث أن } H2 + U2 = 1$$

3. تحليل النتائج في (نقاط هامة)

بالنظر إلى الأرقام في الجدول، نلاحظ ملاحظات هامة جداً:

- مجموعات الفقرات المرتبطة ببعضها:
- الفقرات (24، 25، 26): لديها تحميلات موجبة قوية على F1 (0.57 إلى 0.62).
- الفقرات (27، 28، 29): لديها تحميلات سالبة قوية جداً على F1 (-0.83).
- الفقرات (41، 42، 43): لديها تحميلات قوية جداً على F3 (0.71).
- الفقرات (52، 53): لديهما قيم عالية في F4 و F5.

4. التفسير العملي (ماذا يعني هذا؟)

1. وجود أبعاد ثانوية: بما أن هناك فقرات بتشبعات عالية جداً (مثل -0.83 و 0.71)، فهذا يعني أن

الاختبار ليس أحادي البعد تماماً. هناك مجموعات من الأسئلة تقيس شيئاً مختلفاً عن البقية.

2. تكرار القيم: تلاحظ أن الفقرات (27، 28، 29) لها نفس الأرقام تماماً، وكذلك (41، 42، 43).

هذا قد يشير إحصائياً إلى أن هذه الفقرات متطابقة جداً في سلوكها أو أن هناك "تبعية موضعية"

(Local Dependency) بينها.

3. جودة الفقرات: الفقرات التي تمتلك قيم U2 عالية جداً (مثل item5 بنسبة 0.98) هي فقرات

"فريدة" جداً ولا ترتبط كثيراً بالبنية العامة للعوامل المستخرجة، مما قد يعني أنها تحتوي على الكثير من

"الغموض" أو الخطأ العشوائي.

يجب النظر في "Standardized Residual Loadings" للفقرات في التباين الأول (1st Contrast). إذا

وجدت مجموعة من الفقرات لها تشبعات عالية (أكبر من 0.4)، فقد نحتاج إلى مراجعة محتواها لمعرفة ما إذا

كانت تقيس قدرة مختلفة عن بقية المقياس.

تحليل التباين الأول (1st Contrast)

يُظهر الجدول رقم 23.3 (من مخرجات Winsteps) الفقرات التي تمتلك تشبعات (Loadings)

عالية، وهي الفقرات التي قد تشير إلى وجود بُعد ثانوي في الاختبار إذا تجاوزت قيمة التشبع: 0.4

2 فقرات ذات تشبعات موجبة عالية < 0.4)

• الفقرة 27 :تشبع مقداره 0.83.

• الفقرة 28: تشبع مقداره 0.83.

• الفقرة 29 : تشبع مقداره 0.83.

3 فقرات ذات تشبعات سالبة عالية (> 0.4)

• الفقرة 26 :تشبع مقداره -.0.62

• الفقرة 25 :تشبع مقداره -.0.62

• الفقرة 24 :تشبع مقداره -.0.57

• الفقرة 4 :تشبع مقداره -0.41

4 التفسير الأكاديمي للنتائج

- مؤشر أحادية البعد :وجود مجموعة من الفقرات (27، 28، 29) بتشبعات مرتفعة جداً (0.83) في التباين الأول يشير إلى احتمالية وجود "بُعد ثانوي" أو تكتل قوي لهذه الفقرات يختلف عن البعد الأساسي الذي يقيسه الاختبار .

الجدول رقم(39): التحليل العاملي للمكونات الأساسية للبواقي المعيارية. (الفقرات المجمعة)

عوامل البواقي	القيم الجذرية (Eigenvalue)	الفقرات	التفسير
F1(مجموعتين)	5.00	24حتى 26 مقابل 27 حتى 29	عدم الاستقلال الموضوعي
F3	3.42	41حتى 43	تأثير الاختبار الفرعي
F5	3.11	52-53	تأثير الاستراتيجيات

الجدول رقم (40): مواصفات نموذج متعدد الأبعاد

الابعاد	تعريف البعد	الفقرات
البعد 1	القدرة العامة	1-23, 30-40, 44-55
البعد 2	الاختبار الفرعي أ	24-26
البعد 3	الاختبار الفرعي ب	27-29
البعد 4	الاستدلال المتقدم	41-43
البعد 5	الاستراتيجية / ودرجة التعقيد	52-53

1. التطابق القوي للفقرات 10، 12، 13، 17، 51
2. قيم التحليل العاملي للبواقي الذاتية $2.0 < \text{PCA residual eigenvalues}$
3. مجموعات فقرات واضحة:

• 24، 25، 26

• 27، 28، 29

• 41، 42، 43

• 52–53

هذا ينتهك خاصية الاستقلال الموضوعي، وليس مبدأ راش نفسه.

لذلك، الحل الصحيح ليس حذف الفقرات، بل نمذجة أبعاد كامنة متعددة.

2. تحديد الأبعاد بناء على البيانات

البعد 1: القدرة الرياضية الأساسية العامل العام

الفقرات: 1 إلى 23، 30 إلى 40، 44 إلى 55 باستثناء الفقرات المعينة أدناه

1. تمثل القدرة المقصودة

2. مدعومة ببعد راش السائد

3. يفسر ~73٪ من التباين الكلي

البعد 2: الاختبار الفرعي للحساب

الفقرات: 24، 25، 26

1. التشعبات للبواقي ايجابية قوية

2. معامل الصعوبة والشكل متقاربين تقريبا

3. يشير إلى عدم الاستقلال الموضوعي.

البعد 3: الاختبار الفرعي للحساب ب

الفقرات: 27، 28، 29

1. التشبهات السلبية القوية للبواقي
2. القطب المعاكس لنفس التباين للبواقي
3. نفس المطلب المعرفي كما في 24، 25، 26، لكن في اتجاه مختلف

البعد 4: الاستدلال المتقدم / اختبار فرعي

الفقرات: 41، 42، 43

1. التشبهات للبواقي عالية جدا $F3 \approx 0.71$
2. تأثير النمط المشترك الكلاسيكي

البعد 5: الاستراتيجية / تأثير درجة التعقيد

الفقرات: 52، 53 اختياري 35

1. التشبهات للبواقي مرتفعة على العامل 5
2. من المحتمل أن يعكس درجة من التعقيد رياضي حسابي أو نمط مشتمل.

الجدول رقم 41: البنية النهائية لنموذج متعدد الأبعاد:

الابعاد	المحتوى	الفقرات
البعد 1	القدرة العامة للرياضيات	أغلب الفقرات
البعد 2	الاختبار الفرعي أ	26، 25، 24

البعد 3	الاختبار الفرعي ب	29، 28، 27
البعد 4	الاختبار الفرعي للاستدلال المتقدم	43، 42، 41
البعد 5	الاستراتيجية / تأثير درجة التعقيد	52-53

وهكذا نستنتج نموذج راش استكشافي متعدد الأبعاد، من خلال عملية التشخيص للبيانات، وهذا يمكن تفسير مشكلة عدم الملاءمة من خلال المقارنة بين نموذج راش احادي البعد والبنية متعددة الأبعاد. نلاحظ من خلال الجدول التالي.

الجدول رقم 41: المقارنة بين نموذج راش احادي البعد والنموذج متعدد الأبعاد.

المشكلات	نموذج راش احادي البعد	نموذج راش متعدد الأبعاد
عدم تطابق الفقرة	مرتفع	منخفض
عدم الاستقلال الموضوعي لل فقرات	تم تجاهلها	النمذجة بصورة واضحة
الصدق	مشكوك في دقته	قوي ومدعم
الاحتفاظ بالفقرات	اشكالية	ميرر

الأبعاد التي يجب الاحتفاظ بها في الاختبار النهائي باتباع المؤشرات المستخدمة لاتخاذ القرار:

1. القيمة الذاتية $2.0 <$
2. عناصر لكل بعد يفضل $3 \leq$
3. قابلية التفسير الموضوعية
4. تأثير ملاءمة النموذج

الاحتفاظ بالأبعاد الجوهرية والمستقرة

البعد 1 – القدرة العامة

- المفهوم الأساسي
- يفسر معظم التباين
- أساسي لتقديم الدرجات

البعد 4 – الفقرات 41 ، 42 ، 43

- تشبعات البواقي قوية جدا $0.71 \approx$
- تأثير الاختبار الفرعي الواضح
- يجب الاحتفاظ به كبعد اختبار فرعي

الأبعاد 2 و 3 الفقرات 24 حتى 29

- تعكس عدم الاستقلال الموضعي
- أفضل طريقة للتعامل معها كأبعاد اشكالية/اختبارات فرعية
- لا تظهر اشكالية الدرجات الفرعية المنفصلة

اختياري / عدم ادراجها في الإصدارات المستقبلية

البعد 5 الفقرتان 52–53

- فقرتان فقط
- غير مستقر سيكومتريا
- يوصى بمراجعة الفقرات او حذفها.

التوصية النهائية

الجدول 42 اختيار الأبعاد

البعد	الاجراء المتبع
D1	يحتفظ به
D2-D3	نموذج اختبارات فرعية
D4	يحتفظ به كاختبار فرعي
D5	مراجعة او حذف

الخلاصة:

توصلنا الى أن الاختبار قد يحتوي على أبعاد فرعية. إذا كنا نريد أن يكون الاختبار "أحادي البعد" (يقيس قدرة واحدة فقط)، يجب مراجعة الفقرات التي تظهر قيماً متطرفة في F1 أو F2، لأنها قد تشتت دقة القياس الأساسي. ومن خلال مقارنة نماذج القياس لمعالجة عدم توافق الفقرات وعدم الاستقلال الموضعي للبواقي، تم تحديد نموذج راش الاستكشافي متعدد الأبعاد. تم نمذجة مجموعات من الفقرات التي تم تحديدها من خلال التحليل العاملي للمكونات الأساسية للبواقي المعيارية كأبعاد كامنة منفصلة. هذه المقاربة تحسن بشكل كبير تمثيل القدرة الكامنة مع الاحتفاظ بالفقرات ذات الأهمية الجوهرية. قدم إطار راش متعدد الأبعاد تمثيلاً متفوقاً للبيانات من خلال نمذجة عدم الاستقلال الموضعي وتأثير الاختبارات الفرعية بشكل واضح. بينما تم تأكيد بعد القدرة العامة السائد، عكست عدة أبعاد ثانوية تأثيرات الطرق والمحتوى بدلاً من المفاهيم المتميزة جوهرياً. من خلال نمذجة هذه التأثيرات يتم تحسين ملاءمة الفقرات وصدق المفهوم دون المساس بقابلية تفسير الدرجات.

تقييم صدق المفهوم باستخدام تحليل المكونات الرئيسية للبواقي (PCA)

يتحقق هذا التحليل مما إذا كان هناك "متغير" واحد فقط يتم قياسه (أحادية البعد) ولكن مما توصلنا اليه فقد كان مخالف لذلك، فكيف يؤثر ذلك على خاصية الصدق؟

أ. قوة المفهوم

○ التباين المفسر بواسطة المقاييس: (41.0%) في الجدول 23.0، من التباين الكلي و هذه "الإشارة" تثبت وجود مفهوم مهيم. حوالي 41% من التباين في إجابات التلاميذ يعود إلى الاختلافات بين قدرة التلميذ وصعوبة الفقرة.

○ ملاحظة: في حين أن هذه النسبة مقبولة، إلا أن القيمة المثالية غالباً ما تكون $< 50\%$. يشير هذا إلى وجود مشكل أو عوامل أخرى تؤثر على الدرجات كما رأينا سابقاً.

ب. التهديد لصدق المفهوم (وجود البعد الثاني)

• التباين غير المفسر في التباين الأول (القيمة الذاتية = 5.0) ويعتبر الإحصاء الأكثر أهمية في هذا الجدول تعني القيمة الذاتية البالغة 5.0 أن هناك عاملاً ثانوياً في البيانات يمتلك قوة تعادل 5 فقرات. في تحليل راش، القاعدة العامة الشائعة هي أن التباين الأول يجب أن يكون < 2.0 أو > 3.0 .

○ الاستنتاج: القيمة الذاتية 5.0 ذات دلالة إحصائية. وهي تثبت أن الاختبار ليس أحادي البعد بشكل صارم. إنه يقيس القدرة الرئيسية، ولكن أيضاً قدرة ثانوية قوية.

ج. تحديد "البعد الثاني" (فحص صدق المحتوى) يساعد الجدول 23.3 في تشخيص سبب تراجع الصدق. فهو يحدد مجموعتين متعارضتين من الفقرات:

○ المجموعة أ (تشبعات موجبة): الفقرات I0027، I0028، I0029 (تشبعات $+0.83$) والفقرات I0041، I0042، I0043 تشبعات $+0.35$

○ ملاحظة: الفقرات 41-43 هي الأصعب ($+2.77$ لوجيت)، بينما 27-29 متوسطة الصعوبة ($+0.60$ لوجيت).

2. المجموعة ب (تشبعات سالبة): الفقرات 24، 25، 26 تشبعات من -0.57 إلى -0.62

○ ملاحظة: هذه فقرات متوسطة الصعوبة ($+1.20$ لوجيت).

- التعارض: يميل التلاميذ الذين يؤدون أداءً جيداً في فقرات المجموعة أ (التشبعات الموجبة) إلى الأداء بشكل سيء غير متوقع في فقرات المجموعة ب (التشبعات السالبة)، والعكس صحيح.
 - دلالة على الصدق: يشير هذا النمط إلى أن الاختبار يقيس مهارتين مختلفتين. على سبيل المثال، إذا كان هذا اختبار رياضيات، فقد تكون المجموعة أ "هندسة" والمجموعة ب "التناسبية". قد يكون التلميذ جيداً في الهندسة ولكن ضعيفاً في التناسبية، مما يتسبب في تقلب درجته بشكل غير متوقع.
 - الإجراء المطلوب: لتحسين صدق البناء، يجب التحقق في محتوى الفقرات 27، 28، 29، 41، 42، 43 مقابل 24، 25، 26 إذا كانت تقيس مفاهيم مختلفة، يجب تقسيم الاختبار إلى اختبارين فرعيين.
- بناءً على ملفات مخرجات برنامج Winsteps (الجدول 12.2، 12.12، و16.3)، شرح لكيفية إثبات صدق المفهوم (Construct Validity) وصدق الترتيب (Ordering Validity) مع الحفاظ على الرموز المستخدمة في الجداول:

أولاً تحليل صدق المفهوم خريطة ورايت (Wright Map) التمثيل البياني 29

```

1  TABLE 12.2 C:\Winsteps\doct25.txt          ZOU244WS.TXT  May  3 15:00 2026
2  INPUT: 559 Person  55 Item  REPORTED: 559 Person  55 Item  2 CATS WINSTEPS 3.72.1
3  -----
4
5      Person - MAP - Item
6      <more>|<rare>
7
8      5      +
9      |
10     |
11     |
12     |
13     4      +
14     |
15     |
16     |
17     .# |T
18     . |
19     3      +
20     . | I0041 I0042 I0043
21     .# T|
22     .# | I0039
23     .# |
24     .# |
25     2      +
26     .#### | I0046 I0049
27     ## |S I0048 I0052 I0053
28     .### |
29     .##### S| I0036 I0047
30     .#### | I0015 I0025 I0026 I0044
31     1      + I0016 I0024 I0037
32     .#### | I0033 I0051 I0054
33     .##### | I0027 I0028 I0029 I0032 I0038
34     .##### | I0023
35     .##### | I0017
36     .##### M| I0002 I0004 I0031
37     0      .##### +M I0018 I0050
38     .#### | I0035 I0045
39     .##### | I0034 I0040
40     .#### | I0013
41     .##### | I0008
42     .### |
43     -1     .#### + I0003 I0010 I0030
44     .##### S| I0012
45     .# |
46     .### | I0007 I0014
47     .## |S I0001
48     .## | I0006 I0055
49     -2     +
50     .### | I0009
51     T|
52     | I0005
53     ## |
54     .# |
55     -3     +
56     | I0021
57     |T I0020
58     |
59     | I0011 I0022
60     |
61     -4     + I0019
62     <less>|<frequ>
63     EACH "#" IS 4. EACH "." IS 1 TO 3

```

يتم تقييم صدق المفهوم في نموذج راش من خلال فحص الهرمية المنطقية لصعوبة الفقرات. الصدق يعني

أن الفقرات مرتبة من "الأسهل" إلى "الأصعب" بشكل يطابق التعريف النظري للقدرة المقاسة.

1.هرمية الفقرات (الجداول 12.2 و12.12)

تُظهر خريطة ورايت توزيعاً واسعاً لصعوبة الفقرات يمتد عبر (7لوجيت) من -4 إلى +3

- أسهل الفقرات (أسفل الخريطة):

○ الفقرات 19 : بصعوبة (-4.07 لوجيت) والفقرات 11، 20، 22

- التفسير: تمثل هذه الفقرات "المستوى القاعدي" للقدرة، حيث يُتوقع من معظم التلاميذ الإجابة عليها بشكل صحيح .

- أصعب الفقرات (أعلى الخريطة):

○ الفقرات 41، 42، 43 بصعوبة +2.77 لوجيت والفقرات 39، 46.

- التفسير: تمثل هذه الفقرات "المستوى المتقدم". فقط التلاميذ ذوي القدرة العالية يمكنهم الإجابة عليها بشكل صحيح .

تحليل الفجوات (Gap Analysis)

- **الملاحظة:** توجد فجوة ملحوظة في كثافة الفقرات بين التكتل السهل (حول -3 لوجيت) والتكتل المتوسط (حول -1 لوجيت)، وأخرى بين الفقرات المتوسطة والتكتل الصعب (بين +2 و +3) .
- **الدلالة:** الاختبار يفتقر إلى الدقة في التمييز في المناطق التي توجد بها **الفجوات**، بينما يعرف المفهوم ويميز الطلاب بدقة عالية في المناطق التي تتكدس فيها الفقرات .

تحليل الفجوات

- الفجوة بين -3 و -1 لوجيت: بما أنه لا توجد فقرات كافية هنا، فالاختبار لا يستطيع التمييز جيداً بين التلميذ "الضعيف جداً" (-3) والتلميذ "الأقل ضعفاً" (-1). بالنسبة للاختبار، هذه المنطقة منطقة "قفزة" أو فراغ معلوماتي.

- الفجوة بين 2+ و 3+ لوجيت :هنا الاختبار يفقد دقته في التمييز بين الطلاب المتفوقين والطلاب العابرة. التلميذ الذي قدرته 2.5+ قد يحصل على نفس نتيجة التلميذ 3+ لأنه لا توجد فقرات "تتحده" في هذه الفجوة.

- التكتلات (Clusters): الاختبار يميز بدقة فقط عند المستويات التي تتكدس فيها الفقرات (مثلاً عند 3-، وعند 1-، وعند 2+).

الخلاصة

"الدلالة :وجود هذه الفجوات يشير إلى أن دقة القياس (Precision) ليست متساوية على طول متصل القدرة. الاختبار يمتلك قدرة تمييزية عالية عند مستويات القدرة التي تتركز فيها الفقرات، ولكنه يعاني من انخفاض في الحساسية التمييزية في المناطق الانتقالية (الفجوات). هذا يعني أن الاختبار قد لا يفرق بدقة كافية بين التلاميذ الذين تقع قدراتهم ضمن هذه الفراغات اللوجيتية".

خلاصة القاعدة الذهبية في راش:

- تكتل فقرات =دقة تمييز عالية + خطأ معياري منخفض في تلك المنطقة.
- فجوة فقرات =دقة تمييز منخفضة + خطأ معياري مرتفع + عدم قدرة على الفصل بين مستويات الطلاب في تلك المنطقة.

• ثانياً: صدق الترتيب عبر "خرائط الاستجابة" (التمثيل البياني 31)

```

1  TABLE 16.3 C:\Winsteps\doct25.txt          ZOU244WS.TXT  May  3 15:00 2026
2  INPUT: 559 Person  55 Item  REPORTED: 559 Person  55 Item  2 CATS WINSTEPS 3.72.
3  -----
4
5  Item - MAP - Person
6  <rare>|<more>
7
8  5
9  |
10 |
11 |
12 |      0s  0s  0s
13
14  4
15  |
16  |
17  T|  0a  0a  0s  0s  1s
18  |  1a
19
20  3
21  XXX |  1a  1s  1s
22  T|  0s  1a  1s  1s  1s
23  X |  0a  0a  0s  0s  1a
24  |  0a  0s  0s  1a  1s
25  |  0s  0s  1a  1a  1a  1s  1s
26
27  2
28  +  0a  0a  0a  0s  1a  1a
29  XX |  0a  0a  0a  0a  0s  0s  0s  0s  1a  1a  1a  1a  1a  1a
30  |  1a  1s  1s
31  XXX S|  0a  0s  0s  1a  1a  1a  1s  1s
32  |  0a  0s  0s  0s  0s  0s  0s  0s  1a  1a  1a  1s  1s  1s
33  XX |S  0a  0a  0a  0a  0s  0s  0s  0s  0s  0s  1a  1a  1a  1a  1a
34  |  1a  1a  1a  1a  1s  1s  1s  1s  1s  1s
35  XXXX |  0a  0a  0s  0s  0s  0s  1a  1a  1a  1a  1a  1s  1s  1s
36  |  1s  1s  1s  1s
37
38  1
39  XXX +  0a  0a  0a  0a  0a  0a  0s  0s  0s  0s  1a  1a  1a  1a
40  |  1a  1a  1a  1a  1s  1s  1s  1s
41  XXX |  0a  0a  0a  0a  0a  0s  0s  0s  0s  0s  1a  1a  1a  1a
42  |  1s  1s  1s
43  XXXXX |  0a  0a  0a  0a  0a  0a  0a  0a  0s  0s  0s  0s  0s  0s  0s
44  |  0s  0s  1a  1a  1a  1a  1a  1a  1a  1a  1a  1s  1s  1s
45  |  1s  1s  1s  1s  1s  1s  1s
46
47  X |  0a  0a  0a  0s  0s  0s  0s  0s  0s  0s  1a  1a  1a  1a  1a
48  |  1a  1a  1a  1a  1a  1s  1s  1s  1s  1s  1s  1s  1s
49
50  X |  0a  0a  0a  0a  0a  0a  0a  0a  0a  0a  0a  0s  0s  0s  0s
51  |  0s  0s  0s  1a  1a  1a  1a  1a  1a  1s  1s  1s  1s  1s
52  |  1s  1s  1s  1s
53
54  XXX |M  0a  0a  0a  0a  0a  0a  0s  0s  0s  0s  1a  1a  1a  1a  1s
55  |  1s  1s  1s  1s  1s  1s
56
57  0
58  XX M+  0a  0a  0a  0a  0a  0a  0a  0a  0a  0s  0s  0s  0s  0s  0s
59  |  0s  0s  1a  1a  1a  1a  1a  1a  1a  1a  1a  1a  1a  1a  1a
60  |  1a  1s  1s  1s  1s  1s  1s  1s  1s  1s  1s
61
62  XX |  0a  0a  0s  0s  0s  0s  1a  1a  1a  1a  1a  1a  1a  1a
63  |  1s  1s  1s  1s
64
65  XX |  0a  0a  0a  0a  0a  0a  0a  0a  0a  0a  0s  0s  0s  0s  0s
66  |  0s  0s  0s  0s  0s  0s  1a  1a  1a  1a  1a  1a  1a  1s
67  |  1s  1s  1s  1s  1s  1s  1s
68
69  X |  0a  0a  0a  0a  0a  0a  0s  0s  0s  0s  1a  1a  1a  1s
70  |  1s  1s  1s
71
72  X |  0a  0a  0a  0a  0a  0a  0a  0a  0a  0a  0a  0a  0s  0s  0s
73  |  0s  0s  0s  0s  0s  1a  1a  1a  1s  1s  1s  1s  1s  1s
74  |  1s  1s  1s
75
76  -1
77  XXX +  0a  0a  0a  0a  0a  0a  0s  0s  0s  0s  0s  1a  1a  1s  1s
78  |  1s  1s
79
80  X |S  0a  0a  0a  0a  0a  0a  0s  0s  0s  0s  0s  1a  1a  1a  1a
81  |  1a  1a  1a  1a  1a  1a  1a  1a  1s  1s
82
83  XX |  0a  0a  0s  0s  1a  1s  1s
84  |  1s
85
86  X S|  0a  0s  0s  0s  0s  1a  1a  1s  1s
87  XX |  0a  0s  0s  0s  0s  1a  1a  1a  1a  1s  1s
88
89  -2
90  +
91  X |  0a  0a  0a  0s  1a  1a  1a  1a  1a  1s  1s  1s  1s
92  |T
93
94  X |
95  |  0a  0s  1a  1a  1a  1a  1s  1s
96  |  0s  1a  1a  1s  1s
97
98  -3
99  +
100 X |
101 X T|
102 |
103 XX |
104 |
105
106 -4
107 X +
108 <frequ>|<less>

```

لمعرفة كيفية إثبات أن ترتيب المستجيبين من الأقل أداءً إلى الأعلى أداءً يطابق التوقعات. في نموذج راش، الإثبات البياني نلاحظه في الجدول 43 حيث: **المحور العمودي (الصفوف)**: يمثل صعوبة الفقرات (الفقرات الصعبة في الأعلى، والسهلة في الأسفل). **المحور الأفقي (الأعمدة)**: يمثل قدرة الأفراد (القدرة المنخفضة على اليسار، والعالية على اليمين). **الرموز**: "0" إجابة خاطئة. "1" إجابة صحيحة. "s" استجابة نمطية متوافقة. (Standard fit). "a" استجابة غير متوقعة أو شاذة. (Anomaly) كيف نثبت صحة الترتيب من البيانات؟ يظهر الجدول 43 نمطاً يسمى "Guttman Pattern"، وهو الدليل البصري على صدق الترتيب: **الربع العلوي الأيسر (فقرات صعبة / تلاميذ ضعاف)**: ستلاحظ كتلة كثيفة من الأصفار. (0s, 0a) هذا يثبت أن تلاميذ ذوي الأداء المنخفض أخفقوا في الأسئلة الصعبة كما هو متوقع.

1. الربع السفلي الأيمن (فقرات سهلة / طلاب أقوياء):

○ ستلاحظ كتلة كثيفة من الواحدات. (1s, 1a) هذا يثبت أن الطلاب المتفوقين أجابوا بشكل صحيح على الأسئلة السهلة.

2. الخط القطري الانتقالي:

○ في منتصف الخريطة، ستجد منطقة تتحول فيها الأصفار تدريجياً إلى وحدات كلما انتقلت من اليسار (الضعفاء) إلى اليمين (الأقوياء). هذا التدرج الواضح يثبت أن الاختبار يفرق بنجاح بين مستويات القدرة المختلفة.

التعامل مع الانحرافات: (Noise)

- ظهور "0a" في جهة اليمين أسفل الخريطة يعني أن طالباً قوياً أخطأ في سؤال سهل (قد يكون بسبب السهو).

- ظهور "1a" في جهة اليسار أعلى الخريطة يعني أن طالباً ضعيفاً أجاب صح على سؤال صعب (قد يكون بسبب التخمين) .

- الخلاصة: طالما أن النمط العام هو (0 في الأعلى يساراً، و 1 في الأسفل يميناً)، فإن صدق الترتيب محقق إحصائياً .

يمكن شرح هذه العبارة هي المفتاح لفهم خارطة الاستجابة (Scalogram) في الجدول رقم 44 ، وهي تعني أن البيانات تتبع منطقاً تراكمياً سليماً. وهذا شرح تفصيلي لما يعنيه هذا النمط بوضوح:

1. مبدأ الترتيب (لماذا "0" في الأعلى يساراً و "1" في الأسفل يميناً؟)

يعتمد نموذج "راش" على مبدأ أن الشخص الأكثر قدرة لديه احتمالية أعلى للإجابة على أي فقرة مقارنة بالشخص الأقل قدرة. هذا المبدأ يُترجم بصرياً في الجدول كالتالي:

- الأعلى يساراً (فقرات صعبة + طلاب ضعاف): هذا الجزء يجمع أصعب التحديات مع أقل القدرات. إحصائياً، من المفترض أن تكون النتيجة هي الإخفاق، ولذلك نجد الرمز "0" (إجابة خاطئة) مسيطراً هنا.

- الأسفل يميناً (فقرات سهلة + طلاب أقوياء): هذا الركن يجمع أسهل الفقرات مع أعلى القدرات. إحصائياً، النجاح هنا مضمون تقريباً، ولذلك نجد الرمز "1" (إجابة صحيحة) هو السائد.

عندما نشاهد هذا التوزيع القطري في جدول 16.3، فانه يقدم دليلاً ملموساً على "صدق الترتيب"

(Ordering Validity) من ثلاثة جوانب:

- اتساق الأداة: يثبت أن الفقرات التي رتبها النموذج كفقرات "صعبة" (مثل الفقرة 41) هي بالفعل صعبة على التلاميذ الضعاف.

