مجلة جامعة الملك سعود، م١٠، العلوم التربوية والدراسات الإسلامية (٢)، ص ص ٢٥٩-٢٩٦ (١٤١٨هـ/ ١٩٩٨م).

دراسة مقارنة لبعض طرق تحيز بنود الاختبار

عبدالله بن علي القاطعي

أستاذ مشارك، قسم علم النفس، كلية التربية، جامعة الملك سعود، الرياض، المملكة العربية السعودية

ملخص البحث. يعتبر مفهوم تحيز الاختبارات من المفاهيم الحديثة نسبيا، حيث ارتبط وجوده الفعلي بتفسير درجة الاختبارات والجدل حول صلاحيتها لبعض الفئات التي يُعتقد بأن الاختبار غير عادل نحوها. وقد استحدثت عدة طرق لقياس التحيز، وأجريت بعض الدراسات المقارنة لمعرفة مدى فعالية تلك الطرق في كشف التحيز . إلا أن بعض الطرق الحديثة نسبيا (طرق أنغوف المعدلة، وطريقة كميلي كاي تربيع) لم يتم التأكد من فعاليتها واتخاذ قرار بشأنها .لذا فإن الدراسة الحالية هدفت إلى مقارنة أهم طرق تحيز بنود الاختبار، وإلى مقارنة تعديل اقترحه الباحث على طريقة معامل التمييز ببقية الطرق.

وقد استخدمت بيانات عينة الصدق والثبات لاختبار المعلومات من اختبار وكسلر لذكاء الأطفال المعدل (الصورة السعودية) لهذه المقارنة، حيث تم التقصي عن فعالية طرق السمات الكامنة (نموذج المعْلم الواحد، و نموذج الثلاثة معالم)، وطرق الصعوبة، وطرق مربع كاي، طرق معامل التمييز مقارنة بنموذج المعْلمين في القدرة على التعرف على تحيز بنود الاختبار .

وقد أظهرت النتائج أن طرق السمات الكامنة هي أفضل الطرق لدراسة التحيز، إلا أن بعض الصعوبات العملية تجعل استخدامها في الأحوال العادية غير ممكن . وكان البديل العملي لهذه الطرق هو كميلي كاي تربيع، يليه منتل-هنزل كاي تربيع، ثم طريقة أنغوف المعدلة .كذلك أظهرت طريقة معامل التمييز المعدلة نتائج مشجعة، بيد أن حداثة هذه الطريقة تجعل الاعتماد عليها في الوقت الحاضر غير مستحبا، وقد نُصح بإجراء مزيد من التقصي عن هذه الطريقة .

مقدمة

يعتبر مفهوم التحيز bias في الاختبار من المفاهيم الحديثة نسبيا إذا ما قُورن ببقية المفاهيم المتعلقة بالاختبارات النفسية و التربوية ، إذ أن بدايته الحقيقية لم تكن قبل العقد الثامن من القرن الحالي . ويعود التحيز في نظر الباحثين في هذا المجال إلى " تفضيل مجموعة دون أخرى ليس بناء على الخصائص التي تقيسها أو يفترض أن يقيسها الاختبار ، وإنما بناء على جوانب ليست ذات علاقة بالسمة المقاسة " [١ ، ص٣٧٤] .

ونظرا لكون دراسة تحيز الاختبار لم تعد مطلبا ثانويا ، بل تعد ضرورة لكل الاختبارات النفسية [٢ ؛ ٣]، فإنه تم استحداث عدة طرق إحصائية لدراسة تحيز بنود الاختبارات النفسية و التربوية. بيد أن هذه الطرق تختلف من حيث الثراء الإحصائي لمعادلاتها ومن حيث سهولة الاستخدام وقلة العناء [٤] .ومن أهم هذه الطرق طريقة أنغوف- فورد المعتمدة على صعوبة البنود Transformed Item Difficulty ، وطريقة تمييز البنود ، وطرق مربع كاي ، وطرق الدلالة الفارقة للبند Differential Item Function (DIF) ونظرا لعدم توافر مادة علمية عن هذه الطرق باللغة العربية ، فإنه سيتم استعراض كل طريقة من الطرق السابقة لمعرفة كيفية تحديدها لمعنى التحيز ، ولإثراء المكتبة العربية في هذا المجال.

طريقة الصعوبة المحولة للبند Transformed Item Difficulty Method

يعرف أصحاب هذه الطريقة البند المتحيز على أنه ذلك البند الذي يبدو أنه أكثر صعوبة لمجموعة دون أخرى حين يُقارن ببقية بنود الاختبار [٥] .لذا فإن اختلاف صعوبة البند من فئة لأخرى تعد مقياسا لتحيز البند .ويتم حساب تحيز البند بهذه الطريقة وفقا للخطوات التالية :

١- حساب صعوبة البند لكل مجموعة على حدة في كل بند من بنود الاختبار .

٢- الحصول على قيمة " ز " (الدرجة المعيارية) المقابلة لحاصل طرح قيمة الصعوبة لكل فئة من واحد (١-ص) .

٣- تحويل قيمة "ز" إلى قيمة دلتا Delta value من خلال المعادلة المعروفة($\Delta = 4Z + 13$).

٤- من خلال استخدام مفهوم المحاور السينية والصادية ، يتم رسم قيم دلتا للمجموعتين بحيث تمثل كل نقطة في الرسم قيمة دلتا للمجموعتين على السؤال الواحد. (أي أن النقطة عبارة عن المنطقة التي يتقاطع فيها الخط الممتد من المحور السيني والخط الممتد من المحور الصادي).

٥ - يتم استحداث خط لتمثيل البيانات (كخط معامل الانحدار) بحيث يعتبر هذا الخط بمثابة المعيار لتحديد تحيز البند .فبقدر بُعد البند أو قربه من هذا الخط يتم الحكم على تحيزه من عدمه .ويمكن تحديد هذا من خلال المعادلة التالية:

$$d_j = \frac{aX_i - Y_i + B}{\sqrt{A^2 + 1}} \qquad (١)$$

حيث:

d_i = درجة تحيز البند

$$a = \frac{\left(S_y^2 - S_x^2\right) + \sqrt{\left(S_y^2 - S_x^2\right) + Y4R_{xy}^2 S_x^2 S_y^2}}{2R_{xy}S_xS_y}$$

$$b = \bar{y} - a\bar{x}$$

ويعبر X و Y عن قيم دلتا للمجموعتين S_y^2 و S_x^2 عن تباين المجموعتين ؛ أما R_{xy} فهو معامل الارتباط بين قيم المجموعتين .

