دراسة موازنة ناقدة لنماذج السمات الكامنة، والنماذج الكلاسيكية في القياس النفسي والتربوي

صلاح الدين محمود علام*

* حصل على الدكتوراه في القياس والتقويم والإحصاء النفسي والتربوي من جامعة متشجان الأمريكية ـ آن آربر ١٩٨٠ م .
يعمل مدرسا للقياس والتقويم والإحصاء النفسي والتربوي بقسم علم النفس بكلية الآداب ـ جامعة الكويت .

الملخص

هدفت هذه الدراسة لإجراء موازنة ناقدة بين نماذج السمات الكامنة والنماذج الكلاسيكية في بناء الاختبارات السيكلوجية والتربوية ، في محاولة لتحديد بعض أوجه الاتفاق والاختلاف بين أهم المفاهيم والأسس التي تنطوي عليها ، والمشكلات المتعلقة باستخدام كل من نوعي النماذج . وتناولت الدراسة بالموازنة والتحليل مفاهيم السمة الكامنة ، والدالة المميزة للمفردة ، وصعوبة وتمييز مفردات الاختبار . كما تناولت الأسس المتعلقة بطبيعة هذه النماذج ، وكيفية التقدير الإحصائي لبارامتراتها ، وتقييم مطابقة البيانات لها . وقد اهتم الباحث بإبراز وجهة نظره الخاصة في أسلوب كل من نوعي النماذج ، كما اهتم باقتراح بعض التوجهات العامة التي يمكن أن ترشد الباحثين والمشتغلين في مجال القياس إلى الاستخدامات المناسبة ، والتطبيقات العملية الممكنة لهذه النماذج في مجالات بناء الاختبارات ، وتصميم أنظمة بنوك الأسئلة ، وتفريد قياس التحصيل الدراسي .

مقـدمـة

أسهمت النماذج الكلاسيكية للاختبارات في الخمسين عاما الماضية إسهامات متعددة في بناء الاختبارات السيكلوجية والتربوية ، وتحليل وتفسير نتائجها (Gulliksen, 1950; Lord and Novick, 1968) . وأصبحت المفاهيم والأسس المتعلقة بهذه النماذج في متناول الدارسين والباحثين في مختلف مجالات التربية وعلم النفس .

غير أن جهود كثير من علماء القياس من أمثال (De Gruijter, 1984; Hambleton et al., 1978; Lord, 1980; Samejima, 1974; Wright, 1977) قد اتجهت في الآونة الأخيرة إلى تطوير نماذج سيكومترية جديدة أطلق عليها " نماذج السمات الكامنة / Latent Trait Models (LTM)" ، أو " نماذج الاستجابة للمفردة / Item Response Models (IRM)" .

وعلى الرغم مما تتطلبه هذه النماذج من أسس وعمليات إحصائية ورياضية متعددة ، إلا أن هؤلاء العلماء يرون أن لهذه النماذج مزايا يمكن الإفادة منها في مواجهة كثير من المشكلات التي عجزت عن حلها النماذج الكلاسيكية ، ومعالجة أوجه القصور في تلك النماذج التي تتمثل فيما يلي :

١) جميع الخصائص السيكومترية للاختبارات التي تستند في بنائها على أسس النماذج الكلاسيكية ، من مثل معاملات الصعوبة والتمييز والثبات ، تعتمد على خصائص عينة الأفراد التي يجري عليها الاختبار ، وعلى خصائص عينة المفردات التي يشتمل عليها الاختبار .

فمعاملات صعوبة المفردات تتذبذب وتتغير بتغير سمات أو قدرات أفراد العينة ، وقياس سمات أو قدرات الأفراد يتذبذب ويتغير بتغير مستوى وانتشار درجات هذه السمات أو القدرات ، والدرجات الخام أو المحمولة المستمدة من هذه الاختبارات لا يكون لها معنى أو دلالة في حد ذاتها ، إذ يختلف معناها أو دلالتها باختلاف سهولة أو صعوبة مفردات الاختبار . لذلك عند الموازنة بين مستويي أداء فردين ، أو موازنة أداء نفس الفرد في موقفين مختلفين ، يجب استخدام نفس الاختبار في أي من الحالتين . كذلك إذا اشتمل اختباران على نفس العدد من المفردات التي تقيس سمة أو قدرة معينة لدى مجموعة من الأفراد ، فإن درجة كل فرد سوف تختلف باختلاف صعوبة مفردات كل من الاختبارين .

٢) يفترض أن الدرجات التي يحصل عليها فرد في مفردات الاختبار يمكن جمعها كما لو كانت تمثل ميزانا خطيا Linear Scale ، وأن المفردات المتعلقة بالمتغير المراد قياسه تحمل نفس المعنى لدى جميع المختبرين . غير أن هذا الميزان يكون منحنيا ، لأن الفرق الثابت بين درجتين من درجات الاختبار يختلف معناه بحسب موقع الدرجات على متصل السمة أو القدرة التي يقيسها ذلك الاختبار (Wright, 1980:2) .

٣) يفترض أن درجات الاختبار التي تمثل السمة أو القدرة المقاسة دالة خطية مطردة ، بمعنى أنه كلما زادت درجة الفرد في الاختبار ، دل ذلك على زيادة مقدار السمة أو القدرة لديه . ولكن ربما لا يكون هذا صحيحا في كثير من الحالات ، إذ أن بعض الأفراد من ذوي القدرات المرتفعة يحصلون أحيانا على درجات منخفضة في الاختبارات ، ويحدث العكس أحيانا بالنسبة للأفراد من ذوي القدرات المنخفضة ، (Wright, 1980: 2) .

٤) يتغير تكوين ومعنى مفردات الاختبار بمضي الزمن بالنسبة لمجتمع الأفراد الذي أعد له الاختبار ، فالظروف البيئية تتغير ، والظروف الاختبارية ليست دائما مقننة ، وحذف أو تغيير أي مفردة من مفردات الاختبار يؤدي إلى تغيير في درجات الأفراد تغييرا يصعب التنبؤ به .

وقد بيّنت الدراسات الإحصائية والتجريبية التي أجراها المهتمون بنماذج السمات الكامنة أن الخصائص السيكومترية للاختبارات التي يتم بناؤها وتحليل مفرداتها وفقا لمفاهيم وأسس النماذج الجديدة تكون مستقلة عن عينة الأفراد التي يجري عليها الاختبار ، وعن عينة المفردات التي يشتمل عليها ذلك الاختبار (Hambleton et al., 1978; Hashway, 1977; Lord, 1975; Mislevy and Bock, 1985; Tinsley, 1975; Wright and Douglas, 1977, أ ١٩٨٥ علام) .

غير أن الباحث قد لاحظ من خلال دراساته في هذا المجال (علام ، ١٩٨٢ ب ، ١٩٨٥ أ ، ١٩٨٦) أنه على الرغم من اختلاف المفاهيم والأسس التي يرتكن إليها كل من نوعي النماذج ، إلا أنه إذا تفحصنا هذه المفاهيم والأسس بعناية لوجدنا أن هناك اتفاقاً بينها في جوانب معينة . . ومما لا شك فيه أن تحديد أوجه الاتفاق والاختلاف يمكّننا من تصور الإطار العام الذي يضم كلا النوعين من النماذج ، وبذلك نستطيع الإفادة من ميزات كل منهما في القياس النفسي والتربوي بما يحقق الدقة والفاعلية والمرونة .

الهدف من الدراسة

تمثل الدراسة الحالية محاولة من جانب الباحث لتحديد أوجه الاتفاق والاختلاف بين أهم المفاهيم والأسس التي تنطوي عليها نماذج السمات الكامنة ، والنماذج الكلاسيكية ، في بناء الاختبارات السيكلوجية والتربوية ، من خلال الموازنة بينهما موازنة ناقدة ، واقتراح بعض التوجهات العامة التي ترشد الباحثين والمشتغلين في مجال القياس إلى الاستخدامات المناسبة والتطبيقات العملية الممكنة لهذه النماذج .

العلاقة بين مفهوم السمة الكامنة والدوال المميزة لمفردات الاختبار

لعلّ من الضروري أن نستهل مناقشتنا للنماذج الكلاسيكية ، ونماذج السمات الكامنة بتوضيح

مفهوم السمة الكامنة الذي يعد من المفاهيم الأساسية التي ترتكن إليها كل من هذه النماذج في بناء الاختبارات .

فالاختبارات السيكلوجية والتربوية بعامة تفترض أن هناك سماتٍ أوخصائص معينة يشترك فيها جميع الأفراد ، ولكنهم يختلفون في مقدارها . وعلى الرغم من أن هذه السمات غير منظورة ، إلا أنه يمكن الاستدلال على مقدارها من السلوك الملاحظ للفرد ، الذي يتمثل في استجابته على مفردات الاختبار ، وهذا ما يبرر تسميتها بالسمات الكامنة . فالسمة التي تكمن وراء استجابة الفرد على مفردات اختبار لفظي تختلف عن السمة التي تكمن وراء استجابته على مفردات اختبار عددي أو مكاني . ولكن يمكن أن تكمن سمة واحدة وراء استجاباته على مفردات اختبارين مختلفين متعلقين بنفس المحتوى . لذلك فإن الهدف الأساسي لكلا النوعين من النماذج السيكومترية تحديد العلاقة بين استجابات الأفراد على اختبار معين ، والسمة الكامنة وراء هذه الاستجابات . فالأمر الذي يهمنا في القياس النفسي والتربوي هو تحديد مقدار السمات الكامنة وراء أداء الأفراد في الاختبارات المختلفة ، للإفادة من ذلك في التنبؤ بسلوكهم في مواقف مماثلة ، واتخاذ قرارات معينة بشأنهم في ضوء هذا التقدير الكمي للسمات .

وتتفق كل من النماذج الكلاسيكية ونماذج السمات الكامنة في افتراض وجود متصل للسمة ، وأنه يمكن تقدير احتمال إجابة فرد إجابة صحيحة عن مفردة اختبار إذا علمنا موقعه على هذا المتصل . وبتعبير رياضي يمكن القول بأن الاختبار يعد بمثابة راسم فوقي (دالة) من عينة الأفراد المختبرين إلى متصل السمة الكامنة . كما تتفق هذه النماذج في أن مقدار هذا الاحتمال يكون دالة متزايدة مطردة لمواقع الأفراد على متصل السمة . وهذا يعني أن احتمال توصل الفرد إلى الإجابة الصحيحة عن مفردة اختبار يزداد بازدياد مقدار السمة لديه .

غير أن نماذج السمات الكامنة تختلف عن النماذج الكلاسيكية في تصورها لشكل هذه الدالة . فبينما تفترض النماذج الكلاسيكية أن هذه الدالة تتخذ شكل المنحنى الاعتدالي ، تفترض نماذج السمات الكامنة أنها تتخذ شكل المنحنى اللوغاريتمي . وتسمى هذه الدالة في الحالتين بالدالة المميزة للمفردة ، ويسمى المنحنى الممثل لها بالمنحنى المميز للمفردة .