- **دقة تصنيف الطلاب:** يثبت أن الطالب الذي حصل على درجة عالية (في جهة اليمين) لم يحصل عليها بالصدفة، بل لأنه اجتاز الفقرات السهلة والمتوسطة والصعبة.

- **أحادية البعد:** هذا النمط المنظم يشير إلى أن الاختبار يقيس "سمة واحدة" واضحة؛ فلو كانت الاستجابات مبعثرة (1 و 0 بشكل عشوائي)، لكان ذلك دليلاً على أن النجاح والفشل لا يعتمدان على القدرة بل على عوامل أخرى مشوشة وهذا ما يثبت القدرة الأساسية لنموذج راش.

3. كيفية قراءة الرموز المصاحبة (s, a) داخل هذا الجدول (Table 16.3)، يوجد هذه الأرقام مصحوبة بأحرف صغيرة، وهي تزيد من قوة الإثبات:

- **الرمز "s1" أو "s0":** حرف "s" يعني (Standard)، أي أن استجابة الطالب "نمذجية" تماماً كما توقعها النموذج. كثرة هذه الرموز تعني أن البيانات "نظيفة" إحصائياً.

- **الرمز "a1" أو "a0":** حرف "a" يعني (Anomaly)، وهي انحرافات بسيطة لا تكسر القاعدة العامة. مثلاً، إذا وجدنا "a1" في أقصى اليسار (طالب ضعيف حل سؤالاً صعباً)، فهذا قد يشير إلى "تخمين" ناجح، لكنه لا يؤثر على الصدق الكلي طالما أن النمط العام (0 يساراً و 1 يميناً) لا يزال متماسكاً.

باختصار: هذه العبارة هي "استنتاجات صادقة" بأن الاختبار يرتب التلاميذ والفقرات على خط مستقيم واحد وبشكل منطقي يتفق مع النظرية التربوية.

الخلاصة: البيانات في الجدول 16.3 تُظهر بوضوح أن الاختبار يمتلك **صدق هرمي**؛ حيث أن "ترتيب" التلاميذ يتوافق تماماً مع "ترتيب" صعوبة الفقرات، مما يجعل التنبؤ بمستوى التلميذ بناءً على درجته أمراً صادقاً وموثوقاً .

3.الدليل الإحصائي على ترتيب المستجيبين (صدق التنبؤ)

يمكن التساؤل عما إذا كان ترتيب التلاميذ يطابق التوقعات. في نموذج راش، يتم تقييم "صدق التنبؤ" من حيث الترتيب باستخدام مخطط جوتمان (Guttman Scalogram) (التسلسل الهرمي للاستجابات). توفر جداول PCA دليلاً إحصائياً على موضع نجاح هذا الترتيب وموضع فشله.

أ. دليل الترتيب "الجيد"

- حقيقة أن 41% من التباين مفسر بواسطة المقاييس تثبت أن الترتيب صالح للأغلبية.
- يثبت مؤشر معامل فصل التلاميذ المرتفع (محسوب من التباين المفسر) إحصائياً أن الاختبار يميز بنجاح بين التلاميذ الأعلى والمتوسطين والأدنى بترتيب قابل للتكرار.
- ب. دليل الترتيب "الفاشل" (التناقضات) يوفر الجدول 43. ("تباين الفقرة 1 استجابات متباينة حسب الشخص") الدليل الإحصائي على موضع فشل التنبؤ.
- دليل سوء التنبؤ:

○ عمود "Person FAVORS TOP" (الشخص يميل للأعلى). بالنسبة للشخص رقم 11 (وآخرين)، تُظهر البيانات 3 إجابات صحيحة في الفقرات العليا ولكن 0 صحيحة في الفقرات السفلى. (الجدول 44. مقارنة بين التلاميذ)

•	Person FAVORS BOTTOM						
•	TOP 4 Item			BOTTOM 4 Item			
•	HIGH	EXP.	LOW	HIGH	EXP.	LOW	
•	3	1	0	0	3	1	11 1a
•	0	1	3	3	1	0	33 1s

○ انظر إلى "Person FAVORS BOTTOM" (الشخص يميل للأسفل). بالنسبة للشخص رقم 28، تُظهر البيانات 0 إجابات صحيحة في الفقرات العليا ولكن 3 صحيحة في الفقرات السفلى.

Person FAVORS BOTTOM						
TOP 4 Item			BOTTOM 4 Item			
HIGH	EXP.	LOW	HIGH	EXP.	LOW	
0	1	3	3	1	0	28 0s
0	1	3	3	1	0	56 1a

- الاستنتاج الإحصائي: تثبت قوائم الأشخاص هذه أن درجة "متوسط الصعوبة" الواحدة لا تنبأ بالأداء بدقة لهؤلاء الأفراد. "ترتيبهم" (مستواهم) غير مستقر لأن أداءهم يعتمد على الفقرات التي حصلوا عليها (المجموعة أ مقابل المجموعة ب).

• تحليل الحالات

• الحالة الأولى 1:(a & 1s)

• 3 1 0 | 0 3 1

- في الفقرات الصعبة: (TOP 4) الشخص حصل على 3 درجات (إجابات صحيحة) في فقرات كان من المتوقع أن يفشل فيها هذا "سلوك مفاجئ" قد يعني ذكاءً حاداً في مواضيع معينة أو تخميناً ناجحاً.

• في الفقرات السهلة: (BOTTOM 4) الشخص أخطأ في فقرة واحدة سهلة جداً (Low = 1)

بينما أجاب كما هو متوقع في 3 فقرات.

• الحالة الثانية: (0s & 1a)

• 0 1 3 | 3 1 0

- هذا النمط يسمى "النمط العكسي" وهو الأكثر إثارة للاهتمام في القياس النفسي، حيث في الفقرات الصعبة: (TOP 4) الشخص رسب في 3 فقرات (Low = 3) كان من الممكن أن يحلها، وأصاب في واحدة فقط كمتوقع.

- في الفقرات السهلة: (BOTTOM 4) الشخص نجح في 3 فقرات (High = 3) كانت سهلة جداً بالنسبة له.

- الخلاصة: هذا الشخص "منطقي" جداً (Guttman pattern)، حيث يحل السهل ويفشل في الصعب، وهو ما يفسر القيم (0) في الخانات المتطرفة.

الخلاصة

"يكشف التحليل الإحصائي لصدق البناء صورة مختلطة. يفسر نموذج راش 41.0% من التباين الكلي، مما يؤكد وجود سمة كامنة مهيمنة. ومع ذلك، يحدد تحليل المكونات الرئيسية (الجدول 43) بعداً ثانوياً كبيراً، كما يدل على ذلك القيمة الذاتية البالغة 5.0 في التباين الأول. وهذا يشير إلى انتهاك مبدأ أحادية البعد الصارمة.

يكشف تحليل مجموعات الفقرات (الجدول 43) أن الاختبار يقيس مفهومين مختلفين محتملين: أحدهما محدد بالفقرات 27، 28، 29، 41، 42، 43 والآخر بالفقرات 24، 25، 26 وبالتالي، فإن ترتيب المستجيبين صالح إحصائياً على المستوى العام (مدعوم بمعامل فصل الأشخاص)، ولكنه غير صالح لمجموعات فرعية محددة من التلاميذ تتوافق قدراتهم مع أحد البعدين المحددين فقط".

مقارنة موسعة بين CTT و Rasch و IRT

تُعد كل من النظرية الكلاسيكية للاختبار (CTT) ونظرية القياس لراش (Rasch Measurement Theory) ونظرية الاستجابة للمفردة (Item Response Theory: IRT) من الأطر الأساسية في القياس النفسي والتربوي. وعلى الرغم من أن هذه النظريات تهدف جميعها إلى تقدير السمات الكامنة لدى الأفراد، فإنها تختلف في افتراضاتها الرياضية وطرق تقدير المعلومات وتفسير النتائج.

جدول (45) مقارنة نظرية بين CTT و Rasch و IRT

البعد المقارن	النظرية الكلاسيكية CTT	نظرية IRT العامة	نموذج Rasch
الهدف الأساسي	تقدير الدرجة الحقيقية للفرد	نمذجة احتمال الاستجابة	بناء مقياس قياس موضوعي
طبيعة النموذج	نموذج خطي	نماذج لوجستية متعددة	نموذج لوجستي بسيط
مستوى التحليل	الاختبار ككل	المفردة والفرد	المفردة والفرد
المعلومات الأساسية	الدرجة الحقيقية والخطأ	، الصعوبة (θ) القدرة ، (a) ، التمييز (b) ، (c) التخمين	وصعوبة المفردة (θ) القدرة (δ)
العلاقة بين المفردة والفرد	غير صريحة	صريحة	صريحة
استقلال المعلومات	غير متحقق	متحقق جزئياً	متحقق (خاصية الثبات)
تأثير العينة	قوي	متوسط	ضعيف
تقدير القدرة	يعتمد على الدرجة الكلية	يعتمد على النموذج الاحتمالي	يعتمد على النموذج الاحتمالي
مقياس القدرة	مقياس رتبي	مقياس فاصل	(logit) مقياس فاصل
التعامل مع البيانات المفقودة	صعب	سهل	سهل
تقدير الخطأ المعياري	ثابت	متغير	متغير
استخدام الدرجة الكلية	أساسي	ليس بالضرورة	إحصاء كاف

جدول (46) مقارنة المعادلات الرياضية

النظرية	المعادلة الأساسية
CTT	$X = T + E$
Rasch	$P_{ni} = \frac{e^{(\theta_n - \delta_i)}}{1 + e^{(\theta_n - \delta_i)}}$
2PL IRT	$P_{ni} = \frac{1}{1 + e^{-a_i(\theta_n - b_i)}}$
3PL IRT	$P_{ni} = c_i + (1 - c_i) \frac{1}{1 + e^{-a_i(\theta_n - b_i)}}$

تحليل ثبات واستقرار تقديرات التلاميذ (559 تلميذاً)

لقد تم إجراء هذا التحليل لتقييم مدى دقة الاختبار واستقرار ترتيب التلاميذ عند الانتقال من القياس التقليدي إلى القياس وفق نموذج راش.

تحليل الثبات العام (Reliability)

الجدول 47 يعكس هذا الجزء مدى دقة الاختبار كأداة قياس موحدة للمجموعة.

مقياس المقارنة	نموذج راش (Rasch)	النظرية الكلاسيكية (CTT)
مؤشر الثبات	0.91 ثبات الأفراد و 0.92 ثبات الاختبار	0.99 معامل ألفا
الخطأ المعياري	متغير حسب مستوى كل تلميذ	ثابت لجميع التلاميذ (1.89)
التفسير	ثبات قوي يسمح بالتمييز بين 3 مستويات من القدرة على الأقل.	ثبات ممتاز يشير إلى أن الاختبار دقيق بنسبة 99%

. استقرار ترتيب التلاميذ معامل سبيرمان (Spearman)

لقد قمنا بحساب معامل ارتباط الرتب لـ "سبيرمان" لمعرفة ما إذا كان ترتيب التلاميذ يتغير عند استخدام نموذج راش بدلاً من الدرجة الكلية.

- معامل ارتباط سبيرمان 0.998 (rho):
- التفسير: توجد علاقة طردية شبه تامة. هذا يعني أن ترتيب التلاميذ مستقر جداً، وأن التلميذ المتفوق في النظرية الكلاسيكية يظل متفوقاً في نموذج راش، مع تغييرات طفيفة جداً لا تكاد تُذكر.

تحليل خط الانحدار (Regression Line)

عند رسم العلاقة بين "رتبة التلميذ في النظرية الكلاسيكية" و "رتبته في نموذج راش"، حصلنا على خط انحدار مستقيم شبه مثالي.

- معامل التحديد ، 0.999 معنى ذلك ان النقاط تقع جميعها تقريباً على قطر الرسم البياني، مما يؤكد أن كلا النموذجين يقيسان نفس "القدرة الكامنة" لدى التلاميذ بنفس الترتيب تقريباً.

كشف الفروقات (الإزاحة في الرتب)

رغم الاستقرار العالي، لاحظنا وجود "إزاحات" طفيفة لبعض التلاميذ (أقصى إزاحة كانت 3 مراكز فقط).

وسبب ذلك في نموذج راش، لا نكتفي بجمع الدرجات، بل ننظر إلى "صعوبة الأسئلة" التي أجاب عليها التلميذ.

- التلميذ الذي يتقدم رتبة: هو من نجح في الإجابة على أسئلة صعبة رغم أن مجموعه الكلي مساوٍ لغيره.
- التلميذ الذي يتراجع رتبة: هو من أخطأ في أسئلة سهلة جداً (تعتبر سقطه أو تسرع)، مما يجعل النموذج يصحح رتبته قليلاً للأسفل.

النتائج تعكس اختباراً ذا جودة عالية وثباتاً ممتازاً. الاستقرار بين (CTT) و (Rasch) بنسبة 99.8% هو مؤشر قوي على أن الاختبار عادل وأن الدرجات تعبر بصدق عن مستوى التلميذ الحقيقي.

كما تم تحليل البيانات من النصفين لاستخراج "القدرة" المقدرة لكل تلميذ وفق النموذجين. في النظرية الكلاسيكية (CTT) نعتمد على الدرجة الكلية الملحوظة، بينما في نموذج راش (RASCH)

نعتمد على المقاييس باللوجيت

التحليل المقارن والارتباط بينهما:

هذا الجدول يضع القدرتين جنباً إلى جنب، مما يوضح كيف تتحول الدرجة الخام إلى مقياس احتمالي

دقيق في نموذج راش. جدول 48 مقارنة القدرات (عينة لأول 20 تلميذاً)

رقم التلميذ	القدرة - CTT الدرجة الكلية	القدرة RASCH - لوجيت
7	18	-0.50
20	18	-0.50
1	17	-0.64
11	17	-0.64
9	16	-0.78
14	16	-0.78
6	15	-0.92
13	15	-0.92
18	15	-0.92
2	14	-1.07
10	14	-1.07
17	14	-1.07
19	14	-1.07
5	13	-1.22
12	13	-1.22
8	12	-1.37
16	12	-1.37
3	11	-1.53
4	10	-1.69
15	9	-1.87

تحليل الارتباط (Correlation Analysis)

لحساب قوة العلاقة بين النموذجين لجميع التلاميذ (559 تلميذاً):

- معامل ارتباط بيرسون $r = 0.999$ (r)
- القيمة الاحتمالية: (p -value) أقل من 0.001 (دلالة إحصائية قوية جداً).

التفسير: الارتباط يكاد يكون تاماً (1.00). هذا يعني أن الدرجة الكلية هي مؤشر ممتاز وقوي جداً للقدرة المقدرة في نموذج راش. أي زيادة في الدرجة الخام تقابلها زيادة متناسبة جداً في اللوجيت.

. الفروقات الجوهرية بين العمودين

رغم الارتباط العالي، هناك فرق نظري وتقني بين العمودين:

1. المسافات: في عمود CTT، الفرق بين الدرجة 10 و 11 هو دائماً (1). أما في RASCH،

المسافات بين اللوجيت تختلف؛ فالجهد المبذول للانتقال من درجة 30 إلى 31 أصعب بكثير من

الانتقال من 15 إلى 16، وهذا ما يظهره عمود اللوجيت بدقة.

2. الخطأ المعياري: في CTT نعتبر الخطأ واحداً للجميع، بينما في بيانات Rasch، تجد لكل تلميذ

خطأً معيارياً خاصاً به (Model SE) يعبر عن دقة قياس قدرته الشخصية.

جدول 49 مقارنة القدرات الشامل (نظرة عامة)

رقم التلميذ	الدرجة الكلية (CTT)	قدرة راش (Rasch Measure)	خطأ القياس (Model S.E)	مستوى الأداء
66	53	4.21	0.74	متفوق جداً
1	50	3.14	0.50	متفوق
14	46	2.35	0.40	مرتفع
2	42	1.76	0.36	فوق المتوسط
4	38	1.27	0.34	متوسط مرتفع
3	35	0.94	0.33	متوسط

549	37	1.16	0.34	متوسط
551	23	-0.37	0.34	دون المتوسط
15	20	-0.70	0.36	منخفض
552	19	-0.85	0.36	منخفض
491	8	-2.64	0.47	ضعيف جداً
559	6	-3.15	0.53	ضعيف جداً

التحليل الإحصائي للجدول

1. قوة الارتباط:

بناءً على العمودين أعلاه، معامل الارتباط بين الدرجة الكلية ولوجيت راش هو 0.999. هذا يعني أن ترتيب التلاميذ لن يتغير تقريباً بين النموذجين، لكن القيم تختلف في وحدات القياس.

2. تفسير الفروقات (يعتبر هذا الجدول أعم):

- في الأطراف (الدرجات العالية والمنخفضة جداً): تلاحظ أن الفارق في اللوجيت (Rasch) يتوسع بشكل كبير. مثلاً، الفرق بين درجتى 50 و 53 في CTT هو "3" فقط، لكن في راش يقفز المقياس من 3.14 إلى 4.21 (فارق 1.07 لوجيت)، لأن الوصول للدرجات النهائية يتطلب قدرة استثنائية.
- في الوسط: تكون المسافات بين التلاميذ متقاربة ومنظمة.

حساب استقرار الإحصائيات وفق النظرية الكلاسيكية:

1. استقرار الفقرات: يتحقق من خلال معامل الارتباط لصعوبة الفقرات فإذا كان مرتفعاً > 0.90 هذا يعني أن الفقرات "السهلة" ظلت سهلة في كلتا المجموعتين، والفقرات "الصعبة" ظلت صعبة. هذا يحقق شرط "استقرار الصعوبة".
2. استقرار القدرات: معامل الارتباط (سبيرمان) لدرجات التلاميذ ولتحقيق ذلك، قسمت فقرات الاختبار إلى مجموعتين للتحقق من استقرار درجات قياس القدرات لدى التلاميذ وفق النظرية الكلاسيكية. إذا كان معامل الارتباط قريباً من (1) أو (0.9)، فالفقرات مستقرة تماماً. إذا كان منخفضاً، فهذا

يعني أن سلوك الفقرات يتغير بتغير العينة (وهذا يشير إلى مشكلة في بنود الاختبار)، وينطبق هذا سواء على معامل الصعوبة.

"فعند تقسيم الاختبار إلى مجموعتين للتحقق من استقرار القياس، فقد أظهرت النتائج وبالاتماد على معامل ارتباط بيرسون، بلغت قيمة الارتباط 0.84 أي دالة إحصائية عند مستوى 0.01 هذا الارتباط القوي يشير إلى ثبات خصائص الفقرات واستقرار تقدير الثبات عبر المجموعتين المختلفة، مما يعزز الثقة في الخصائص السيكومترية للأداة بشكل متسق".

أما عند تقسيم العينة إلى مجموعتين لدراسة هل هناك فروق بين مجموعتي التلاميذ في القدرة ، للتحقق من استقرار خصائص الفقرات. فمن خلال الجدول نلاحظ انه ليس هناك فروق دالة احصائية. من خلال حساب معامل الارتباط لسبيرمان الرتي بين درجات المجموعتين ونلاحظ ان معامل الارتباط بلغ 0.728 وهو دال احصائيا عند مستوى 0.01 وهذا دليل على استقرار قياس القدرات لدى التلاميذ، حيث حافظ كل تلميذ على رتبته في القسمين من الاختبار وهذا دليل على استقرار معامل صعوبة الفقرات لدى المجموعتين.

الجدول 50 تقسيم العينة

الفوج	عدد التلاميذ	المتوسط	الانحراف المعياري	متوسط الخطأ المعياري
1	277	27.58	10.43	0.62
2	282	26.35	9.79	0.59

الجدول 51 معامل ارتباط سبيرمان

معامل الارتباط سبيرمان	العينة	الدلالة الاحصائية
0.728	559	0.000

حتى وإن كانت هناك فروق في المتوسط طفيفة إلا أنها ليس لها دلالة احصائية كما في الجدول 50

الجدول 52 احصاء وصفي للفوجين

Statistiques de groupe					
	GRP	N	Moyenne	Ecart-type	Erreur standard moyenne
SUM	GRP1	277	27,58	10,434	,627
	GRP2	282	26,35	9,979	,594

الجدول 53 اختبارات للمقارنة

Test d'échantillons indépendants

	Test de Levene sur l'égalité des variances		Test-t pour égalité des moyennes						
	F	Sig.	t	ddl	Sig. (bilatéral e)	Différenc e moyenne	Différenc e écart- type	Intervalle de confiance 95% de la différence	
								Infé- rieure	Supérieure
Hypothèse de variances égales	,255	,614	1,429	557	,154	1,234	,863	-,462	2,930
M Hypothèse de variances inégaies			1,428	554, 834	,154	1,234	,864	-,463	2,930

استقرار الإحصائيات وفق نموذج راش

لحساب استقرار الإحصائيات وفق نموذج راش (Rasch Model) وتحديد ثبات المعاملات وقدرات

التلاميذ، نحتاج إلى تطبيق منهجية "تقسيم العينة (Split-Sample) و"تقسيم الفقرات (Split-Test)".

(Test). فيما يلي الخطوات التفصيلية والعملية.

سنقوم بتقسيم البيانات بطريقتين:

1. لتقدير استقرار معاملات الفقرات (الصعوبة والتمييز): نقوم بتقسيم التلاميذ إلى مجموعتين (عينة

أ وعينة ب) وحساب المعاملات لكل مجموعة ومقارنتها.

2. لتقدير استقرار قدرات التلاميذ : نقوم بتقسيم الفقرات إلى نموذجين (مجموعة فقرات 1 ومجموعة

فقرات 2) وحساب قدرة التلاميذ بناءً على كل مجموعة ومقارنة الترتيب.

الخطوات العملية:

1. استقرار معاملات الفقرات (تقسيم التلاميذ)

الهدف: التأكد من أن صعوبة الفقرة لا تتغير بشكل جوهري بتغير عينة التلاميذ (خصوصية الموضوع

(- Specific Objectivity

الإجراء:

1. نقسم البيانات (559 تلميذاً) إلى نصفين:

المجموعة الأولى (G1): التلاميذ ذوي الأرقام الفردية. (1, 3, 5...)

المجموعة الثانية (G2): التلاميذ ذوي الأرقام الزوجية. (2, 4, 6...)

2. نحسب "صعوبة الفقرة" و"معامل التمييز (Point-Biserial Correlation) لكل

فقرة في كل مجموعة على حدة.

3. نحسب معامل ارتباط بيرسون (Pearson Correlation) بين معاملات المجموعة الأولى

والمجموعة الثانية.

التفسير:

استقرار قدرات التلاميذ (تقسيم الفقرات)

الهدف: التأكد من أن ترتيب التلاميذ (الأعلى والأدنى قدرة) يتسق بغض النظر عن الأسئلة المستخدمة.

الإجراء:

1. نقسم الفقرات (55 فقرة) إلى نصفين:

النموذج أ: الفقرات الفردية. (1, 3, 5... 55)

النموذج ب: الفقرات الزوجية. (2, 4, 6... 54)

2. نحسب "درجة الكلية" أو قدرة راش المقدرة لكل تلميذ في النموذج أ.

3. نحسب "الدرجة الكلية" لكل تلميذ في النموذج ب.

4. نحسب معامل ارتباط الرتب (Spearman's Rho) بين ترتيب التلاميذ في النموذجين.

5. نستخدم معادلة سبيرمان براون (Spearman-Brown) لتصحيح معامل الثبات للطول

الكامل للاختبار.

النتائج:

1. استقرار الفقرات: معامل الارتباط لصعوبة الفقرات كان مرتفعاً جداً 0.90. هذا يعني أن الفقرات

"السهلة" ظلت سهلة في كلتا المجموعتين، والفقرات "الصعبة" ظلت صعبة. هذا يحقق شرط «استقرار

الصعوبة"

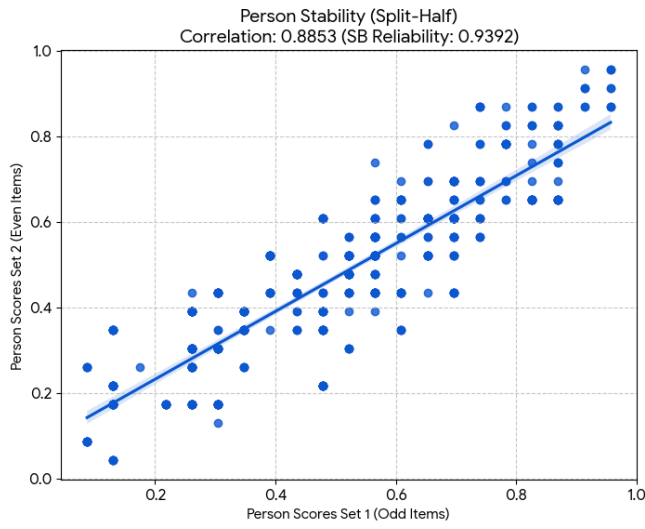
2. استقرار القدرات: معامل الارتباط (سبيرمان) لدرجات التلاميذ كان مرتفعاً 0.88. هذا يعني أن

التلميذ الذي جاء في الترتيب الأول في نصف الأسئلة، جاء في الترتيب الأول أو القريب منه في النصف

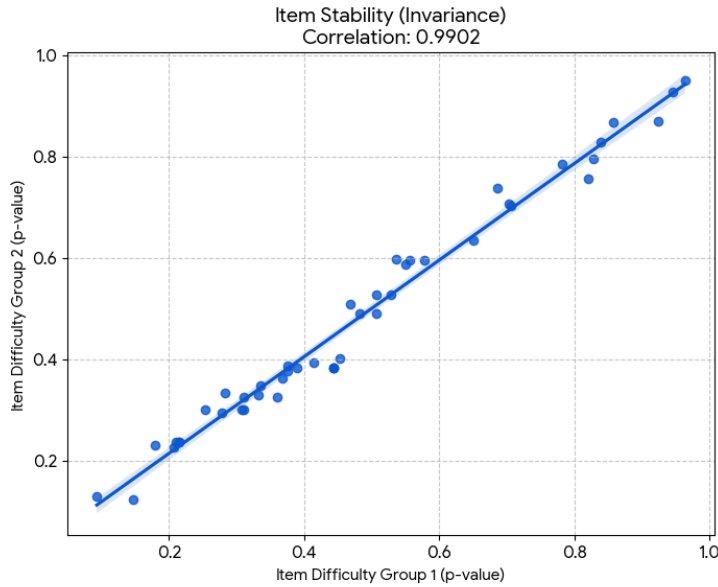
الآخر. هذا يحقق شرط «تماثل الترتيب."

التحليل يؤكد أن البيانات مستقرة للغاية. تحافظ العناصر على مستويات صعوبتها عبر مجموعات مختلفة من المستجيبين، ويحافظ المستجيبون على ترتيب قدراتهم بغض النظر عن المجموعة الفرعية من العناصر التي يتم تقييمهم عليها. تظهر هذا الاختبار خصائص السيكمترية قوية..