الطريقة المعدلة لصعوبة البند المحولة Modified Item Difficulty Method

نظرا لاعتماد طريقة أنغوف-فورد على صعوبة البند فقط فقد اقترح أنغوف بعض التعديل على هذه الطريقة [٦]، إلا أن هذا التعديل لم يكن أفضل من الطريقة الأساسية [٧]، لذا فقد اقترح Shepard و آخرون [٧] تعديلا على هذه الطريقة يقوم على أساس اختزال معامل التمييز من قيم دلتا وذلك على النحو التالي:

١- استخدام معدلة الانحدار لحساب القيم المتنبأ بها لدلتا عن طريق قيم الارتباط الثنائي الحقيقي point-biserial correlation للبنود (للمجموعتين معا) .

٢- حساب البواقي residuals عن طريق طرح القيمة الفعلية لدلتا من القيمة المتنبأ بها .

٣- حساب قيمة التحيز - المشار إليها في الفقرة الخامسة من الطريقة الأساسية - من خلال استخدام قيم البواقي بدلا من قيم دلتا الأساسية .

طريقة معامل التمييز Item Discrimination Method

استخدم Green و Draper [٨] طريقة معامل الارتباط الثنائي الحقيقي point biserial بين درجة البند والدرجة الكلية لتقويم تحيز البند. وقد بُنيت هذه الطريقة على افتراض أن قدرة البند على التمييز بين ذوي الدرجات العالية والمنخفضة في الاختبار ينبغي أن تكون متماثلة لدى المجموعات المختلفة .ولحساب تحيز البند فإنه ينبغي اتباع الخطوات التالية :

١- حساب معامل الارتباط الثنائي الحقيقي للبند لكل مجموعة .

٢- تصنيف قيمة معامل الارتباط الثنائي الحقيقي - بناء على قيمة الوسيط - إلى تمييز جيد و تمييز غير جيد .

٣- يُعد البند متحيزا حين يكون ضمن البنود ذات التمييز الجيد لمجموعة و ضمن البنود ذات التمييز غير الجيد للمجموعة الأخرى .

٤- تحسب نسبة البنود المتحيزة من خلال قسمة عدد البنود المتحيزة على نصف عدد أسئلة الاختبار [٦ ،ص١٠٩] .

وللحصول على قيمة متصلة للتحيز في هذه الطريقة فقد استخدمت الفروق المطلقة لحساب معامل الارتباط الثنائي الحقيقي للمجموعتين لكل بند [٩]؛ لذا فإنه كلما زادت الفروق المطلقة بين القيمتين كان البند متحيزا .

طريقة كميلي كاي تربيعCamilli Chi-square Method

نظرا لأن شانمن كاي تربيع [١٠] يعتمد على الإجابات الصحيحة فقط ، ونظرا للعيوب المنطقية لمثل هذه الطريقة ، فقد دعا Camilliإلى استخدام كاي تربيع المعتاد Full Chi-square ، والذي يتم حسابه بناء على الإجابات الصحيحة والخاطئة معا. ويُعرَّف عدم التحيز حسب هذه الطريقة على أنه " تساوي نسبة الإجابات الصحيحة للمجموعتين في البند عندما تتساوى درجاتهم في الاختبار الذي يحتوي ذلك البند " [١١، ص ١٤٥]. وتقوم هذه الطريقة على افتراض أن الاختبار ثابت وصادق ومتجانس. ويحسب التحيز بهذه الطريقة من خلال خطوتين هما :

١-تحديد فئات القدرة على درجات الاختبار، وذلك بتقسيم درجات الاختبار إلى فئات .بيد أنه ينبغي عند تحديد الفئات مراعات الأمور التالية :

أ - أن لا تكون نسبة الإجابة الصحيحة في كل فئة ١،٠٠ .

ب- أن يكون في كل خلية على الأقل "١٠" إلى "٢٠" استجابة .

جـ- أن لا يقل الحد الأدنى للتكرار المتوقع في كل خلية عن ٥٪.

٢ - حساب قيمة كاي تربيع لكل فئة قدرة وجمعها لكل بند ثم اختبار دلالة الفروق [١١ ، ص ١٤٥]، ولمزيد من الإيضاح يمكن الرجوع إلى القاطعي [١] .

طريقة منتل-هنزل كاي تربيع Mantel-Haenzel Chi-square

يقارن منتل-هنزل كاي تربيع أداء مجموعة الأكثرية بأداء الأقلية على البند reference and focal groups في مستويات القدرة المختلفة [١٢]. وحين يكون البند غير متحيز فإن أداء المجموعتين سيكون متماثلا في ذلك البند .ويوضح شكل رقم ١ التصميم الإحصائي لمنتل- هنزل كاي تربيع لمستوى قدرة معين. إذ نجد أن A_j تمثل عدد أفراد مجموعة الأكثرية الذين أجابوا عن البند إجابة صحيحة في مستوى القدرة j، بينما تمثل B_j عدد أفراد مجموعة الأكثرية الذين أجابوا عن البند إجابة خاطئة في نفس مستوى القدرة j، وأن N_{Rj} هي حاصل جمع A_j و B_j. أما C_j و D_j ، فتعبران عن عدد من أجابوا إجابة صحيحة (C_j) ؛ وخاطئة (D_j) لمجموعة الأقلية ، وتمثل N_{Fj} حاصل جمع C_j و D_j. كذلك فإن N_{1j} و N_{0j} تمثل عدد من أجابوا عن البند صح أو خطأ ، على التوالي، في مستوى القدرة j .أما N_j فهي عبارة عن مجموع عدد المجموعتين .لذا فإنه سيكون عدد الجداول المماثلة لشكل رقم ١ مساويا لعدد مستويات القدرة .

شكل رقم ١. التصميم الإحصائي لمنتل- هنزل كاي تربيع لمستوى قدرة معين.