فالنماذج الكلاسيكية للاختبارات تفترض أن الدالة المميزة للمفردة لا تختلف باختلاف مفردات الاختبار ، وهذه الدالة تمثل بالمنحنى الاعتدالي التراكمي لمتغير عشوائي . فإذا كانت ل (س) ترمز إلى احتمال الإجابة الصحيحة عن مفردة اختبار لفرد تمثل قدرته بالنقطة س على متصل السمة ، فإن :

$$ل(س) = (٢\ ط)^{-\frac{١}{٢}} \int_{-\infty}^{س} e^{-\frac{١}{٢} م^{٢}}\ د م$$

(١) ٠٠٠٠٠٠٠٠

حيث ط ترمز إلى النسبة التقريبية

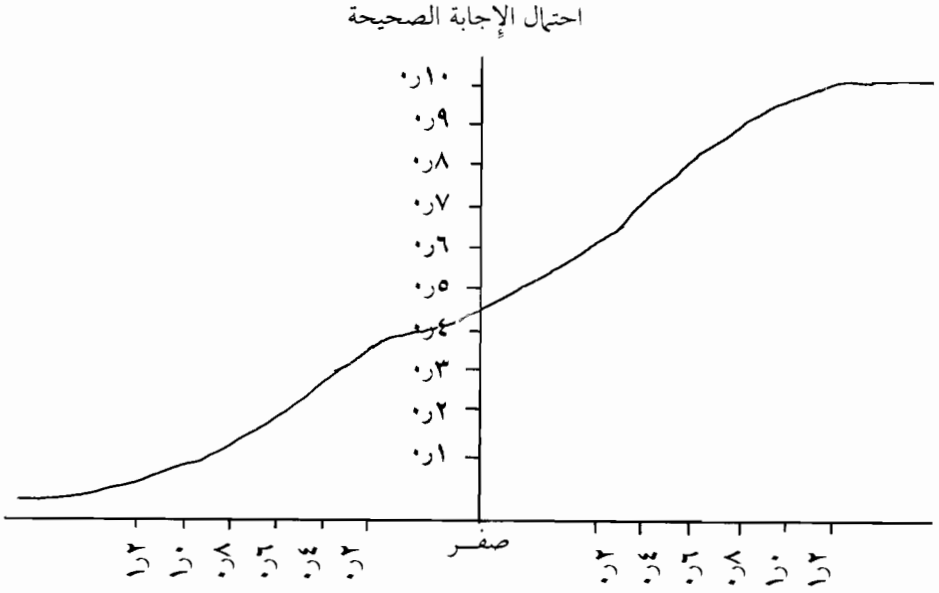
م ترمز إلى المتغير الرمزي للتكامل*
e ترمز إلى الأساس اللوغاريتمي الطبيعي .

ويمكن الحصول على قيمة تقريبية للمقدار ل (س) بإجراء عملية التكامل المحدود للطرف الأيسر لهذه المعادلة باستخدام أساليب التحليل العددي .

والعلاقة بين المتغيرين س ، ل (س) موضحة بالشكل (١) التالي (Hashway, 1977) :

شكـــل (١)

المنحنى المميز لمفردة اختبار وفقا لأسس النظرية الكلاسيكية



* يشير الرمز $\int_{-\infty}^{س}$ إلى إجراء عملية تكامل محدود للدالة على متصل القدرة بين القيمتين ـ ∞ ، س .

ويفترض في هذا المنحنى أن ل (س) أي احتمال الإجابة الصحيحة عن مفردة اختبار دالة تزايدية مطردة لموقع الفرد على بعد السمة الكامنة ، وأن احتمالات الإجابات الصحيحة لفرد عن جميع مفردات الاختبار تكون متساوية .

ونماذج السمات الكامنة لا تتفق مع النماذج الكلاسيكية للاختبارات في هذه الفرضية . إذ

تفترض وجود دالة مميزة خاصة بكل مفردة من مفردات الاختبار تتخذ كل منها شكل المنحنى اللوغاريتمي التراكمي / Logistic Ogive Curve .

ويرى الباحث أنه من الناحية العملية الواقعية يمكن أن تتباين صعوبة المفردات المختلفة في اختبار معين بالنسبة للفرد الواحد ، وبذلك يكون احتمال الإجابات الصحيحة عن مجموعة من مفردات اختبار دالة لكل من موقع الفرد وموقع المفردة على متصل السمة الكامنة التي تقيسها هذه المفردات ، وهذا يؤدي إلى اختلاف الدوال المميزة لكل مفردة منها . لذلك ربما تكون نماذج السمات الكامنة أكثر تمثيلا لواقع عملية القياس النفسي والتربوي من هذه الناحية .

ومما لا شك فيه أن اختلاف هذا التوجه أدى إلى بعض الاختلافات في معالجة كل من نوعي النماذج لمشكلة تحديد الخصائص السيكومترية للاختبارات ، وهذا ما سنتناوله بالحوار بعد عرضنا لأنواع نماذج السمات الكامنة .

أنواع نماذج السمات الكامنة

تهدف نماذج السمات الكامنة كما ذكرنا إلى تحديد علاقة بين أداء الفرد في اختبار معين وبين السمات أو القدرات التي تكمن وراء هذا الأداء وتفسره ، لذلك فإن هذه النماذج هي دوال رياضية احتمالية .

وتنقسم هذه النماذج إلى مجموعتين إحداهما النماذج الاستاتيكية ، والأخرى النماذج الديناميكية . وتهتم النماذج الاستاتيكية بالقياس في فترة زمنية واحدة ، كما تهتم بتحديد العمليات التي ينطوي عليها الأداء في الاختبارات السيكلوجية والتربوية . أما النماذج الديناميكية فتهتم بمشكلة قياس التغير في السمات النامية المختلفة في فترات زمنية متباعدة . غير أن النماذج الاستاتيكية هي التي نالت كثيرا من الاهتمام من جانب المشتغلين بالقياس النفسي والتربوي . لذلك سنقتصر على توضيح الصور الرياضية لهذه النماذج .

النماذج الاستاتيكية للسمات الكامنة

ذكرنا أن احتمال الإجابة الصحيحة عن إحدى مفردات اختبار يكون دالة للفرق بين موقعي المفردة والفرد على بعد السمة الكامنة التي يقيسها الاختبار . وتختلف هذه الدالة باختلاف نماذج السمات الكامنة التي أشرنا إليها فيما سبق . ففي حالة النموذج أحادي البارامتر (نموذج راش Rasch Model) تكون هذه الدالة كالتالي (Hambleton and Traub, 1973: 196) :

$$ل = \frac{١}{١ + e^{ص_م - ق}} \qquad (٢) \cdots\cdots$$

م = ١ ، ٢ ، ٣ ، ... ، ن

حيث ل ترمز إلى احتمال أن يجيب فرد إجابة صحيحة عن مفردة م .
ص م ترمز إلى صعوبة المفردة م .
ق ترمز إلى مقدار السمة لدى الفرد كما تقاس بالاختبار .
e ترمز إلى الأساس اللوغاريتمي الطبيعي .

ويتطلب تطبيق هذا النموذج تساوي مقاديـر تمييز جميع مفردات الاختبار ، أي أن تكون جميع المنحنيات المميزة لمفردات الاختبار غير متقاطعة ، وتختلف عن بعضها البعض في نقط التقائها بالمحور الأفقي الذي يمثل متصل السمة الكامنة ، أي تختلف في الصعوبة فقط . (علام ، ١٩٨٥ ، ١٩٨٦) ، (Whitely, 1980, Wright and Stone, 1979) .

وفي حالة النموذج ثنائي البارامتر (نموذج لورد Lord's Model) تكون هذه الدالة كالتالي (Hambleton and Traub, 1973: 196) ، (علام ، ١٩٨٦ : ١١٤) :

$$ل = \frac{١}{١ + e^{ت_م (ص_م - ق)}} \qquad (٣) \cdots\cdots$$

م = ١ ، ٢ ، ٣ ، ... ، ن

حيث ت م ترمز إلى تمييز المفردة (م) ، وتعرّف بقية الرموز كما في الدالة (٢) . ويلاحظ أن لورد Lord أضاف بارامترا جديدا إلى نموذج راش وهو بارامتر التمييز لكل مفردة . أي أنه يفترض في نموذج لورد أن مفردات الاختبار تختلف في صعوبتها وقدرتها على التمييز .

ولذلك يمكن أن تتقاطع المنحنيات المميزة لمفردات الاختبار الذي يتم إعداده وفقا لهذا النموذج . وقد أضاف لورد هذا البارامتر الجديد لأنه من الصعب إيجاد مجموعة من المفردات تميز بدرجة واحدة بين مستويات السمة أو القدرة التي يقيسها الاختبار ، وهو الافتراض الذي اعتمد عليه نموذج راش (Lord, 1968) .

وفي حالة النموذج ثلاثي البارامتر (نموذج بيرنبوم Birnbaum's Model) تكون هذه الدالة كالتالي (Lord and Novick, 1968: 404) ، (علام ، ١٩٨٦ : ١١٦) :

$$ل = ى_م + \frac{١ - ى_م}{١ + e^{ت_م (ص_م - ق)}} \qquad (٤) \cdots\cdots$$

(م = ١ ، ٢ ، ٣ ، ... ، ن

حيث ي ترمز إلى بارامتر الخط التقاربي الأدنى للمنحنى المميز للمفردة (٢) . وتعرّف بقية الرموز كما في الدالة (٢) . وقد أضاف بيرنبوم Birnbaum هذا البارامتر الثالث ، لأن بعض مفردات الاختبارات تسمح أحيانا للأفراد من ذوي القدرة أو السمة المنخفضة جداً بالتوصل إلى الإجابة الصحيحة عن طريق التخمين . والقيمة التقديرية لهذا البارامتر تمثل احتمال توصل هؤلاء الأفراد إلى الإجابة الصحيحة عن المفردة . فعندئذ يجب أن يؤخذ ذلك بعين الاعتبار عند مطابقة البيانات المستمدة من الاختبار للنموذج ، وتكون المنحنيات المميزة للمفردات في هذه الحالة متقاطعة ، أي تختلف عن بعضها البعض في كل من الجزء الذي تلتقي فيه بالمحور الأفقي الذي يمثل متصل السمة الكامنة (بارامتر الصعوبة) ، والميل (بارامتر التمييز) ، والخط التقاربي الأدنى ، وهو الخط الذي يلتقي بالطرف الأسفل للمنحنى في اللانهاية (بارامتر التخمين ، أو ما يسمى عادة ببارامتر شبه الصدفة (Lord and Novick, 1968), (Psuedo - Chance Level (علام ، ١٩٨٦) .