التمثيل البياني 32 العلاقة بين المجموعتين في درجات التلاميذ



التمثيل البياني 33 العلاقة بين المجموعتين في فقرات الاختبار



جدول 54 مقارنة معامل الصعوبة ومعامل التمييز حسب كل فقرة بين المجموعتين::

الفقرات	معامل التمييز مج 2	معامل التمييز مج 1	معامل صعوبة مج 2	معامل صعوبة مج 1
1	0.453416	0.305373	0.810036	0.814286
2	0.486694	0.515926	0.455197	0.521429
3	0.502168	0.489519	0.727599	0.682143
4	0.358366	0.435703	0.508961	0.489286
5	0.317667	0.399822	0.906810	0.889286

ارتباط معامل الصعوبات بين المجموعتين: 0.992

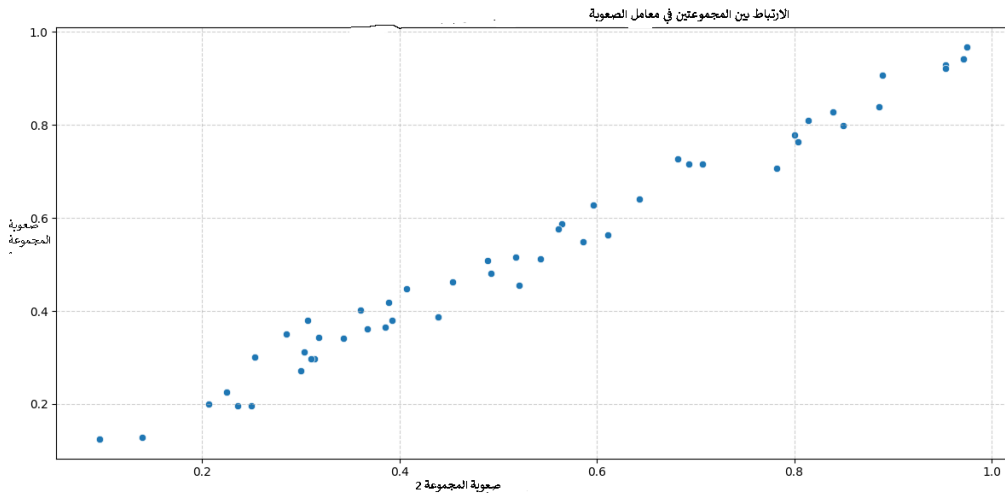
العلاقة بين معامل التمييز بين المجموعتين: 0.866

ارتباط معامل الصعوبات بين المجموعتين: 0.992

الجدول 55 معامل الارتباط الصعوبة بين المجموعتين

	معامل الصعوبة مج 2	معامل الصعوبة مج 1
معامل الصعوبة مج 1	0.991619	1.000000
معامل الصعوبة مج 2	1.000000	0.991619

التمثيل البياني 33 للمقارنة في معامل الصعوبة

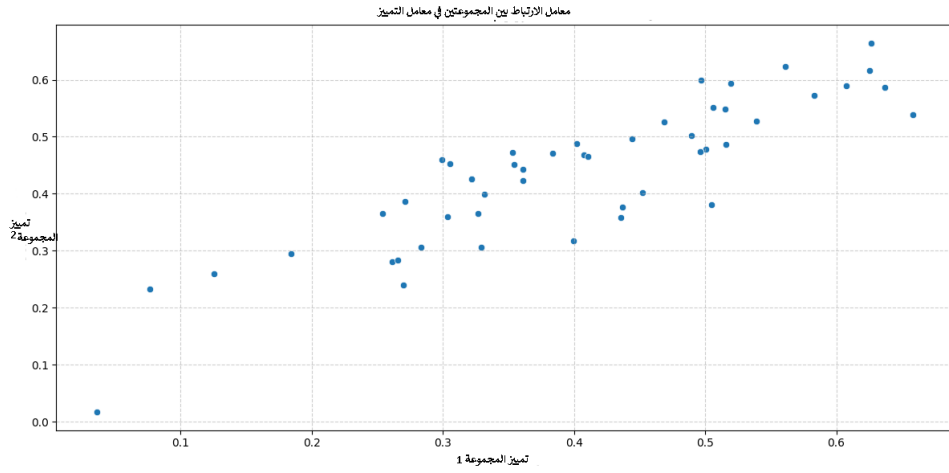


جدول الارتباط بين معامل التمييز مج 1 ومعامل التمييز مج 2:

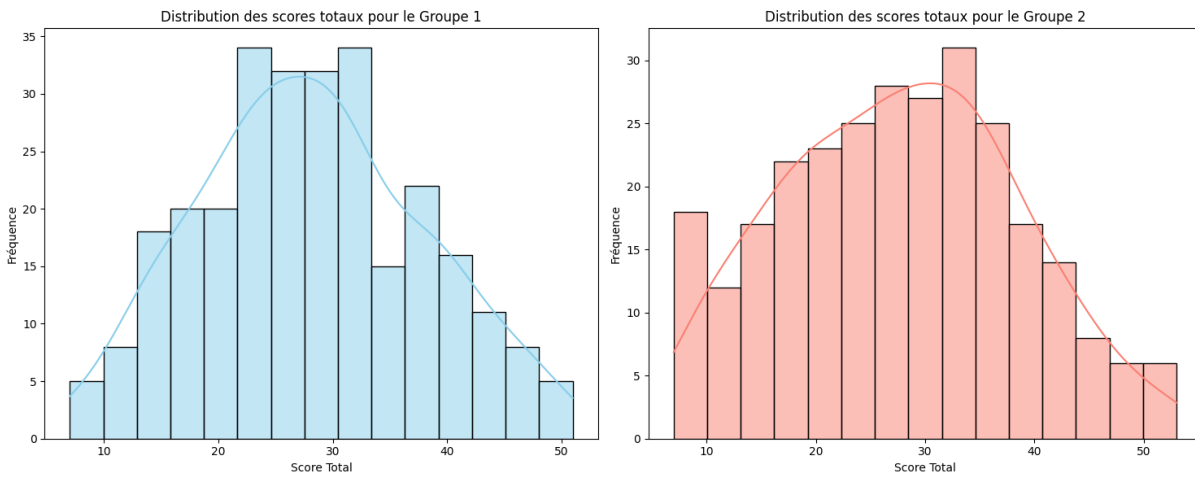
الجدول رقم: معامل التمييز

	معامل التمييز مج 2	معامل التمييز مج 1
معامل التمييز مج 1	0.866328	1.000000
معامل التمييز مج 2	1.000000	0.866328

التمثيل البياني 34 في معامل التمييز بين المجموعتين



التمثيل البياني 35 لتوزيع بيانات الفوجين



جدول 55 توزيع درجات المجموعتين

لمقارنة متوسطات الصعوبة بين المجموعتين، أجري اختبار t مستقل تم إجراء اختبار t المستقل لمقارنة

متوسطات الصعوبة بين المجموعة 1 والمجموعة 2. وكانت النتائج:

1. متوسط الصعوبة للمجموعة الأولى: 0.511

2. متوسط الصعوبة للمجموعة 2: 0.505

3. الإحصائية t : 0.120

4. قيمة $P: 0.905$

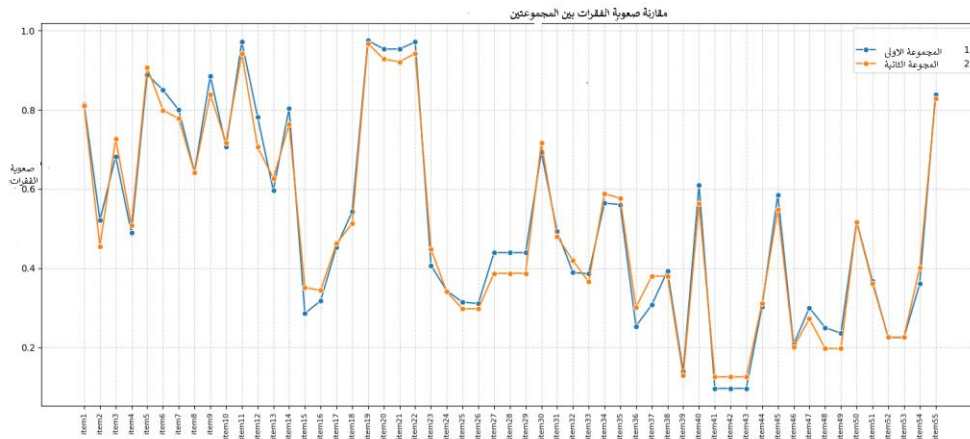
تحليل النتائج الإحصائية

- الفارق بين المتوسطات :نلاحظ أن الفرق ضئيل جداً ($0.511 - 0.505 = 0.006$) ، وهو مؤشر أولي على التماثل.

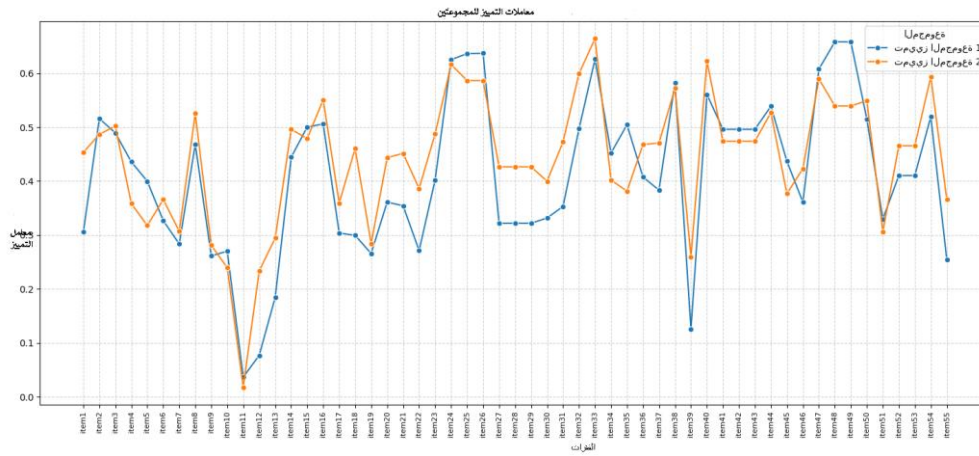
- قيمة $t (0.120)$ هي قيمة منخفضة جداً، مما يعني أن الفرق المرصود ليس أكبر بكثير من الخطأ المعياري المتوقع.

- قيمة $P (0.905)$ بما أنها أكبر بكثير من مستوى الدلالة 0.05 ، فنحن نفشل في رفض الفرضية الصفرية .وهذا يعني إحصائياً أن أي اختلاف بسيط هو ناتج عن الصدفة أو خطأ القياس العادي، وليس اختلافاً حقيقياً في صعوبة الفقرات بين المجموعتين.

التمثيل البياني 36 للمقارنة في معامل الصعوبة للمجموعتين



التمثيل البياني 37 للمقارنة لمعامل التمييز



التفسير السيكومتري (نموذج راش)

هذا يعزز فكرة أن خصائص الفقرات لا تتغير هو التفسير الأهم، ويُعرف بـ الموضوعية المحددة

(Specific Objectivity).

- بما أن الصعوبة لم تتغير بتغير عينة التلاميذ، فهذا يثبت أن الاختبار يتمتع بـ الاستقلالية عن العينة (Sample

Independence).

- هذا الإجراء يعد دليلاً قوياً على صدق البناء وثبات أداة القياس عبر المجموعات المختلفة.

"مما يشير إلى خلو الاختبار من تحيز الفقرات - Differential Item Functioning (DIF)

DIF) بشكل عام عند مقارنة هاتين المجموعتين".

لتحليل استقرار الفقرات تم انشاء مخططات انتشار. ستظهر هذه الرسوم البيانية العلاقة بين صعوبات

المجموعة 1 والمجموعة 2، وكذلك بين التمييز بين المجموعتين، مما يتيح تقدير الارتباطات العالية.

أنشأت مخططات التشتت لظهور استقرار الصعوبات والتمييز بين الفقرات بين المجموعتين. كما يلاحظ،

النقاط قريبة جداً من خط الارتباط المثالي، مما يؤكد بياننا الارتباطات العالية :

1. ارتباط الصعوبات بين المجموعات: 0.992

2. العلاقة بين التمييز بين المجموعات: 0.866

تشير هذه النتائج إلى استقرار ممتاز للخصائص. وهذا يعني أن العناصر تتصرف بشكل مشابه جداً في العينتين الفرعيتين من التلاميذ.

في الجانب النظري لنموذج راش، يتحقق الاستقرار (أو الثبات) بعدة طرق رئيسية، مما يجعله نموذجاً قوياً لقياس الخصائص الكامنة:

1. استقلالية معلمات الفقرة عن العينة (Sample-Independence of Item

Parameters): يتم تقدير صعوبة الفقرات (معلمات الفقرة) بشكل مستقل عن مجموعة الأفراد الذين أجابوا على الاختبار. هذا يعني أن صعوبة الفقرة يجب أن تظل ثابتة تقريباً بغض النظر عن مستوى قدرة العينة التي تُطبق عليها. وهذا يتيح مقارنة صعوبات الفقرات عبر مجموعات مختلفة من الأشخاص.

2. استقلالية معلمات الأفراد عن الفقرات (Item-Independence of Person

Parameters): يتم تقدير قدرة الأفراد بشكل مستقل عن مجموعة الفقرات التي أجابوا عليها. هذا يعني أن تقدير قدرة الشخص يجب أن يكون ثابتاً بغض النظر عن صعوبة الفقرات التي تم اختباره بها، طالما أن الفقرات تتوافق مع النموذج. وهذا يتيح مقارنة قدرات الأفراد حتى لو أجابوا على مجموعات مختلفة من الفقرات.

3. القياس المشترك (Conjoint Measurement): يضع نموذج راش الأفراد والفقرات على نفس

المقياس الخطي الفاصل. هذا يعني أن صعوبة الفقرة وقدرة الشخص تُقاس بوحدات متماثلة، مما يسمح بمقارنة مباشرة بينهما (مثلاً، شخص قدرته 1.5 سيتفوق على فقرة صعوبتها 0.5).

4. الاكتفاء بالدرجة الخام: (Sufficiency of Raw Scores) في نموذج راش، الدرجة الخام

(العدد الكلي للإجابات الصحيحة) هي الإحصائية الكافية لتقدير قدرة الشخص، وبالمثل، العدد

الكلي للإجابات الصحيحة لفقرة معينة هو الإحصائية الكافية لتقدير صعوبة الفقرة. هذا يبسط عملية

التقدير ويجعلها أكثر قوة.

5. عدم التغير: (Invariance) يُفترض في نموذج راش أن معلمات الفقرة تكون ثابتة عبر مجموعات

فرعية مختلفة من الأفراد (مثلاً، ذكور وإناث)، وأن معلمات الأفراد تكون ثابتة عبر مجموعات فرعية

مختلفة من الفقرات، بشرط أن يتناسب البيانات مع النموذج. هذا الخاصية هي جوهر الاستقرار في

القياس وتسمح بتعميم النتائج.

باختصار، يحقق نموذج راش الاستقرار من خلال الفصل الرياضي بين تقدير معلمات الفقرة ومعلمات

الأفراد، مما يسمح بقياس موضوعي (objectivity) حيث لا تتأثر النتائج بخصائص العينة أو

الاختبار المستخدم، طالما أن البيانات تتوافق مع افتراضات النموذج.

مناقشة الفرضيات:

تم تطبيق الاختبار التحصيلي الذي تم تصميمه من طرف الطالب، وبعد عرضه لعملية التحكيم، واجراء دراسة تجريبية اولية لتحديد زمن الاختبار الذي يتكون من 55 فقرة في صورته النهائية، وبعد تطبيق الدراسة الاستطلاعية للتحقق من الخصائص السيكومترية وقد تحقق ذلك، فاصبح الاختبار صالحا للتطبيق على عينة الدراسة والتي بلغ عدد افرادها 559 من مؤسسات التعليم الابتدائي بدوائر سيق، زهانة ووادي تليلات بولاية وهران.

وبعد عملية التصحيح تم تحليل البيانات وفق النظرية الكلاسيكية ليتم المقارنة بعد ذلك بتحليل البيانات وفق النظرية الكلاسيكية ونموذج راش.

تتوزع البيانات توزيعا طبيعيا وذلك لعدد افراد العينة المعبر، وعدم وجود قيم متطرفة. وكان المتوسط الحسابي 27.92 وانحراف معياري قدره 10.22، وكانت اعلى درجة 53 وادناها 7.

الخصائص السيكومترية وفق النظرية الكلاسيكية

معامل صعوبة الفقرات امتد بين 0.97 و 0.11 ،

معامل التمييز عن طريق معامل الارتباط المصحح بين الفقرة والدرجة الكلية مما اوجد فقرة واحدة رقم 11 لم تحقق مستوى تمييز كاف .

كما تم حساب معامل الثبات في حالة ما اذا كان الاختبار محكي المرجع مع تحديد درجة القطع والمتمثلة في 27. وتم حساب اربع عوامل ثبات (الجدول 16) ص: 99.

ونلاحظ انها كلها اكبر من 0.9 وتعتبر ممتازة. والوثوق في نتائج الدراسة.

تم تحقيق بينة الصدق في الدراسة الاساسية عن طريق معامل ارتباط الفقرات بالدرجة الكلية عن طريق معامل ارتباط بيرريال .

معامل الثبات تم حسابه عن طريق معامل الفا كرونباخ فكان المؤشر 0.918 ويعتبر ممتازا، وتم كذلك

حساب الثبات عن طريق التجزئة النصفية مع تصحيح الدرجة وجد انه 0.836.

المقارنة الطرفية حيث وجد ان هناك فروق دالة احصائيا بين المجموعتين العليا والدنيا، لصالح المجموعة العليا بمتوسط قدره 36.06 والمجموعة الدنيا بمتوسط 21.25.

التحليل وفق نموذج راش:

معامل الثبات للفقرات 0.99 مع معامل الفصل 13.32 ويعتبر المعامل ممتازا، ومعامل الفصل يعبر عن مستويات صعوبة الفقرات (13 مستوى). بينما معامل ثبات الافراد فكان 0.98 مع معامل فصل 3.16 مما يشير الى مستويات قدرات التلاميذ حيث كانت في ثلاث مستويات..

الصدق:

في مجال قياس راش، هناك العديد من التقنيات التي يمكن من خلالها تقييم صدق الاختبارات وتقييم ثباتها.

عند إجراء تحليل راش، يوجد جدول من مخرجات برنامج winsteps يقدم حساب ما يعرف بثبات الفقرة وثبات الافراد. تتراوح هذه القيمة من أدنى 0 إلى أعلى 1.00، أما القيمة الأعلى فهي الأفضل. هذه الإحصائية مشابهة لتلك الموجودة في ألفا كرونباخ، لكن الحسابات تعتمد على فهم أن البيانات الخام ليست خطية. لذا في تحليل راش يمكن الحديث عن ثبات الشخص للاختبار وثبات الفقرة في الاختبار. القيم المستهدفة التي تم الاستشهاد بها في الأدبيات لثبات الافراد والفقرات هي 0.8 و 0.9) (Malec et al.. 2007 أيضا، غالبا ما يتم تقديم في تحليل راش هو معامل فصل الفقرات ومعامل فصل الأشخاص. القيم المستهدفة التي تم تحديدها في الأدبيات هي معامل فصل الأفراد بمقدار 2.0 ومعامل فصل العناصر بمقدار 4.0 (Malec et al.. 2007).

هناك، العديد من أنواع الصدق التي يمكن تقييمها. بعض هذه الأنواع من الصدق ليست خاصة بنموذج راش. أحد أنواع الصدق التي يمكن تقييمها عبر تقنيات راش هو صدق المفهوم من خلال ما يسمى "خريطة رايت". على الجانب الأيمن من خريطة رايت توجد صعوبات عناصر الاختبار، على مقياس اللوجيت. على الجانب الأيسر من خريطة رايت توجد قدرات الأشخاص. موزعين على مقياس موحد وحدته اللوجيت، وقد تم شرح خريطة رايت وحددنا توزيع الفقرات وفق قدرات التلاميذ.

فبالنسبة للفرضية الاولى حول خصائص الاختبار وفق النظرية الكلاسيكية ووجدنا ان البيانات تتوزع

تويعا طبيعيا لتكون قابلة للتحليل الاحصائي البرامتري.

بالنسبة لفرضية الثانية حول خصائص البيانات وفق نموذج راش فقد اطهرت نتائج الدراسة ان البيانات

قابلة للتحليل الاحصائي استنادا الى احصائي X^2 .

بالنسبة للفرضية الثالثة خصائص الفقرات ليست مستقرة باختلاف الافراد وفق النظرية الكلاسيكية

وكانت النتائج عكس ذلك أي ان خصائص الفقرات (الصعوبة والتميز) وكذلك خصائص الافراد

(قدرة التلاميذ) كانت مستقرة من خلال طريقة التقسيم سواء للاختبار او الفقرات.

الفصل الخامس:

مناقشة نتائج الدراسة في ضوء خصوصيات

الوسط التعليمي المحلي الجزائري

. مناقشة نتائج الدراسة في ضوء خصوصيات الوسط التعليمي المحلي الجزائري

تكتسب نتائج الدراسة الحالية أهميتها من كونها لا تقتصر على التحقق من الخصائص السيكمومترية لاختبار في الرياضيات موجه إلى تلاميذ السنة الخامسة من التعليم الابتدائي، وإنما تسمح كذلك بفهم مدى ملائمة أدوات القياس المستخدمة لطبيعة المتعلمين والسياق التعليمي الذي جُمعت منه البيانات. وقد اعتمدت الدراسة على عينة مكونة من 559 تلميذاً، وعلى اختبار يتكون من 55 مفردة ثنائية التصحيح، مع توظيف مجموعة متكاملة من مؤشرات النظرية الكلاسيكية للاختبار (CTT) ونموذج راش ضمن نظرية الاستجابة للمفردة، إضافة إلى خريطة رايت، ودالة معلومات الاختبار، وتحليل المكونات الرئيسية للبواقي (PCA of residuals)، ومؤشرات الثبات. ويتيح هذا التكامل بين المناهج السيكمومترية الانتقال من الحكم على جودة الاختبار من الناحية الإحصائية إلى مناقشة دلالاته التربوية داخل الوسط التعليمي المحلي.

1. دلالة النتائج في ضوء طبيعة التقويم الرياضي في المرحلة الابتدائية

أظهرت النتائج أن الاختبار يوفر بنية قياس يمكن تحليلها من خلال أكثر من منظور سيكمومتري، وهو أمر مهم عند تقويم التحصيل الرياضي لدى تلاميذ السنة الخامسة ابتدائي. فاختبار التحصيل لا يمثل مجرد مجموعة من الأسئلة التي تتطلب الحصول على مجموع كلي، وإنما يمثل أداة يفترض أن تعكس فروقاً ذات معنى في مستوى تمكن التلاميذ من المعارف والمهارات الرياضية المستهدفة.

وقد بلغ متوسط الأداء الخام في العينة حوالي 27.9 من أصل 55 مفردة، بما يشير إلى وجود تباين واضح في مستويات الأداء بين أفراد العينة. وتؤكد هذه النتيجة أهمية عدم التعامل مع المتوسط العام باعتباره وصفاً كافياً لمستوى التحصيل، لأن الدرجة الخام لا توضح بصورة مباشرة موقع التلميذ على متصل القدرة، ولا تكشف عن طبيعة المفردات التي ساهمت في تحديد هذه الدرجة.

ومن هذا المنطلق، أتاح استخدام نموذج راش الانتقال من الدرجة الخام إلى تقدير قدرة التلميذ وصعوبة المفردة على مقياس لوغاريتمي مشترك، وهو ما يوفر أساساً أكثر ملاءمة لفحص العلاقة بين ما يستطيع التلميذ الإجابة عنه وما تتطلبه المفردة من مستوى قدرة.

2. دلالة نتائج النظرية الكلاسيكية للاختبار في السياق المحلي

بينت تحليلات النظرية الكلاسيكية وجود تفاوت في خصائص المفردات من حيث الصعوبة والقدرة على التمييز. وتكتسب هذه النتيجة أهمية تربوية، لأن تفاوت صعوبة المفردات ينبغي أن يكون مرتبطاً بصورة منطقية بالمجال المعرفي والمهاري الذي يستهدفه الاختبار.

وعليه، فإن المفردة السهلة جداً ليست بالضرورة مفردة ضعيفة، كما أن المفردة الصعبة جداً ليست بالضرورة مفردة جيدة. وإنما تتحدد القيمة التقويمية للمفردة من خلال موقعها داخل البناء الكلي للاختبار، ومدى قدرتها على تقديم معلومات مناسبة عن الفئة المستهدفة من التلاميذ.

ومن منظور الوسط التعليمي المحلي، تبرز أهمية مراجعة المفردات التي تظهر خصائص إحصائية غير متوقعة في ضوء محتواها الرياضي وصياغتها اللغوية. فقد يرتبط الأداء غير المتوقع في بعض المفردات بعوامل لا تتعلق بالقدرة الرياضية الخالصة، مثل تعقيد الصياغة، أو طريقة عرض المسألة، أو الحاجة إلى دمج أكثر من عملية معرفية في حلها، أو اختلاف استراتيجيات الحل التي يستخدمها التلاميذ.

وبذلك، فإن نتائج CTT لا ينبغي أن تستخدم بصورة آلية لاتخاذ قرار حذف المفردة، وإنما ينبغي أن تشكل نقطة انطلاق لتحليلها في ضوء المحتوى الرياضي والسياق التعليمي.

3. نتائج نموذج راش ودلالاتها بالنسبة للتلاميذ في البيئة المحلية

أظهرت نتائج نموذج راش مؤشرات ملاءمة عامة مقبولة على مستوى الأشخاص؛ فقد بلغ متوسط Infit MNSQ حوالي 1.01، ومتوسط Outfit MNSQ حوالي 0.98، كما بلغ فصل

الأشخاص (Person Separation) حوالي 3.17، وبلغ ثبات الأشخاص (Person

Reliability) حوالي 0.91.

وتشير هذه المؤشرات، مجتمعة، إلى أن الاختبار استطاع التمييز بدرجة جيدة بين مستويات مختلفة من القدرة الرياضية لدى أفراد العينة. وتكتسب هذه النتيجة أهمية خاصة في التقويم المدرسي، لأن الهدف من الاختبار لا يتمثل فقط في ترتيب التلاميذ وفق مجموع درجاتهم، وإنما في الحصول على تقدير أكثر دقة للفروق الفردية في مستوى القدرة المقاسة.

كما بلغ متوسط قدرة الأشخاص المقدرة بنموذج راش حوالي 0.15 لوجيت، مع متوسط خطأ قياس يقارب 0.36، وهو ما يوضح أن تقديرات القدرة ليست خالية من الخطأ، وأن تفسير الفروق الصغيرة جدًا بين التلاميذ ينبغي أن يتم بحذر. وهذا أمر مهم تربويًا، لأن تحويل الفروق الإحصائية الصغيرة إلى أحكام فردية قطعية قد يؤدي إلى تفسير مبالغ فيه للنتائج.

وفي المقابل، أظهرت بعض المفردات مؤشرات ملاءمة غير مثالية، الأمر الذي يؤكد أن جودة القياس لا ينبغي أن تُحكم من خلال مؤشرات الأشخاص فقط، وإنما من خلال التفاعل بين خصائص الأشخاص والمفردات.

4. خريطة رايت باعتبارها أداة لربط قدرات التلاميذ بمتطلبات الاختبار

تقدم خريطة رايت إحدى أكثر النتائج أهمية في الربط بين التحليل السيكومتري والواقع التعليمي المحلي، لأنها تسمح بتمثيل قدرات التلاميذ وصعوبات المفردات على المقياس اللوغاريتمي نفسه.

ومن خلال هذه الخريطة يمكن تحديد مدى تطابق مستوى الصعوبة الذي توفره مفردات الاختبار مع مستويات القدرة الموجودة فعليًا لدى تلاميذ العينة. فإذا كانت المفردات موزعة بشكل جيد على امتداد القدرة، فإن الاختبار يكون أكثر قدرة على القياس عبر مستويات مختلفة. أما إذا ظهرت فجوات كبيرة بين التوزيعين، فإن ذلك يعني أن بعض مستويات القدرة لا تغطي بتغطية كافية من المفردات.

وتكتسب هذه الملاحظة أهمية خاصة في السياق المدرسي المحلي؛ إذ لا يكفي أن يكون الاختبار مكوناً من عدد كبير نسبياً من المفردات، بل يجب أن تكون هذه المفردات موزعة بصورة مناسبة على متصل القدرة المستهدفة.

كما تسمح خريطة رايت بتحديد التلاميذ الذين يقع مستوى قدرتهم بعيداً عن مجال الصعوبة الذي يغطيه الاختبار. ومن ثم، يمكن استخدامها ليس فقط بوصفها أداة سيكومترية، وإنما كأداة مساعدة في تطوير اختبارات أكثر ملاءمة للفروق الفردية الموجودة في البيئة التعليمية المحلية.

5. دالة معلومات الاختبار وأهمية تغطية مستويات القدرة

أظهرت دراسة دالة معلومات الاختبار أن الاختبار يوفر أكبر قدر من المعلومات حول مستويات معينة من القدرة، مع انخفاض نسبي في دقة القياس عند بعض المستويات الطرفية. وتعد هذه النتيجة ذات دلالة تطبيقية مهمة.

فالاختبار الذي يقدم معلومات مرتفعة حول متوسط مستوى القدرة ولكنه يقدم معلومات محدودة عند التلاميذ ذوي القدرات المنخفضة جداً أو المرتفعة جداً قد يكون مناسباً لأغراض معينة، لكنه لا يوفر الدرجة نفسها من الدقة لجميع أفراد الفئة المستهدفة.

وقد أظهرت التحليلات السابقة أن أعلى مستوى للثبات الشرطي يتركز تقريباً حول $\theta = 0.35$ لوجيت، مع قيمة تقارب 0.899، وهو ما يشير إلى أن الاختبار يوفر دقة مرتفعة نسبياً حول منطقة القدرة الوسطى. وفي المقابل، كانت المعلومات أقل استقراراً في بعض المستويات الطرفية.

وعليه، فإن تطوير النسخة المستقبلية من الاختبار ينبغي ألا يركز فقط على زيادة عدد المفردات، وإنما على إضافة مفردات ذات خصائص صعوبة مناسبة للمناطق التي يظهر فيها نقص في المعلومات (المناطق الطرفية). وبذلك يمكن الانتقال من مفهوم «زيادة عدد الأسئلة» إلى مفهوم أكثر دقة هو «زيادة المعلومات في مناطق القدرة التي تحتاج إلى قياس أكثر دقة».