المجموع	الدرجة على البند		مجموعة	
	0	1		
N_{Rj}	A_j B_j		R	الأكثرية
N_{Fj}	C_j D_j		F	الأقلية
N_j	N_{0j}	A_{1j}		المجموع

ويتم حساب ألفا لهذا الأسلوب من خلال المعادلة التالية :

$$\overline{\alpha} = \frac{\Sigma a_j d_j n_j}{\Sigma b_j c_j n_j} \qquad (٢)$$

ويمكن أن تكون قيمة ألفا بين صفر و ∞ ، وسيكون البند غير متحيز عندما تكون قيمة ألفا واحدا [١٣] .ونظرا لعدم واقعية ألفا فقد اقترح Mantel و Haenzel مربع كاي لاختبار ما إذا كان ألفا الفعلي observed يختلف عن "١." و يتم حسابه من خلال المعادلة التالية :

$$X = \frac{\left(\left|\Sigma a_j \Sigma E(A_j)\right| - .5\right)^2}{\Sigma VAR(a_j)} \qquad (٣)$$

ويتم الجمع هنا (Σ) لكل مستويات القدرة أما $E(A_j)$ هي القيمة لـ A_j ، و $VAR(A_j)$ هي تباين A_j . ويمكن الحصول على القيمة المتوقعة و التباين بالمعادلات التالية :

$$E(a_j) = (N_{Rj})(N_{lj}) / N_j$$

$$VAR(a_j) = \frac{N_{Rj} N_{Fj} N_{lj} N_{oj}}{N_j^2 (N_j - 1)} \qquad (٤)$$

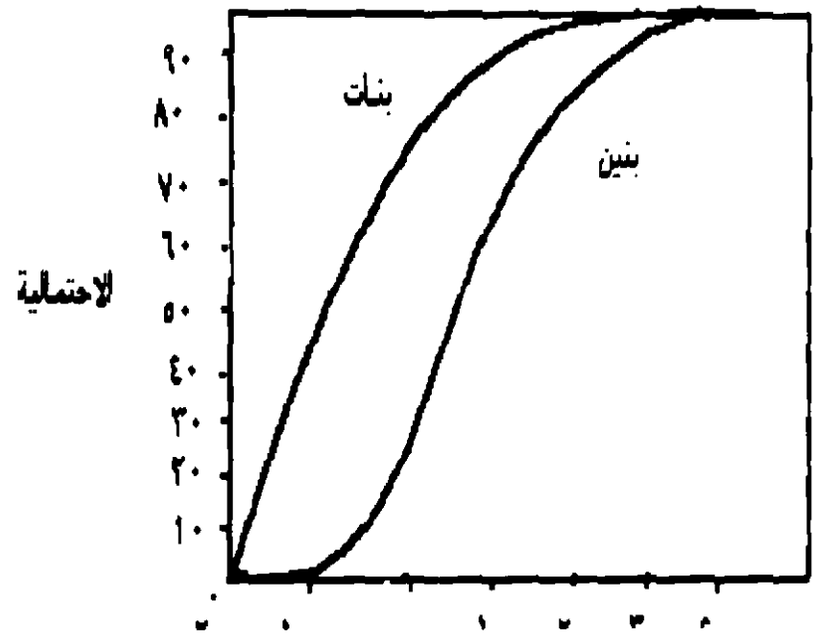
لذا فإن مربع كاي الموضح في المعادلة رقم (٣) يستخدم لاختبار مدى اختلاف قيمة ألفا الفعلي عن "١." وتكون درجة الحرية لمربع كاي "١" [١٤ ، ص ص٢-٣].

طريقة السمات الكامنة Item Characteristic Curve Method

تقوم هذه الطريقة على التحري عن العلاقة بين قدرة الأشخاص واحتمالية الإجابة عن البند إجابة صحيحة .وتُمثل هذه العلاقة بأنواع من المنحنيات منها ما يُعرف بـ normal ogive curve.

وهناك نماذج رياضية استحدثت للتعبير عن تلك العلاقة و تعتمد على معلْم parameter واحد - هو الصعوبة - Rasch Model ، أو معلْمين - هما الصعوبة والتمييز- أو ثلاثة معالِم - هي الصعوبة و التمييز و التخمين Birnbaum Model.

بناء على هـذه الطريقة "يكـون البنـد غـير متحـيز عندمـا تكـون احتماليـة الإجابة الصحيحة متماثلة للأشخاص المتساوية قدراتهم حتى وإن كانوا من مجموعات مختلفة [١٥ ، ص٢٢٣] .ويمكن توضيح ذلك من خلال شكل رقم ٢.



شكل رقم ٢. رسم توضيحي لبند متحيز. (فمثلا نجد أن احتمالية الإجابة الصحيحة لمستوى القدرة (أ) لكل من البنين والبنات مختلفة.

ويمكن حساب لـورد كـاي تربيع [١٦] مـن خـلال المعادلـة التاليـة لنمـوذج راوش Rasch Model-IPM.

$$X^2 = \frac{(b_1 - b_2)^2}{VAR(b_1) + VAR(b_2)} \qquad (٥)$$

وهنـا b_j و b_2 معْلمـي الصعوبـة للمجموعـة الأولى و الثانيـة، و $VAR(b_1)$ و $VAR(b_2)$ تباين معْلمي الصعوبة الأول والثاني على التوالي .

أما حساب لورد كاي تربيع لنموذج المعلْمين 2PM ، فيتم من خلال المعادلة التالية :

$$X^2 = \left(\underline{\zeta_1} - \underline{\zeta_2}\right)' \Sigma^{-1} \left(\underline{\zeta_1} - \underline{\zeta_2}\right)' \qquad (٦)$$

حيث إن $\underline{\zeta_1} = (a_1 b_1)'$ ؛ $\underline{\zeta_2} = (a_2 b_2)'$ ؛ و $\underline{\Sigma}$ عبارة عن مصفوفة للتشتت مكونة من٢×٢ وبالصيغة التالية $\underline{\Sigma} = \underline{\Sigma_1} + \underline{\Sigma_2}$ ، حيث إن $\underline{\Sigma_1}$ و $\underline{\Sigma_2}$ عبارة عن مصفوفة "التباين-التباين المشترك " للمجموعتين. فمثلا نجد أن هذه المصفوفة للمجموعة الأولى تكون بالشكل التالي :

$$\underline{\Sigma_1} = \begin{pmatrix} \text{var}(a_1) & \text{cov}(a_1, b_1) \\ \text{cov}(a_1, b_1) & \text{var}(b_1) \end{pmatrix} \qquad (٧)$$

حيث إن a_1 و b_1 هما معْلما الصعوبة والتمييز للمجموعة الأولى على التوالي.

أما اختباري المساحة ، فيمكن الرجوع إليهما في Raju [١٧ ؛ ١٨]، نظرا لكثرة المعدلات وسعيا وراء التبسيط .