ونود هنا أن نشير إلى أننا سوف نناقش هذه الدوال وكيفية الحصول على قيم تقديرية لبارامتراتها في الفقرة التالية .

وعلى الرغم من اختلاف النماذج الثلاثة في تركيبها البارامتري الذي يفسر استجابات الأفراد لمفردات الاختبارات والمقاييس ، إلا أنها تتفق في مجموعة من الافتراضات التي يجب أن تتحقق في البيانات قبل استخدام أي منها في بناء هذه الاختبارات أو المقاييس . ومن بين هذه الافتراضات أن تكون مفردات الاختبار من النوع الثنائي (أي تتطلب إما إجابة صحيحة أو إجابة خطأ) ، ولا تعتمد إجابة أي منها على إجابة الأخرى ، وأن يكون الاختبار من نوع اختبارات القوة ، أي لا تلعب السرعة دوراً في الإجابة عن مفرداته ، ولا تتأثر الإجابة بعامل التخمين ، وأن تقيس جميع مفردات الاختبار أو المقياس سمة أو قدرة أحادية البعد .

وعدم تحقق أي من هذه الافتراضات يؤدي إلى عدم مطابقة البيانات المستمدة من الاختبار للنموذج المستخدم بدرجة أو بأخرى . لذلك يجب على المشتغلين بالقياس العناية باختيار النموذج الذي يتلاءم وطبيعة السمة المراد قياسها ، ونوع البيانات المستمدة من الاختبار الذي يبنى لقياس هذه السمة بحيث تحقق هذه البيانات فروض النموذج المستخدم .

الخصائص السيكومترية لمفردات الاختبارات في إطار كل من نوعي النماذج

من الجدير بالذكر أن الخصائص السيكومترية لمفردات الاختبار ، من مثل الصعوبة والتمييز هي التي تحدد العلاقة بين السمة الكامنة التي يهدف الاختبار إلى قياسها ، ودرجات الأفراد في ذلك الاختبار كما هو واضح في الدوال الممثلة لهذه النماذج (Lord and Novick, 1968: 359) .

وإذا أجرينا موازنة بين النماذج الكلاسيكية ونماذج السمات الكامنة في هذا الشأن ، لوجدنا أنه

على الرغم من اتفاقها حول هذه الخصائص ، إلا أن هناك اختلافا بينهما في كيفية معالجتهما للمفاهيم المختلفة المتعلقة بهذه الخصائص .

فمفهوم «صعوبة المفردة» في إطار النماذج الكلاسيكية يدل على نسبة عدد أفراد المجموعة الذين أجابوا إجابة صحيحة عن هذه المفردة . أو بعبارة أخرى يدل على متوسط احتمالات الإجابة الصحيحة عن المفردة لجميع أفراد المجموعة . (Lord and Novick, 1968: 328, Nitko, 1983 : 288) .

وتقدير صعوبة مفردة اختبار لا يرتكن إلى فروض معينة تتعلق بالمفردة ، وتتغير قيمته بتغير أفراد المجموعة اعتمادا على متوسط مقدار السمة المراد قياسها لديهم . ومعرفتنا بالقيمة التقديرية لصعوبة مفردة اختبار بالنسبة لمجموعة معينة من الأفراد لا يفيدنا كثيراً في معرفة درجة صعوبة المفردة بالنسبة لواحد من أفراد المجموعة . فلتحديد احتمال إجابة فرد إجابة صحيحة عن مفردة معينة باستخدام قيمة هذا الإحصاء ، يجب أن يكون معلوما لدينا توزيع درجات السمة لمجموعة الأفراد ، والمركز النسبي للفرد في هذه المجموعة . ويتطلب ذلك إجراء افتراض فرضيات معينة تتعلق بالسمة المراد قياسها وعلاقتها بالأداء في مفردة الاختبار .

أما مفهوم «تمييز المفردة» فيشير إلى مدى فاعلية مفردة الاختبار في التمييز بين المستويات المختلفة للسمة المراد قياسها لدى المختبرين . وتوجد طرق إحصائية مختلفة في إطار النماذج الكلاسيكية لتقدير ذلك ، أكثرها استخداما يعتمد على إيجاد معامل الارتباط بين درجات الأفراد في إحدى المفردات ودرجاتهم الكلية في الاختبار (Lord and Novick, 1968: 331) .

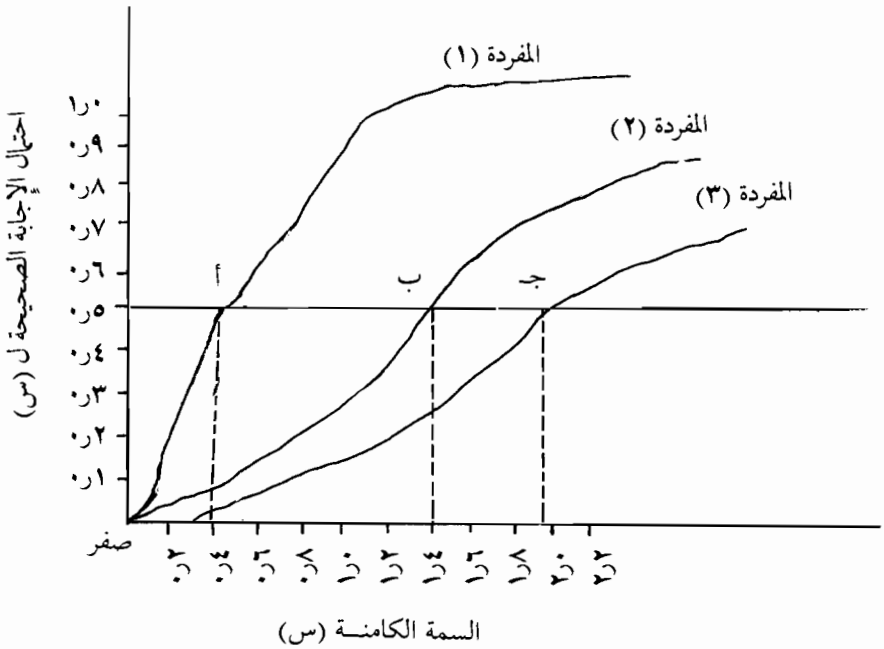
وتعتمد قيمة هذا المعامل على درجة صعوبة المفردة وعلى خصائص مجموعة الأفراد المختبرين .

ويتخذ هذان المفهومان معنى مختلفا في إطار نماذج السمات الكامنة . ويرجع ذلك إلى افتراض هذه النماذج وجود دالة مميزة لكل مفردة من مفردات الاختبار تتخذ شكل المنحنى اللوغاريتمي التراكمي . وتتميز هذه الدّالة اللوغاريتمية التراكمية بخصائص يمكن الإفادة منها في جعل ميزان درجات الاختبار خطيا ، وفي تقدير الخصائص السيكومترية للاختبارات تقديرا مستقلا عن خصائص عيّنة الأفراد المختبرين .

وسوف نوضح فيما يلي كيفية معالجة هذه النماذج لمفهومي الصعوبة والتمييز لمفردة الاختبار في ضوء المنحنى الممثل لهذه الدالة الذي أشرنا إليه فيما سبق باسم «المنحنى المميز للمفردة» .

لذلك نفترض أن لدينا ثلاث مفردات يختلف موقع كل منها على متصل السمة الكامنة . كما هو مبين بالشكل (٢) التالي المشتق عن هاشواي (Hashway, 1977) بعد إجراء بعض التعديلات عليه :

شكـــل (٢)
المنحنيات المميزة لمفردات اختبار وفقا لأسس أحد نماذج السمات الكامنة



ويتضح من هذا الشكل أن المفردات الثلاث تقع عند النقط ٤ر٠ ، ١ر٢ ، ٢ر٢ ، على متصل السّمة الكامنة . فالأفراد الذين يقعون عند إحدى هذه النقط على نفس المتصل يكون احتمال إجابتهم إجابة صحيحة عن المفردة المناظرة لنفس الموقع ٥٠ر٠ ، لذلك فإن صعوبة المفردة في إطار نماذج السمات الكـــامنـــة ليست نسبـة عدد الأفـراد الـذين يجيبـون عنهـا إجـابـة صحيحـة كما هو الحـال في النـمـاذج الكلاسيكية ، وإنما هي النقطة التي على متصل السمة الكامنة التي نتوقع عندها أن يكون احتمال إجابة فرد إجابة صحيحة عن المفردة بدون تخمين ٥٠ر٠ .

فالفـرد الـذي يقع عند النقطة ٤ر٠ على متصل السمة الكامنة يكون مقدار احتمال إجابته إجابة صحيحـة عن المفـردة (١) يسـاوي ٥٠ر٠ ، بينما لا يكون الحال كذلك عند محاولته الإجابة عن أي من المفـــردتـــين (٢) أو (٣) مثـلاً، وذلـك لأن المفـردة (١) أقـل صعـوبـة من كل من المفـردتـين (٢) ، (٣) بالنسبـة لدرجته في السّمة الكامنة . وهذا يدل أيضا على أن المفردات الثلاث تختلف في تمييزها بين المستويات المختلفـة للسمّـة . فتمييـز المفـردة في إطـار نماذج السـمات الكـامنـة يختلف أيضـا عن نظيره في النماذج الكلاسيكية . فلكي نحدد معامل تمييز كل من المفردات الثلاث في شكل (٢) يجب أن نرسم خطا أفقيا عنـد النقطـة ٥٠ر٠ على المحـور الـرأسي التي تناظر احتمال الإجابة الصحيحة بحيث يقطع المنحنيات

المميزة لهذه المفردات . ومعامل تمييز كل مفردة عندئذ هو ميل المنحنى المميز لها عند نقطة التقاطع (نقطة انقلاب المنحنى) . فإذا نظرنا إلى الشكل نجد أن المفردة (١) أكثر تمييزا من المفردة (٢) ، والمفردة (٢) أكثر تمييزا من المفردة (٣) ، وذلك لأن ميل المنحنيات الثلاثة عند النقط أ ، ب ، جـ يزيد عن الواحد الصحيح بالنسبة للمفردة (١) ، ويقل عن الواحد الصحيح بالنسبة للمفردتين (٢) ، (٣) . فمعامل التمييز المرتفع يدل على أن الفرد الذي تفوق سمته الكامنة صعوبة المفردة على نفس متصل السمة يكون احتمال إجابته إجابة صحيحة عن هذه المفردة كبيرا ، والفرد الذي تقل سمته الكامنة عن صعوبة المفردة يكون احتمال إجابته إجابة خطأ عن المفردة كبيرا .

ومعامل التمييز المنخفض يدل على أن احتمال الإجابة الصحيحة عن المفردة منتظم ، وأن المفردة لا تميز جيدا بين مستويات السمة المعينة .

ولقد أجريت العديد من الدراسات للموازنة بين نتائج تحليل مفردات الاختبارات باستخدام كل من نوعي النماذج (Hashway, 1977; Tinsley and Dawis, 1975) ، (علام ، ١٩٨٥ أ) .