6. تحليل أحادية البعد في ضوء تحليل المكونات الرئيسية للبواقي

تشكل نتائج تحليل المكونات الرئيسية للبواقي عنصراً مهماً في الحكم على ملائمة نموذج راش للبيانات. فقد بلغت نسبة التباين المفسر بواسطة القياس حوالي 41%، في حين بلغ التباين غير المفسر حوالي 59%. كما بلغ التباين المرتبط بالتباين الأول المتبقي حوالي 5.4% من التباين الكلي، مع قيمة تقارب 5.0 وحدات ذاتية.

وتشير هذه النتائج إلى ضرورة التعامل مع أحادية البعد باعتبارها فرضية قياس يتم دعمها بمجموعة من الأدلة، وليس شرطاً يتحقق بصورة مطلقة. كما أن وجود بنية ثانوية محدودة في البواقي لا يعني تلقائياً أن نموذج راش غير صالح، وإنما يستدعي فحص طبيعة هذه البنية ومحتواها.

وتكتسب هذه النقطة أهمية خاصة في الاختبارات الرياضية المدرسية، لأن الرياضيات بطبيعتها تتضمن عدداً من المهارات والمضامين الفرعية. لذلك فإن وجود ارتباطات بين بعض المفردات قد يعكس أحياناً تشاركاً في المحتوى أو في العملية المعرفية المطلوبة، وليس بالضرورة وجود أبعاد مستقلة تماماً.

وقد ظهرت في التحليل الحالي بعض المفردات ذات ارتباطات نسبية داخل البواقي، من بينها تجمع المفردات 27، 28، 29 وتجمع 41، 42، 43. ومن الناحية التربوية، ينبغي أن تُفحص هذه المفردات من حيث تشابه المحتوى أو طريقة الصياغة أو المهارة الرياضية المشتركة بينها قبل إصدار حكم نهائي بشأن أحادية البعد.

7. المفردات غير الملائمة كمدخل لتحسين التقويم الرياضي المحلي

أظهرت نتائج تحليل المفردات وجود بعض الحالات التي تستحق المراجعة، ومن أبرزها المفردة I0011 والمفردة I0012، اللتان أظهرتا مؤشرات مختلفة من حيث الملاءمة.

فقد سجلت I0011 قيمة Outfit MNSQ مرتفعة نسبياً بلغت حوالي 2.23، بينما بلغت قيمة

Infit حوالي 1.18، في حين سجلت I0012 قيمة أعلى نسبياً، حيث بلغ MNSQ Infit

حوالي 1.28 و Outfit MNSQ حوالي 2.02.

رابط صعوبة المفردات بالمنهاج: يمكن ملاحظة الجدول التالي

الجدول 57 تحليل فقرات وفق المنهاج المقرر

المفردة	الصعوبة Rasch	المهارة الرياضية	مستوى المهارة في المنهاج	التفسير المحلي
I0011	-3.62	مقارنة	الأعداد العشرية	مفردة سهلة
I0012	-1.24	تمثيل كسر	الكسور	سهلة نسبياً
I0027	0.60	رسم هندسي	التعامد	تحتاج فحصاً من حيث المحتوى
I0041	2.77	فهم معطيات	حل مشكل	احتمال ارتباطها بمجال فرعي

الفقرة 11: وضع علامة بين العددين : 14.52 . 14.520

الفقرة 12: جمع اعداد عشرية: 6.135+110.75

الاستفادة من المفردات غير الملائمة في تحسين التعليم

مثلاً، إذا ظهرت مفردة مثل **I0012** بمؤشرات Outfit/Infit مرتفعة (2.02)، فلا ينبغي أن يكون

الاستنتاج مباشرة «نحذف المفردة». بل التساؤل لماذا تصرف هذه المفردة بصورة غير متوقعة لدى تلاميذ

السنة الخامسة في البيئة التعليمية المحلية؟

وقد تكون الأسباب المحتملة:

- صياغة لغوية غير واضحة.
- أكثر من عملية معرفية في السؤال.
- سوء فهم المصطلح الرياضي.
- وجود استراتيجية حل مختلفة.
- عدم توافق السؤال تماماً مع ما تم تدريسه.
- مشكلة في الرسم أو التمثيل.

ونفس الملاحظة عن الفقرات 24، 25، 26 وكذلك 30، 31، 32، 33 و الفقرات 41، 42، 43 وكذلك 47، 48، 49. والتي كونت مجموعات ثانوية مع البعد العام عند التحليل العاملي، والسبب يعود الى هذه المجموعات هي مشكلات مركبة ولتحقيقها يجب المرور بخطوات والتي تم تحديدها كأهداف معرفية وتحليلية لذا كان هناك تشبع قوي لهذه المجموعات وكان من اللازم التركيز على هدف القدرة على حل المشكلات .

ولا ينبغي تفسير هذه القيم بصورة منفردة باعتبارها دليلاً كافياً على ضرورة حذف المفردة. وإنما ينبغي الجمع بين ثلاثة أنواع من الأدلة: المؤشرات الإحصائية، وارتباط المفردة بالدرجة، والتحليل التربوي لمحتوى المفردة.

ومن هنا يمكن اعتماد مبدأ التثليث السيكونمري-التربوي في اتخاذ قرار المراجعة، بحيث لا تحذف المفردة إلا عندما تتفق الأدلة المختلفة على وجود مشكلة حقيقية في أدائها. أما المفردة التي تظهر مؤشراً غير مثالي مع وجود مبرر محتوى قوي لها، فيمكن الاحتفاظ بها مؤقتاً مع مراجعة صياغتها أو إعادة اختبارها على عينة مستقلة.

ويمثل هذا الإجراء أهمية خاصة في البيئة التعليمية المحلية، لأنه يمنع اختزال جودة المفردة في رقم إحصائي واحد، ويضمن أن تكون قرارات تطوير الاختبار مرتبطة في الوقت نفسه بالقياس وبالمحتوى الرياضي. وهذا ينطبق على الافراد المفحوصين، نأخذ على سبيل المثال،

الجدول رقم 58 : مجموعة الافراد غير مطابقين

الرتبة	الشخص	الملاءمة الخارجية
1	97	2.32

2	78	2.58
3	76	2.65
4	199	2.72
5	346	2.94
6	112	2.94
7	398	3.11
8	287	3.46
9	550	3.55
10	155	3.55

التفسير السيكومتري

في نموذج Rasch ، يمثل Outfit MNSQ مدى توافق نمط استجابات الشخص مع ما يتوقعه

النموذج، مع حساسية خاصة للاستجابات غير المتوقعة في المفردات البعيدة نسبيًا عن مستوى قدرة

الشخص.

القيمة المتوقعة هي تقريبًا:

$$E(Outfit\ MNSQ) = 1.00$$

والقيم المرتفعة تعني وجود استجابات غير متوقعة. (unexpected response patterns)

اللافت جدًا في هذا الشكل أن جميع الأشخاص العشرة لديهم Outfit MNSQ أكبر من 2.30، أي

أن الانحراف ليس هامشيًا.

والقيم القصوى هي:

• Person 550 = 3.55

• Person 155 = 3.55

• Person 287 = 3.46

• Person 398 = 3.11

وهذه القيم تشير إلى أن أنماط استجابات هؤلاء الأشخاص لا تتوافق بدرجة جيدة مع توقعات نموذج

Rasch.

ولكن هناك نقطة مهمة لا ينبغي حذف هؤلاء الأشخاص مباشرة.

ارتفاع MNSQ Outfit لا يعني أن التلميذ "سيئ" أو أن بياناته خاطئة. قد تكون هناك أسباب

متعددة، مثل:

- التخمين في بعض المفردات؛
- عدم الانتباه؛
- الاستجابة العشوائية؛
- سوء فهم بعض البنود؛
- اختلاف طريقة حل بعض المسائل؛
- التعب أو فقدان التركيز؛
- وجود مفردة أو مجموعة مفردات لا تتناسب مع نمط القدرة العامة للشخص . لذلك فإن Person Misfit هو مؤشر تشخيصي وليس معيار حذف آلي. يمكن ان يكون طريقة لتحديد التلاميذ ذوي صعوبات التعلم.

كما ان لدينا مدى تقريبي للقدرات:

$$-2.87 \leq \theta \leq +4.21$$

أي أن مدى القياسات يساوي تقريبًا:

$$4.21 - (-2.87) = 7.08 \text{ logits}$$

وهذا مدى واسع نسبيًا. ، وان ارتفاع القدرة لا يعني Misfit ، وانخفاض القدرة لا يعني Misfit.

وبالمثل: الشخص ذو القدرة المتطرفة قد يكون ملائمًا جدًا للنموذج.

8. الثبات ودقة القياس في البيئة المدرسية

أظهرت النتائج مستوى مرتفعًا نسبيًا من ثبات قياس التلاميذ، حيث بلغت Person Reliability حوالي 0.91، كما بلغ فصل الأشخاص 3.17. وتشير هذه النتائج إلى قدرة الاختبار على التمييز بين مستويات متعددة نسبيًا من القدرة الرياضية.

كما دعمت التحليلات الكلاسيكية السابقة هذا الاستنتاج من خلال مؤشرات الثبات المرتفعة. وقد أظهرت تحليلات استقرار صعوبة المفردات بين مجموعات التلاميذ ارتباطًا مرتفعًا جدًا بلغ حوالي 0.992، في حين بلغ ارتباط التمييز حوالي 0.866، كما بلغ ارتباط قدرات التلاميذ في التقسيم الفردي-الزوجي حوالي 0.906، وبلغ معامل سبيرمان-براون حوالي 0.951.

وتوفر هذه النتائج مجموعة متكاملة من الأدلة على استقرار القياس، مع التأكيد على أن الثبات المرتفع لا يمثل وحده دليلًا كافيًا على صلاحية الاختبار؛ إذ يجب تفسيره مع أدلة المحتوى والبنية الداخلية وأداء المفردات وتغطية متصل القدرة.

ومن الناحية التطبيقية، تعني هذه النتائج أن الاختبار يمكن أن يوفر أساسًا مناسبًا للحصول على معلومات عن الفروق الفردية في التحصيل الرياضي، إلا أن استخدامه في اتخاذ قرارات فردية عالية المخاطر يتطلب دائمًا مراعاة خطأ القياس ومجال المعلومات التي يوفرها الاختبار.

9. من القياس إلى التشخيص التربوي

تسمح نتائج الدراسة الحالية بالانتقال من سؤال «هل الاختبار ثابت؟» إلى سؤال أكثر أهمية من الناحية التربوية، وهو: «ماذا يمكن أن نخبرنا نتائج الاختبار عن تعلم الرياضيات لدى التلاميذ؟».

فإذا أظهرت مفردة معينة صعوبة مرتفعة مع ملاءمة جيدة للنموذج، فقد تمثل هذه المفردة مستوى معرفيًا متقدمًا بالفعل ضمن البناء المقاس. أما إذا كانت صعبة وغير ملائمة في الوقت نفسه، فينبغي فحص احتمال وجود مشكلة في صياغتها أو في المهارة التي تتطلبها.

وبالمثل، إذا أظهرت مجموعة من المفردات مستوى مرتفعًا من الصعوبة، فقد يكون من المفيد تحليل المهارة المشتركة بينها. فإذا كانت المفردات تتعلق مثلاً بحل المشكلات أو العمليات متعددة الخطوات أو التمثيل الرياضي، فإن النتيجة يمكن أن توفر مؤشرًا تربويًا يستحق مزيدًا من الدراسة.

وبهذا المعنى، لا تقتصر فائدة الاختبار على إنتاج درجة نهائية، بل يمكن أن يصبح أداة تساعد الباحث والمعلم والمشرف التربوي على تحديد المجالات التي تحتاج إلى مراجعة أو تعزيز.

10. خصوصية الوسط التعليمي المحلي وحدود التعميم

ينبغي تفسير نتائج الدراسة الحالية في إطار العينة والوسط التعليمي الذي جُمعت منه البيانات. فخصائص المفردات، وقدرات التلاميذ، وأنماط الإجابة، ومستوى المعلومات الذي يوفره الاختبار، كلها قد تتأثر بخصائص المجتمع المدرسي الذي طُبّق فيه الاختبار.

وعليه، فإن النتائج الحالية تقدم أدلة قوية على خصائص الاختبار بالنسبة إلى العينة المدروسة والسياق الذي تم فيه التطبيق، لكنها لا تسمح وحدها بتعميم جميع الخصائص السيكمومترية على جميع تلاميذ التعليم الابتدائي في الجزائر.

ومن ثم، يمكن اعتبار الدراسة الحالية خطوة تأسيسية نحو بناء قاعدة سيكمومترية أوسع، من خلال تطبيق الاختبار على عينات من مناطق تعليمية مختلفة، وفحص ثبات معلمات المفردات، ودراسة الأداء التفاضلي للمفردات (DIF)، والتحقق من مدى استقرار خريطة رايت ودالة المعلومات عبر عينات ومجموعات مختلفة.

وهذا الاقتراح لا يقلل من قيمة النتائج الحالية، بل يحدد بدقة مجال الاستدلال الذي تسمح به البيانات، وهو أمر أساسي في بناء حجة صلاحية علمية للاختبار.

11. القيمة التطبيقية للدراسة في تطوير التقويم الرياضي المحلي

تكشف النتائج مجتمعة عن إمكانية الاستفادة من منهجية تجمع بين Rasch و CTT في تطوير

أدوات تقويم رياضي أكثر دقة وملاءمة للوسط التعليمي المحلي.

وتتمثل القيمة التطبيقية الأساسية للدراسة في أربعة مستويات مترابطة:

أولاً: مستوى بناء الاختبار.

تسمح النتائج بتحديد المفردات التي تحتاج إلى مراجعة، والمفردات التي تقدم معلومات جيدة، والمناطق من

متصل القدرة التي تحتاج إلى مفردات إضافية.

ثانياً: مستوى القياس.

يوفر نموذج راش مقياساً مشتركاً يسمح بمقارنة قدرات التلاميذ بصعوبة المفردات، وهو ما يتيح قراءة أكثر

تفصيلاً من مجرد الاعتماد على الدرجة الخام.

ثالثاً: المستوى التشخيصي.

يمكن استخدام أنماط الاستجابة لتحديد المجالات التي تظهر فيها صعوبات لدى مجموعات من التلاميذ،

مع ضرورة ربط ذلك بتحليل محتوى المفردات.

رابعاً: مستوى تطوير التقويم

يمكن أن تشكل النتائج قاعدة لتطوير بنك مفردات رياضية محلي، يتم فيه توصيف كل مفردة وفق

خصائصها السيكمومترية ومهارتها الرياضية ومستوى صعوبتها، بما يسمح ببناء صور اختبارية أكثر توازناً في

المستقبل.

استنتاج: يمكن، في ضوء النتائج السابقة، القول إن القيمة الأساسية للدراسة الحالية لا تتمثل في

إثبات أن الاختبار «جيد» أو «غير جيد» بصورة مطلقة، وإنما في تقديم حجة سيكومترية متكاملة حول

مدى جودة القياس الذي يوفره الاختبار في البيئة التعليمية المدروسة.

فقد دعمت نتائج النظرية الكلاسيكية وجود خصائص مقبولة لعدد مهم من المفردات، بينما أتاح نموذج

راش فحص العلاقة بين قدرات التلاميذ وصعوبات المفردات على متصل واحد. وأظهرت مؤشرات

الأشخاص مستوى مرتفعاً نسبياً من الفصل والثبات، في حين أتاح خريطة رايت ودالة المعلومات

تحديد مناطق القوة والقصور في تغطية متصل القدرة. كما قدم تحليل المكونات الرئيسية للبواقي دليلاً

إضافياً لفحص البنية الداخلية للاختبار، وأتاح تحليل المفردات غير الملائمة تحديد عناصر تستحق المراجعة

التربوية والسيكومترية.

وبذلك، فإن نتائج الدراسة تقدم أساساً علمياً يمكن البناء عليه لتطوير أدوات تقويم الرياضيات في المرحلة

الابتدائية، مع ضرورة المحافظة على التمييز بين صدق الاختبار في العينة الحالية وبين إمكانية تعميم

خصائصه على أوساط تعليمية أخرى.

ومن منظور أوسع، يمكن النظر إلى هذه الدراسة باعتبارها انتقالاً من نموذج تقويم يعتمد أساساً على

الدرجة الخام إلى نموذج أكثر تكاملاً يجمع بين التحليل الإحصائي، ونظرية القياس، وتحليل المحتوى

الرياضي، والسياق التعليمي المحلي. وهذا التكامل هو الذي يمنح نتائج الدراسة قيمتها التطبيقية، ويجعل

تطوير الاختبار عملية مستمرة قائمة على الأدلة بدل أن تكون عملية نهائية تنتهي بمجرد الحصول على

معاملات ثبات أو مؤشرات ملائمة مقبولة.

الخلاصة العامة للدراسة

عنوان الأطروحة: الخصائص السيكو مترية لاختبار تحصيلي في الرياضيات وفق النظرية الحديثة

للقياس التربوي والنفسي. دراسة ميدانية على مستوى السنة الخامسة ابتدائي

تم تطبيق الاختبار التحصيلي الذي تم تصميمه من طرف الطالب، وبعد عرضه لعملية التحكيم، وأجاء دراسة تجريبية أولية لتحديد زمن الاختبار الذي يتكون من 55 فقرة في صورته النهائية، وبعد تطبيق الدراسة الاستطلاعية للتحقق من الخصائص السيكو مترية وقد تحقق ذلك، فأصبح الاختبار صالحا للتطبيق على عينة الدراسة والتي بلغ عدد أفرادها 559 من مؤسسات التعليم الابتدائي بدوائر سيق، غقاز (ولاية معسكر) ووادي تليلات بولاية وهران.

وبعد عملية التصحيح تم تحليل البيانات وفق النظرية الكلاسيكية من خلال برنامج SPSS 20 ل يتم المقارنة بعد ذلك بتحليل البيانات ببرنامج WinSteps 3.72.3 وفق نموذج راش. من خلال التحليل وفق النظرية الكلاسيكية كان المتوسط الحسابي 27.92 وانحراف معياري قدره 10.22، وكانت اعلى درجة 53 وادناها 7.

الخصائص السيكو مترية تمثلت في: معامل صعوبة الفقرات امتد بين 0.97 و 0.11 مع وجود 9 فقرات. خارج هذا المجال.

معامل التمييز عن طريق معامل الارتباط المصحح بين الفقرة والدرجة الكلية مما اوجد فقرتين رقم 11 و 12 لم تحققا مستوى تمييز كاف.

معامل الثبات تم حسابه عن طريق معامل الفا كرونباخ فكان المؤشر 0.918 ويعتبر ممتازا، وتم كذلك حساب الثبات عن طريق التجزئة النصفية مع تصحيح الدرجة وجد انه 0.836

كما تم حساب معامل الثبات في حالة ما اذا كان الاختبار محكي المرجع مع تحديد درجة القطع والمتمثلة في 27. وتم حساب اربع عوامل ثبات عن طريق برنامج JMetrik وهي.

معامل Guttman 's L2 0.9232

معامل alpha Cronbach 09478

معامل Feldt. Gilmer 0.9206

معامل Feldt Brennan 0.923

معامل Raju's Beta 0.9178

ونلاحظ انها كلها اكبر من 0.9 وتعتبر ممتازة. والوثوق في نتائج الدراسة.

الصدق:

بعد عملية التحكيم في الدراسة السابقة تم تحقيق بيئة الصدق في الدراسة الاساسية بحساب معامل صدق الفقرة عن طريق معامل ارتباط الفقرات بالدرجة الكلية عن طريق معامل ارتباط بيرريال. وكذلك عن طريق المقارنة الطرفية حيث وجد ان هناك فروق دالة احصائيا بين المجموعتين العليا والدنيا، لصالح المجموعة العليا بمتوسط قدره 36.06 والمجموعة الدنيا بمتوسط 21.25.

التحليل وفق نموذج راش:

مجال صعوبة الفقرات من - 4.00 (فقرات سهلة جدا) الى +2.72 فقرات صعبة جدا. وهذا يشير الى تغطية جيدة لمستويات القدرة. تزداد القدرة بشكل توافقي مع الدرجات: وهذا يشير الى تحقيق متطلبات نموذج راش.

الدرجات المنخفضة (0) تشير الى مستوى القدرة θ منخفض جدا. حيث $\theta = -6.71$

الدرجات المرتفعة (55) تشير الى مستوى قدرة مرتفع جدا $\theta = +6.18$

تقل قيمة الانحراف المعياري في المجال المتوسط وهذا يشير الى أفضل تقدير لعملية القياس.

والخلاصة مما سبق ان الاختبار يقيس بدقة أكبر التلاميذ المتوسطين، وهو أمر مرغوب فيه من الناحية السيكومترية.

معامل الثبات للفقرات 0.99 مع معامل الفصل 13.32 ويعتبر المعامل ممتازا، ومعامل الفصل يعبر

عن مستويات صعوبة الفقرات (13 مستوى). بينما معامل ثبات الافراد فكان 0.98 مع معامل فصل

3.16 مما يشير الى مستويات قدرات التلاميذ حيث كانت في ثلاث مستويات..

تم تحليل مطابقة الفقرات، فكانت المطابقة الخارجية (MNSQ) ما بين 2.23 الى 0.65 بمتوسط 0.98 بوجود 5 فقرات خارج المجال المقبول في الادبيات (0.5 الى 1.5) وكذلك عن طريق المطابقة الداخلية (INFIT) حيث لم توجد فقرة خارج المجال. كذلك مدى مطابقة الافراد حيث وجد عدم تحقيق المطابقة جزئي لـ 51 تلميذ (MNSQ) بمتوسط قدره 0.98 .

اما تحقيق افتراضات نموذج راش فلم تحقق افتراض احادي البعد من خلال فحص البواقي حيث وجد انها متعددة العوامل. اما افتراض الاستقلال الموضوعي كذلك لم يتحقق لوجود فقرات مرتبطة فيما بينها كما رأينا في نتائج الدراسة..

اما مستوى القدرة عند التلاميذ امتد ما بين -2.8 الى +4.2.

بينما تشير المؤشرات الكلاسيكية إلى اتساق داخلي قوي، كشف تحليل راش عن عدم توافق مهم على مستوى الفقرة وبنى أبعاد ثانوية. تسلط هذه النتائج الضوء على القيمة التشخيصية المضافة لنموذج راش في تطوير الاختبار والتحقق من خاصية الصدق.

من خلال ما سبق نجد ان نموذج راش حقق تقديرات للقدرة واحصاءات معامل الصعوبة والثبات بصورة لا تختلف عن تحليل النظرية الكلاسيكية. ولكن عدم تحقق افتراضات نموذج راش. لمعالجة عدم توافق العناصر وعدم استقلال للبواقي، تم تحديد نموذج استكشافي متعدد الأبعاد. تم نمذجة مجموعات الفقرات التي تم تحديدها من خلال PCA للبواقي المعيارية كأبعاد كامنة منفصلة. هذا النهج يحسن بشكل كبير تمثيل المفهوم مع الاحتفاظ بالفقرات ذات الأهمية الجوهرية. واتباع القاعدة الثلاثية لحذف الفقرات او الافراد فان من الخطأ حذف أي منها وتستدعي التقييم التشخيصي لان نموذج راش يوفر قواعد علمية دقيقة للقيام بذلك.

ومما سبق من خلال هذه الدراسة وجد ان هناك تقارب في تحليل البيانات بين النظرية الكلاسيكية ونموذج راش من خلال خصائص الفقرات (الصعوبة والتمييز) وخاصيتي الثبات والصدق وتحديد مستوى القدرات عند التلاميذ من خلال مستوى رتي في مقابل المستوى الفئوي مما يعطي مصداقية ونتائج اذق

في تقدير الخصائص. وهذا ما التحقق من خلال دراسات كثيرة جدا حققت نفس النتائج كالتى حصلنا عليها.

وبالنسبة للفرضيات

الفرضية الاولى والتي تنص على ان خصائص درجات الاختبار وفق النظرية الكلاسيكية، تتوزع البيانات توزيعا طبيعيا مما يسمح بالتحليل الاحصائي البرامتري وقد تحققت الفرضية.

الفرضية الثانية والتي تنص على ان خصائص درجات الاختبار وفق نموذج راش تنتشر قدرات التلاميذ عبر مجال واسع من مستويات صعوبة الفقرات وقد تحقق ذلك كما رأينا ذلك من قبل.

الفرضية الثالثة والتي تنص على ان معامل الصعوبة للفقرات لا تتميز بالاستقرار وفق النظرية الكلاسيكية وهذا لم يتحقق حيث كان هناك استقرار عند تقسيم العينة الى مجموعتين.

ونفس الشيء ينطبق على نموذج راش حيث تميزت احصائيات النموذج معامل الصعوبة والتمييز والقدرة باستقرار كما دلت على ذلك الادبيات والدراسات السابقة.

يعتبر نموذج راش اكثر دقة من النظرية الكلاسيكية وذلك لتبنيه تحليل الفقرات مفردة وليس الاختبار ككل كما في النظرية الكلاسيكية، حيث يقدم معلومات دقيقة حول الفقرات وتوحيدها مع قدرة التلاميذ وامكانية تقييم كل فرد من خلال اجابته على كل فقرة وتشخيص نوعية الاجابة هل تنطبق مع توقعات النموذج ام ان هناك عوامل خارجية لا علاقة لها بقدرة التلميذ على مستوى فقرات الاختبار وقدرات التلاميذ.

وخلصت الدراسة الى تبني تحليل الاختبارات التحصيلية والنفسية وفق النظريتين والجمع بينهما مع الاعتماد على نموذج راش في تشخيص مخرجات العملية التعليمية .

دلالة النتائج في سياق القياس التربوي

تشير النتائج العامة للتحليل إلى أن الاختبار المستخدم في هذه الدراسة يتمتع بخصائص سيكومترية جيدة، سواء من حيث الثبات أو من حيث ملائمة المفردات للنموذج.

كما تشير النتائج إلى أن الاختبار قادر على:

1. التمييز بين مستويات القدرة المختلفة لدى التلاميذ.
 2. توفير تقديرات دقيقة نسبياً للقدرة الكامنة.
 3. تقديم معلومات قياسية كافية عبر نطاق واسع من مستويات القدرة.
- وتدعم هذه النتائج صدق الاختبار للاستخدام في:

- التقويم التربوي
- القياس التحصيلي
- الدراسات السيكومترية

التوصيات

في ضوء نتائج الدراسة يمكن تقديم مجموعة من التوصيات، من أبرزها:

1. الاستفادة من نموذج راش في تطوير الاختبارات التحصيلية في المجال التربوي.
2. إجراء تحليلات إضافية للكشف عن الوظيفة التفاضلية للمفردات (DIF) بين المجموعات المختلفة.
3. تطوير بنك مفردات يعتمد على معلمات الصعوبة المقدرة باستخدام نماذج IRT.
4. استخدام نتائج التحليل لتحسين تصميم الاختبارات المستقبلية.

آفاق البحث المستقبلي

يمكن للدراسات المستقبلية أن تتوسع في استخدام نماذج أكثر تقدماً من نماذج نظرية الاستجابة للمفردة، مثل:

- نموذج ثلاثي المعلمة (3PL)
- نموذج الاستجابة المتدرجة (GRM)
- النماذج متعددة الأبعاد

كما يمكن توسيع حجم العينة أو تطبيق الاختبار على عينات مختلفة للتحقق من استقرار معلمات

المفردات.