طريقة معامل التمييز المعدلة

تعتمد طريقة Green و Draper في قياس التحيز على كون معامل التمييز في النصف الأعلى لمجموعة وفي النصف الأدنى للمجموعة الأخرى .كما تعتمد مقارنة هذه الطريقة ببقية الطرق على الفروق المطلقة لمعامل التمييز لمجموعتي المقارنة [٤ ؛ ٧] .وعلى الرغم من أن البند قد يكون في النصف الأعلى للمجموعتين ، ويكون هناك فروق كبيرة بين المعاملين (كأن يكون معامل تمييز البند للمجموعة الأولى ٠.٩٥ ، وللمجموعة الثانية ٠.٦٠) ، لذا فإن التعريف السابق لمفهوم التحيز قد لا يكون حساسا لمثل هذه الفروق مما يحد من قدرته على اكتشاف التحيز في مثل هذه البنود .لذا فإنه ليس بغريب أن تكون هذه الطريقة من أقل الطرق فعالية في دراسات التحيز (سيرد ذكرها) .وتلافيا لمثل هذا

الضعف، فإنه تم استحداث طريقة جديدة لقياس التحيز بناء على معامل التمييز تمثلت خطواتها في ما يلي :

١- يتم حساب معامل التمييز بالطرق الإحصائية المعروفة point-biserial لكل مجموعة على حدة (مثلا للبنين على حدة وللبنات على حدة) .

٢- تحول معاملات التمييز لكل مجموعة من المجموعات إلى مقابلاتها من الدرجة المعيارية بالطريقة المتعارف عليها إحصائيا (Z-transformation) .

٣- تختبر دلالة الفروق للدرجات المعيارية من خلال المعادلة التالية :

$$Z = \frac{Z_{r_1} - Z_{r_2}}{\sqrt{\frac{1}{n_1 - 3} + \frac{1}{n_2 - 3}}} \qquad (٨)$$

حيث Z_{r_1} هو قيمة الدرجة المعيارية لمعامل تمييز المجموعة الأولى على البند، هو قيمة الدرجة المعيارية لمعامل تمييز المجموعة الثانية على البند، و n_1 عدد المجموعة الأولى و n_2 عدد المجموعة الثانية .

٣- تختبر دلالة الفروق و ذلك بالرجوع إلى جدول توزيع المنحنى الاعتدالي ومقارنة قيمة Z بالقيمة الجدولية المقابلة لمستوى الدلالة الذي يحدده الباحث .فمثلا نجد أن البند سيكون متحيزا عند مستوى الدلالة ٠,٠٥ حين تكون قيمة Z مساوية أو أكثر من ١.٩٦ درجه معيارية .ونظرا لأن قيمة (Z) غير مستقلة عن حجم العينة، فإنه يُنصح باستخدام مستوى دلالة منخفض جدا خصوصا عندما يكون حجم العينة كبيرا .

٤- تستخدم قيمة (Z) المستخرجة للبنود كأساس لحساب معامل الارتباط الرتبي Spearman rank order correlation بين هذه الطريقة وطرق التحيز الأخرى .

٥- تحسب نسبة الاتفاق بناء على عدد البنود المتحيزة حسب ما ورد في الفقرة "٣." لذا فإن البند يعتبر غير متحيز-في هذه الطريقة- عندما لا يكون الاختلاف بين معامل تمييز المجموعة الأولى ومعامل تمييز المجموعة الثانية دالا إحصائيا .

هدف الدراسة

خصت جميع الدراسات المقارنة لطرق تحيز البنود طرق الوظيفة الفارقة للبند - Differential Item Function - DIF-1 , DIF-2 , DIF-3 ، وأنغوف، ومعامل التمييز، وكاي تربيع سواء كانت مجتمعة أو متفرقة بالمقارنة، لذا فإن طريقة منتل-هنزل، والتعديل الذي طرأ على طريقة أنغوف، لم يتم مقارنتها - حسب علم الباحث - ببقية الطرق .لذا نجد أن Raju وآخرين [١٤] يقولون إنه " على الرغم من أن الأساس النظري لطريقة منتل-هنزل يعد قويا، إلا أن إمكانية استخدامه لتقدير تحيز البنود أو لتحديد الأداء الفارق للبند DIF يحتاج إلى مزيد من التحقق الامبريقي " [١٤ ، ص٤].

كذلك اعتمدت مقارنة معامل التمييز بالطرق الأخرى على مفهوم متعارض مع تعريف هذه الطريقة للتحيز .فبيانات المقارنة لهذه الطريقة تعتمد على الفروق المطلقة بين معاملات التمييز للفئات المقارنة [٩]، في حين عرف البند المتحيز بأنه ذلك البند الذي يكون معامل تمييزه في النصف الأعلى لإحدى المجموعات وفي النصف الأدنى للمجموعة الأخرى .لذا فإنه عندما تكون الفروق كبيرة بين معاملي التمييز للفئتين حتى لو كان المعاملان في النصف الأعلى للمجموعتين، فسوف ينظر إلى البند على أنه متحيز في الدراسات المقارنة وعلى أنه غير متحيز بموجب التعريف .ويؤكد هذا التصور ضعف نسبة الاتفاق مع الطرق الأخرى، كما سنرى .

لكل ماسبق، فإن الدراسة الحالية تهدف إلى مقارنة طرق تحيز البنود الأكثر شهرة مضافا إليها طريقة منتل - هنزل وتعديل Shepard وآخرين [٧] لطريقة أنغوف،

والتعديل المقترح في هذه الدراسة لطريقة معامل التمييز، وعليه فإن البحث الحالي يحاول الإجابة عن الأسئلة التالية :

١- ما مدى العلاقة بين نموذج المعلمين DIF-2 وكل طريقة من طرق تحيز البنود الأخرى ؟

٢-ما مدى علاقة بقية الطرق الأخرى ببعضها البعض ؟

٣- ما مدى الاتفاق بين نموذج المعلمين DIF-2 وكل طريقة من طرق تحيز البنود الأخرى ؟

٤- ما مدى الاتفاق بين طريقة منتل-هنزل وبقية الطرق الأخرى ؟

٥-ما مدى الاتفاق بين طريقة أنغوف المعدلة وبقية الطرق الأخرى ؟

٦-ما مدى الاتفاق بين طريقة معامل التمييز وبقية الطرق الأخرى ؟

مقارنة طرق تحيز بنود الاختبار

تركز دراسات المقارنة بين الطرق المختلفة على علاقة الطرق ببعضها، ومدى الاتفاق بينها في التعرف على التحيز .ونظرا لأن دراسة التحيز تعد من المواضيع الحديثة نسبيا، وكذلك الطرق المستخدمة لقياسه، فإن الدراسات المقارنة في هذا الجانب تعد قليلة نسبيا، إلا أن الاستعراض الحالي سيتم للدراسات التمثلية simulated studies أولا، ثم للدراسات الإمبريقية empirical studies ثانيا .

الدراسات التمثلية Simulated studies

تستخدم الدراسات التمثلية طريق منتكارلو Monte Carlo procedure لاستحداث البيانات إذ يمكن أن تحدد طبيعة ودرجة التحيز في البيانات مسبقا .لذا فإن هذه الطريقة يمكن أن تقوم ليس مجرد الاتفاق بين الطرق، بل التحيز الوهمي الذي يتراءى لكل طريقة من الطريق [١٩].