وبيّنت معظم هذه الدراسات أن المفردات التي يتم انتقاؤها وفقا لكل من نوعي النماذج تختلف إذا كان مدى تمييز مفردات الاختبار متسعا . فالنماذج الكلاسيكية تستبعد فقط المفردات ذات التمييز المنخفض ، أما نموذج راش الذي يعد أقل نماذج السمات الكامنة تعقيدا ، وأكثرها استخداما في بناء الاختبارات العقلية والتحصيلية ، فإنه يستبعد المفردات التي يكون تمييزها متطرفا أي منخفضا أو مرتفعا .

ويرى الباحث أن هذا التفاوت في انتقاء المفردات ربما يرجع إلى اختلاف الهدف من تحليل المفردات في كل من الحالتين . فالهدف في حالة النماذج الكلاسيكية يكون موجها نحو الكشف عن مفردات الاختبار غير الصالحة ، وتقدير البارامترات المتعلقة بالمفردة (من مثل ، الصعوبة ، والارتباط بمحك ، والارتباط بدرجة الاختبار الكلية) ، واستخدام القيم التقديرية لهذه البارامترات في تقدير الخصائص السيكومترية للاختبار ككل . أما الهدف في حالة نماذج السمات الكامنة فيكون موجها نحو الحصول على تدريج مفردات الاختبار Item Calibration ، أي تقدير صعوبة مفرداته ، وتقييم حسن مطابقة درجات مفردات الاختبار للنموذج المعين باستخدام قيم مربع كاي مثلا (علام ، ١٩٨٦) . والإفادة من ذلك في تقدير سمات الأفراد . فإذا ما بينت النتائج عدم مطابقة درجات مفردات معينة للنموذج تستبعد هذه المفردات من التجمع النهائي لمفردات الاختبار . وهنا مكمن الخطر في انتقاء المفردات على أساس إحصائي سواء باستخدام أسس النماذج الكلاسيكية أو نماذج السمات الكامنة . فاستبعاد المفردات على هذا الأساس الإحصائي يجعل العيّنة النهائية من المفردات التي يشتمل عليها الاختبار غير ممثلة للنطاق السلوكي الذي يقيسه ذلك الاختبار ، وتزداد حدة هذه المشكلة إذا كنا بصدد بناء اختبار من النوع المحَكّي المرجع Criterion - Referenced Test الذي يستخدم في تقدير أداء الفرد بالنسبة إلى نطاق سلوكي محدد من المعارف والمهارات المرتبطة بمحتوى معين ، بغض النظر عن موازنة أدائه بأداء الأفراد الآخرين (علام ، ١٩٨١ ، ١٩٨٢ ، أ ، ١٩٨٤) . وإذا أجري تعديل على

المفـردات غير الصـالحـة ، فإنـه ربـما يجب إعـادة عمليـة تحليل المفردات ، وهذا أمر ضروري في حالة استخـدام النـماذج الكـلاسيكيـة . ولكن المشكلة هنـا هي أن تقدير الخصائص السيكومتر ية للاختبار تعتمـد على القيم التقـديـريـة الأصليـة لبـارامترات المفردات ، وهذه القيم ليست مستقلة عن مجتمع المفردات التي أجرى التحليل على عينة ممثلة له .

أما في حالة إجراء تحليل المفردات باستخدام أحد نماذج السمات الكامنة ، فإن إعادة التحليل لا تكـون ضروريـة ، وذلـك لأن المفـردات الصـالحـة المتبقية بعد استبعاد المفردات التي لم تطابق النموذج تكـون كافيـة للحصـول على قيم تقديرية للسمة المراد قياسها لدى الأفراد . وهذه القيم تكون مستقلة عن عينة المفردات التي اختبر بها الفرد (Tinsley and Dawis, 1975, Rentz, 1976 ، علام ، ١٩٨٦) .

وعلى الـرغم من أن هذه الخـاصيـة تعـد إحـدى ميـزات هذه النـماذج ولا تتـوافـر في النـماذج الكلاسيكية ، فإن الباحث يرى أن هناك عدة مشكلات متعلقة بها :

١) إن استقـلاليـة بارامـترات نماذج السـمات الكـامنـة ، أي عدم تبـاين invariance تقديرات قدرة الأفـراد بتبـاين عيّنـة المفـردات التي يختبر ون بها ، وعدم تباين الخصائص السيكومتر ية للمفردات بتباين عيّنة الأفراد التي أجري عليها الاختبار ، ليس لها ما يبر رها في الواقع العملي الفعلي ، ولكنها ممكنـة فقـط في إطـار نظـري إحصائي بحت . لذلك أكدت كثير من الدراسات من مثل (Bejar, 1983; Hambleton and Murray, 1983) ، أهميـة التحقق بصـورة دوريـة من عدم حدوث تغـيرات ربـما تطـرأ على قيم بارامـترات نماذج السـمات الكـامنـة ، التي يتم تقديرها نتيجة لتغير خصائص أفراد المجتمع المستهدف ، أو تغير طبيعة أو محتوى أو معنى المفردات بالنسبة إليهم بمرور الزمن .

٢) إن نماذج السمات الكامنة تتطلب قدرا كبيرا من العمليات الإحصائية ، التي تتطلب برامج خاصة للحاسب الآلي ، تعتمد على أساليب التحليل العددي . وهذه البرامج ربما لا تكون متاحة لكثير من البـاحثـين والمشتغلين بالقيـاس النفسي والتر بوي ، وبخاصة في منطقتنا العربية . إضافة إلى ضرورة توافـر عيّنـة من الأفـراد كبيرة الحجم نسبيا للحصول على قيم تقديرية مستقرة لبارامترات هذه النماذج .

٣) إن الكشف عن المفـردات غير الصـالحـة في الاختبار في ضوء محّك عدم مطابقتها للنموذج المناسب باستخـدام مقيـاس مربـع كاي يؤدي إلى مشكلات أخرى أهمها أن قيم مربع كاي تعتمد اعتمادا كبيرا على حجم العينة . ويؤدي كبر حجم العينة بدرجة كافية عادة إلى رفض الفرض الصفري المتعلق بمطـابقـة النمـوذج لدرجات مفردات الاختبار ، حيث تكون القيم الاحتمالية عادة أقل من القيم الحرجة التي يحددها الباحث . كما أن تفسير نتائج حسن مطابقة المفردات للنموذج المناسب يتطلب مهارة من جانب مصمم الاختبار .

ونظـراً لأن تحليـل مفـردات الاختبـارات باستخـدام نماذج السمات الكامنة يعتمد اعتماداً أساسيا على تقدير بارامترات النموذج المستخدم ، وتقييم حسن مطابقة الدرجات المستمدة من الاختبار له ، فإنـه ربـما يكـون من المنـاسب أن نعـرض فيما يلي بعض الطرق والملاحظات التي يمكن أن يسترشد بها المشتغلون بالقياس في هذا الشأن .

بعض الطرق والملاحظات المتعلقة بتقدير البارامترات

يتبـين لنـا من الحـوار السـابق أن نماذج السـمات الكـامنـة هي دوال ريـاضية تتضمن بارامترات إحصـــائيـة بعضهـا يتعلق بمفـردات الاختبـار ، والبعض الآخـر يتعلق بسـمات الأفـراد التي يقيسهـا الاختبـار . ويعـد التـوصـل إلى تقديرات دقيقة ومتسقة وغير متحيزة لهذه البارامترات أحد المشكلات الـرئيسـة عنـد استخـدام هذه النـماذج في بناء الاختبارات . فنماذج السمات الكامنة تفترض أن احتمال الاستجـابـة الصحيحـة لفـرد على مفـردة اختبـار يمكن أن يتحدد مقداره إذا علمنا قدرة الفرد (السمة الكـامنـة) ، وخصـائص المفـردة . لذلـك فإن الـدالـة المميـزة للمفردة التي تحدد علاقة بين مقدار هذا الاحتـمال وبـين البـارامـترات المنـاظـرة لقـدرة الفـرد ، وخصـائص المفردة ، هي العمود الفقري لهذه النماذج ، وما يترتب عليها من تطبيقات عملية في مختلف مجالات القياس .

ونظراً لأن الدرجات الخام التي يحصل عليها الفرد في اختبار سيكلوجي أو تربوي يشوبها الخطأ ، فإن تحديد قيم هذه البارامترات باستخدام هذه الدرجات تصبح مشكلة «تقدير إحصائي Statistical Estimation » . وقـولنـا إن التقـديـر الإحصـائي يعـد «مشكلة» إنـما يرجـع إلى صعـوبـة الأساليب الإحصـائيـة المستخـدمـة في التـوصل إلى هذا التقدير ، وتعدد الشروط التي يجب أن يتميز بها إذا أردنا الإفادة منه في عينات أخرى من الأفراد والمفردات .

ولتـوضيـح ذلـك نفـترض أننـا وجـدنا أن النموذج ثنائي البارامتر (نموذج لورد Lord) مناسبا ، واخترنا عدداً قدره (ن) من الأفراد بمفردات اختبار عددها (م) . فعندئذ يجب الحصول على تقديرات إحصـائيـة عددهـا (٢ م) لبـارامـترات المفـردات (أي الصعـوبة ، والتمييز) ، (ن) لبارامترات الأفراد (القدرة أو السمة) .

وكلما زاد عدد الأفـراد والمفـردات زاد عدد البارامترات التي يجب تقديرها إحصائيا . وهذا يؤدي بالطبـع إلى مشكـلات إحصـائيـة وعـدديـة ينبغي مواجهتهـا . فالمشكـلات الإحصائية تتعلق بطبيعة وخصـائص هذه التقـديـرات ، أمـا المشكـلات العـدديـة فتتعلق بكيفية إيجاد حلول لمعادلات التقدير Estimation Equations ، ومـا يرتبـط بها من مشكلة إمكـانيـة التقـارب Convergence عنـد استخدام برنامج معين للحاسب الآلي لحل هذه المعادلات بواسطة أساليب التحليل العددي .

لذلـك نجـد في أدبيـات نماذج السـمات الكـامنـة العـديد من الطرق والأساليب المقترحة لتقدير بارامـــترات هذه النـــماذج بعضهـــا يعتمـد على دالـة الأرجحيـة القصـوى Maximum Likelihood

Methods ، والبعض الآخر يعتمد على إحصاء بييز Bayesian Methods . وقد تجمعت هذه الطرق والأساليب من الدراسات والبحوث في هذا المجال في غضون العقدين الماضيين .

طرق تعتمد على دالة الأرجحية القصوى

يتطلب استخدام هذه الطرق في تقدير بارامترات نماذج السمات الكامنة توافر عيّنة كبيرة من الأفراد المختبرين ، وعدد كبير من مفردات الاختبار ، لكي تكون القيم التقديرية الناتجة غير متحيّزة ، كما يتطلب برامج خاصة للحاسب الآلي لإجراء عمليات التحليل العددي اللّازمة .