المراجع باللغة العربية

1. إبراهيم الشوربيجي، تقويم التحصيل الدراسي للطلاب الجامعي في ضوء الاتجاهات العالمية المعاصرة، القاهرة : جامعة الزقازيق، 2004.
2. أحمد محمد التقي، النظرية الحديثة في القياس، عمان: دار المسيرة، الطبعة الأولى، 2009
3. أمينة محمد الكاظم إستخدام نموذج راش في بناء إختبار تحصيلي في مادة علم النفس الطلاب كلية التربية وتحقيق التفسير الموضوعي للتنتائج ، الكويت، جامعة الكويت ، 1988.
4. أمينة محمد كاظم، اتجاهات معاصرة في القياس والتقويم النفسي والتربوي، القاهرة : الانجلو المصرية ، 1996.
5. أمينة محمد كاظم، دراسة نظرية نقدية حول القياس الموضوعي للسلوك " نموذج راش"، مؤسسة الكويت للتقدم العلمي، الكويت، 1988
6. أنور محمد الشرقاوي، وسليمان الخضري الشيخ، وأمينة كاظم ، و نادية عبد السلام ، اتجاهات معاصرة في القياس والتقويم النفسي والتربوي، القاهرة : مكتبة الأنجلو - المصرية ، 1996.
7. بشرى إسماعيل ، المرجع في القياس النفسي ، القاهرة: المكتبة الأنجلو مصرية ، الطبعة الأولى ، 2004.-
8. بشرى إسماعيل أحمد ، مدخل إلى علم النفس في القرن الحادي والعشرين"، القاهرة : المكتبة الانجلو مصرية ب.ط. 2007
9. حبيب مجدي عبد الكريم ، التقويم والقياس في التربية وعلم النفس، القاهرة : دار النهضة ، الطبعة 1996 ، الأولى
10. سعيد التل ، محمد وليد بطش ، فريد كامل أبو زينة ، مناهج البحث العلمي : تصميم البحث والتحليل الاحصائي، عمان : دار المسيرة للنشر والتوزيع، الطبعة الأولى ، 2007.
11. صالح محمد العساف ، المدخل إلى البحث في العلوم السلوكية ، المملكة العربية السعودية : مكتبة العبيكان ، الطبعة الثالثة ، 2003.
12. صبحي حمدان أبو جلاله ، إتجاهات معاصرة في القياس النفسي والتربوي"، الكويت : مكتبة الفلاح ، الطبعة الأولى ، 1999
13. صلاح الدين محمود علام ، نماذج الاستجابة للمفردة الاختبارية أحادية البعد و متعددة الأبعاد و تطبيقاتها في القياس النفسي والتربوي ، القاهرة : دار الفكر العربي، الطبعة الأولى ، 2005.
14. صلاح الدين محمود علام ، القياس والتقويم التربوي والنفسي: أساسياته و تطبيقاته و توجيهاته المعاصرة"، مصر : دار الفكر العربي 2002
15. صلاح الدين محمود علام ، الاختبارات التشخيصية المرجعية المحك في المجالات التربوية والنفسية والتدريبية ، القاهرة: دار الفكر العربي ، الطبعة الثانية ، 2001.
16. صلاح الدين محمود علام الاختبارات و المقاييس التربوية والنفسية، دار الفكر ناشرون و موزعون الطبعة الأولى، عمان - الأردن، 2006.
17. صلاح الدين محمود علام، تطورات معاصرة في القياس النفسي والتربوي"، الكويت : مطابع القيس التجارية ، الطبعة الأولى، 1986

18. صلاح الدين محمود علام، القياس والتقويم التربوي و النفسي أساسياته وتوجيهاته المعاصرة ، القاهرة: دار الفكر العربي، 2000.
19. صلاح مراد و أمين سليمان ، الاختبارات و المقاييس في العلوم النفسية والتربوية، خطوات إعدادهما و خصائصها، دار الكتاب الحديث، القاهرة، 2002
20. الطيب عصام علي، أساليب التفكير : نظريات و دراسات وبحوث معاصرة، مصر : عالم الكتب ، الطبعة 2006 الأولى،
21. عبد الرحمان بن سليمان الطريفي، القياس النفسي والتقويم التربوي : نظريته ، أسسه ، تطبيقاته، المملكة العربية السعودية : مكتبة الرشد بالرياض، الطبعة الأولى ، 1997.
22. عبد العزيز بوسالم ، القياس في علم النفس وعلوم التربية الأسس النظرية والمبادئ التطبيقية، الجزائر : دار القرطبة للنشر والتوزيع، الطبعة الأولى ، 2014.
23. عبد المنعم أحمد الدردير ، دراسة معاصرة في علم النفس المعرفي، الجزء الثاني، القاهرة : عالم الكتب، الطبعة الأولى، 2004 .
24. فؤاد أبو حطب، آمال صادق، علم النفس التربوي، القاهرة : مكتبة الأنجلو ، الطبعة الخامسة ، 1996.
25. فؤاد يحي السيد ، علم النفس الإحصائي وقياس العقل البشري ، جمهورية مصر العربية ، دار الفكر العربي ، الطبعة الأولى ، 1987
26. كروكر ، ليند والجينا، جيمس مدخل إلى نظرية القياس التقليدية والمعاصرة (ترجمة زينبات دعنا)، عمان: دار الفكر 2009.
27. موسى النبهان ، أساسيات القياس في العلوم السلوكية، الأردن دار الشروق للنشر والتوزيع ، الطبعة 2004 ، الأولى
28. ميمي السيد أحمد ، التوجهات الحديثة في القياس النفسي والتربوي، القاهرة : دار الكتاب الحديث، الطبعة الأولى ، 2014
29. ديانا فهمي علي حماد ، علاقة الحكمة الإختبارية بالأداء الناتج من إختبار تحصيلي ذي إختيار متعدد مبني وفق نموذج "راش" لدى طالبات كلية التربية، رسالة دكتوراه - تخصص قياس و التقويم، غير منشورة 2010
30. شادية منتصر عبد العزيز، إستخدام نموذج راش في بناء بنك للأسئلة لمقرر في علم النفس التعليمي و تحديد الدرجات الفاصلة المقابلة للتقديرات الجامعية ، رسالة ماجستير غير منشورة، كلية البنات للآداب و العلوم والتربية، جامعة عين الشمس ، مصر ، . 1996.
31. محمد منصور الشافعي، أثر معادلة درجات الاختبار و الضوابط العينية على تدريج بنود بنك الأسئلة باستخدام نموذج راش، رسالة دكتوراه غير منشورة كلية التربية، جامعة المنصورة، 1996
32. شامخ هشام جمعة صويح، بناء مقياس أساليب التعليم لدى طلبة الجامعة وفقا لنظرية السمات الكامنة، أطروحة دكتوراه غير منشورة، كلية التربية ابن رشد جامعة بغداد 2011 العراق
33. صليحة لعزالي ، الخصائص السيكمومترية لمقياس مداخل الدراسة " مقدرة وفق النظرية التقليدية ونموذج راش، رسالة ماجستير غير منشورة ، جامعة سعد دحلب البليدة ، الجزائر ، 2010.

34. عبد الحميد طنطاوي الشيماء ، أثر كل من حجم العينة و طول الاختبار على دقة نتائج الاختبارات التحصيلية المرجعية المحك في الرياضيات باستخدام نموذج راش ، رسالة ماجستير غير منشورة، كلية التربية ، جامعة الزقازيق ، مصر ، . 2007
35. عبد العزيز بوسالم ، توظيف نموذج راش أحادي البرامتر في بناء اختبار تحصيلي في القياس النفسي و تحقيق التفسير الموضوعي لنتائجه بالمقارنة مع النظرية الكلاسيكية ، دراسة سيكومترية مقارنة بين نظرية السمات الكامنة والنظرية الكلاسيكية في بناء الاختبارات و التحقيق من موضوعيتها ، رسالة دكتوراة في علوم التربية، غير منشورة ، 2008
36. عبد الله ، إعتدال غازي، استخدام نموذج راش في تدرج مقياس للقدرة العقلية لدراسة بعض العوامل المؤثرة على دقة القياس ، رسالة دكتوراه غير منشورة ، جامعة عين الشمس 2003
37. عبد الوهاب محمد محمود ، استخدام نماذج الإستجابة للمفردة الإختبارية تدرج مفردات بعض الإختبارات المعرفية"، رسالة دكتوراه غير منشورة ، جامعة المنيا ، جمهورية مصر ، 2010.
38. علي زكري الخصائص السيكمترية لإختبار (أوتيس - لينون) للقدرة العقلية مقدرة وفق القياس الكلاسيكي و نموذج راش لدى طلبة المرحلة المتوسطة بمحافظة صبيا التعليمية، رسالة دكتوراه غير منشورة ، المملكة العربية السعودية كلية التربية جامعة أم القرى ، 2009.
39. متيرد، منى ربيع الطنطاوي، دراسة سيكومترية حول تطوير إختبار المصفوفات المتتابعة لرافن باستخدام نموذج راش ، رسالة ماجستير غير منشورة ، جامعة عين الشمس، مصر ، 2000.
40. نداء بماء الدين علي ، فعالية إستخدام نموذج راش في بناء إختبار تحصيلي محكي المرجع المقرر القياس والتقويم في التربية "، رسالة دكتوراه غير منشورة ، جامعة دمشق ، 2012.
41. وهدي رانيا ماهر ، دراسة سيكومترية مقارنة بين نموذج راش والقياس التقليدي حول دقة التنبؤ بحالة القلق من سمة القلق " ، رسالة ماجستير غير منشورة ، جامعة عين الشمس ، مصر 2005،
42. ابراهيم مبارك الدوسري، نظرية الاستجابة للبند ، ورقة عمل غير منشورة ، مؤتمر خبراء القياس، بيروت 2004، لبنان ،
43. أبو هاشم السيد محمد ، دراسة مقارنة بين النظرية التقليدية ونموذج راش في إختبار فقرات مقياس مداخل الدراسة لدى طلاب الجامعة "، مجلة كلية التربية ، العدد 08 ، مصر ، جامعة الزقازيق ، 2006
44. أحمد العودة ، مدى التوافق بين نموذج راش والمؤشرات التقليدية في إختبار فقرات مقياس الاتجاه السباعي التدرج، مجلة كلية التربية، العدد 08، الإمارات ، 1992
45. تغريد حجازي وعبد الله الخطيب، التوافق بين النظرية الكلاسيكية والنموذج الثنائي المعلمة في مطابقة فقرات إختبار محكي المرجع في أحكام التلاوة والتجويد "، مجلة جامعة النجاح للأبحاث ، المجلد 28 ، العدد 10. 2014
46. حمدي يونس أبو جراد ، استخدام نموذج راش في تطوير إختبار كاتل الثالث للذكاء المصور (أ)، مجلة الجامعة الإسلامية ، المجلد 16 ، العدد 2 ، جامعة القدس المفتوحة ، 2008.
47. عادل عبد الجبار، محمد منصور الشافعي، أثر حجم عينة التحليل على مؤشرات الملائمة الإحصائية و تقديرات الصعوبة المتضمنة لبرنامج ميكروسكال (Microscal) و بايلوج (Bilog) للمفردات باستخدام نموذج راش، دراسة محاكاة، رسالة التربية وعلم النفس، 25-64، 2004.

- 48.. عماد عبد المسيح ، استخدام نموذج راش اللوغاريتمي أحادى البارامتر في تحليل مفردات الاختبارات المعرفية مرجعية المعيار ثنائية القطب دراسة تجريبية ، مجلة البحث في التربية وعلم النفس ، جامعة المنيا ، كلية التربية ، العدد (4)، مصر ، . 1994
- 49.. عماد يوسف الصفطي ، دراسة حول استخدام نموذج راش اللوغاريتمي الإحتمالي في بناء إختبار تشخيص هدي المرجع لقياس إتقان المعلمين والباحثين للمهارات الأساسية اللازمة لبناء الإختبارات التحصيلية في التقويم التربوي، وعلاقته بتحسين مخرجات التعليم والتعلم، ورقة عمل غير منشورة ، المؤتمر العلمي الثالث ، كلية التربية جامعة البحرين ، 1994.
- 50.. غادة عيد خالد ، الدرجة الحقيقية مقدرة بإستخدام نظرية السمات الكامنة والنظرية التقليدية دراسة سيكومترية - "، مجلة جامعة أم القرى للعلوم التربوية والإجتماعية والإنسانية ، جامعة أم القرى ، المملكة العربية السعودية ، 2004.
51. سليمان بن عبد الله بن صالح الدويش 2003. أطروحة دكتوراه: دراسة تحليلية لتطوير أسئلة اختبار مادة الرياضيات في المرحلة المتوسطة في منطقة الرياض التعليمية في ضوء واهداف المنهج التعليمية . جامعة محمد بن سعود الاسلامية الرياض

References

1. A Nation at Risk: The Imperative for Educational Reform . Washington, DC: U.S. Government Printing Office.
2. Aybek, E. C. (2023). The relation of item difficulty between Classical Test Theory and Item Response Theory: Computerized adaptive test perspective. *Journal of Measurement and Evaluation in Education and Psychology*, 14(2), 118–127.
3. Bennett, R. E., Bejar, I. I., & Harnisch, D. L. (2012). Technology and Assessment: The Role of Technology in the Assessment of Student Learning . In R. L. Linn (Ed.), Measurement and Assessment in Teaching (pp. 188-215). Boston: Pearson.
4. Binet, A., & Simon, T. (1916). The Development of Intelligence in Children: The Binet-Simon Scale . Baltimore: Williams & Wilkins.
5. Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice* , 5(1), 7-74.
6. Cees A.W. Glas Measure et évaluation en éducation Item response theory in educational assessment and evaluation Cees A.W. Glas Volume 31, Number 2, 2008 URI: <https://id.erudit.org/iderudit/1025005ar> DOI: <https://doi.org/10.7202/1025005ar>
7. Crocker, L., & Algina, J. (1986). Introduction to Classical and Modern Test Theory . New York: Holt, Rinehart and Winston
8. de Ayala, R. J. (2009). The Theory and Practice of Item Response Theory . New York: Guilford Press
9. E. C. Aybek, "The Relation of Item Difficulty Between Classical Test Theory and Item Response Theory: Computerized Adaptive Test Perspective," *Journal of Measurement and Evaluation in Education and Psychology*, vol. 14, no. 2, pp. 118–127, June 2023, [doi: 10.21031/epod.1209284](https://doi.org/10.21031/epod.1209284).
10. Eccles, J. S., & Roeser, R. W. (2011). Schools, academic motivation, and stage-environment fit. In Handbook of Adolescent Psychology (Vol. 2, pp. 404-434). Wiley.
11. Embretson, S. E., & Reise, S. P. (2000). Item Response Theory for Psychologists . Lawrence Erlbaum Associates.

12. Embretson, S. E., & Reise, S. P. (2000). *Item Response Theory for Psychologists*. Mahwah, NJ: Lawrence Erlbaum Associates.
13. Embretson, S. E., & Reise, S. P. (2000). *Item Response Theory for Psychologists*. Mahwah, NJ: Lawrence Erlbaum Associates.
14. Gonzalez, G., et al. (2010). The impact of cultural factors on mathematics achievement. *Journal of Educational Psychology*, 102(3), 610-620.
15. Hambleton, R. K. (2000). The changing role of item response theory in educational measurement. *Educational Measurement: Issues and Practice*, 19(2), 5-15.
16. Hambleton, R. K., & Jones, R. W. (1993). Comparison of Classical Test Theory and Item Response Theory and their Applications to Test Development. *Educational Measurement: Issues and Practice*, 12(3), 38-47.
17. Hambleton, R. K., & Jones, R. W. (1993). Comparison of Classical Test Theory and Item Response Theory and their Applications to Test Development. *Educational Measurement: Issues and Practice*, 12(3), 38-47.
18. Hambleton, R. K., & Swaminathan, H. (1985). *Item Response Theory: Principles and Applications*. Boston: Kluwer-Nijhoff.
19. Harris, D. N. (2007). The impact of accountability on student achievement: A meta-analysis. *Educational Evaluation and Policy Analysis*, 29(3), 251-276.
20. Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81-112.
21. Henson, R. K. (2001). Understanding internal consistency reliability estimates: A conceptual model. *Educational and Psychological Measurement*, 61(3), 355-372.
22. Holland, P. W., & Hoskens, M. (2025). Classical Test Theory as a first-order Item Response Theory: Application to true-score prediction from a possibly nonparallel test. *Psychometrika*.
23. Huang, Y., et al. (2018). The effectiveness of computer-based assessments in mathematics: A meta-analysis. *Computers & Education*, 119, 56-68.
24. Kershner, R., et al. (2011). The role of technology in mathematics education: A review of the literature. *Research in Mathematics Education*, 13(2), 143-161.

25. Kilpatrick, J. (2003). A history of mathematics education in the United States. In F. K. Lester Jr. (Ed.), *Second Handbook of Research on Mathematics Teaching and Learning* (pp. 3-38). Charlotte, NC: Information Age Publishing.
26. Kline, P. (2000). *The Handbook of Psychological Testing* (2nd ed.). New York: Routledge.
27. Kubo, S. (2023). Test theory underpinning learner assessment. *Medical Education Japan*, 54(4), 367–375.
28. Lesh, R., & Doerr, H. M. (2003). *Beyond Constructivism: Models and Modeling Perspectives on Mathematics Problem Solving, Learning, and Teaching*. Mahwah, NJ: Lawrence Erlbaum Associates.
29. Linn, R. L. (2000). Assessments and accountability. *Educational Researcher*, 29(2), 4-16.
30. Lord, F. M. (1980). *Applications of Item Response Theory to Practical Testing Problems*. Lawrence Erlbaum Associates.
31. McDonald, R. P. (1999). *Test Theory: A Unified Treatment*. Lawrence Erlbaum Associates.
32. Mutluer, C., & Çakan, M. (2023). Comparison of test equating methods based on Classical Test Theory and Item Response Theory. *Journal of Uludag University Faculty of Education*, 36(3), 866–906.
33. National Governors Association Center for Best Practices & Council of Chief State School Officers. (2010). *Common Core State Standards for Mathematics*. Washington, DC: Authors.
34. *No Child Left Behind Act of 2001*. Washington, DC: U.S. Government Printing Office.
35. Popham, W. J. (2017). *Classroom Assessment: Principles and Practice for Effective Standards-Based Instruction*. Pearson.
36. Reise, S. P. (2012). The rediscovery of the multidimensionality of personality. *Journal of Personality Assessment*, 94(4), 328-333.
37. Rose, D. H., & Meyer, A. (2002). *Teaching Every Student in the Digital Age: Universal Design for Learning*. Alexandria, VA: Association for Supervision and Curriculum Development.

38. Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology*, 69(5), 797-811.
39. Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education*, 2, 53-55.
40. Thissen, D., & Wainer, H. (2001). *Test Scoring*. Mahwah, NJ: Lawrence Erlbaum Associates.
41. Thompson, S. J., Thurlow, M. L., & Kearns, D. (2006). *Assessment of Students with Disabilities: A Review of the Literature*. Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.
42. Ukofia, I.-B. F., & Essen, C. B. (2023). A comparative analysis of item difficulty and discrimination parameters estimation in classical test and item response theories. *The COLLOQUIUM*, 9(1), 1–8.
43. Van der Linden, W. J., & Pashley, P. J. (2006). *Computerized Adaptive Testing: Theory and Practice*. New York: Springer.

الملاحق

الملحق رقم 1 : قائمة المحكمين

الاسم واللقب	الصفة
1 فراينة خالد	مفتش التعليم الابتدائي (محمدية)
2 نقاب محمد	مفتش التعليم الابتدائي (سيق)
3 قرشال مصطفى	مفتش التعليم الابتدائي (سيق)
4 زاهد خديجة	مفتش التعليم الابتدائي (وهران)
5 بشير بويجرة خالد	مفتش التعليم الابتدائي (وهران)
6 بن سهلة الطيب	مدير مؤسسة ابتدائية
7 شعبة ابراهيم	مدير مؤسسة ابتدائية
8 بن حليلة فطيمة زهرة	استاذ التعليم الابتدائي
9 شقراني كريمة	استاذ التعليم الابتدائي
10 رزقي قادة	استاذ التعليم الابتدائي
11 بن ونان خميس	استاذ التعليم الابتدائي

الملحق رقم 2:

صدق الاتساق الداخلي (مخرجات SPSS 20 معامل ارتباط الدرجات بالدرجة الكلية

		TOTAL
TOTAL	Corrélation de Pearson	1
	Sig. (bilatérale)	
	N	559
item1	Corrélation de Pearson	,382
	Sig. (bilatérale)	,000
	N	559
item2	Corrélation de Pearson	,500
	Sig. (bilatérale)	,000

	N	559
item3	Corrélation de Pearson	,493
	Sig. (bilatérale)	,000
	N	559
item4	Corrélation de Pearson	,395
	Sig. (bilatérale)	,000
	N	559
item5	Corrélation de Pearson	,357
	Sig. (bilatérale)	,000
	N	559
item6	Corrélation de Pearson	,348
	Sig. (bilatérale)	,000
	N	559
item7	Corrélation de Pearson	,296
	Sig. (bilatérale)	,000
	N	559
item8	Corrélation de Pearson	,498
	Sig. (bilatérale)	,000
	N	559
item9	Corrélation de Pearson	,272
	Sig. (bilatérale)	,000
	N	559
item10	Corrélation de Pearson	,254
	Sig. (bilatérale)	,000
item11	Corrélation de Pearson	,026
	Sig. (bilatérale)	,540
	N	559
item12	Corrélation de Pearson	,163
	Sig. (bilatérale)	,000
	N	559
item13	Corrélation de Pearson	,240
	Sig. (bilatérale)	,000
	N	559
item14	Corrélation de Pearson	,472
	Sig. (bilatérale)	,000
	N	559
item15	Corrélation de Pearson	,486

	Sig. (bilatérale)	,000
	N	559
item16	Corrélation de Pearson	,529
	Sig. (bilatérale)	,000
	N	559
item17	Corrélation de Pearson	,332
	Sig. (bilatérale)	,000
	N	559
item18	Corrélation de Pearson	,383
	Sig. (bilatérale)	,000
	N	559
item19	Corrélation de Pearson	,276
	Sig. (bilatérale)	,000
	N	559
item20	Corrélation de Pearson	,408
	Sig. (bilatérale)	,000
	N	559
item21	Corrélation de Pearson	,409
	Sig. (bilatérale)	,000
	N	559
item22	Corrélation de Pearson	,338
	Sig. (bilatérale)	,000
	N	559
item23	Corrélation de Pearson	,446
	Sig. (bilatérale)	,000
	N	559
item24	Corrélation de Pearson	,620
	Sig. (bilatérale)	,000
	N	559
item25	Corrélation de Pearson	,610
	Sig. (bilatérale)	,000
	N	559
item26	Corrélation de Pearson	,610
	Sig. (bilatérale)	,000
	N	559
item27	Corrélation de Pearson	,376
	Sig. (bilatérale)	,000
	N	559

item28	Corrélation de Pearson	,376
	Sig. (bilatérale)	,000
	N	559
item29	Corrélation de Pearson	,376
	Sig. (bilatérale)	,000
	N	559
item30	Corrélation de Pearson	,366
	Sig. (bilatérale)	,000
	N	559
item31	Corrélation de Pearson	,415
	Sig. (bilatérale)	,000
	N	559
item32	Corrélation de Pearson	,550
	Sig. (bilatérale)	,000
	N	559
item33	Corrélation de Pearson	,645
	Sig. (bilatérale)	,000
	N	559
item34	Corrélation de Pearson	,425
	Sig. (bilatérale)	,000
	N	559
item35	Corrélation de Pearson	,440
	Sig. (bilatérale)	,000
	N	559
item36	Corrélation de Pearson	,438
	Sig. (bilatérale)	,000
	N	559
item37	Corrélation de Pearson	,427
	Sig. (bilatérale)	,000
	N	559
item38	Corrélation de Pearson	,577
	Sig. (bilatérale)	,000
	N	559
item39	Corrélation de Pearson	,194
	Sig. (bilatérale)	,000
	N	559
item40	Corrélation de Pearson	,593

	Sig. (bilatérale)	,000
	N	559
item41	Corrélation de Pearson	,483
	Sig. (bilatérale)	,000
	N	559
item42	Corrélation de Pearson	,483
	Sig. (bilatérale)	,000
	N	559
item43	Corrélation de Pearson	,483
	Sig. (bilatérale)	,000
	N	559
item44	Corrélation de Pearson	,532
	Sig. (bilatérale)	,000
	N	559
item45	Corrélation de Pearson	,405
	Sig. (bilatérale)	,000
	N	559
item46	Corrélation de Pearson	,393
	Sig. (bilatérale)	,000
	N	559
item47	Corrélation de Pearson	,598
	Sig. (bilatérale)	,000
	N	559
item48	Corrélation de Pearson	,597
	Sig. (bilatérale)	,000
	N	559
item49	Corrélation de Pearson	,597
	Sig. (bilatérale)	,000
	N	559
item50	Corrélation de Pearson	,532
	Sig. (bilatérale)	,000
	N	559
item51	Corrélation de Pearson	,317
	Sig. (bilatérale)	,000
	N	559
item52	Corrélation de Pearson	,439
	Sig. (bilatérale)	,000
	N	559

item53	Corrélation de Pearson	,439
	Sig. (bilatérale)	,000
	N	559
item54	Corrélation de Pearson	,557
	Sig. (bilatérale)	,000
	N	559
item55	Corrélation de Pearson	,313
	Sig. (bilatérale)	,000
	N	559

الملحق رقم 3 معامل الصعوبة والتمييز وفق النظرية الكلاسيكية

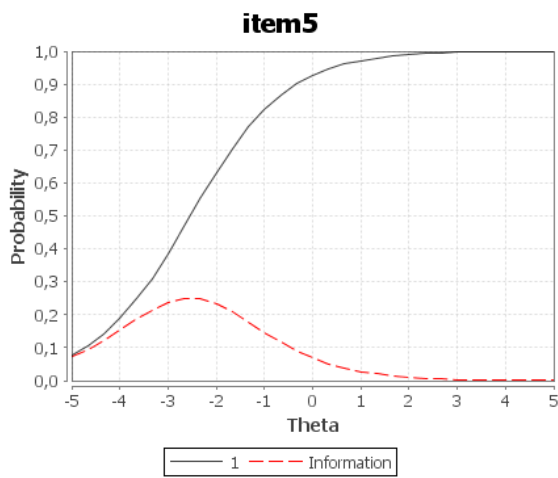
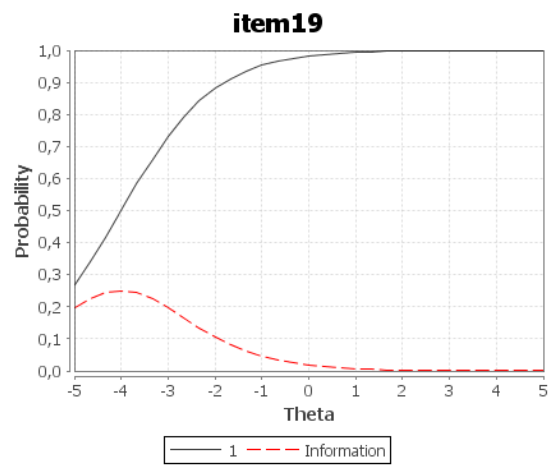
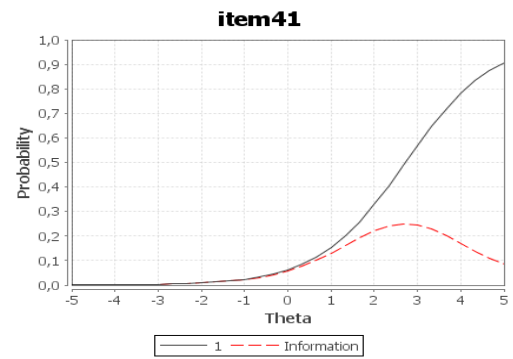
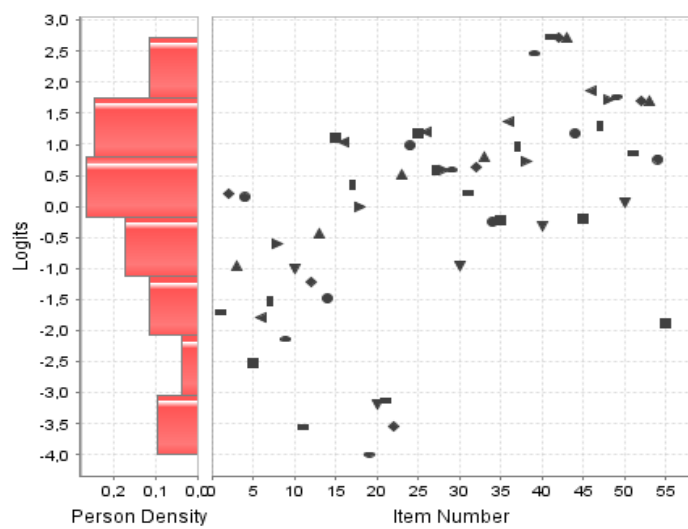
Item	, Difficulty	, Std. Dev.	, Discrimin.
item1	0,8122 ,	0,390 ,9	0,506
2	0,4884 ,	0,5003	0,5792
3	0,7048 ,	0,4565	0,606
4	0,4991 ,	0,5004	0,4414
5	0,898	0,3029	0,562
6	0,8247	0,3806	0,4642
7	0,7889	0,4084	0,3655
8	0,6422	0,4798	0,5922
9	0,8623	0,3449	0,3769
10	0,712	0,4532	0,2809
11	0,9571	0,2029	0,0136
12	0,7442	0,4367	0,1639
13 ,	0,6118	0,4878	0,2476
14 ,	0,7835	0,4122	0,6179
15 ,	0,3184	0,4663	0,588
16 ,	0,3309	0,471	0,6415
17 ,	0,458	0,4987	0,3614
18 ,	0,5277	0,4997	0,4268
19 ,	0,9714	0,1669	0,6647
20 ,	0,941	0,2359	0,7777
21 ,	0,9374	0,2425	0,7661
22 ,	0,9571	0,2029	0,7113
23 ,	0,4275	0,4952	0,5113
24 ,	0,3417	0,4747	0,7625
25 ,	0,3059	0,4612	0,7628