درس Rudner وآخرون [١٥] طريقة أنغوف، وكاي تربيع بمستويات مختلفة من القدرة، ونموذج راوش DIF-1 ، وطريقة الثلاثة معالم DIF-3 . وقد تم استحداث عدد مختلف من مستويات التحيز في الصعوبة والتمييز (أربعة في كل منها)، وعدد مختلف من البنود (٢٠، ٤٠، ٦٠، ٨٠ بندا) .وأوجدت العلاقة بين التحيز المستحدث و التحيز المكتشف لبنود الاختبار بالطرق المختلفة، فكان نموذج الثلاثة معالم DIF-3 أعلاها قيمة ٠,٧٩ ، يليه مربع كاي ٠,٧٢ ، ثم طريقة أنغوف ٠,٦٧ ، فنموذج راوش DIF-1 ٠,٥٥ .

كما كان معامل الارتباط بين التحيز المستحدث في صعوبة البند وكل من طريقة أنغوف ومربع كاي ٠,٨٧ و ٠,٨٤ على التوالي، وتعدان أفضل من نموذج راوش .أما عندما كان التحيز في تمييز البند، فقد كان مربع كاي أفضل الطرق ٠,٨١ ،بعد طريقة الثلاثة معالم، يليه نموذج راوش ٠,٧١ .بيد أن طريقة أنغوف تعد أقل الطرق حساسية للتحيز الناتج عن تمييز البند، إذ كانت قيمة معامل الارتباط ٠,٤٧ .

وقد خلص Rudner وآخرون إلى القول إن طريقة الثلاثة معالم DIF-3 ومربع كاي (ذو خمسة مستويات للقدرة) وطريقة أنغوف تعد طرقا جيدة لدراسة التحيز، بيد أن طريقة الثلاثة معالم تعتمد على دقة تقدير المعالم، في حين أن مربع كاي يعتمد على عدد مستويات القدرة .أما طريقة أنغوف، فليست حساسة للتحيز الناتج عن تمييز البند [١٩ ،ص١٦٣] .

أما Merz و Grossen [٢٠]، فقد درسا العلاقة بين التحيز المستحدث في مستويين للصعوبة (٦٠٪، ٨٠٪) والتحيز المكتشف بكل طريقة من طرق التحيز التالية :نموذج الثلاثة معالم، ونموذج راوش، وكاي تربيع، ومعامل التمييز، وأنغوف .وقد أظهرت النتائج أن معاملات الارتباط تراوحت بن ٠,٠٧ و ٠,٩٨ و أن طريقة أنغوف كانت أفضل الطرق (تراوحت معاملات ارتباطها بين ٠,٩٥ و ٠,٩٨)، يليها مربع كاي، حيث تراوحت معاملات الارتباط بين ٠,٧٥ و ٠,٩١ ؛ أما طريقة التمييز، فلم تكن ذات فعالية جيدة .

من خلال الاستعراض السابق اتضح أن طريقة الثلاثة معالم قد أظهرت فعالية أكثر من بقية الطرق .إلا أن هذه الطريقة هي نفسها التي استخدمت لاستحداث التحيز في المقام الأول، لذا فإنة من المتوقع أن لا تكون فعالية بقية الطرق مماثلة لفعالية طريقة الثلاثة

معالم .بل إن Burrill اعتبر أنه "حين نظن أننا قد استنتجنا الطريقة المثلى لدراسة التحيز من خلال الدراسات التمثلية نوعا من التفكير الأجوف" [١٩ ، ص١٦٦] نظرا لأننا نستحدث البيانات و نحتكم إلى طريقة واحدة هي طريقة الثلاثة معالم .

الدراسات الإمبريقية Empirical studies

قارن Nugester [٢١] طريقة أنغوف وشانمن كاي تربيع وطريقة فشباين لمعرفة مدى تحيز اختبار فلوردا المحلي للصف التاسع للجنس .و تراوحت معاملات الارتباط الرتبية بين ٠,٦٥ و ٠,٧٧ للأسئلة المتعلقة باللغة ، وبين ٠,٨٧ و ٠,٩٨ لأسئلة الرياضيات. وأوضح أن نتائج الطرق الثلاث تكاد تكون متماثلة .

كذلك قارنت Instasuwan [٢٢] طريقة شانمن كاي تربيع المعدل ونماذج المعلم الواحد و الثلاثة معالم على بيانات عينات أمريكية و إنجليزية ونيوزيلاندية مستقاة من اختبار فهم المقروء التابع للجمعية العالمية لتقويم التحصيل .وقد تراوحت معاملات الارتباط بين ٠,٥١ و ٠,٩٨ ، وقد كان أعلى معاملات الارتباط بين طريقة المعلم الواحد ومربع كاي ٠,٩٨ ، يليه معامل الارتباط بين طريقة الثلاثة معالم ومربع كاي ٠,٥٧ ، ثم بين طريقتي المعلم الواحد و الثلاثة معالم ٠,٥١ .وقد خلصت إلى القول إن وجود تقارب بين هذه الطرق الثلاث يعزز إمكانية استخدام مربع كاي أو راوش حين تكون العينات صغيرة ،بيد أن هناك اختلافا بين هاتين الطريقتين وطريقة الثلاثة معالم، مما يعني أن لديها حساسية نحو بعض الفروق التي قد تكون مختلفة عن تلك التي تكتشفها طريقة الثلاثة معالم [٢٢ ، ص٩٥] .

أما Ironson و Subkoviak [٤]، فقد قارنا نموذج الثلاثة معالم، وطريقة أنغوف ومعامل التمييز ومربع كاي لدراسة تحيز عينة من البيض والسود في بطارية الدراسة الطولية الوطنية .وعلى الرغم من أن معاملات الارتباط منخفضة بشكل عام بين الطرق المختلفة عندما لايؤخذ اتجاه التحيز في الحسبان (unsigned) ، إلا أن أعلى معاملات الارتباط كانت بين مربع كاي وطريقة الثلاثة معالم ٠,٤٩ ، يليها الارتباط بين طريقتي

مربع كاي وأنغوف ٠,٣٧ .أما عندما يوضح لصالح أية فئة كان التحيز signed ، فإن معاملات الارتباط ارتفعت قيمتها بشكل عام، فكان معامل الارتباط بين مربع كاي وطريقة أنغوف ٠,٦٣ ، وبين كاي تربيع وطريقة الثلاثة معالم ٠,٥٨ ، وبين طريقة أنغوف وطريقة الثلاثة معالم ٠,٤٩ .أما معامل التمييز ، فلم يكن ارتباطه عاليا في كلتي الحالتين (حين يوضح أو عندما لا يوضح لصالح من كان التحيز)، إذ لم يتعد ٠,١٤ في أفضل أحواله.