فإذا اختبرنا عيّنة تشتمل على (ن) من الأفراد بعدد قدره (م) من مفردات اختبار يقيس قدرة أو سمة معينة ، فإنه يمكن التعبير عن دالة الأرجحية Likelihood Function للتوزيع الاحتمالي المشترك بين أنماط إجابات الأفراد الذين تكون مستويات السمة لديهم ق ١ ، ق ٢ ، ق ٣ ، . . . ، ق ن بالصورة الرياضية التالية (Hambleton et al., 1978: 477) :

$$ل = \prod_{ك=١}^{ن} \prod_{هـ=١}^{م} ل_{هـ}(ق_ك)^{و_{هـ}} \; ل'_{هـ}(ق_ك)^{١-و_{هـ}} \qquad \cdots\cdots (٥)$$

حيث ل ترمز إلى دالة الأرجحية أو الاحتمال المشترك

، $\prod$ ترمز إلى حاصل ضرب نواتج الدالة للأفراد أو المفردات .

، ل (ق) ترمز إلى احتمال أن يجيب الفرد (ك) إجابة صحيحة عن المفردة (هـ) .
هـ ك

، ل' (ق) = ١ ـ ل (ق)
هـ ك هـ ك

، و هـ = (١ أو صفر) تبعا لما إذا كانت الإجابة عن المفردة (هـ) صحيحة أو خطأ على الترتيب .

فإذا استخدمنا النموذج ثنائي البارامتر في هذه الحالة نجد أنه سيكون لدينا (ن) من مستويات القدرة أو السمة ، (٢ م) من بارامترات المفردات . ويصبح لدينا (ن + ٢ م) من البارامترات المراد تقدير قيمة كل منها . وكلما زاد عدد الأفراد أو عدد المفردات زاد عدد هذه البارامترات . بينما إذا استخدمنا النموذج أحادي البارامتر (نموذج راش) يصبح عدد هذه البارامترات ن + م فقط .

ويمكن التوصل إلى القيم التقديرية النهائية لهذه البارامترات بأن نفترض قيما مبدئية لها ثم نجري عليها عمليات تحليل عددي متكرر بغرض الوصول إلى قيم تجعل دالة الأرجحية نهاية عظمى . ولمعرفة كيفية إجراء ذلك إحصائيا يمكن الرجوع إلى :

أنـدرسـون (Anderson, 1970, 1972) ، بوك (Bock, 1972) ، لورد (Lord, 1974, 1975) .

ومن الجدير بالذكر أن هناك نوعين من دالة الأرجحية القصوى ، أحدهما يسمى دالة الأرجحية القصوى المشروطة Conditional Maximum Likelihood (GML) ، والآخر يسمى دالة الأرجحية القصوى الهامشية أو غير المشروطة Marginal Maximum Likelihood (MML) . فطرق التقدير التي تعتمد على الأرجحية القصوى المشروطة تفترض أن قدرات أو سمات الأفراد كميات ثابتة غير معلومة ، وليست كميات يمكن معاينتها . أي أن إحدى مجموعات البارامترات تكون غير معلومة ، والمجموعة الأخرى معلومة . وتقدر مستويات السمة لدى الأفراد بمعلومية بارامترات مفردات الاختبار .

أما الطرق التي تعتمد على الأرجحية القصوى الهامشية أو غير المشروطة ، فتفترض أن الأفراد يمثلون عيّنة عشوائية مسحوبة من مجتمع يكون توزيع السمة المراد قياسها لديهم معلوما . أي أن التقدير غير المشروط يفترض افتراضات معينة بشأن توزيع إحدى مجموعات البارامترات (عادة تكون بارامترات السمة الكامنة) . ثم تستخدم الطرق الإحصائية في عزل أثر بارامترات السمة الكامنة من تقديرات البارامترات المتعلقة بمفردات الاختبار . وقد صمّم علماء القياس العديد من برامج الحاسبات الآلية التي يمكن الاستعانة بها في تقدير بارامترات نماذج السمات الكامنة باستخدام دوال الأرجحية القصوى المشروطة وغير المشروطة ، ومن أمثلتها برامج UCON, ICON, FCON التي أعدّها رايت ، ودوجلاس (Wright and Douglas, 1977) لتقدير بارامترات نموذج راش (Rasch Model) ، وبرنامجا NORMOG, LOGOG اللذان أعدهما كولاكاوسكي وبوك (Kolakowski and Bock, 1973) لتقدير بارامترات نموذج لورد الثنائي البارامتر ، وبرنامج LOGIST الذي أعده وود ، ولورد (Wood and Lord, 1976) لتقدير بارامترات نموذج بيرنبوم الثلاثي البارامتر .

ويستغرق تنفيذ برامج تقدير بارامترات نموذج راش وقتا أقل من الوقت الذي تستغرقه برامج تقدير بارامترات النموذجين الآخرين للوصول إلى التقارب المطلوب Convergence أثناء عمليات التحليل العددي للأرجحية القصوى . فهذه البرامج الأخيرة تؤدي أحيانا إلى قيمة قصوى غير محددة لدالة الأرجحية ، مما يجعل عملية تقدير بارامترات أي من النموذجين في هذه الحالة غير ممكن ، وبخاصة إذا كان عدد مفردات الاختبار أقل من ٢٠ مفردة . كما أن البارامترات التي يتم تقديرها في حالة نموذج راش تتميز بالاتساق (أي عدم التحيز) ، والفاعلية (أي تباين تقديرات الأرجحية القصوى يكون أقل ما يمكن) ، والكفاية (أي تكون الدرجات الكلية للأفراد في الاختبار إحصاء كافيا للحصول على القيم التقديرية للسمة لدى كل منهم) ، بينما تفتقر تقديرات بارامترات النماذج الأخرى إلى بعض هذه الميزات ، مما يقلل أحيانا من الثقة في نتائجها بناء على البرامج المتاحة في الوقت الحاضر .

ومما هو جدير بالذكر أنه يمكن تقدير بارامترات نموذج راش بدون استخدام برامج الحاسب الآلي ، فهناك طريقة PROX التي أعدّها رايت وستون (Wright and Stone, 1979: 28 - 45) لهذا

الغرض . وقد استخدم الباحث هذه الطريقة في تحليل مفردات أحد الاختبارات العقلية الشائعة الاستخدام في المنطقة العربية (علاّم ، ١٩٨٥ أ) .

ولاحظ الباحث أن هذه الطريقة تستغرق وقتا وجهدا كبيرين ، ولكنها تؤدي إلى نتائج مماثلة تقريبا لنتائج تحليل المفردات باستخدام النماذج الكلاسيكية من حيث فاعلية مفردات الاختبار المستخدم في قياس القدرة العامة لدى أفراد العينة . غير أن القيم التقديرية لبارامترات النموذج التي نتجت عن التحليل باستخدام نموذج راش تميزت بخاصيّة عدم التباين Invariance . وقد تحقق الباحث من ذلك بموازنة هذه القيم التقديرية لكل من المجموعة العليا والمجموعة الدنيا من أفراد عينة الدراسة . وهذا لا يحدث في حالة تحليل المفردات باستخدام النماذج الكلاسيكية .

بعض الملاحظات المتعلقة بتقييم حسن المطابقة

إن إحدى الميزات الأساسية لاستخدام نماذج السمات الكامنة من وجهة نظر الباحث هي أنها تجعل مصمم الاختبار لا يقتصر على مجرد تجميع أو صياغة عيّنة من المفردات التي تقيس سمة معينة ، وإنما تجعله أكثر استبصارا بطبيعة البيانات المستمدة من الاختبار . إذ أنها تمكنه منه الفهم المستنير لما تنطوي عليه كل مفردة من مفردات الاختبار ، ومدى اتفاقها مع بقية المفردات في قياس السمة المعينة . وربما يكون صحيحا أن النماذج الكلاسيكية تحقق نفس الغرض ، ولكن نماذج السمات الكامنة تسمح لمصمم الاختبار بدراسة أكثر عمقا للنمط العام لاستجابات الأفراد على كل مفردة من مفردات الاختبار ، وكذلك النمط العام لاستجابات كل فرد على عيّنة المفردات التي يشتمل عليها الاختبار من خلال تقييم حسن مطابقة البيانات الاختبارية للنموذج المستخدم . وربما كان هذا هو السبب في اعتبار المدخل الذي تؤكده هذه النماذج هو المدخل الفردي في تحليل المفردات ، أو المدخل الذي يأخذ الفروق الفردية بعين الاعتبار (Rasch, 1966: 49) . فمصمم الاختبار في هذه الحالة لا يكتفي بالتحقق من مطابقة بعض المفردات للنموذج ، وإنما يقوم أيضا بتحليل بواقي فروق عدم المطابقة Misfit Analysis . فمثل هذا التحليل لا يفيد فقط في التعرف على المفردات التي يجب مراجعتها أو إعادة النظر فيها ، كما هو الحال في النماذج الكلاسيكية ، وإنما أيضا في التعرف على أنماط الاستجابات التي يبدو أنها لا تتسق مع توقعات النموذج . وربما يعيد مصمم الاختبار تحليل البيانات بعد تطبيق الاختبار على جماعات مختلفة من حيث النوع أو الجنس للتعرف على مدى تحيز بعض مفرداته نحو بعض أفراد هذه الجماعات . كما أن تقييم حسن المطابقة يفيد مصمم الاختبار الذي استخدم أكثر من نموذج في تحليل بيانات نفس الاختبار في اتخاذ قرار بشأن النموذج الأكثر ملاءمة لهذه البيانات . ومما لا شك فيه أن هذا التحليل يتيسر بدرجة كبيرة إذا استعان مصمم الاختبار بأحد برامج الحاسب الآلي التي ربما تكون متاحة لديه ، وبخاصة فيما يتعلق بأكثر نماذج السمات الكامنة استخداما في الوقت الحاضر وهو نموذج راش Rasch Model .