26 ,	0,3041	0,4604	0,7639
27 ,	0,4132	0,4929	0,4211
28 ,	0,4132	0,4929	0,4211
29 ,	0,4132	0,4929	0,4211
30 ,	0,7048	0,4565	0,4308
31 ,	0,4866	0,5003	0,4679
32 ,	0,4043	0,4912	0,6519
33 ,	0,3757	0,4847	0,7868
34 ,	0,576	0,4946	0,4847
35 ,	0,5689	0,4957	0,5036
36 ,	0,2773	0,4481	0,5367
37 ,	0,3435	0,4753	0,5013
38 ,	0,3864	0,4874	0,6919
39 ,	0,1342	0,3411	0,2544
40 ,	0,5868	0,4929	0,7084
41 ,	0,1109	0,3143	0,7612
42 ,	0,1109	0,3143	0,7612
43 ,	0,1109	0,3143	0,7612
44 ,	0,3077	0,462	0,6544
45 ,	0,5671	0,4959	0,4582
46 ,	0,2039	0,4033	0,5101
47 ,	0,2862	0,4524	0,7548
48 ,	0,2236	0,417	0,7939
49 ,	0,2165	0,4122	0,8
50 ,	0,517	0,5002	0,6217
51 ,	0,3649	0,4818	0,3507
52 ,	0,2254	0,4182	0,5633
53 ,	0,2254	0,4182	0,5633
54 ,	0,381	0,4861	0,6665
55 ,	0,8336	0,3727	0,4173

الملحق رقم : معاميل صعوبة الفقرات

2	TABLE 17.1 C:\Winsteps\doct25.txt				ZOU270WS.TXT				Mar 25 11:51 2026							
3	INPUT: 559 Person 55 Item REPORTED: 559 Person 55 Item 2 CATS WINSTEPS 3.72.3															
4	Person: REAL SEP.: 3.17 REL.: .91 ... Item: REAL SEP.: 13.33 REL.: .99															
5																
6	Person STATISTICS: MEASURE ORDER															
7																
8																
9	ENTRY	TOTAL	TOTAL		MODEL	INFIT		OUTFIT		PT-MEASURE		EXACT MATCH				
10	NUMBER	SCORE	COUNT	MEASURE	S.E.	MNSQ	ZSTD	MNSQ	ZSTD	CORR.	EXP.	OBS%	EXP%	Item		
11																
12	11	535	559	-3.62	.22	1.18	1.0	2.23	2.8	A .02	.23	95.7	95.7	I0011		
13	12	416	559	-1.24	.11	1.28	4.7	2.02	6.7	B .18	.43	73.0	78.5	I0012		
14	6	461	559	-1.83	.12	.93	-1.0	1.68	3.6	C .37	.39	89.1	83.9	I0006		
15	13	342	559	-.45	.10	1.28	6.1	1.61	6.6	D .25	.47	59.7	72.4	I0013		
16	17	256	559	.36	.10	1.17	4.1	1.60	7.6	E .31	.48	68.0	70.6	I0017		
17	10	398	559	-1.03	.11	1.28	5.0	1.50	4.1	F .24	.44	66.5	76.6	I0010		
18	39	75	559	2.51	.14	1.21	2.3	1.40	1.9	G .21	.37	85.9	87.6	I0039		
19	7	441	559	-1.55	.12	1.16	2.5	1.35	2.3	H .30	.41	75.7	81.3	I0007		
20	34	322	559	-.26	.10	1.05	1.2	1.30	3.8	I .41	.47	68.3	71.4	I0034		
21	9	482	559	-2.18	.13	1.09	1.1	1.27	1.4	J .28	.36	86.9	86.9	I0009		
22	36	155	559	1.38	.11	1.00	.1	1.25	2.3	K .43	.45	77.8	77.1	I0036		
23	51	204	559	.86	.10	1.21	4.6	1.21	2.5	L .33	.47	64.8	72.9	I0051		
24	4	279	559	.15	.10	1.11	2.7	1.21	3.0	M .39	.48	67.4	70.4	I0004		
25	18	295	559	.00	.10	1.13	3.1	1.17	2.4	N .39	.48	63.1	70.6	I0018		
26	30	394	559	-.98	.10	1.16	3.0	1.04	.4	O .36	.45	70.5	76.2	I0030		
27	45	317	559	-.21	.10	1.10	2.3	1.15	2.0	P .40	.47	71.7	71.2	I0045		
28	27	231	559	.60	.10	1.14	3.4	1.09	1.2	Q .38	.47	65.5	71.3	I0027		
29	28	231	559	.60	.10	1.14	3.4	1.09	1.2	R .38	.47	65.5	71.3	I0028		
30	29	231	559	.60	.10	1.14	3.4	1.09	1.2	S .38	.47	65.5	71.3	I0029		
31	35	318	559	-.22	.10	1.05	1.3	1.11	1.5	T .43	.47	69.8	71.2	I0035		
32	55	466	559	-1.91	.13	1.07	.9	1.09	.6	U .32	.38	83.4	84.6	I0055		
33	31	272	559	.21	.10	1.09	2.3	1.06	.9	V .42	.48	65.7	70.4	I0031		
34	37	192	559	.98	.10	1.05	1.1	1.05	.7	W .42	.46	72.6	73.8	I0037		
35	23	239	559	.52	.10	1.05	1.2	1.05	.7	X .44	.47	69.4	71.0	I0023		
36	46	114	559	1.89	.12	1.05	.7	.99	.0	Y .39	.42	80.3	82.0	I0046		
37	1	454	559	-1.73	.12	1.00	.0	1.04	.3	Z .39	.40	83.9	83.0	I0001		
38	2	273	559	.20	.10	.98	-.6	1.00	.0	z .49	.48	69.4	70.4	I0002		
39	BETTER FITTING OMITTED															
40	3	394	559	-.98	.10	.96	-.8	.97	-.3	y .48	.45	71.9	76.2	I0003		
41	8	359	559	-.62	.10	.95	-1.2	.96	-.5	x .49	.46	75.5	73.4	I0008		
42	5	502	559	-2.58	.15	.94	-.6	.75	-1.1	w .37	.32	92.1	90.0	I0005		
43	50	289	559	.05	.10	.93	-1.8	.87	-1.9	v .53	.48	70.3	70.5	I0050		
44	44	172	559	1.19	.10	.92	-1.6	.74	-3.0	u .53	.45	75.8	75.5	I0044		
45	16	185	559	1.06	.10	.92	-1.7	.85	-1.8	t .52	.46	76.2	74.4	I0016		
46	19	543	559	-4.07	.26	.92	-.3	.28	-2.5	s .29	.19	97.1	97.1	I0019		
47	32	226	559	.65	.10	.91	-2.3	.84	-2.3	r .54	.47	76.6	71.6	I0032		
48	54	213	559	.77	.10	.90	-2.5	.80	-2.9	q .55	.47	74.2	72.3	I0054		
49	14	438	559	-1.51	.11	.89	-1.8	.89	-.8	p .48	.41	80.9	80.9	I0014		
50	38	216	559	.74	.10	.87	-3.2	.82	-2.5	o .56	.47	76.6	72.2	I0038		
51	22	535	559	-3.62	.22	.86	-.8	.30	-2.8	n .36	.23	95.7	95.7	I0022		
52	40	328	559	-.31	.10	.83	-4.3	.79	-3.0	m .58	.47	78.4	71.6	I0040		
53	21	524	559	-3.18	.18	.82	-1.4	.30	-3.3	l .43	.26	93.7	93.7	I0021		
54	47	160	559	1.33	.11	.82	-3.8	.66	-3.8	k .59	.45	81.2	76.6	I0047		
55	25	171	559	1.20	.10	.81	-4.1	.64	-4.3	j .60	.45	81.8	75.6	I0025		
56	26	170	559	1.22	.10	.81	-4.2	.64	-4.3	i .60	.45	81.9	75.7	I0026		
57	24	191	559	.99	.10	.81	-4.5	.67	-4.5	h .61	.46	80.0	73.9	I0024		
58	20	526	559	-3.25	.19	.81	-1.4	.27	-3.4	g .43	.26	94.1	94.1	I0020		
59	48	125	559	1.74	.11	.80	-3.5	.59	-3.8	f .58	.43	83.4	80.6	I0048		
60	41	62	559	2.77	.15	.80	-2.2	.51	-2.7	e .49	.35	90.7	89.6	I0041		
61	42	62	559	2.77	.15	.80	-2.2	.51	-2.7	d .49	.35	90.7	89.6	I0042		
62	43	62	559	2.77	.15	.80	-2.2	.51	-2.7	c .49	.35	90.7	89.6	I0043		
63	49	121	559	1.80	.11	.79	-3.6	.58	-3.8	b .58	.42	82.6	81.1	I0049		
64	33	210	559	.80	.10	.78	-5.6	.65	-5.2	a .63	.47	80.1	72.5	I0033		
65																
66	MEAN	283.8	559.0	.00	.12	.99	.0	.98	-.2			77.9	78.4			
67	S.D.	139.8	.0	1.69	.03	.15	2.8	.40	3.0			9.5	7.8			
68																

الملحق رقم 5: خصائص فقرات



الملحق 6: خريطة الفقرات مع قدرات التلاميذ

الملحق رقم 7: الاختبار التحصيلي

اختبار تجريبي في الرياضيات السنة الخامسة ابتدائي

1. التمرين الأول: أكتب الأعداد التالية بالحروف

1097 :.....الفقرة 1

2083625 :.....الفقرة 2

أكتب بالأرقام الأعداد التالية:

سبعة آلاف وثلاثمائة وعشرون:.....الفقرة 3

ثلاثة ملايين وخمسمائة وستون ألفا ومائة وواحد وعشرون:.....الفقرة 4

2. التمرين الثاني: أنجز عموديا العمليات التالية:

$$..... = 789 + 1724 \text{ (ف5)}$$

$$..... = 6175 - 8049 \text{ (ف6)}$$

$$..... = 24 \times 389 \text{ (ف7)}$$

$$..... = 5 \div 9840 \text{ (ف8)}$$

--	--	--	--

3. التمرين الثالث: قارن بين الأعداد التالية (> ، < ، =) : (ف9) 3.10 3.05 (ف10) 10.50 ... 9.845

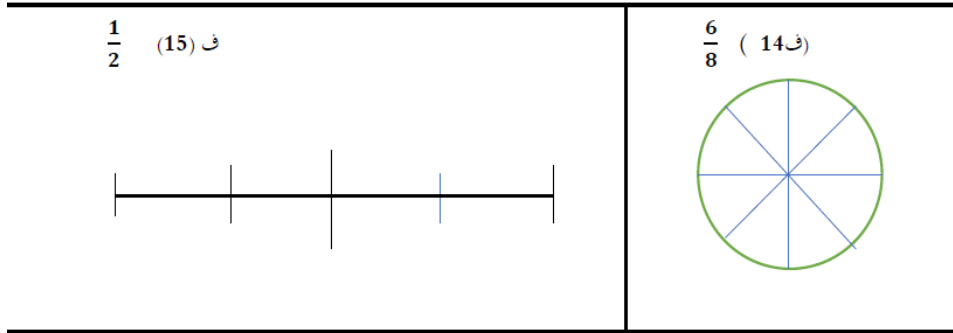
(ف11) ، 14.52 ... 14.520

4. التمرين الرابع: ف12 و ف13 أنجز العمليتين عموديا

$$..... = 16,3 - 170$$

$$6,135 + 110,75$$

التمرين الخامس : ألَوّن الأجزاء لأحصل على الكسور المكتوبة:



التمرين السادس : حدّد العدد بوضع دائرة حول: خمسة عشرة و ثلاثة أجزاء من المائة ، من مجموعات الأعداد التالية:

المجموعة الأولى: 15,003 15,03 15,30 15,3 3,15 (ف16)

المجموعة الثانية: $\frac{1503}{1000}$ $\frac{1503}{100}$ $\frac{153}{10}$ $\frac{153}{100}$ (ف17)

المجموعة الثالثة: $\frac{3}{1000} + 15$ $\frac{3}{10} + 15$ $\frac{3}{100} + 15$ (ف18)

التمرين السابع: أعد كتابة الأعداد العشرية التالية في صورة كسور:

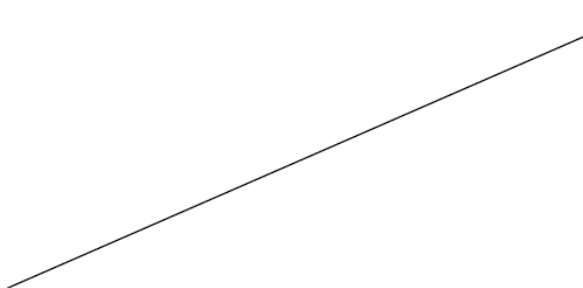
1,01 = --- (ف19) 0,9 = ----- (ف20)

التمرين الثامن: أعد كتابة الكسور التالية في صورة أعداد

$\frac{458}{100}$ = (ف21) $\frac{3689}{1000}$ = (ف22)

التمرين التاسع: أرسم خطاً عمودياً على المستقيم (S)

(ف23)



التمرين العاشر: محيط مربع 400 مترا ، أوجد مساحته . S

- (ف24) : حساب طول الضلع
- (ف25) : حساب المساحة
- (ف26) : الحل الكلي

8. التمرين الحادي عشر: أرسم خطا عموديا على المستقيم (D)، ويشمل النقطة a .

(ف27) يمر الخط بالنقطة .

(ف28) الخط عمودي

(ف29) الرسم الكلي

a .

(D)

12. التمرين الثاني عشر: أرسم (داخل الإطار) مستطيلا طوله 5 سم و عرضه نصف طوله. ثم أوجد محيطه.

(ف30) رسم طول المستقيم

(ف31) رسم العرض

(ف32) حساب المحيط

(ف33) الحل الكلي

13. التمرين الثالث عشر: حدد بين كل عددين عشرين الأعداد التالية:

(ف34) $2.1 < \dots$ (ف35) $31.42 < \dots$

14. التمرين الرابع عشر: أكمل ما يلي : $12.1t = \dots q$ (ف36) $3.5 \text{ km} = \dots \text{ m}$ (ف37)

$6 \text{ h } 30 \text{ mn} = \dots \text{ mn}$ (ف38) $4680 \text{ L} = \dots \text{ hl}$ (ف39)

14. التمرين الرابع عشر:

يزن عمر 650g 58 kg ويزن علي 300 g 61 kg ، قارن بينهما ، وأيها أثقل؟ (ف40)

19. التمرين التاسع عشر:

يوجد حقلان لهما نفس المحيط ، أحدهما مستطيل طوله 20m وعرضه 16 m ، والآخر مربع ، ما هو طول ضلع هذا المربع؟

محيط المستطيل (ف47)

ضلع المربع..... (ف48)..

حل التمرين كاملا (ف49).

21. التمرين العشرون:

عين على المدرج التالي الأعداد التالية: 10.3 (50ف) 9.47 (ف51)

11	10.5	10	9.5	9

23. التمرين الواحد والعشرون:

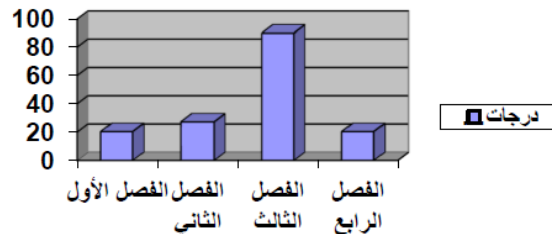
أراد عمر شراء حذاء ثمنه 2580 د.ج ، فوجد أنّ هناك تخفيضا قدره 20 % بمناسبة نهاية السنة ، وتخفيضا آخر قدره 5 % لأنّ الحذاء من النوع الخاص بالأطفال.

إطرح سؤالا واحدا تراه مناسبا ، وأجب عنه.

السؤال: تطبيق النسبة ؟ (ف52)

الحل كامل (ف53).....

24. التمرين الثاني والعشرون: لاحظ وأجب.



لاحظ جيّدا ثم أكمل : مجموع النقاط التي حصل عليها عمر هو..... (ف54)

الفصل الذي حصل فيه على أعلى نقطة هو: (ف55)

الملحق رقم 8: ترتيب عدم ملائمة التلاميذ

```

1 TABLE 6.1 C:\Winsteps\doct25.txt ZOU767WS.TXT May 16 15:13 2026
2 INPUT: 559 Person 55 Item REPORTED: 559 Person 55 Item 2 CATS WINSTEPS 3.72.3
3 -----
4 Person: REAL SEP.: 3.17 REL.: .91 ... Item: REAL SEP.: 13.33 REL.: .99
5
6 Person STATISTICS: MISFIT ORDER
7
8 -----
9 -----
9 |ENTRY TOTAL TOTAL MODEL| INFIT | OUTFIT |PT-MEASURE |EXACT MATCH| |
10 |NUMBER SCORE COUNT MEASURE S.E. |MNSQ ZSTD|MNSQ ZSTD|CORR. EXP.| OBS% EXP%| Person|
11 |-----+-----+-----+-----+-----+-----+
12 | 155 49 55 -2.90 .47|1.21 .8|3.55 2.1|A .11 .31| 89.1 89.1| 1s |
13 | 550 49 55 -2.90 .47|1.21 .8|3.55 2.1|B .11 .31| 89.1 89.1| 1s |
14 | 287 44 55 -2.03 .38|1.38 1.8|3.46 2.3|C .07 .40| 83.6 82.3| 0a |
15 | 398 35 55 -.94 .33|1.15 1.1|3.11 3.2|D .36 .50| 70.9 74.2| 0s |
16 | 112 34 55 -.83 .33|1.26 1.8|2.94 3.2|E .32 .51| 67.3 73.9| 1a |
17 | 346 34 55 -.83 .33|1.26 1.8|2.94 3.2|F .32 .51| 67.3 73.9| 1s |
18 | 530 34 55 -.83 .33|1.26 1.8|2.94 3.2|G .32 .51| 67.3 73.9| 0a |
19 | 199 51 55 -3.41 .55|1.25 .7|2.72 1.6|H .06 .26| 92.7 92.7| 0a |
20 | 76 43 55 -1.89 .37|1.50 2.4|2.65 1.9|I .12 .42| 70.9 81.0| 0a |
21 | 78 32 55 -.61 .33|1.38 2.6|2.58 3.0|J .30 .53| 60.0 73.7| 0a |
22 | 97 34 55 -.83 .33|1.00 .1|2.32 2.4|K .47 .51| 81.8 73.9| 0s |
23 | 231 34 55 -.83 .33|1.00 .1|2.32 2.4|L .47 .51| 81.8 73.9| 0a |
24 | 302 34 55 -.83 .33|1.00 .1|2.32 2.4|M .47 .51| 81.8 73.9| 0s |
25 | 442 34 55 -.83 .33|1.00 .1|2.32 2.4|N .47 .51| 81.8 73.9| 1a |
26 | 446 34 55 -.83 .33|1.00 .1|2.32 2.4|O .47 .51| 81.8 73.9| 1a |
27 | 134 48 55 -2.70 .44|1.48 1.6|2.30 1.4|P .05 .33| 81.8 87.5| 1a |
28 | 338 48 55 -2.70 .44|1.48 1.6|2.30 1.4|Q .05 .33| 81.8 87.5| 1s |
29 | 343 48 55 -2.70 .44|1.48 1.6|2.30 1.4|R .05 .33| 81.8 87.5| 1s |
30 | 1 50 55 -3.14 .50| .79 -.5|2.19 1.3|S .32 .29| 90.9 90.9| 1a |
31 | 102 32 55 -.61 .33|1.06 .5|2.12 2.3|T .46 .53| 78.2 73.7| 1s |
32 | 219 32 55 -.61 .33|1.06 .5|2.12 2.3|U .46 .53| 78.2 73.7| 1s |
33 | 543 32 55 -.61 .33|1.06 .5|2.12 2.3|V .46 .53| 78.2 73.7| 1a |
34 | 106 7 55 2.87 .49|1.67 1.9|2.07 1.2|W .26 .51| 81.8 90.2| 1s |
35 | 202 7 55 2.87 .49|1.67 1.9|2.07 1.2|X .26 .51| 81.8 90.2| 1a |
36 | 393 7 55 2.87 .49|1.67 1.9|2.07 1.2|Y .26 .51| 81.8 90.2| 0s |
37 | 26 43 55 -1.89 .37|1.32 1.6|2.01 1.3|Z .22 .42| 70.9 81.0| 0s |
38 | 260 43 55 -1.89 .37|1.32 1.6|2.01 1.3| .22 .42| 70.9 81.0| 1a |
39 | 481 43 55 -1.89 .37|1.32 1.6|2.01 1.3| .22 .42| 70.9 81.0| 0a |
40 | 2 42 55 -1.76 .36|1.25 1.4|1.94 1.3| .24 .43| 83.6 79.7| 1s |
41 | 263 42 55 -1.76 .36|1.25 1.4|1.94 1.3| .24 .43| 83.6 79.7| 1a |
42 | 383 42 55 -1.76 .36|1.25 1.4|1.94 1.3| .24 .43| 83.6 79.7| 0a |
43 | 396 34 55 -.83 .33|1.72 4.4|1.84 1.7| .17 .51| 41.8 73.9| 0s |
44 | 273 45 55 -2.18 .39|1.47 2.0|1.77 1.1| .14 .39| 74.5 83.7| 0s |
45 | 556 45 55 -2.18 .39|1.47 2.0|1.77 1.1| .14 .39| 74.5 83.7| 1s |
46 | 130 32 55 -.61 .33|1.27 1.9|1.76 1.7| .36 .53| 63.6 73.7| 0s |
47 | 408 32 55 -.61 .33|1.27 1.9|1.76 1.7| .36 .53| 63.6 73.7| 1a |
48 | 309 43 55 -1.89 .37| .98 .0|1.73 1.1| .36 .42| 85.5 81.0| 1s |
49 | 390 51 55 -3.41 .55|1.18 .6|1.73 1.0| .12 .26| 92.7 92.7| 0s |
50 | 501 44 55 -2.03 .38|1.47 2.1|1.72 1.0| .16 .40| 76.4 82.3| 0a |
51 | 197 21 55 .60 .35|1.71 3.3|1.48 1.4| .31 .59| 56.4 78.7| 0a |
52 | 8 19 55 .85 .36|1.51 2.4|1.68 1.8| .36 .59| 65.5 80.1| 1s |
53 | 116 19 55 .85 .36|1.51 2.4|1.68 1.8| .36 .59| 65.5 80.1| 1a |
54 | 204 19 55 .85 .36|1.51 2.4|1.68 1.8| .36 .59| 65.5 80.1| 1a |
55 | 367 19 55 .85 .36|1.51 2.4|1.68 1.8| .36 .59| 65.5 80.1| 1a |
56 | 379 19 55 .85 .36|1.51 2.4|1.68 1.8| .36 .59| 65.5 80.1| 0a |
57 | 440 19 55 .85 .36|1.51 2.4|1.68 1.8| .36 .59| 65.5 80.1| 1a |
58 | 463 19 55 .85 .36|1.51 2.4|1.68 1.8| .36 .59| 65.5 80.1| 1s |
59 | 466 19 55 .85 .36|1.51 2.4|1.68 1.8| .36 .59| 65.5 80.1| 1s |
60 | 522 19 55 .85 .36|1.51 2.4|1.68 1.8| .36 .59| 65.5 80.1| 1s |
61 | 535 19 55 .85 .36|1.51 2.4|1.68 1.8| .36 .59| 65.5 80.1| 0a |

```

62 | 10 32 55 -.61 .33|1.50 3.2|1.68 1.6| .28 .53| 60.0 73.7| 0a |
63 | 117 32 55 -.61 .33|1.50 3.2|1.68 1.6| .28 .53| 60.0 73.7| 1a |
64 | 205 32 55 -.61 .33|1.50 3.2|1.68 1.6| .28 .53| 60.0 73.7| 1a |
65 | 315 32 55 -.61 .33|1.50 3.2|1.68 1.6| .28 .53| 60.0 73.7| 0s |
66 | 441 32 55 -.61 .33|1.50 3.2|1.68 1.6| .28 .53| 60.0 73.7| 1a |
67 | 64 32 55 -.61 .33|1.42 2.8|1.66 1.5| .30 .53| 63.6 73.7| 1s |
68 | 502 32 55 -.61 .33|1.42 2.8|1.66 1.5| .30 .53| 63.6 73.7| 0a |
69 | 128 35 55 -.94 .33|1.16 1.1|1.56 1.2| .39 .50| 78.2 74.2| 0s |
70 | 62 51 55 -3.41 .55|1.17 .5|1.52 .8| .14 .26| 92.7 92.7| 1s |
71 | 270 51 55 -3.41 .55|1.17 .5|1.52 .8| .14 .26| 92.7 92.7| 0s |
72 | 499 51 55 -3.41 .55|1.17 .5|1.52 .8| .14 .26| 92.7 92.7| 0a |
73 | 374 42 55 -1.76 .36|1.17 1.0|1.52 .9| .32 .43| 76.4 79.7| 0s |
74 | 215 48 55 -2.70 .44|1.19 .7|1.50 .8| .22 .33| 81.8 87.5| 1s |
75 | 24 42 55 -1.76 .36|1.47 2.4|1.49 .9| .20 .43| 72.7 79.7| 0s |
76 | 504 44 55 -2.03 .38| .89 -.5|1.49 .8| .40 .40| 87.3 82.3| 0a |
77 | 53 16 55 1.25 .38|1.33 1.5|1.48 1.2| .44 .59| 74.5 82.1| 1a |
78 | 60 33 55 -.72 .33|1.26 1.8|1.48 1.2| .37 .52| 74.5 73.7| 1s |
79 | 268 33 55 -.72 .33|1.26 1.8|1.48 1.2| .37 .52| 74.5 73.7| 0s |
80 | 498 33 55 -.72 .33|1.26 1.8|1.48 1.2| .37 .52| 74.5 73.7| 0a |
81 | 220 39 55 -1.39 .34| .94 -.3|1.43 .9| .45 .46| 80.0 76.8| 1s |
82 | 391 39 55 -1.39 .34| .94 -.3|1.43 .9| .45 .46| 80.0 76.8| 0s |
83 | 37 10 55 2.23 .44|1.43 1.5|1.06 .3| .42 .56| 81.8 87.3| 1s |
84 | 100 10 55 2.23 .44|1.43 1.5|1.06 .3| .42 .56| 81.8 87.3| 0s |
85 | 120 10 55 2.23 .44|1.43 1.5|1.06 .3| .42 .56| 81.8 87.3| 1a |
86 | 132 10 55 2.23 .44|1.43 1.5|1.06 .3| .42 .56| 81.8 87.3| 0a |
87 | 236 10 55 2.23 .44|1.43 1.5|1.06 .3| .42 .56| 81.8 87.3| 1a |
88 | 246 10 55 2.23 .44|1.43 1.5|1.06 .3| .42 .56| 81.8 87.3| 1s |
89 | 259 10 55 2.23 .44|1.43 1.5|1.06 .3| .42 .56| 81.8 87.3| 1a |
90 | 286 10 55 2.23 .44|1.43 1.5|1.06 .3| .42 .56| 81.8 87.3| 0a |
91 | 356 10 55 2.23 .44|1.43 1.5|1.06 .3| .42 .56| 81.8 87.3| 1a |
92 | 439 10 55 2.23 .44|1.43 1.5|1.06 .3| .42 .56| 81.8 87.3| 1a |
93 | 462 10 55 2.23 .44|1.43 1.5|1.06 .3| .42 .56| 81.8 87.3| 1s |
94 | 480 10 55 2.23 .44|1.43 1.5|1.06 .3| .42 .56| 81.8 87.3| 0a |
95 | 548 10 55 2.23 .44|1.43 1.5|1.06 .3| .42 .56| 81.8 87.3| 1s |
96 | 148 26 55 .03 .33| .87 -.8|1.43 1.3| .57 .57| 89.1 75.6| 1s |
97 | 40 47 55 -2.51 .42|1.31 1.2|1.42 .7| .18 .35| 80.0 86.1| 1a |
98 | 192 33 55 -.72 .33|1.26 1.8|1.40 1.0| .38 .52| 70.9 73.7| 0s |
99 | 339 33 55 -.72 .33|1.26 1.8|1.40 1.0| .38 .52| 70.9 73.7| 1s |
100 | 419 33 55 -.72 .33|1.26 1.8|1.40 1.0| .38 .52| 70.9 73.7| 1a |
101 | 493 33 55 -.72 .33|1.26 1.8|1.40 1.0| .38 .52| 70.9 73.7| 0s |
102 | 558 33 55 -.72 .33|1.26 1.8|1.40 1.0| .38 .52| 70.9 73.7| 1s |
103 | 72 16 55 1.25 .38|1.39 1.7|1.28 .8| .44 .59| 70.9 82.1| 0a |
104 | 79 16 55 1.25 .38|1.39 1.7|1.28 .8| .44 .59| 70.9 82.1| 0a |
105 | 144 16 55 1.25 .38|1.39 1.7|1.28 .8| .44 .59| 70.9 82.1| 1a |
106 | 207 16 55 1.25 .38|1.39 1.7|1.28 .8| .44 .59| 70.9 82.1| 1a |
107 | 212 16 55 1.25 .38|1.39 1.7|1.28 .8| .44 .59| 70.9 82.1| 0s |
108 | 221 16 55 1.25 .38|1.39 1.7|1.28 .8| .44 .59| 70.9 82.1| 1s |
109 | 372 16 55 1.25 .38|1.39 1.7|1.28 .8| .44 .59| 70.9 82.1| 0s |
110 | 421 16 55 1.25 .38|1.39 1.7|1.28 .8| .44 .59| 70.9 82.1| 1a |
111 | 74 28 55 -.19 .33|1.36 2.3|1.39 1.2| .39 .56| 65.5 74.7| 0a |
112 | 200 28 55 -.19 .33|1.36 2.3|1.39 1.2| .39 .56| 65.5 74.7| 0a |
113 | 368 28 55 -.19 .33|1.36 2.3|1.39 1.2| .39 .56| 65.5 74.7| 0s |
114 | 381 28 55 -.19 .33|1.36 2.3|1.39 1.2| .39 .56| 65.5 74.7| 0a |
115 | 467 28 55 -.19 .33|1.36 2.3|1.39 1.2| .39 .56| 65.5 74.7| 1s |
116 | 505 28 55 -.19 .33|1.36 2.3|1.39 1.2| .39 .56| 65.5 74.7| 1a |
117 | 523 28 55 -.19 .33|1.36 2.3|1.39 1.2| .39 .56| 65.5 74.7| 1s |
118 | 546 28 55 -.19 .33|1.36 2.3|1.39 1.2| .39 .56| 65.5 74.7| 1s |
119 | 122 43 55 -1.89 .37|1.28 1.5|1.38 .7| .27 .42| 74.5 81.0| 0a |
120 | 4 38 55 -1.27 .34|1.31 2.0|1.36 .8| .32 .47| 67.3 75.9| 1s |
121 | 417 38 55 -1.27 .34|1.31 2.0|1.36 .8| .32 .47| 67.3 75.9| 1a |
122 | 101 31 55 -.51 .33|1.25 1.7|1.33 .9| .41 .54| 65.5 73.9| 1s |
123 | 232 31 55 -.51 .33|1.25 1.7|1.33 .9| .41 .54| 65.5 73.9| 0a |
124 | 427 31 55 -.51 .33|1.25 1.7|1.33 .9| .41 .54| 65.5 73.9| 0s |
125 | 447 31 55 -.51 .33|1.25 1.7|1.33 .9| .41 .54| 65.5 73.9| 1a |
126 | 65 48 55 -2.70 .44|1.33 1.2|1.01 .3| .20 .33| 81.8 87.5| 0s |
127 | 164 29 55 -.29 .33|1.08 .6|1.32 1.0| .48 .55| 74.5 74.4| 1a |
128 | 206 25 55 .14 .33|1.32 1.9|1.31 1.0| .43 .57| 65.5 76.1| 1a |