وقد كانت نسبة الاتفاق بين الطرق في البنود الـ ٢٤ الأكثر تحيزا - حسب كل طريقة - متقاربة في نتائجها مع نتائج معاملات الارتباط، حيث كان الاتفاق بين طريقة الثلاثة معالم ومربع كاي ٥٤٪ ، يليها الاتفاق بين مربع كاي وطريقة أنغوف ٣٨٪ ، ثم طريقة أنغوف وطريقة الثلاثة معالم ٣٣٪ .لذا فقد أشار Ironson و Subkoviak إلى أنه يمكن استخدام طريقة كاي تربيع أو طريقة أنغوف نظرا لعدم حاجتهما لعينات كبيرة أو برامج إحصائية متقدمة [٤ ، ص٢٢٢].

كذلك قارن Shepard وآخرون [٩] طريقة أنغوف، وطريقتي كاي تربيع (شائمن و كميلي)، ومعامل التمييز، وطريقتي المعلم الواحد والثلاثة معالم لعينة من البيض والسود وذوي الأصول المكسيكية على اختبار لورج-ثورندايك للقدرة العقلية .وقد استخدمت المحكات الداخلية لتحديد القدرة (الدرجة الكلية للاختبار) والمحكات الخارجية (اختبار المصفوفات لريفن) .كما استخدم مفهوم التحيز المطلق والتحيز لصالح فئة معينة signed and unsigned procedures وقد تراوحت معاملات الارتباط بين -٠,٠٤ و ٠,٦٨ حين كان المحك داخليا .وقد كانت أهم النتائج تلك المرتبطة بعلاقة كميلي كاي تربيع بطريقة الثلاثة معالم وبطريقة المعلم الواحد، إذ كان الارتباط مع الأولى ٠,٦٨ ، ومع الثانية ٠,٥٩ .أما طريقة أنغوف فقد كان ارتباطها بطريقة الثلاثة معالم ٠,٥١ ، وبطريقة المعلم الواحد ٠,٩٨ ، في حين كان ارتباط طريقة معامل التمييز بطريقة الثلاثة

معالم ٠,٣٩ ، وبطريقة المعلم الواحد ٠,٢٦ .أما شانمن كاي تربيع ، فقد كان ارتباطه بطريقة الثلاثة معالم ٠,٥٨ ، وبطريقة المعلم الواحد ٠,٥٢ .وعلى الرغم من أن طريقتي كاي تربيع تُعد أفضل من طريقة أنغوف ، وأن كميلي كاي تربيع يُعد أفضلها جميعا - نظرا لارتباطه العالي بطريقة الثلاثة معالم - إلا أن بعض المشاكل المنهجية في هذه الدراسة تجعل التحفظ على قبولها مُفضلا. فقد استخدمت طريقة الثلاثة معالم كمحك للكحم على بقية الطرق في حين أن العينة أقل مما يمكن قبوله لهذه الطريقة [١٩].

كما قارن Burrill [١٩] سبع طرق لدراسة التحيز ، منها أربع طرق تتعلق بصعوبة البنود وطريقتين للتمييز وطريقة كاي تربيع على بيانات عينات متماثلة وعشوائية من البيض والسود على اختبار Metropolitan Readiness . وقد كانت معاملات الارتباط عالية بين الطرق المتعلقة بالصعوبة .كما كان ارتباط مربع كاي بطريقة أنغوف ومثيلاتها جيدا (بين ٠,٦٠ و ٠,٦٩ للعينة العشوائية ، و بين ٠,٧٠ و ٠,٧٩ للعينة المتماثلة) .كما كانت معاملات الارتباط بين طرق معامل التمييز عالية ، فقد كان الوسيط لمعاملات الارتباط فوق الـ ٠,٩٢ للعينات المختلفة .وعلى الرغم من أن ارتباط معاملات التمييز بطريقة الصعوبة أو مربع كاي كان متوسطا إلا أن ارتباطها بمربع كاي كان أعلى من ارتباطها بمعاملات الصعوبة .وقد خلص Burrill إلى القول إن " طريقة كاي تربيع تستطيع التعرف على جوانب لا يمكن التعرف عليها من خلال طرق الصعوبة " [١٩ ، ص١٧٢] .

وقد درس Raju و Normand [٢٣] إمكانية استخدام أسلوب الانحدار الإحصائي لدراسة تحيز البنود و الاختبار وقارناه ببعض الطرق الأخرى على بيانات عينة من البيض والسود على اختبار جمعية البحث العلمي (سلسلة الاختبارات التحصيلية) .وقد تراوحت معاملات الارتباط للرياضيات بين -٠,٠٣٩ و ٠,٩٢ ، حيث كان ارتباط المساحة لطريقة الثلاثة معالم بمساحة طريقة المعْلمين ٠,٣٤ ، وبمساحة نموذج راوش ٠,٢٢ ، وبفرق الصعوبة لنموذج راوش ٠,٤٥ ، وبكميلي كاي تربيع ٠,٤٠ ، وبمعامل التمييز ٠,١٩ ،

وبطريقة أنغوف ٠,٤٨ .أما طريقة المعلمين فقد كانت علاقتها بطريقة أنغوف ٠,٦١ ، وبطريقة معامل التمييز ٠,١٩ ، وبطريقة كميلي كاي تربيع ٠,٥٣ ، وبنموذج راوش (الصعوبة) ٠,٣٤ . أما ارتباط طريقة كميلي بطريقة أنغوف فكان ٠,٦٨ ، وبمعامل التمييز ٠,٤٧ .وقد كانت الارتباطات على اختبار اللغة مماثلة إلى حد كبير لتلك المستعرضة لاختبار الرياضيات.