بعض التوجهات المتعلقة بالتطبيقات العملية المناسبة لكل من نوعي النماذج

أبرزنا في الحوار السابق بعض أوجه الاتفاق والاختلاف بين كل من النماذج الكلاسيكية ونماذج

السمات الكامنة في مجال بناء الاختبارات وتحليل مفرداتها . وسوف نقدم في هذه الفقرة بعض التوجهات العامة التي يمكن أن يسترشد بها الباحثون والمشتغلون بالقياس إلى الاستخدامات المناسبة والتطبيقات العملية الممكنة لهذه النماذج في بعض مجالات القياس . فلقد استخدمت النماذج الكلاسيكية منذ أوائل القرن العشرين في بناء مختلف أنواع الاختبارات السيكلوجية والتربوية المتعلقة بالتوجيه المهني والتربوي ، والتشخيصي الاكلينيكي ، وانتقاء الأفراد ، وتقويم الطلاب . غير أن اهتمام خبراء القياس ببناء اختبارات ومقاييس تكون خصائصها السيكومترية مستقلة احصائيا عن خصائص عينات الأفراد التي يجري عليها الاختبار ، وتقدير قدرات أو سمات الأفراد تقديرا مستقلا إحصائيا عن عيّنة المفردات التي يختبرون بها ، أدى إلى كثير من التطبيقات لنماذج السمات الكامنة التي أصبحت تزخر بها أدبيات القياس النفسي والتربوي في الآونة الأخيرة . ولا يدعي الباحث القول بأن هذه النماذج قد عالجت أهم المشكلات التي عجزت عن حلها النماذج الكلاسيكية علاجا حاسما كما يذكر مؤيدوها ، وإنما نستطيع القول بأن أسلوب النمذجة الرياضية والإحصائية التي استندت إليه ، بما يتميز به من خصائص دقيقة ، يمكن أن تجعل عملية القياس أكثر مرونة وفاعلية ، إذا استطاع المشتغلون بالقياس بناء اختباراتهم بطريقة تضمن مطابقة البيانات المستمدة منها لأحد هذه النماذج . غير أن الباحث قد لمس من خلال دراساته في هذا المجال صعوبة تحقيق ذلك في الاختبارات السيكلوجية متعددة الأبعاد ، ولكنه يكون ممكنا في بعض الاختبارات العقلية أحادية البعد ، والاختبارات التحصيلية التي تقيس أهدافا تعليمية متجانسة في محتواها (علام ، ١٩٨٥ أ ، ١٩٨٥ ب) .

وربما يؤكد ذلك ما ذكرته وايتلي (Whitely, 1980: 97) من أن نماذج السمات الكامنة لم تؤثر على طرق دراسة الذكاء وصدق التكوين الفرضي لاختبارات القدرات العقلية حتى وقتنا الحاضر إلا تأثيرا محدودا ، مما يتطلب المزيد من الدراسات والبحوث في هذا المجال .

ومع هذا لا نستطيع إغفال الجهود الكبيرة التي يبذلها علماء القياس في تطوير هذه النماذج وتبسيطها للمشتغلين بالقياس بحيث يمكنهم استخدامها في تصميم وبناء مختلف أنواع الاختبارات ، وفي غير ذلك من التطبيقات السيكولوجية والتربوية .

وسوف نعرض فيما يلي بعض التطبيقات العملية التي تناسب هذه النماذج بدرجة أكبر من النماذج الكلاسيكية . ولكن ينبغي أن نوجه نظر المشتغلين بالقياس إلى أن هذه التطبيقات لا تكون ممكنة ، ولا يستفاد بمزايا هذه النماذج ، إلا إذا انطبقت البيانات على النموذج المستخدم انطباقا جيدا . ومما لا شك فيه أن إجراء ذلك يتطلب الكثير من الوقت والجهد والمتابعة .

تصميم وبناء الاختبارات

إن الاختلاف الرئيسي بين النماذج الكلاسيكية ونماذج السمات الكامنة في مجال تصميم وبناء الاختبارات يبرز أثناء تحليل مفردات الاختبار ، وانتقاء مفرداته وتقدير ثبات البيانات المستمدة منه . فباستخدام نماذج السمات الكامنة يمكن الحصول على قيم تقديرية لبارامترات المفردات ـ التي سبق أن

أشـرنـا إليهـا ـ تتميز بخاصية عدم التباين Invariance بالمفهوم الإحصائي الذي وضحناه ، وهو ما لا يحدث في حالة استخدام النماذج الكلاسيكية . فبناء الاختبارات يمر عادة بعدة خطوات أهمها : إعداد مواصفـات الاختبـار ، وكتـابـة أو تجميـع مفـرداته ، وتحديد الخصائص السيكومترية لهذه المفردات ، وانتقاء المفردات الصالحة ، وإعداد معايير الاختبار إذا كان جماعي المرجع Norm - Referenced ، أو تحديـد درجـات القطـع Cutting Scores إذا كان محكّي المرجع Criterion - Referenced ، ثم تقدير ثبات البيانات المستمدة منه ، والتحقق من صدق الاختبار . وتختلف نماذج السمات الكامنة عن النـماذج الكـلاسيكيـة في الخطـوات المتعلقة بتحديد خصائص المفردات ، وانتقاء المفردات الصالحة ، وتقدير الثبات .

ولكي نعطي للمشتغلين بالقياس صورة أوضح لهذا الاختلاف نعرض فيما يلي تصميما عملياً عن كيفيـة بنـاء اختبـار تحصيـلي باستخـدام أبسط نماذج السمات الكامنة وهو نموذج راش Rasch Model (أنظر الفقرة المتعلقة بأنواع هذه النماذج) .

فهذا النموذج يقتصر مثل بقية نماذج السمات الكامنة على قياس المتغيرات (السمات أو القدرات) أحـاديـة البعـد . ولـذلـك يجب على القـائم بتصميم الاختبار أن تكون لديه فكرة واضحة عن محتوى المتغيـر الـذي سوف تعتمـد عليـه القيـاسـات . وتحديد مكونات المتغير المراد قياسه تعد الخطوة الأولى الأساسية في بناء الاختبار ، وتكون الخطوة التالية كتابة أو تجميع عيّنة مناسبة من المفردات التي يمكن أن تستدعي السلوك المرتبط بذلك المتغير . ويجب العناية بانتقاء وصياغة هذه المفردات ، إذ أنها تعد بمثابة تعريف إجرائي للمتغير . وبعد تطبيق الاختبار على عينة مناسبة (يفضل أن لا تقل عن ١٠٠ فرد تمثل المجتمع المستهدف) وتصحيحه ، ينبغي إعداد مصفوفة بيانات خاصة بمفرداته . وتبين هذه المصفوفة توزيع الدرجات التي حصل عليها كل فرد من أفراد العينة في كل مفردة (الدرجات إما واحد صحيح أو صفـر) ، ثم يستبعـد الأفـراد الـذين أجـابـوا عن جميع المفردات إما إجابة صحيحة أو إجابة خطأ ، كما تستبعد المفردات التي أجاب عنها كل فرد من أفراد العينة إما إجابة صحيحة أو إجابة خطأ لأنها لا تصلح لتقـديـر قدرات الأفـراد . وتـرتب بعـد ذلك الدرجات الكلية للأفراد في الاختبار . وهنا ربما يكون من المنـاسب فحص أنـماط الاستجـابـات للتحقق مما إذا كانت تتفق مع توقعـات مصمم الاختبـار . فمن المنطقي توقـع وجـود اتّسـاق بين أنـماط استجـابـات الفـرد ، وترتيب صعوبة المفردات بالنسبة للمتغير المطلوب قيـاسـه . فإذا لم يحدث ذلـك في واقـع الأمـر ، فإنـه يجب فحص المفردات ، أو مراجعة أنماط الاستجابات قبل استخدام هذه البيانات في إيجاد القيم التقديرية لقدرة كل فرد من أفراد العينة ، بهدف التعرف على الأسباب التي ربما تكون قد أدّت إلى عدم الاتساق . وبعد التحقق من ذلك تحوّل هذه المجموعات من الدرجات الكلية إلى ترجيحات لوغاريتمية Log Odds أو Logits ، وذلك باستخراج اللوغـاريتم الطبيعي لخارج قسمـة النسب المئـويـة للإجـابـات الخطأ على النسب المئـويـة للإجـابات الصحيحـة لكـل مفردة ، وكذلك استخراج اللوغاريتم الطبيعي لخارج قسمة النسب المئوية للإجابات الخطأ على النسب المئـويـة للإجـابـات الصحيحة لكل مفردة ، وكذلك استخراج اللوغاريتم الطبيعي لخارج قسمـة النسب المئـويـة للأفـراد الذين أجابوا إجابة خطأ عن كل مفردة . وبذلك تتحول النسب المئوية التي تنحصر بين صفر وواحد صحيح إلى ميزان خطّي جديد Linear Scale للسمة الكامنة يمتد

تدريجه من ـ ∞ إلى + ∞ . ثم يحسب متوسط وتباين كل من توزيعات الترجيحات اللوغاريتمية ، ويعتبر متوسط الترجيح اللوغاريتمي للمفردة هو المركز الذي يناظر الصفر على ميزان تدريج المفردات . ويستخدم تباين كل من الترجيح اللوغاريتمي للمفردات ، والترجيح اللوغاريتمي للأفراد في حساب معاملي تعديل أحدهما للمفردات والآخر للأفراد . وهذان المعاملان يجعلان تقدير صعوبة كل مفردة متحررا من أثر اختلاف قدرة عينة الأفراد ، كما يجعلان تقدير قدرة كل فرد متحرراً من أثر اختلاف صعوبة المفردات . ثم تحسب الأخطاء المعيارية لهذه التقديرات لمعرفة مدى دقتها ، وبالطبع تقل هذه الأخطاء كلما كانت صعوبة المفردات مناسبة لقدرات الأفراد .

وللتحقق من أن النمط العام لتوزيع الدرجات في مصفوفة البيانات الأصلية يتفق مع النمط المتوقع طبقاً للنموذج ، يوجد الفرق بين كل من القيمة التقديرية للقدرة ، والقيم التقديرية المناظرة لصعوبة المفردة التي أجاب عنها الفرد إجابة غير متوقعة . ونقصد بالإجابة غير المتوقعة تلك الإجابة التي لا تتفق ومستوى قدرة الفرد ، فنحن نتوقع أن يجيب الطالب الذي قدرته منخفضة إجابة صحيحة عن المفردات التي تقل في صعوبتها عن مستوى قدرته ، ويجيب إجابة خطأ عن المفردات التي تفوق صعوبتها مستوى قدرته . لذلك يجب تحديد المفردات التي يحدث فيها عكس ذلك ، وإيجاد الفرق المذكور لكل منها ، وتحول هذه الفروق إلى بواقي معيارية وذلك برفع كل من هذه الفروق إلى القوة e (الأساس اللوغاريتمي الطبيعي) ، وإيجاد مجموعها لكل صف ولكل عمود في مصفوفة البيانات الأصلية .

ولإيجاد حسن مطابقة كل من صعوبة المفردات وقدرة الفرد للنموذج ، تقسم كل من هذه البواقي على درجات الحرية المناظرة لها ، وتستخرج الدلالات الإحصائية لهذه المتوسطات من جدول توزيع النسبة الفائية ، وبناء على هذه النتائج تحدد المفردات التي انطبقت على النموذج ، وكذلك يحدد الأفراد الذين اتفقت استجاباتهم مع النمط المتوقع وفقا للنموذج . وعندئذ تكون هذه العينة المتبقية من المفردات لها خصائص سيكومترية مستقلة إحصائيا عن عينة الأفراد التي استمدت منها هذه الخصائص .