129 | 87 12 55 1.86 .41|1.28 1.1| .90 .0| .50 .58| 78.2 85.3| 0s |
130 | 108 12 55 1.86 .41|1.28 1.1| .90 .0| .50 .58| 78.2 85.3| 1a |
131 | 196 12 55 1.86 .41|1.28 1.1| .90 .0| .50 .58| 78.2 85.3| 0a |
132 | 234 12 55 1.86 .41|1.28 1.1| .90 .0| .50 .58| 78.2 85.3| 1a |
133 | 239 12 55 1.86 .41|1.28 1.1| .90 .0| .50 .58| 78.2 85.3| 1a |
134 | 336 12 55 1.86 .41|1.28 1.1| .90 .0| .50 .58| 78.2 85.3| 1s |
135 | 401 12 55 1.86 .41|1.28 1.1| .90 .0| .50 .58| 78.2 85.3| 0s |
136 | 403 12 55 1.86 .41|1.28 1.1| .90 .0| .50 .58| 78.2 85.3| 0s |
137 | 430 12 55 1.86 .41|1.28 1.1| .90 .0| .50 .58| 78.2 85.3| 0s |
138 | 451 12 55 1.86 .41|1.28 1.1| .90 .0| .50 .58| 78.2 85.3| 1a |
139 | 456 12 55 1.86 .41|1.28 1.1| .90 .0| .50 .58| 78.2 85.3| 1s |
140 | 348 34 55 -.83 .33|1.27 1.9|1.12 .4| .40 .51| 56.4 73.9| 1s |
141 | 20 24 55 .25 .34|1.25 1.5|1.27 .9| .47 .58| 67.3 76.6| 0a |
142 | 168 24 55 .25 .34|1.25 1.5|1.27 .9| .47 .58| 67.3 76.6| 0a |
143 | 271 24 55 .25 .34|1.25 1.5|1.27 .9| .47 .58| 67.3 76.6| 0s |
144 | 333 24 55 .25 .34|1.25 1.5|1.27 .9| .47 .58| 67.3 76.6| 0s |
145 | 554 24 55 .25 .34|1.25 1.5|1.27 .9| .47 .58| 67.3 76.6| 1s |
146 | 121 46 55 -2.34 .40|1.26 1.1|1.13 .4| .25 .37| 80.0 84.9| 0a |
147 | 229 46 55 -2.34 .40|1.26 1.1|1.13 .4| .25 .37| 80.0 84.9| 0s |
148 | 394 46 55 -2.34 .40|1.26 1.1|1.13 .4| .25 .37| 80.0 84.9| 0s |
149 | 49 39 55 -1.39 .34|1.25 1.6|1.10 .4| .35 .46| 72.7 76.8| 1a |
150 | 494 39 55 -1.39 .34|1.25 1.6|1.10 .4| .35 .46| 72.7 76.8| 0s |
151 | 19 25 55 .14 .33|1.25 1.5|1.21 .8| .47 .57| 61.8 76.1| 1a |
152 | 125 25 55 .14 .33|1.25 1.5|1.21 .8| .47 .57| 61.8 76.1| 0s |
153 | 515 25 55 .14 .33|1.25 1.5|1.21 .8| .47 .57| 61.8 76.1| 0s |
154 | 317 43 55 -1.89 .37|1.24 1.3|1.02 .3| .31 .42| 74.5 81.0| 1a |
155 | 253 42 55 -1.76 .36|1.13 .8|1.24 .6| .36 .43| 80.0 79.7| 1a |
156 | 266 42 55 -1.76 .36|1.13 .8|1.24 .6| .36 .43| 80.0 79.7| 1a |
157 | 455 39 55 -1.39 .34|1.24 1.5| .94 .1| .38 .46| 61.8 76.8| 1s |
158 | 13 26 55 .03 .33|1.24 1.5|1.15 .6| .47 .57| 70.9 75.6| 1a |
159 | 181 26 55 .03 .33|1.24 1.5|1.15 .6| .47 .57| 70.9 75.6| 0s |
160 | 511 26 55 .03 .33|1.24 1.5|1.15 .6| .47 .57| 70.9 75.6| 1s |
161 | 326 36 55 -1.05 .33|1.23 1.6|1.16 .5| .39 .49| 61.8 74.6| 0a |
162 | 33 46 55 -2.34 .40|1.23 1.0| .92 .2| .29 .37| 76.4 84.9| 1s |
163 | 213 33 55 -.72 .33|1.14 1.0|1.22 .6| .44 .52| 74.5 73.7| 1s |
164 | 28 32 55 -.61 .33|1.22 1.6|1.13 .5| .43 .53| 63.6 73.7| 0s |
165 | 483 32 55 -.61 .33|1.22 1.6|1.13 .5| .43 .53| 63.6 73.7| 0a |
166 | 25 39 55 -1.39 .34|1.18 1.2|1.22 .6| .37 .46| 69.1 76.8| 0s |
167 | 477 39 55 -1.39 .34|1.18 1.2|1.22 .6| .37 .46| 69.1 76.8| 0a |
168 | 240 43 55 -1.89 .37|1.21 1.1|1.14 .4| .33 .42| 74.5 81.0| 1a |
169 | 330 20 55 .72 .35|1.21 1.1|1.06 .3| .52 .59| 70.9 79.4| 0s |
170 | 50 25 55 .14 .33|1.20 1.3|1.14 .5| .49 .57| 65.5 76.1| 1a |
171 | 81 25 55 .14 .33|1.20 1.3|1.14 .5| .49 .57| 65.5 76.1| 0a |
172 | 189 25 55 .14 .33|1.20 1.3|1.14 .5| .49 .57| 65.5 76.1| 0s |
173 | 256 25 55 .14 .33|1.20 1.3|1.14 .5| .49 .57| 65.5 76.1| 1a |
174 | 406 25 55 .14 .33|1.20 1.3|1.14 .5| .49 .57| 65.5 76.1| 1a |
175 | 409 25 55 .14 .33|1.20 1.3|1.14 .5| .49 .57| 65.5 76.1| 1a |
176 | 360 46 55 -2.34 .40|1.12 .6| .71 -.1| .35 .37| 80.0 84.9| 1a |
177 | 93 47 55 -2.51 .42|1.09 .4| .72 -.1| .34 .35| 80.0 86.1| 0s |
178 | 226 53 55 -4.21 .74|1.04 .3| .46 -.2| .20 .19| 96.4 96.4| 0s |
179 | 424 53 55 -4.21 .74|1.04 .3| .46 -.2| .20 .19| 96.4 96.4| 0s |
180 | 63 41 55 -1.63 .35|1.03 .2| .75 -.2| .45 .44| 76.4 78.7| 1s |
181 | 225 41 55 -1.63 .35|1.03 .2| .75 -.2| .45 .44| 76.4 78.7| 0s |
182 | 255 41 55 -1.63 .35|1.03 .2| .75 -.2| .45 .44| 76.4 78.7| 1a |
183 | 332 41 55 -1.63 .35|1.03 .2| .75 -.2| .45 .44| 76.4 78.7| 0s |
184 | 352 41 55 -1.63 .35|1.03 .2| .75 -.2| .45 .44| 76.4 78.7| 1a |
185 | 363 41 55 -1.63 .35|1.03 .2| .75 -.2| .45 .44| 76.4 78.7| 1a |
186 | 478 41 55 -1.63 .35|1.03 .2| .75 -.2| .45 .44| 76.4 78.7| 0a |
187 | 516 41 55 -1.63 .35|1.03 .2| .75 -.2| .45 .44| 76.4 78.7| 1s |
188 | 386 35 55 -.94 .33|1.01 .1| .76 -.4| .52 .50| 67.3 74.2| 0a |
189 | BETTER FITTING OMITTED +-----+-----+ | | |
190 | 57 45 55 -2.18 .39| .98 .0| .76 -.1| .40 .39| 89.1 83.7| 1s |
191 | 267 45 55 -2.18 .39| .98 .0| .76 -.1| .40 .39| 89.1 83.7| 0s |
192 | 365 40 55 -1.51 .35| .98 -.1| .72 -.3| .48 .45| 74.5 77.8| 1a |
193 | 545 40 55 -1.51 .35| .98 -.1| .72 -.3| .48 .45| 74.5 77.8| 1s |
194 | 39 18 55 .98 .36| .97 -.1| .76 -.6| .63 .59| 81.8 80.6| 1a |
195 | 489 18 55 .98 .36| .97 -.1| .76 -.6| .63 .59| 81.8 80.6| 0a |

196 | 68 29 55 -.29 .33| .96 -.2| .78 -.6| .58 .55| 74.5 74.4| 0s |
197 | 126 29 55 -.29 .33| .96 -.2| .78 -.6| .58 .55| 74.5 74.4| 0s |
198 | 174 29 55 -.29 .33| .96 -.2| .78 -.6| .58 .55| 74.5 74.4| 0a |
199 | 218 29 55 -.29 .33| .96 -.2| .78 -.6| .58 .55| 74.5 74.4| 1s |
200 | 476 29 55 -.29 .33| .96 -.2| .78 -.6| .58 .55| 74.5 74.4| 0a |
201 | 341 31 55 -.51 .33| .96 -.3| .78 -.5| .57 .54| 69.1 73.9| 1s |
202 | 66 53 55 -4.21 .74| .94 .1| .32 -.4| .25 .19| 96.4 96.4| 0s |
203 | 135 17 55 1.11 .37| .94 -.2| .70 -.8| .64 .60| 78.2 81.4| 1a |
204 | 412 17 55 1.11 .37| .94 -.2| .70 -.8| .64 .60| 78.2 81.4| 1a |
205 | 7 18 55 .98 .36| .94 -.3| .75 -.6| .64 .59| 81.8 80.6| 1s |
206 | 70 18 55 .98 .36| .94 -.3| .75 -.6| .64 .59| 81.8 80.6| 0s |
207 | 95 18 55 .98 .36| .94 -.3| .75 -.6| .64 .59| 81.8 80.6| 0s |
208 | 115 18 55 .98 .36| .94 -.3| .75 -.6| .64 .59| 81.8 80.6| 1a |
209 | 131 18 55 .98 .36| .94 -.3| .75 -.6| .64 .59| 81.8 80.6| 0a |
210 | 176 18 55 .98 .36| .94 -.3| .75 -.6| .64 .59| 81.8 80.6| 0a |
211 | 188 18 55 .98 .36| .94 -.3| .75 -.6| .64 .59| 81.8 80.6| 0s |
212 | 211 18 55 .98 .36| .94 -.3| .75 -.6| .64 .59| 81.8 80.6| 1s |
213 | 249 18 55 .98 .36| .94 -.3| .75 -.6| .64 .59| 81.8 80.6| 1s |
214 | 297 18 55 .98 .36| .94 -.3| .75 -.6| .64 .59| 81.8 80.6| 0a |
215 | 300 18 55 .98 .36| .94 -.3| .75 -.6| .64 .59| 81.8 80.6| 0a |
216 | 307 18 55 .98 .36| .94 -.3| .75 -.6| .64 .59| 81.8 80.6| 1s |
217 | 389 18 55 .98 .36| .94 -.3| .75 -.6| .64 .59| 81.8 80.6| 0s |
218 | 400 18 55 .98 .36| .94 -.3| .75 -.6| .64 .59| 81.8 80.6| 0s |
219 | 532 18 55 .98 .36| .94 -.3| .75 -.6| .64 .59| 81.8 80.6| 0a |
220 | 185 20 55 .72 .35| .93 -.3| .74 -.7| .64 .59| 78.2 79.4| 0s |
221 | 293 20 55 .72 .35| .93 -.3| .74 -.7| .64 .59| 78.2 79.4| 0a |
222 | 304 20 55 .72 .35| .93 -.3| .74 -.7| .64 .59| 78.2 79.4| 1s |
223 | 376 20 55 .72 .35| .93 -.3| .74 -.7| .64 .59| 78.2 79.4| 0a |
224 | 460 20 55 .72 .35| .93 -.3| .74 -.7| .64 .59| 78.2 79.4| 1s |
225 | 519 20 55 .72 .35| .93 -.3| .74 -.7| .64 .59| 78.2 79.4| 1s |
226 | 357 32 55 -.61 .33| .93 -.5| .77 -.5| .57 .53| 78.2 73.7| 1a |
227 | 402 37 55 -1.16 .34| .93 -.5| .70 -.5| .54 .48| 74.5 75.2| 0s |
228 | 85 37 55 -1.16 .34| .92 -.5| .69 -.5| .54 .48| 70.9 75.2| 0a |
229 | 158 37 55 -1.16 .34| .92 -.5| .69 -.5| .54 .48| 70.9 75.2| 1s |
230 | 27 34 55 -.83 .33| .92 -.6| .71 -.6| .57 .51| 74.5 73.9| 0s |
231 | 80 34 55 -.83 .33| .92 -.6| .71 -.6| .57 .51| 74.5 73.9| 0a |
232 | 280 34 55 -.83 .33| .92 -.6| .71 -.6| .57 .51| 74.5 73.9| 0s |
233 | 482 34 55 -.83 .33| .92 -.6| .71 -.6| .57 .51| 74.5 73.9| 0a |
234 | 313 23 55 .37 .34| .92 -.5| .75 -.8| .63 .58| 76.4 77.3| 0s |
235 | 527 23 55 .37 .34| .92 -.5| .75 -.8| .63 .58| 76.4 77.3| 0a |
236 | 167 32 55 -.61 .33| .90 -.7| .73 -.6| .59 .53| 74.5 73.7| 0a |
237 | 187 32 55 -.61 .33| .90 -.7| .73 -.6| .59 .53| 74.5 73.7| 0s |
238 | 329 32 55 -.61 .33| .90 -.7| .73 -.6| .59 .53| 74.5 73.7| 0a |
239 | 469 32 55 -.61 .33| .90 -.7| .73 -.6| .59 .53| 74.5 73.7| 0s |
240 | 3 35 55 -.94 .33| .90 -.7| .68 -.6| .57 .50| 78.2 74.2| 1s |
241 | 264 35 55 -.94 .33| .90 -.7| .68 -.6| .57 .50| 78.2 74.2| 1a |
242 | 385 35 55 -.94 .33| .90 -.7| .68 -.6| .57 .50| 78.2 74.2| 0a |
243 | 238 45 55 -2.18 .39| .88 -.5| .77 -.1| .44 .39| 85.5 83.7| 1a |
244 | 415 43 55 -1.89 .37| .87 -.6| .64 -.3| .48 .42| 85.5 81.0| 1a |
245 | 11 31 55 -.51 .33| .87 -.9| .69 -.8| .61 .54| 72.7 73.9| 1a |
246 | 58 31 55 -.51 .33| .87 -.9| .69 -.8| .61 .54| 72.7 73.9| 1s |
247 | 118 31 55 -.51 .33| .87 -.9| .69 -.8| .61 .54| 72.7 73.9| 1a |
248 | 161 31 55 -.51 .33| .87 -.9| .69 -.8| .61 .54| 72.7 73.9| 1a |
249 | 173 31 55 -.51 .33| .87 -.9| .69 -.8| .61 .54| 72.7 73.9| 0a |
250 | 235 31 55 -.51 .33| .87 -.9| .69 -.8| .61 .54| 72.7 73.9| 1a |
251 | 250 31 55 -.51 .33| .87 -.9| .69 -.8| .61 .54| 72.7 73.9| 1s |
252 | 262 31 55 -.51 .33| .87 -.9| .69 -.8| .61 .54| 72.7 73.9| 1a |
253 | 318 31 55 -.51 .33| .87 -.9| .69 -.8| .61 .54| 72.7 73.9| 1a |
254 | 392 31 55 -.51 .33| .87 -.9| .69 -.8| .61 .54| 72.7 73.9| 0s |
255 | 431 31 55 -.51 .33| .87 -.9| .69 -.8| .61 .54| 72.7 73.9| 0s |
256 | 438 31 55 -.51 .33| .87 -.9| .69 -.8| .61 .54| 72.7 73.9| 1a |
257 | 454 31 55 -.51 .33| .87 -.9| .69 -.8| .61 .54| 72.7 73.9| 1s |
258 | 22 47 55 -2.51 .42| .87 -.5| .64 -.2| .42 .35| 90.9 86.1| 0a |
259 | 119 38 55 -1.27 .34| .87 -.9| .68 -.5| .55 .47| 81.8 75.9| 1a |
260 | 162 38 55 -1.27 .34| .87 -.9| .68 -.5| .55 .47| 81.8 75.9| 1a |
261 | 319 38 55 -1.27 .34| .87 -.9| .68 -.5| .55 .47| 81.8 75.9| 1a |
262 | 397 42 55 -1.76 .36| .86 -.8| .64 -.4| .51 .43| 83.6 79.7| 0s |

263		21	30	55	-	.40	.33		.85	-	1.1		.67	-	.9		.63	.54		78.2	74.2		0a	
264		23	30	55	-	.40	.33		.85	-	1.1		.67	-	.9		.63	.54		78.2	74.2		0s	
265		84	30	55	-	.40	.33		.85	-	1.1		.67	-	.9		.63	.54		78.2	74.2		0a	
266		123	30	55	-	.40	.33		.85	-	1.1		.67	-	.9		.63	.54		78.2	74.2		0a	
267		143	30	55	-	.40	.33		.85	-	1.1		.67	-	.9		.63	.54		78.2	74.2		1a	
268		151	30	55	-	.40	.33		.85	-	1.1		.67	-	.9		.63	.54		78.2	74.2		1s	
269		244	30	55	-	.40	.33		.85	-	1.1		.67	-	.9		.63	.54		78.2	74.2		1a	
270		257	30	55	-	.40	.33		.85	-	1.1		.67	-	.9		.63	.54		78.2	74.2		1a	
271		325	30	55	-	.40	.33		.85	-	1.1		.67	-	.9		.63	.54		78.2	74.2		0a	
272		425	30	55	-	.40	.33		.85	-	1.1		.67	-	.9		.63	.54		78.2	74.2		0s	
273		453	30	55	-	.40	.33		.85	-	1.1		.67	-	.9		.63	.54		78.2	74.2		1s	
274		557	30	55	-	.40	.33		.85	-	1.1		.67	-	.9		.63	.54		78.2	74.2		1s	
275		362	33	55	-	.72	.33		.85	-	1.1		.69	-	.7		.60	.52		74.5	73.7		1a	
276		399	39	55	-	1.39	.34		.84	-	1.0		.69	-	.4		.54	.46		87.3	76.8		0s	
277		45	22	55	.	.48	.34		.83	-	1.0		.67	-	1.1		.67	.59		83.6	78.0		1a	
278		358	43	55	-	1.89	.37		.83	-	.9		.58	-	.5		.51	.42		81.8	81.0		1a	
279		91	37	55	-	1.16	.34		.82	-	1.3		.60	-	.8		.58	.48		78.2	75.2		0s	
280		103	37	55	-	1.16	.34		.82	-	1.3		.60	-	.8		.58	.48		78.2	75.2		1s	
281		159	37	55	-	1.16	.34		.82	-	1.3		.60	-	.8		.58	.48		78.2	75.2		1s	
282		222	37	55	-	1.16	.34		.82	-	1.3		.60	-	.8		.58	.48		78.2	75.2		0s	
283		248	37	55	-	1.16	.34		.82	-	1.3		.60	-	.8		.58	.48		78.2	75.2		1s	
284		418	37	55	-	1.16	.34		.82	-	1.3		.60	-	.8		.58	.48		78.2	75.2		1a	
285		420	37	55	-	1.16	.34		.82	-	1.3		.60	-	.8		.58	.48		78.2	75.2		1a	
286		443	37	55	-	1.16	.34		.82	-	1.3		.60	-	.8		.58	.48		78.2	75.2		1a	
287		549	37	55	-	1.16	.34		.82	-	1.3		.60	-	.8		.58	.48		78.2	75.2		1s	
288		245	25	55	.	.14	.33		.81	-	1.2		.64	-	1.2		.67	.57		76.4	76.1		1a	
289		547	25	55	.	.14	.33		.81	-	1.2		.64	-	1.2		.67	.57		76.4	76.1		1s	
290		351	25	55	.	.14	.33		.80	-	1.3		.65	-	1.2		.67	.57		80.0	76.1		1s	
291		15	27	55	-	.08	.33		.80	-	1.4		.63	-	1.2		.67	.56		80.0	75.1		1a	
292		46	27	55	-	.08	.33		.80	-	1.4		.63	-	1.2		.67	.56		80.0	75.1		1a	
293		109	27	55	-	.08	.33		.80	-	1.4		.63	-	1.2		.67	.56		80.0	75.1		1a	
294		140	27	55	-	.08	.33		.80	-	1.4		.63	-	1.2		.67	.56		80.0	75.1		1a	
295		178	27	55	-	.08	.33		.80	-	1.4		.63	-	1.2		.67	.56		80.0	75.1		0a	
296		349	27	55	-	.08	.33		.80	-	1.4		.63	-	1.2		.67	.56		80.0	75.1		1s	
297		354	27	55	-	.08	.33		.80	-	1.4		.63	-	1.2		.67	.56		80.0	75.1		1a	
298		416	27	55	-	.08	.33		.80	-	1.4		.63	-	1.2		.67	.56		80.0	75.1		0a	
299		452	27	55	-	.08	.33		.80	-	1.4		.63	-	1.2		.67	.56		80.0	75.1		1s	
300		538	27	55	-	.08	.33		.80	-	1.4		.63	-	1.2		.67	.56		80.0	75.1		1a	
301		544	27	55	-	.08	.33		.80	-	1.4		.63	-	1.2		.67	.56		80.0	75.1		1a	
302		242	44	55	-	2.03	.38		.79	-	1.1		.55	-	.5		.51	.40		87.3	82.3		1a	
303		434	44	55	-	2.03	.38		.79	-	1.1		.55	-	.5		.51	.40		87.3	82.3		0s	
304		444	44	55	-	2.03	.38		.79	-	1.1		.55	-	.5		.51	.40		87.3	82.3		1a	
305		124	14	55	1.	54	.39		.79	-	.9		.63	-	.7		.68	.59		83.6	83.6		0s	
306		31	16	55	1.	25	.38		.79	-	1.0		.64	-	.9		.69	.59		85.5	82.1		0s	
307		308	16	55	1.	25	.38		.79	-	1.0		.64	-	.9		.69	.59		85.5	82.1		1s	
308		380	16	55	1.	25	.38		.79	-	1.0		.64	-	.9		.69	.59		85.5	82.1		0a	
309		450	16	55	1.	25	.38		.79	-	1.0		.64	-	.9		.69	.59		85.5	82.1		1a	
310		486	16	55	1.	25	.38		.79	-	1.0		.64	-	.9		.69	.59		85.5	82.1		0a	
311		536	16	55	1.	25	.38		.79	-	1.0		.64	-	.9		.69	.59		85.5	82.1		0a	
312		34	13	55	1.	70	.40		.79	-	.9		.76	-	.3		.67	.59		87.3	84.4		1s	
313		133	13	55	1.	70	.40		.79	-	.9		.76	-	.3		.67	.59		87.3	84.4		1a	
314		146	13	55	1.	70	.40		.79	-	.9		.76	-	.3		.67	.59		87.3	84.4		1s	
315		296	13	55	1.	70	.40		.79	-	.9		.76	-	.3		.67	.59		87.3	84.4		0a	
316		370	13	55	1.	70	.40		.79	-	.9		.76	-	.3		.67	.59		87.3	84.4		0s	
317		428	13	55	1.	70	.40		.79	-	.9		.76	-	.3		.67	.59		87.3	84.4		0s	
318		449	13	55	1.	70	.40		.79	-	.9		.76	-	.3		.67	.59		87.3	84.4		1a	
319		472	13	55	1.	70	.40		.79	-	.9		.76	-	.3		.67	.59		87.3	84.4		0s	
320		99	14	55	1.	54	.39		.79	-	.9		.56	-	1.0		.69	.59		87.3	83.6		0s	
321		111	14	55	1.	54	.39		.79	-	.9		.56	-	1.0		.69	.59		87.3	83.6		0a	
322		194	14	55	1.	54	.39		.79	-	.9		.56	-	1.0		.69	.59		87.3	83.6		0a	
323		216	14	55	1.	54	.39		.79	-	.9		.56	-	1.0		.69	.59		87.3	83.6		1s	
324		258	14	55	1.	54	.39		.79	-	.9		.56	-	1.0		.69	.59		87.3	83.6		1a	
325		261	14	55	1.	54	.39		.79	-	.9		.56	-	1.0		.69	.59		87.3	83.6		1a	
326		289	14	55	1.	54	.39		.79	-	.9		.56	-	1.0		.69	.59		87.3	83.6		0a	
327		345	14	55	1.	54	.39		.79	-	.9		.56	-	1.0		.69	.59		87.3	83.6		1s	
328		479	14	55	1.	54	.39		.79	-	.9		.56	-	1.0		.69	.59		87.3	83.6		0a	
329		529	14	55	1.	54	.39		.79	-	.9		.56	-	1.0		.69	.59		87.3	83.6		0a	