كما قارنت Perlman وآخرون [٢٤] طريقة أنغوف وطريقة منتل-هنزل ونموذج المعلم الواحد وطريقة أنغوف المعدلة من حيث العلاقة و الثبات على بيانات عينة من البنين والبنات وفئات مختلفة الحجم من أعراق مختلفة في اختبار شيكاغو لإتقان المهارات. وقد أظهرت النتائج أن معاملات ارتباط نموذج المعلم الواحد بالطرق الأخرى تراوحت بين ٠,٨٥ و ١,٠٠ ؛ فقد كان أعلى ارتباطاتها مع منتل-هنزل ، وأقلها مع طريقة أنغوف المعدلة .أما عندما زيد حجم العينة فقد كان ارتباط طريقة راوش بطريقة أنغوف المعدلة ٠,٨٢ و بطريقة منتل-هنزل ١,٠٠ .أما ثبات الطرق السابقة فقد تراوح بين ٠,٨٣ و٠,٧٦ للعينات الكبيرة (ن=١٠٠٠) ، حيث كانت طريقة راوش ومنتل-هنزل أعلى الطرق ثباتا. وقد انخفضت قيمة الثبات للعينات الصغيرة ، بيد أن هذه الطرق حافظت على نفس الترتيب السابق من حيث قيمة الثبات (تراوحت قيمة الثبات بين ٠,٥٩ و ٠,٤٨ عندما كانت ن=٣٠٠) .

أما DeMauro [٢٥] فقد درس أثر نسبة البنين إلى البنات على طريقة منتل-هنزل. وأظهرت نتائجه عدم وجود تفاعل بين مستوى الصعوبة وحجم عينة الأقلية ، كما أظهرت- أيضا - أنه عندما تكون قيمة منتل-هنزل بالسالب ، فإن حجم العينة يؤثر على صعوبة البند في كل العينات المدروسة ، مما جعل DeMauro يقول بضرورة تماثل العينة المستخدمة من المجتمع الأساسي في القدرة لكل من الأقلية و الأكثرية ، وفي التمثيل العام ، وفي تمثيل المستويات المختلفة من القدرة حتى نحصل على تقدير جيد لمعلم هذا الإحصائي.

من خلال الاستعراض السابق يتضح أن أفضل الطرق لدراسة التحيز هي طرق السمات الكامنة. بيد أن هذه الطرق تحتاج إلى عينات كبيرة وطرق إحصائية معقدة مما يجعل الاستفادة منها محدودة. أما بالنسبة لبقية الطرق الأخرى فإنه يمكن القول إن طرق كاي تربيع وبالذات طريقة كميلي تعد أفضل الطرق نظرا لعلاقتها الجيدة بطرق السمات الكامنة يليها طريقة أنغوف. أما طرق معامل التمييز فلم تكن من بين الطرق الجيدة لقياس التحيز.

الإجراءات

العينة

استخدمت بيانات عينات اختيار البنود وتقدير ثبات وصدق اختبار وكسلر لذكاء الأطفال المعدل (الصورة السعودية) كأساس لتقدير تحيز البنود. وقد تكونت العينة من ٦١٦ طفلا و٥٦٠ طفلة موزعة على الأعمار من ٦ سنوات إلى ١٦ سنة. وقد تم اختيار العينة عشوائيا بناء على توزيع الأعمار على المناطق المختلفة لمدينة الرياض [٢٦].

الأداة

استخدمت بيانات أحد اختبارات وكسلر لذكاء الأطفال المعدل (الصورة السعودية) وهو اختبار المعلومات كأساس لمقارنة طرق تحيز البنود .ويتكون الاختبار من ٣٠ سؤالا في المعلومات العامة ، حيث يطلب من المفحوص تقديم الإجابة عن كل سؤال. ويعد الاختبار ثابتا إذ تراوح ثباته بين ٠.٨٠ و ٠.٩٣ للأعمار المختلفة بمتوسط قدره ٠.٨٨ .وعن صدق الاختبار فقد تشبع-مع بقية الاختبارات- على العامل العام وعلى العامل اللفظي الذي يندرج تحته من الناحية النظرية[٢٦] .وعليه يمكن القول إن الاختبار

يتمتع بقدر عال من الثبات وأن مؤشرات الصدق تدل على أنه يتمتع بقدر جيد من الصدق.

كما أجري تحليل عاملي لبنود الاختبار وقد فسر العامل الأول ٢١.٢٪ من التباين، كما فسرت العوامل مجتمعة ٦٦.٤٪ من التباين .لذا فإن الاختبار قد استوفى الحد الأدنى اللازم لضمان تقدير مستقر لمعالم نماذج السمات الكامنة الذي حدده Reckase [٢٧] وهو ٢٠٪ .

التصميم الإحصائي

تم حساب معامل أنغوف، وأنغوف المعدل، وطريقة معامل التمييز، وطريقة معامل التمييز المعدلة، وكميلي كاي تربيع، ومنتل-هنزل كاي تربيع، وطرق الوظيفة الفارقة للبند DIF-1 , DIF-2 , DIF-3 بناء على الخطوات الموضحة سابقا لتقدير تحيز بنود اختبار المعلومات لمجموعتي البنين والبنات .

طرق السمات الكامنة

اعتبر نموذج المعلمين DIF-2 محكا يحتكم إليه بدلا من نموذج الثلاثة معالم DIF-3 نظرا "لصعوبة تقدير معلم التخمين (C) حتى مع توافر عينة كبيرة " [٩، ص٣٢٥]، ونظرا لأنه ليس هناك بدائل يختار من بينها المفحوص في اختبار المعلومات وليست الأسئلة من نوع الصح والخطأ، وعلى المفحوص أن يقدم الإجابة بنفسه .لذا فإن فرصة التخمين في الإجابة عن الأسئلة نادرة إن لم تكن معدومة، وبهذا يكون استخدام نموذج المعلمين أكثر ملاءمة لطبيعة هذا الاختبار من نموذج الثلاثة معالم .أما إدراج نموذج الثلاثة معالم فلم يكن لاعتماده محك لبقية الطرق ولكن لمقارنته بها، وقد حسب معلم

التخمين لهذه الطريقة على العينة الكلية (البنين والبنات معا)، وتم تثبيته بعد ذلك في المعادلة - بناء على نتائج العينة الكلية - لكل من البنين والبنات [٩].

وقد تم استخدام برنامج Bilog-3 [٢٨] لتقدير معْلمي الصعوبة والتمييز لبنود الاختبار - لطرق السمات الكامنة. ونظرا لعدم تماثل وحدات القياس للمعالم المستخرجة من بيانات المجموعات المختلفة، فإنه تم توحيد وحدات القياس للمجموعتين من خلال استخدام طريقة Loyd و Hoover mean and sigma equating method [٢٩]. وتعد هذه الطريقة مماثلة في الفعالية لبقية الطرق حين تكون العينة كبيرة (أكثر من ٥٠٠) [٣٠؛ ٣١].

وقد حُسبت الوظيفة الفارقة للبند DIF من خلال استخدام برنامج IRTDIF [٣٢]. ويقدم هذا البرنامج :

١ - لورد كاي تربيع لاختبار دلالة الفروق لمعْلم الصعوبة DIF-1 ، أو معْلمي الصعوبة والتمييز DIF-2 and DIF-3 with fixed C parameter.