ويلاحظ أن نموذج راش شأنه كشأن بقية نماذج السمات الكامنة ، قد استعاض عن مفهوم الثبات وما يرتبط به من مفهوم الخطأ المعياري للقياس في إطار النماذج الكلاسيكية بمفهوم الخطأ المعياري للقيم التقديرية لكل من صعوبة مفردات الاختبار وقدرة الأفراد . وهذه الأخطاء المعيارية تعد مؤشراً لمدى دقة هذه القيم ، ويمكن أن تختلف من مفردة إلى أخرى ، ومن فرد إلى آخر ، على عكس الخطأ المعياري التقليدي الذي تكون له قيمة واحدة للاختبار ككل . كما استعاض عن معايير الاختبارات Norms المتعارف عليها في إطار النماذج الكلاسيكية بالقيم التقديرية لقدرات الأفراد المستقلة عن خصائص عيّنة المفردات .

ولمزيد من التوضيح لكل من الخطوات السابقة يمكن الرجوع إلى (علام ، ١٩٨٥ أ ، ١٩٨٦) .

تصميم أنظمة بنوك الأسئلة

إن بنـاء أنظمـة اختبارية يمكن الإفادة منها في قياس التحصيل الدراسي للطلاب يتطلب إعداد بنـوك للأسئلة . ويتضمن بنـك الأسئلة مجمـوعـة كبـيرة من المفـردات أو الأسئلة التي تتميز بخصائص سيكـومـتر يـة محددة ، وتقيس جميعهـا قدرة أو سمـة كامنـة معيّنة . ولكي يحقق هذا النظام الاختباري أهـدافـه في مختلف المـواقف التعليميـة ينبغي أن يتميـز بالاتسـاق والفـاعليـة والمـرونـة ، ويكون متاحا للمعلمين والمهتمين ببناء الاختبارات الصفية . وعلى الرغم من أن النماذج الكلاسيكية قد استخدمت في العقدين الماضيين لتصميم مثل هذه الأنظمة ، إلا أن الإفادة منها كانت محدودة . ويرجع ذلك إلى اعتماد تحليل مفردات أو أسئلة البنك في إطار هذه النماذج على خصائص عينة الأفراد المستخدمة في هذا التحليـل كما سبق أن ذكـرنـا . لذلك يفضل مواجهة هذه المشكلة اعتمادا على نماذج السمات الكامنة ، حيث أن بارامـتراتهـا (صعـوبـة المفردات وتمييزها بين المستويات المختلفة للقدرة أو السمة المراد قياسها) تتميز بخاصية عدم التباين بالمفهوم الإحصائي الذي أشرنا إليه .

وقـد اهتمت بعض دول العـالم من مثـل الولايات المتحدة الأمريكية وانجلترا بتنفيذ أنظمة بنوك الأسئلة وفقا لمفاهيم وأسس هذه النماذج وبخاصة نموذج راش ، وأصبحت هذه الأنظمة مكونة رئيسة من مكونات أنظمتها التعليمية (Choppin, 1978; Wright, 1979 ، علام ، ١٩٨٢ أ) .

ويستهـدف بنـك الأسئلة تقـدير قدرة أو سمة معينة لدى الطالب باستخدام اختبار يشتمل على عيّنـة من المفـردات أو الأسئلة المسحـوبـة من البنك ، والموازنة بين مستويات هذه القدرة لدى الطلاب على الرغم من اختلاف المفردات أو الأسئلة التي يختبر بها كل منهم . وهذا يتطلب توحيد تداريج ميزان تقـديـر صعـوبـة المفـردات . ويمكن إجراء ذلك عن طريق تقسيم مجتمع المفردات الذي يشتمل عليه البنك إلى مجموعات تشكل اختبارات ، بحيث يشترك كل اختبارين منها في عدد معين من المفردات . وتطبق كل من هذه الاختبارات على عيّنة مختلفة من نفس مجتمع الطلاب المستهدف ، ثم يُجرى تحليل للبيـانـات المستمـدة من كل اختبـار باستخدام نموذج راش ، وذلك باتباع الخطوات التي أشرنا إليها في الفقـرة السـابقـة . فإذا كانت هنـاك مطابقة جيدة بين البيانات والنموذج ، فإننا نتوقع أن تختلف القيم التقديرية لصعوبة المفردات المشتركة بين كل اختبارين منها بمقدار ثابت . عندئذ يجب إيجاد قيمة هذا المقدار الثابت وإضافته إلى جميع القيم التقديرية لصعوبة مفردات أحد هذين الاختبارين ، وبذلك يتم توحيد تداريج ميزان صعوبة كل اختبارين من خلال المفردات المشتركة بينهما .

وبـالطبـع تنشأ بعض المشكلات أثناء إجراء هذا التحليل تتعلق بكيفية انتقاء المفردات المشتركة وعدد هذه المفردات ، وكيفية التحليل إذا لم يكن الفرق بين القيم التقديرية لصعوبة المفردات المشتركة ثابتـا ، ومـدى تأثـير ذلـك على تقـديـر قدرة الأفـراد باستخدام الاختبارات المختلفة التي تشتمل على مجمـوعـات متبـاينـة من مفـردات أو أسئلة البنك . هذا بالإضافة إلى ما يتطلبه إجراء هذا التحليل من عمليات إحصائية متعددة .

ويـمـكـن الـرجـوع إلى (Lord, 1980, Wright and Stone 1979 ، علام ، ١٩٨٦) ، للاطلاع على كيفية مواجهة هذه المشكلات .

تفريد قياس التحصيل الدراسي

إن القيـاس التقليـدي للتحصيل الدراسي يتم بشكل جماعي ، حيث يطبق على جميع الطلاب نفس الاختبـار أو صور مكـافئـة له على الرغم من تباين مستويات قدراتهم . فلكي تكون للاختبارات الجماعية المقننة فائدة عملية في إطار النماذج الكلاسيكية ، يجب أن تشتمل على عدد كبير من المفردات المنـاسبـة للأفـراد من ذوي القـدرة المتـوسطـة . ولكن يترتب على ذلـك عدم فاعليتهـا في قياس جميع مستـويـات قدرة الأفـراد بنفس الدقة . وربما يظن البعض أن إضافة بعض المفردات السهلة والمفردات الصعبـة يعـد حلا لهذه المشكلة ، ولكن هذا ليس صحيحـا . إذ أن المفـردات السهلة ليس لها فائدة في تقدير مستوى الأفراد من ذوي القدرة المرتفعة ، والمفردات الصعبة لا تفيد في تقدير مستوى الأفراد من ذوي القدرة المنخفضة . فتأثير عامل الملل في الحالة الأولى ، وعامل التخمين في الحالة الثانية يزيد من الأخطـاء العشـوائيـة للقيـاس الناجمة عن إضافة هذه المفردات . كما أن اعتماد الخصائص السيكومتر ية لمثـل هذه الاختبـارات على خصـائص عيّنـة الأفراد للمختبر ين تجعل الموازنة بينهم ممكنة إذا اختبر كل منهم باختبار مختلف يقيس نفس القدرة .

وتفريد قياس التحصيل الدراسي يتطلب أن تكون صعوبة مفردات الاختبار مناسبة لمستوى قدرة الفرد الذي يختبر بها . وهذا بدوره يتطلب توافر مجموعة كبيرة من المفردات ذات خصائص سيكومتر ية معلومـة بحيث يمكن سحب عيّنـات منهـا تناسب المستويات المختلفة لقدرة الأفراد . واستخدام نماذج السـمات الكـامنـة يسمـح بتقـدير مستوى قدرة كل فرد تقديرا مستقلا عن خصائص عيّنة المفردات التي اختـبر بها ، كما يسمـح بالمـوازنـة بين مستـوى قدرة فردين اختبر كل منهما بعينة مختلفة من المفردات التي تقيس القدرة المعينة .

لذلـك اتجـه خبراء القيـاس إلى الإفـادة من مزايـا هذه النـماذج في بنـاء اختبـارات أطلق عليها «الاختبـارات التفـريدية Tailored Tests» . وهناك طرق متعددة لتصميم وبناء هذه الاختبارات ، غير أنهـا جميعـا تعتمـد على إعداد بنك للمفـردات أو للأسئلة باستخدام أحد نماذج السمات الكامنة كما سبق أن ذكرنا ، وتصميم برامج معينة للحاسب الآلي (Urry, 1977; Weis, 1974) ، تجري بعض العمليـات الإحصـائيـة المتعلقة بانتقاء أفضل المفردات المناسبة لقدرة الطالب في كل مرحلة من مراحل العمليـة الاختبـاريـة ، للوصـول إلى أفضـل قيمة تقديرية لقدرته ، بحيث يكون الخطأ المعياري لهذه القيمة أقل ما يمكن .

ويصعب بالطبـع تنفيـذ نظـام تفـريـد قياس التحصيل الدراسي بدون حاسب آلي . ولمزيد من التفصيل في هذا المجال يمكن الرجوع إلى (علام ، ١٩٨٦) . ونظراً لأن هذا النظام ما زال في مراحل التطـور الأولى ، فإن هنـاك بعض المشكـلات الفنيـة المتعلقة بتنفيذه على نطاق واسع ، يسعى خبراء

القياس إلى التغلب عليها ، لكي يصبح مألوفا لدى المشتغلين بالقياس في المستقبل القريب . ومن أهم هذه المشكلات في الوقت الحاضر الكلفة الكبيرة نسبيا التي يتطلبها تنفيذ هذا النظام ، وبخاصة ما يتعلق منها بفنيّات تجهيز معدات الحاسبات الآلية Hardware ، وإعداد البرامج الجاهزة المساعدة Software اللازمة لتنظيم شبكات بنوك الأسئلة ، وإدارة العملية الاختبارية بأسلوب ييّسر على الفرد والمختبر فهم التعليمات المطلوبة أثناء كل مرحلة من مراحل هذه العملية .

غير أن هذا النظام يساعد على جعل عملية قياس التحصيل الدراسي أكثر مرونة وفاعلية ودقة بشرط توافر بنك جيد للأسئلة وفقا للمعايير التي سبق أن أشرنا إليها ، إضافة إلى إسهامه في جعل العملية الاختبارية أكثر تقنيناً مما هي عليه في إطار الممارسات الحالية ، وإمكانية الحصول على بيانات تتعلق بالتقدم الدراسي للفرد في وقت قصير . وهذا يعتمد بالدرجة الأولى على توافر اختصاصيين مدرّبين تدريبا عاليا في مجالات القياس المستحدثة .

لذلك يوصي الباحث بالاهتمام في برامج الدراسات العليا بكليات الآداب والتربية في الجامعات بإعداد هؤلاء الاختصاصيين ، وتوجيه البحوث والدراسات نحو الجوانب المختلفة لهذه المجالات .