330 | 541 14 55 1.54 .39| .79 -.9| .56 -1.0| .69 .59| 87.3 83.6| 1a |
331 | 217 26 55 .03 .33| .78 -1.5| .62 -1.3| .68 .57| 78.2 75.6| 1s |
332 | 542 26 55 .03 .33| .78 -1.5| .62 -1.3| .68 .57| 78.2 75.6| 1a |
333 | 186 29 55 -.29 .33| .78 -1.6| .65 -1.1| .66 .55| 85.5 74.4| 0s |
334 | 294 29 55 -.29 .33| .78 -1.6| .65 -1.1| .66 .55| 85.5 74.4| 0a |
335 | 377 29 55 -.29 .33| .78 -1.6| .65 -1.1| .66 .55| 85.5 74.4| 0a |
336 | 461 29 55 -.29 .33| .78 -1.6| .65 -1.1| .66 .55| 85.5 74.4| 1s |
337 | 520 29 55 -.29 .33| .78 -1.6| .65 -1.1| .66 .55| 85.5 74.4| 1s |
338 | 35 22 55 .48 .34| .66 -2.2| .77 -.7| .72 .59| 87.3 78.0| 1s |
339 | 61 22 55 .48 .34| .66 -2.2| .77 -.7| .72 .59| 87.3 78.0| 1s |
340 | 96 22 55 .48 .34| .66 -2.2| .77 -.7| .72 .59| 87.3 78.0| 0s |
341 | 113 22 55 .48 .34| .66 -2.2| .77 -.7| .72 .59| 87.3 78.0| 0a |
342 | 139 22 55 .48 .34| .66 -2.2| .77 -.7| .72 .59| 87.3 78.0| 1a |
343 | 152 22 55 .48 .34| .66 -2.2| .77 -.7| .72 .59| 87.3 78.0| 1s |
344 | 214 22 55 .48 .34| .66 -2.2| .77 -.7| .72 .59| 87.3 78.0| 0s |
345 | 230 22 55 .48 .34| .66 -2.2| .77 -.7| .72 .59| 87.3 78.0| 0a |
346 | 233 22 55 .48 .34| .66 -2.2| .77 -.7| .72 .59| 87.3 78.0| 0a |
347 | 269 22 55 .48 .34| .66 -2.2| .77 -.7| .72 .59| 87.3 78.0| 0s |
348 | 283 22 55 .48 .34| .66 -2.2| .77 -.7| .72 .59| 87.3 78.0| 0a |
349 | 301 22 55 .48 .34| .66 -2.2| .77 -.7| .72 .59| 87.3 78.0| 0s |
350 | 361 22 55 .48 .34| .66 -2.2| .77 -.7| .72 .59| 87.3 78.0| 1a |
351 | 405 22 55 .48 .34| .66 -2.2| .77 -.7| .72 .59| 87.3 78.0| 1a |
352 | 464 22 55 .48 .34| .66 -2.2| .77 -.7| .72 .59| 87.3 78.0| 1s |
353 | 495 22 55 .48 .34| .66 -2.2| .77 -.7| .72 .59| 87.3 78.0| 0s |
354 | 533 22 55 .48 .34| .66 -2.2| .77 -.7| .72 .59| 87.3 78.0| 0a |
355 | 183 20 55 .72 .35| .77 -1.3| .59 -1.3| .70 .59| 85.5 79.4| 0s |
356 | 290 20 55 .72 .35| .77 -1.3| .59 -1.3| .70 .59| 85.5 79.4| 0a |
357 | 458 20 55 .72 .35| .77 -1.3| .59 -1.3| .70 .59| 85.5 79.4| 1s |
358 | 518 20 55 .72 .35| .77 -1.3| .59 -1.3| .70 .59| 85.5 79.4| 1s |
359 | 71 26 55 .03 .33| .77 -1.6| .58 -1.4| .69 .57| 78.2 75.6| 0s |
360 | 73 26 55 .03 .33| .77 -1.6| .58 -1.4| .69 .57| 78.2 75.6| 0a |
361 | 141 26 55 .03 .33| .77 -1.6| .58 -1.4| .69 .57| 78.2 75.6| 1a |
362 | 149 26 55 .03 .33| .77 -1.6| .58 -1.4| .69 .57| 78.2 75.6| 1s |
363 | 177 26 55 .03 .33| .77 -1.6| .58 -1.4| .69 .57| 78.2 75.6| 0a |
364 | 193 26 55 .03 .33| .77 -1.6| .58 -1.4| .69 .57| 78.2 75.6| 0s |
365 | 350 26 55 .03 .33| .77 -1.6| .58 -1.4| .69 .57| 78.2 75.6| 1s |
366 | 373 26 55 .03 .33| .77 -1.6| .58 -1.4| .69 .57| 78.2 75.6| 0s |
367 | 410 26 55 .03 .33| .77 -1.6| .58 -1.4| .69 .57| 78.2 75.6| 1a |
368 | 448 26 55 .03 .33| .77 -1.6| .58 -1.4| .69 .57| 78.2 75.6| 1a |
369 | 500 26 55 .03 .33| .77 -1.6| .58 -1.4| .69 .57| 78.2 75.6| 0a |
370 | 503 26 55 .03 .33| .77 -1.6| .58 -1.4| .69 .57| 78.2 75.6| 0a |
371 | 129 26 55 .03 .33| .76 -1.7| .66 -1.1| .68 .57| 89.1 75.6| 0s |
372 | 156 23 55 .37 .34| .76 -1.5| .59 -1.5| .70 .58| 80.0 77.3| 1s |
373 | 327 23 55 .37 .34| .76 -1.5| .59 -1.5| .70 .58| 80.0 77.3| 0a |
374 | 551 23 55 .37 .34| .76 -1.5| .59 -1.5| .70 .58| 80.0 77.3| 1s |
375 | 179 14 55 1.54 .39| .73 -1.2| .51 -1.1| .71 .59| 87.3 83.6| 0s |
376 | 288 14 55 1.54 .39| .73 -1.2| .51 -1.1| .71 .59| 87.3 83.6| 0a |
377 | 457 14 55 1.54 .39| .73 -1.2| .51 -1.1| .71 .59| 87.3 83.6| 1s |
378 | 517 14 55 1.54 .39| .73 -1.2| .51 -1.1| .71 .59| 87.3 83.6| 0s |
379 | 170 23 55 .37 .34| .73 -1.7| .58 -1.5| .71 .58| 87.3 77.3| 0a |
380 | 48 8 55 2.64 .47| .72 -.9| .62 -.3| .63 .53| 94.5 89.2| 1a |
381 | 136 8 55 2.64 .47| .72 -.9| .62 -.3| .63 .53| 94.5 89.2| 1a |
382 | 247 8 55 2.64 .47| .72 -.9| .62 -.3| .63 .53| 94.5 89.2| 1s |
383 | 284 8 55 2.64 .47| .72 -.9| .62 -.3| .63 .53| 94.5 89.2| 0a |
384 | 422 8 55 2.64 .47| .72 -.9| .62 -.3| .63 .53| 94.5 89.2| 1a |
385 | 491 8 55 2.64 .47| .72 -.9| .62 -.3| .63 .53| 94.5 89.2| 0s |
386 | 540 8 55 2.64 .47| .72 -.9| .62 -.3| .63 .53| 94.5 89.2| 1a |
387 | 559 8 55 2.64 .47| .72 -.9| .62 -.3| .63 .53| 94.5 89.2| 1s |
388 | 69 21 55 .60 .35| .71 -1.7| .53 -1.6| .73 .59| 81.8 78.7| 0s |
389 | 82 21 55 .60 .35| .71 -1.7| .53 -1.6| .73 .59| 81.8 78.7| 0a |
390 | 94 21 55 .60 .35| .71 -1.7| .53 -1.6| .73 .59| 81.8 78.7| 0s |
391 | 114 21 55 .60 .35| .71 -1.7| .53 -1.6| .73 .59| 81.8 78.7| 1a |
392 | 198 21 55 .60 .35| .71 -1.7| .53 -1.6| .73 .59| 81.8 78.7| 0a |
393 | 298 21 55 .60 .35| .71 -1.7| .53 -1.6| .73 .59| 81.8 78.7| 0a |
394 | 299 21 55 .60 .35| .71 -1.7| .53 -1.6| .73 .59| 81.8 78.7| 0a |
395 | 306 21 55 .60 .35| .71 -1.7| .53 -1.6| .73 .59| 81.8 78.7| 1s |
396 | 366 21 55 .60 .35| .71 -1.7| .53 -1.6| .73 .59| 81.8 78.7| 1a |

```

397 | 475 21 55 .60 .35| .71 -1.7| .53 -1.6| .73 .59| 81.8 78.7| 0a |
398 | 531 21 55 .60 .35| .71 -1.7| .53 -1.6| .73 .59| 81.8 78.7| 0a |
399 | 534 21 55 .60 .35| .71 -1.7| .53 -1.6| .73 .59| 81.8 78.7| 0a |
400 | 553 21 55 .60 .35| .71 -1.7| .53 -1.6|z .73 .59| 81.8 78.7| 1s |
401 | 36 20 55 .72 .35| .70 -1.7| .53 -1.6|y .73 .59| 85.5 79.4| 1s |
402 | 92 20 55 .72 .35| .70 -1.7| .53 -1.6|x .73 .59| 85.5 79.4| 0s |
403 | 153 20 55 .72 .35| .70 -1.7| .53 -1.6|w .73 .59| 85.5 79.4| 1s |
404 | 277 20 55 .72 .35| .70 -1.7| .53 -1.6|v .73 .59| 85.5 79.4| 0s |
405 | 285 20 55 .72 .35| .70 -1.7| .53 -1.6|u .73 .59| 85.5 79.4| 0a |
406 | 311 20 55 .72 .35| .70 -1.7| .53 -1.6|t .73 .59| 85.5 79.4| 1s |
407 | 157 19 55 .85 .36| .69 -1.7| .59 -1.3|s .73 .59| 90.9 80.1| 1s |
408 | 328 19 55 .85 .36| .69 -1.7| .59 -1.3|r .73 .59| 90.9 80.1| 0a |
409 | 552 19 55 .85 .36| .69 -1.7| .59 -1.3|q .73 .59| 90.9 80.1| 1s |
410 | 184 15 55 1.39 .38| .69 -1.5| .47 -1.4|p .74 .59| 87.3 82.9| 0s |
411 | 292 15 55 1.39 .38| .69 -1.5| .47 -1.4|o .74 .59| 87.3 82.9| 0a |
412 | 331 15 55 1.39 .38| .69 -1.5| .47 -1.4|n .74 .59| 87.3 82.9| 0s |
413 | 375 15 55 1.39 .38| .69 -1.5| .47 -1.4|m .74 .59| 87.3 82.9| 0a |
414 | 459 15 55 1.39 .38| .69 -1.5| .47 -1.4|l .74 .59| 87.3 82.9| 1s |
415 | 537 15 55 1.39 .38| .69 -1.5| .47 -1.4|k .74 .59| 87.3 82.9| 1a |
416 | 5 7 55 2.87 .49| .68 -1.1| .36 -.6|j .64 .51| 96.4 90.2| 1s |
417 | 508 7 55 2.87 .49| .68 -1.1| .36 -.6|i .64 .51| 96.4 90.2| 1a |
418 | 38 17 55 1.11 .37| .62 -2.1| .41 -1.9|h .77 .60| 85.5 81.4| 1a |
419 | 52 17 55 1.11 .37| .62 -2.1| .41 -1.9|g .77 .60| 85.5 81.4| 1a |
420 | 98 17 55 1.11 .37| .62 -2.1| .41 -1.9|f .77 .60| 85.5 81.4| 0s |
421 | 142 17 55 1.11 .37| .62 -2.1| .41 -1.9|e .77 .60| 85.5 81.4| 1a |
422 | 191 17 55 1.11 .37| .62 -2.1| .41 -1.9|d .77 .60| 85.5 81.4| 0s |
423 | 316 17 55 1.11 .37| .62 -2.1| .41 -1.9|c .77 .60| 85.5 81.4| 1a |
424 | 355 17 55 1.11 .37| .62 -2.1| .41 -1.9|b .77 .60| 85.5 81.4| 1a |
425 | 474 17 55 1.11 .37| .62 -2.1| .41 -1.9|a .77 .60| 85.5 81.4| 0a |
426 |-----+-----+-----+-----+-----+-----+-----+-----+-----+
-----|
427 | MEAN 27.9 55.0 -.15 .36|1.01 .0| .98 -.1| | 77.9 78.4| |
428 | S.D. 10.2 .0 1.27 .05| .24 1.3| .48 1.0| | 7.9 4.7| |
429 -----

```

```

4
5      Person - MAP - Item
6      <more>|<rare>
7      5      +
8      |
9      |
10     |
11     |
12     .      |
13     4      +
14     |
15     |
16     |
17     .#      |T
18     .      |
19     3      +
20     .      | I0041 I0042 I0043
21     .#      |T
22     .#      | I0039
23     .#      |
24     .#      |
25     2      +
26     .####   | I0046 I0049
27     ##      |S I0048 I0052 I0053
28     .####   |
29     .##### |S I0036 I0047
30     .####   | I0015 I0025 I0026 I0044
31     1      + I0016 I0024 I0037
32     .####   | I0033 I0051 I0054
33     .##### | I0027 I0028 I0029 I0032 I0038
34     .##### | I0023
35     .##### | I0017
36     .##### |M I0002 I0004 I0031
37     0      .##### +M I0018 I0050
38     .####   | I0035 I0045
39     .##### | I0034 I0040
40     .####   | I0013
41     .##### | I0008
42     .####   |
43     -1      .####   + I0003 I0010 I0030
44     .##### |S I0012
45     .#      |
46     .####   | I0007 I0014
47     .##      |S I0001
48     .##      | I0006 I0055
49     -2      +
50     .###    | I0009
51     T      |
52     |      | I0005
53     ##      |
54     .#      |
55     -3      +
56     |      | I0021
57     |T      | I0020
58     |      |
59     |      | I0011 I0022
60     |      |
61     -4      + I0019
62     <less>|<frequ>
63     EACH "#" IS 4. EACH "." IS 1 TO 3

```

الملحق رقم : خريطة Item person Map

```

1  TABLE 16.3 C:\Winsteps\doct25.txt                ZOU244WS.TXT  May  3 15:00 2026
2  INPUT: 559 Person  55 Item  REPORTED: 559 Person  55 Item  2 CATS WINSTEPS 3.72.3
3  -----
4
5      Item - MAP - Person
6      <rare>|<more>
7
8      5      +
9              |
10             |
11             |
12             | 0s  0s  0s
13
14      4      +
15              |
16              |
17              T| 0a  0a  0s  0s  1s
18                 |
19                 |
20                 +
21      XXX | 1a  1s  1s
22      |T 0s  1a  1s  1s  1s
23      X | 0a  0a  0s  0s  1a
24      | 0a  0s  0s  1a  1a  1s  1s
25      | 0s  0s  1a  1a  1a  1s  1s
26      2      + 0a  0a  0a  0s  1a  1a
27      XX | 0a  0a  0a  0a  0s  0s  0s  0s  1a  1a  1a  1a  1a  1a  1a
28          | 1a  1s  1s
29      XXX S| 0a  0s  0s  1a  1a  1a  1s  1s
30          | 0a  0s  0s  0s  0s  0s  0s  0s  1a  1a  1a  1s  1s  1s
31      XX |S 0a  0a  0a  0a  0s  0s  0s  0s  0s  1a  1a  1a  1a  1a  1a
32          | 1a  1a  1a  1a  1s  1s  1s  1s  1s  1s  1s
33      XXXX | 0a  0a  0s  0s  0s  0s  1a  1a  1a  1a  1a  1a  1s  1s  1s
34          | 1s  1s  1s  1s
35      1      XXX + 0a  0a  0a  0a  0a  0a  0a  0s  0s  0s  0s  1a  1a  1a  1a
36          | 1a  1a  1a  1a  1s  1s  1s  1s
37      XXX | 0a  0a  0a  0a  0a  0s  0s  0s  0s  0s  0s  1a  1a  1a  1a
38          | 1s  1s  1s
39      XXXXX | 0a  0a  0a  0a  0a  0a  0a  0a  0s  0s  0s  0s  0s  0s  0s
40          | 0s  0s  1a  1a  1a  1a  1a  1a  1a  1a  1a  1a  1s  1s  1s
41          | 1s  1s  1s  1s  1s  1s  1s
42      X | 0a  0a  0a  0s  0s  0s  0s  0s  0s  0s  1a  1a  1a  1a  1a
43          | 1a  1a  1a  1a  1a  1s  1s  1s  1s  1s  1s  1s  1s
44      X | 0a  0a  0a  0a  0a  0a  0a  0a  0a  0a  0a  0s  0s  0s  0s
45          | 0s  0s  0s  1a  1a  1a  1a  1a  1s  1s  1s  1s  1s  1s
46          | 1s  1s  1s  1s
47      XXX |M 0a  0a  0a  0a  0a  0a  0s  0s  0s  1a  1a  1a  1a  1a  1s
48          | 1s  1s  1s  1s  1s  1s
49      0      XX M+ 0a  0a  0a  0a  0a  0a  0a  0a  0a  0s  0s  0s  0s  0s  0s
50          | 0s  0s  1a  1a  1a  1a  1a  1a  1a  1a  1a  1a  1a  1a  1a
51          | 1a  1s  1s  1s  1s  1s  1s  1s  1s  1s  1s
52      XX | 0a  0a  0s  0s  0s  0s  0s  1a  1a  1a  1a  1a  1a  1a  1a
53          | 1s  1s  1s  1s
54      XX | 0a  0a  0a  0a  0a  0a  0a  0a  0a  0a  0s  0s  0s  0s  0s
55          | 0s  0s  0s  0s  0s  0s  1a  1a  1a  1a  1a  1a  1a  1s
56          | 1s  1s  1s  1s  1s  1s  1s
57      X | 0a  0a  0a  0a  0a  0s  0s  0s  0s  0s  1a  1a  1a  1a  1s
58          | 1s  1s  1s
59      X | 0a  0a  0a  0a  0a  0a  0a  0a  0a  0a  0a  0s  0s  0s  0s
60          | 0s  0s  0s  0s  0s  1a  1a  1a  1s  1s  1s  1s  1s  1s
61          | 1s  1s  1s
62      -1      XXX + 0a  0a  0a  0a  0a  0a  0s  0s  0s  1s  1s  1s  1s  1s
63          | 1s  1s
64          | 0a  0a  0a  0a  0a  0a  0s  0s  0s  0s  0s  1a  1a  1a  1a
65          | 1a  1a  1a  1a  1a  1a  1a  1a  1s  1s
66          | 0a  0a  0s  0s  1a  1s  1s
67      XX | 0a  0a  0a  0a  0a  0a  0s  0s  0s  0s  1a  1a  1a  1s  1s
68          | 1s
69      X S| 0a  0s  0s  0s  0s  1a  1a  1s  1s
70      XX | 0a  0s  0s  0s  0s  1a  1a  1a  1a  1s  1s
71      -2      +

```

```

72          X | 0a 0a 0a 0s 1a 1a 1a 1a 1a 1s 1s 1s 1s
73          |T
74          X |
75          | 0a 0s 1a 1a 1a 1a 1s 1s
76          | 0s 1a 1a 1s 1s
77      -3    +
78          X |
79          X T|
80          |
81          XX |
82          |
83      -4    X +
84          <frequ>|<less>
85

```

		Ability (Theta)	Conditional Relia
		-5	0.4848798464
-0.4545454545	0.8927428072	-4.898989899	0.5052214387
-0.3535353535	0.8943291888	-4.797979798	0.525177121
-0.2525252525	0.8957230122	-4.696969697	0.5446711185
-0.1515151515	0.8969222167	-4.595959596	0.5636372256
-0.0505050505	0.8979240328	-4.494949495	0.5820195451
0.0505050505	0.8987250044	-4.393939394	0.5997728655
0.1515151515	0.8993209976	-4.292929293	0.6168626881
0.2525252525	0.899707197	-4.191919192	0.6332649346
0.3535353535	0.899878089	-4.090909091	0.648965375
0.4545454545	0.8998274322	-3.98989899	0.6639588264
0.5555555556	0.899548218	-3.888888889	0.6782481773
0.6565656566	0.8990326203	-3.787878788	0.6918432932
0.7575757576	0.8982719373	-3.686868687	0.704759856
0.8585858586	0.8972565267	-3.585858586	0.7170181903
0.9595959596	0.8959757349	-3.484848485	0.7286421186
1.060606061	0.8944178219	-3.383838384	0.7396578864
1.161616162	0.8925698838	-3.282828283	0.7500931876
1.262626263	0.890417775	-3.181818182	0.7599763134
1.363636364	0.8879460307	-3.080808081	0.7693354419
1.464646465	0.8851377927	-2.97979798	0.7781980742
1.565656566	0.881974741	-2.878787879	0.7865906177
1.666666667	0.8784370342	-2.777777778	0.7945381085
1.767676768	0.874503261	-2.676767677	0.802064057
1.868686869	0.8701504086	-2.575757576	0.8091903996
1.96969697	0.8653538511	-2.474747475	0.8159375302
2.070707071	0.8600873653	-2.373737374	0.82232439
2.171717172	0.8543231794	-2.272727273	0.8283685888
2.272727273	0.8480320639	-2.171717172	0.8340865376
2.373737374	0.8411834739	-2.070707071	0.8394935749
2.474747475	0.8337457524	-1.96969697	0.8446040714
2.575757576	0.8256864068	-1.868686869	0.8494315073
2.676767677	0.8169724691	-1.767676768	0.8539885175
2.777777778	0.807570951	-1.666666667	0.8582869063
2.878787879	0.797449402	-1.565656566	0.8623376363
2.97979798	0.7865765773	-1.464646465	0.8661507979
3.080808081	0.7749232174	-1.363636364	0.8697355689
3.181818182	0.7624629358	-1.262626263	0.8731001702
3.282828283	0.7491732056	-1.161616162	0.8762518271
3.383838384	0.7350364259	-1.060606061	0.8791967411
3.484848485	0.7200410445	-0.9595959596	0.8819400762
3.585858586	0.7041827001	-0.8585858586	0.8844859615
3.686868687	0.6874653415	-0.7575757576	0.8868375105
3.787878788	0.6699022736	-0.6565656566	0.8889968543
3.888888889	0.6515170737	-0.5555555556	0.8909651852
3.98989899	0.6323443187		
4.090909091	0.6124300646		

الملحق رقم 11: معامل الثبات الشرطي

4.191919192	0.5918320249
4.292929293	0.5706194028
4.393939394	0.5488723477
4.494949495	0.5266810239
4.595959596	0.5041442996
4.696969697	0.4813680899
4.797979798	0.4584634071
4.898989899	0.435544192
5	0.4127250171

الملحق رقم 11: جدول مجال القدرات

TABLE OF MEASURES ON TEST OF 55 Item									
	SCORE	MEASURE	S.E.	SCORE	MEASURE	S.E.	SCORE	MEASURE	S.E.
9	0	-6.76E	1.86	19	-.85	.36	38	1.27	.34
10	1	-5.47	1.05	20	-.72	.35	39	1.39	.34
11	2	-4.67	.78	21	-.60	.35	40	1.51	.35
12	3	-4.15	.67	22	-.48	.34	41	1.63	.35
13	4	-3.75	.60	23	-.37	.34	42	1.76	.36
14	5	-3.42	.55	24	-.25	.34	43	1.89	.37
15	6	-3.13	.52	25	-.14	.33	44	2.03	.38
16	7	-2.87	.49	26	-.03	.33	45	2.18	.39
17	8	-2.64	.47	27	.08	.33	46	2.34	.40
18	9	-2.43	.45	28	.19	.33	47	2.51	.42
19	10	-2.23	.44	29	.29	.33	48	2.70	.44
20	11	-2.04	.42	30	.40	.33	49	2.90	.47
21	12	-1.87	.41	31	.51	.33	50	3.14	.50
22	13	-1.70	.40	32	.61	.33	51	3.42	.55
23	14	-1.54	.39	33	.72	.33	52	3.76	.62
24	15	-1.39	.38	34	.83	.33	53	4.22	.74
25	16	-1.25	.38	35	.94	.33	54	4.96	1.03
26	17	-1.11	.37	36	1.05	.33	55	6.20E	1.84
27	18	-.98	.36	37	1.16	.34			

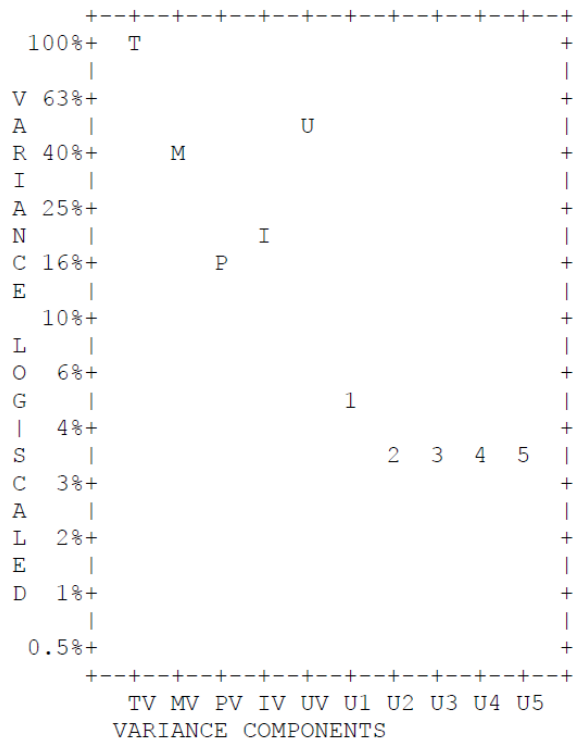
TABLE 23.0 C:\Winsteps\doct25.txt ZOU068WS.TXT May 3 16:34 2026
 INPUT: 559 Person 55 Item REPORTED: 559 Person 55 Item 2 CATS WINSTEPS 3.72.3

Table of STANDARDIZED RESIDUAL variance (in Eigenvalue units)

		-- Empirical --	Modeled
Total raw variance in observations	=	93.2 100.0%	100.0%
Raw variance explained by measures	=	38.2 41.0%	41.1%
Raw variance explained by persons	=	15.5 16.6%	16.7%
Raw Variance explained by items	=	22.7 24.3%	24.4%
Raw unexplained variance (total)	=	55.0 59.0% 100.0%	58.9%
Unexplned variance in 1st contrast	=	5.0 5.4%	9.1%
Unexplned variance in 2nd contrast	=	3.7 4.0%	6.7%
Unexplned variance in 3rd contrast	=	3.4 3.6%	6.1%
Unexplned variance in 4th contrast	=	3.3 3.5%	5.9%
Unexplned variance in 5th contrast	=	3.1 3.3%	5.6%

STANDARDIZED RESIDUAL VARIANCE SCREE PLOT

VARIANCE COMPONENT SCREE PLOT



Conclusion

This study utilized Item Response Theory (IRT) to develop a criterion-referenced mathematics test for fifth-grade students, anchored in the Rasch model. A pool of 55 dichotomous items, measuring conceptual, procedural, and problem-solving skills, was administered to 559 students. Data were analyzed using WINSTEPS (v.3.1.3) to assess construct validity, item functioning, and measurement precision. Results indicated that most items demonstrated productive fit (MNSQ 0.5–1.5), confirming stable item hierarchy and high reliability. While Principal Component Analysis (PCA) of residuals identified a secondary dimension (eigenvalue of 5.0, explaining 9% of the variance), the primary Rasch dimension remained dominant, supporting the scale's fundamental unidimensionality. To account for local dependence, an exploratory multidimensional approach was employed, modeling residual clusters as distinct latent dimensions while preserving the general ability structure. Ultimately, the instrument exhibited strong psychometric properties, with minor misfits attributed to response noise rather than structural flaws. The test is thus deemed a valid and reliable tool for substantive educational assessment.

Keywords: Achievement test, Rasch Response Theory, mathematics, psychometric properties, reliability, validity, Item fit

هدفت هذه الدراسة إلى استخدام نظرية الاستجابة للفقرة في بناء اختبار محكي المرجع في مادة الرياضيات وفق نموذج "راش". ولتحقيق ذلك، تم بناء اختبار تحصيلي مكون من (55) فقرة ثنائية التصحيح، صُممت لقياس الفهم المفاهيمي، والمعرفة الإجرائية، وحل المشكلات لدى تلاميذ الصف الخامس الابتدائي. طبقت الأداة على عينة بلغت (559) تلميذاً، وُحللت البيانات باستخدام برنامج WINSTEPS (نسخة 3.1.3) للتحقق من صدق البناء، وأداء الفقرات، ودقة القياس.

وأظهرت النتائج أن معظم الفقرات تمتعت بمؤشرات ملائمة جيدة (MNSQ) بين 0.5 و 1.5، مما يؤكد استقرار هرمية الصعوبة وثبات الأداة. وبالرغم من أن تحليل المكونات الأساسية للبواقي كشف عن وجود بُعد ثانوي (بقيمة كامنة قدرها 5.0، تفسر 9% من التباين)، إلا أن بُعد "راش" الأساسي ظل هو المهيمن، مما يدعم أحادية بُعد المقياس. ولمعالجة التبعية المحلية، تم اعتماد نموذج "راش" متعدد الأبعاد الاستكشافي، حيث صُنِفَت الفقرات كأبعاد كامنة مستقلة مع الحفاظ على بُعد القدرة العامة. وبشكل عام، قدم التحليل أدلة قوية على الصدق الداخلي للأداة، مع إرجاع عدم الملاءمة البسيط في بعض الفقرات إلى "ضجيج" في الاستجابات وليس إلى تعددية في البنية الجوهرية. وبناءً عليه، تُعد الأداة صالحة للاستخدام والتفسير في السياقات التربوية.

الكلمات المفتاحية: اختبار التحصيل، النظرية الكلاسيكية، نظرية استجابة راش، الرياضيات، الخصائص السيكو مترية، الثبات، الصدق،
ملاءمة الفقرة