المراجع العربية

علام ، صلاح الدين محمود :
- «استراتيجيات جديدة لبناء الاختبارات التحصيلية » ، القاهرة : **صحيفة التربية** ، العدد الرابع ، السنة الثانية والثلاثون ، مايو ١٩٨١ .
- **اختبار تشخيصي مرجعي الميزان لقياس اتقان المعلمين للمهارات الأساسية في بناء الاختبارات المدرسية** ، القاهرة : مكتبة الانجلو المصرية ، ١٩٨٢ (أ) .
- «استخدام نماذج السمات الكامنة في بناء بنوك الأسئلة» ، القاهرة : **صحيفة التربية** ، العدد الثاني ، السنة الثالثة والثلاثون ، يناير ١٩٨٢ (ب) .
- «تطورات معاصرة في بنوك الأسئلة» ، جامعة عين شمس : **مجلة كلية التربية** ، العدد الخامس ، ١٩٨٢ (جـ) .
- «بعض الاتجاهات العالمية المعاصرة في مجال تقويم الطلاب» ورقة مقدمة في ندوة وورشة تطوير أساليب تقويم التعليم العام بدول الخليج العربي ، **المركز العربي للبحوث التربوية لدول الخليج العربي** ، ديسمبر ١٩٨٤ .
- «تحليل بيانات الاختبارات العقلية باستخدام نموذج راش اللوغاريتمي الاحتمالي : دراسة تجريبية» ، الكويت : **المجلة العربية للعلوم الانسانية** ، العدد السابع عشر ، المجلد الخامس ، (شتاء) ١٩٨٥ (أ) .
- «استخدام النموذج ذي الحدين في تقدير درجة القطع لاختبار محكيّ المرجع (دراسة احصائية وتجريبية)» الكويت : **المجلة العربية للعلوم الانسانية** ، العدد التاسع عشر ، المجلد الخامس ، صيف ١٩٨٥ .
- **تحليل البيانات في البحوث النفسية والتربوية** ، القاهرة : دار الفكر العربي ، ١٩٨٥ (حـ) .
- **تطورات معاصرة في القياس النفسي والتربوي** ، الكويت : جامعة الكويت ، ١٩٨٦ .

المراجع الأجنبية

Ahman, J. and Glock. M.	**Evaluating Pupil Growth.** Boston: Allyn and Bacon, 1967, Third Edition.
Allam, S.	"Minimum Measurement Competencies for Classroom Teachers. No. 010686," U.S.A., New Jersey: **Educational Testing Service,** 1980.
Anderson, E.	"Asymptotic Properties of Conditional Maximum Likelihood Estimates", **Journal of the Royal Statistical Society** (Series B), (1970), 32, 283-301. - "The Solution of a set of Conditional Estimation Equations". **Journal of the Royal Statistical Society,** (1972), 34. 42-54. -"A Goodness of Fit Test for the Rasch Model". **Psychometrica,** (1973) 38, 123-140.
Anderson, E. and Madsen, M	"Estimating the Parameters of the Latent Population Distribution". **Psychometrica,** (1977), 42, 357-374.
Bejar. I.	**Achievement Testing.** Calif.: Sage, 1983.
Birnbaum, A.	"Some Latent Trait Models and Their Use in Inferring an Examinee's Ability" In F. Lord and M.Novick, **Statistical Theories of Mental Test Scores.** Mass.: Addition - Wesley, 1968.
Bock. R.	"Estimating Item Parameters and Latent Ability When Responses Are Scored in Two or More Nominal Categories". **Psychometrica,** (1972), 37. 29-51.
Bock, R. and Lieberman, M.	"Fitting a Response Model for Dichotomously Scored Items". **Psychometrica, (1070), 35,179-197.**
Brink, N.	"Effect of Item Discrimination in the Rasch Model". **The Proceedings 79th Annual Convention,** APA, 1971.
Choppin, B.	**Item Banks and the Monitoring of Achievement.** London: NFER: Research In Progress Series, 1978.
Coughlin, J.	**A Multidimensional Extension of the Normal Ogive, Logistic, and Linear Latent Trait Models.** Unpublished Thesis, University of Toronto, 1974.
Cronbach, L.	"Coefficient Alpha and the Internal Structure of Tests". **Psychometrica,** (1951), 16, 297-334.
De Gruijter, D.	"A Comment on 'Some Standard Errors in Item Response Theory". **Psychometrica,** (1984), 49, 269-272.
Green,D. and Droper J.	"Exploratory Studies of Bias in Achievement Tests. **Paper presented at the Annual Meeting of APA,** Los Angeles, New York: McGrow-Hill, August 1972.
Guilford. J.	**Psychometric Methods.** New York: McGraw-Hill, 1954, Second Edition.
Gulliksen,H.	**Theory of Mental Tests.** New York: Wiley, 1950.
Gustafasson,J.	"Testing and Obtaining Fit to Data to the Rasch Model". **British Journal of Mathematical and Statistical Psychology,** (1980), 33, 205-233.
Hambleton, R.	"Latent Ability Scales, Interpretations and Uses". In S. Mayo (Ed.) **New Directions for Testing and Measurement.** San Francisco: Jossey-Bass, 1980.
Hambleton,R. and Murray, L.	**Some Goodness of Fit Investigations for Item Response Models.** Amherst, MA: University of Massachusetts, 1982.
Hambleton, R. and Traub, R.	"Analysis of Empirical Data Using Two Logistic Latent Trait Models." **Journal of Mathematical and Statistical Psychology,** (1973), 26, 196.
Hambleton, R.: Swaminathan,H.; Cook L.	"Developments in Latent Trait Theory". **Review of Educational Research,** (1978), 48, 467-510.
Harris, B.	**Theory of Probability: Reading,** Mass. Addison Wesley, 1966.
Hashway, R.	**A Comparison of Tests Derived Using Rasch and Traditional Psychometric Paradigms.**ph.D. Dissertation, Mass.: Boston College, 1977.
Ironson, G.	"Use of Chi-Square and Latent Trait Approaches for Detecting Item Bias". In R. Berk (Ed.) **Handbook of Methods for Detecting Test Bias.** Baltimore: John Hopkins University Press, 1982.
Kolakowski, D. and Book, R	**LOGOG:** Maximum Likelihood Item Analysis and Test Scoring, Logistic Model for Multiple Item Responses. Ann Arbor: National Educational Resources, 1973.

Lord. F. "Theory of Test Scores". **Psychometeric Monograph,** (1952), No. 7.
- "The Relation of Test Scores to the Trait Underlying the Test". **Educational and Psychological Measurement** (1953), 13, 517-548.
- "Some Test Theory for Tailored Testing". In W.H. Holtzman (Ed.) Computer Assisted Instruction. New York: Harper and Row, 1970.
- Estimation of Latent Ability and Item Parameters when there are Omitted Responses. **Psychometrica, (1974), 39, 247-264.**
- "Evaluation with Artificial Data of a Procedure for Estimating Ability and Item Characteristic Curve Parameters". Research Bulletin 75-33, Princeton, New Jersey: ETS, 1975.
- **Applications of Item Response Theory to Practical Testing Problems.** Hillsdale, New Jersey: Laurence Erlbaum, 1980.
-"Some How and Which for Practical Tailored Testing". In L. Van Der Kamp, **Psychometrics for Educational Debates.** New Jersey: Wiley, 1980.

Lord, F. and Novick M. **Statistical Theories of Mental Scores Readings.** Mass: Addison Wisley, 1968.

Mislevy, R. and Bock R. **Stability of Item Parameters in the Grade 3 California Assessment.** Chicago: IES, 1985.

Nitko, A. **Educational Tests and Measurement: An Introduction.** New York: Harcourt Jovanovich, Inc., 1983.

Nunnally, J. **Psychometric Theory.** New York: McGraw - Hill, 1967.

Pine, s. "Applications of Item Response Theory to the Problem of Test Bias". In D. Weiss (Ed.) **Applications of Computerized Adaptive Testing.** Minn.: University of Minnesota, 1977.

Rasch, G. "An Item Analysis which Takes Individual Difference Into Account". **The British Journal of Mathematical and Statistical Psycholog** (1966), 19, (pt-1), 49-57.

Rentz, C. "Rasch Model Invariance as a Function of the shape of the sample Distribution and Degree of Model-Data Fit. Paper Presented at the**Meeting of the Florida Educational Research Association,** Orlando, Florida, 1976.

Rudner, L. An Approach to Biased Item Identification Using Latent Trait Measurement Theory". Paper Presented at the **Annual Meeting of the AERA,** New York, April.

Samegima, F. - "Estimation of Latent Ability Using a Response Pattern of Graded Scores". **Psychometrica,** (1969). 34, (pt-2), 1-97 (Monograph).
- "Normal - Ogive Model on the Continuous Response Level in Multidinesional Latent Space". **Psychometrica,** (1974), 39, 111-121.

Tinsley, H. and Dawis, R. "An Investigation of the Rasch Simple Logistic Model". **Journal of Educational and Psychological Measurement,** (1975), 35, 325-339.

Urry, V. "A Five-Year Quest: Is Computarized Adaptive Testing Feasible?" **Proceedings of the First Conference on Computarized Adaptive Testing.** Washington, D.C. Government Printing Office, 1976.

Weiss, D. "Strategies of Adaptive Ability Measurement". **Research Report** 74-5, Minn.: University of Minnesota,1974.

Whitely, S. "Multicomponent Latent Trait Models for Ability Tests". **Psychometrica,** (1976), 41, 405-40.
-"Latent Trait Models in the Study of Intelligence". **Intelligence.** (1980), 4, 97-132.

Wingersky, M.Barton, M. and Lord, F. **LOGIST Users Guide.** Princeton, New Jersey: ETS, 1982.

Wood, R. "Response - Contigent Testing". **Review of Educational Research,** (1973), 13, 34-84.

Wood,R. and Lord,F. "A User's Guide to LOGIST" **Research Memorandum** 76-4. Princeton, New Jersey: ETS, 1976.

Wright,B. and Bell,S. **Fair and Useful Testing with Item Banks.** Chicago: MESA, Research Memo No. 23, 1980.

Wright, B. and Douglass, G. "Conditional Varsus Unconditional Procedures for Sample-Free Item Analysis". **Journal of Educational and Psychological Measurement** (1977), 37,3,573-586.
-"Better Procedures for Sample - Free Item Analysis". **Applied Psychological Measurement,** (1977),1 281-293.

wright, B. and Masters, G. **The Measurement of Knowledge and Attitude.** Chicago: MESA Psychometric Laboratory, 1981..

Wright,B and Mead,B. **CALFIT: Sample Free Calibration with a Rasch Measurement Model.** Research Memo, University of Chicago, 1975.

Wright, B. and Stone,M. **Best Test Design.** Chicago: MESA, 1979.

Wright,B. Mead,R. and Draba, R. **Detecting and Correcting Item Bias with a Logistic Response Model.** Research Memo No.22. Chicago: University of Chicago, 1976.