٢ - المساحة المؤشرة بين منحنيي المجموعتين signed area .

٣- المساحة المطلقة بين منحنيي المجموعتين unsigned area .

وقد أعتبر البند متحيزا لاختباري المساحة (المطلق والمؤشر signed and unsigned area عندما تكون قيمة الدرجة المعيارية (Z) ± ٣ أخذا بنصيحة Raju [١٨] نظرا "لأن الانحراف المعياري لكلي الاختبارين يعتمد على حجم العينة " [١٨ ، ص٢٠٢].

طريقة أنغوف

تم حساب معامل الصعوبة لبنود الاختبار لكل من البنين والبنات على حدة، وحسبت قيمة دلتا، وحسبت قيمة (d_i) بناء على المعادلة الموضحة سابقا .

جدول رقم ١ . معاملات ارتباط الرتب لطرق تحيز بنود الاختبار .

	نموذج الثلاثة معالم		نموذج المعلمين			راوش		أنغوف		كاي تربيع		معامل	معامل
	CSA-3	CUA-3	DIF-2	SA-2	UA-2	DIF-1	SA-1	أنغوف المعدلة		كميلي هنزل	منتـــل-	التمييز	التمييز المعدل
نموذج الثلاثة معالم:													
DIF-3 و	٠,٢٠	-٠,١٠	٠,٦٥	٠,٥٧	٠,٢١	٠,٦٤	٠,٦٧	٠,٥٢	٠,٣١	٠,٥٣	٠,٥٣	٠,٤٠	٠,٦١
CSA-3 و			٠,١٧	٠,٤٠	-٠,٤	٠,٢٠	٠,٢٣	٠,٠٠	٠,٢٦	٠,٣٣	٠,٣٣	٠,١٤	٠,٠٧
CUA-3 و		٠,٥٢	٠,٣٥	٠,٣٧	-٠,١٩	٠,٥٠	٠,٣٤	٠,٠٠	٠,٤٦	٠,٥٣	٠,٥٤	٠,٤٠	٠,٣٥
نموذج المعلمين :													
DIF-2 و													
SA-2 و			٠,٨٧			٠,٧٣		٠,٨١	٠,٣٩	٠,٨٤	٠,٨٣	٠,٣٨	٠.٤٦
UA-2 و			٠,٠٢	-٠,٠٢		-٠,٠٨		٠,٢	-٠,١٤	-٠,٠٣	-٠,٠٣	-٠,١٦	-٠,١٦
راوش :													
DIF-1 و			٠,٨٦										
SA-1 و			٠,٦٤	٠,٧١	٠,٠٤	٠,٨٣		٠,٦٤	٠,٥٩	٠,٧٩	٠,٧٩	٠,٦٠	٠,٦٦
أنغوف			٠,٤٢			٠,٦٤		٠,٢٧					
أنغوف المعدله			٠,٧٧			٠,٥٧							
كاي تربيع :													
كميلي			٠,٨١			٠,٨٤		٠,٧٠	٠,٧٣				
منتل-هنزل			٠,٨١			٠,٨٤		٠,٦٩	٠,٧٤	٠,٩٩			
معامل التمييز			٠,٥٣			٠,٦٨		٠,٢٨	٠,٧٦	٠,٦٠	٠,٦٠		
معامل التمييز المعدل			٠,٦٢			٠,٧٧		٠,٤٠	٠,٦٩	٠,٦٨	٠,٦٨	٠,٨١	

طريقة أنغوف المعدلة

تم حساب قيم دلتا كما في طريقة أنغوف، ثم حُسبت قيمة البواقي و اعتبرت أساسا لحساب قيمة (d_i) وقد أعتبر البند متحيزا في هذه الطريقة وفي الطريقة الأساسية عندما تكون قيمة d_i للمقارنة بين البنين و البنات أعلى من قيمته لمقارنة عينتين من البنين، حيث أوصى Angoff [٦] باتباع هذا الإجراء نظرا لأن المجموعات المتشابهة ينبغي أن تمثل قيمة d_i فيها أساس المقارنة لمعرفة تحيز البند .

طريقة كميلي كاي تربيع

لتحديد مستويات القدرة، تمت مراعات الأمور الموضحة سابقا وشروط كاي تربيع، ونتج عن ذلك ثلاثة مستويات للقدرة .وقد تم حساب كاي تربيع لكل مستوى من هذه المستويات، وتم جمع قيمة كاي تربيع لهذه المستويات الثلاثة بعد تحديد اتجاه التحيز في كل مستوى من مستويات القدرة، وذلك بوضع إشارة ناقص عندما يكون البند متحيزا لصالح البنين، وإشارة زائد عندما يكون البند متحيزا لصالح البنات [٣٣، ص١٢٧]، ويطلق على هذا الاختبار كاي تربيع المؤشر Signed chi-square test.

طريقة منتل–هنزل

تم تحديد مستويات القدرة بناء على شروط كاي تربيع الموضحة آنفا، وقد تم التوصل إلى ثلاثة مستويات للقدرة - كما هو الحال في طريقة كميلي كاي تربيع .وقد كانت عينة البنين هي مجموعة الأكثرية reference group، وعينة البنات هي مجموعة الأقلية focal group.

طريقة معامل التمييز

حُسب معامل تمييز البنود لكل من البنين والبنات، واعتبر البند متحيزا إذا كانت قيمة معامل التمييز أكثر من قيمة الوسيط لإحدى المجموعات، وأقل من قيمة الوسيط للمجموعة الأخرى، علما بأنه تم حساب الوسيط بناء على بيانات العينة الكلية (البنين

والبنات معا) -كما ذُكر سابقا - واستخدمت الفروق بين قيم معاملات التمييز أساسا للمقارنة ببقية الطرق.

طريقة معامل التمييز المعدلة

حُسبت قيمة معامل التمييز كما في الطريقة الأساسية، وحولت إلى علامات معيارية، وحُسبت دلالة الفروق بين العلامات المعيارية للبنين والبنات على كل بند من البنود، بالطريقة المُشار إليها سابقا .واعتبر البند متحيزا عندما تكون قيمة Z (العلامة المعيارية) المطلقة تساوي ٢,٥٨ أو أكثر .

النتائـــــــج

يوضح جدول رقم ١ معاملات ارتباط سبيرمان للرتب بين طرق التحيز المختلفة .وقد تراوحت معاملات الارتباط بين -٠,٤٠ و ٠,٩٩ .ونظرا لاستخدام طريقة نموذج المعْلمين كمحك لبقية الطرق، فإنه سيتم استعراض علاقة هذه الطريقة ببقية الطرق أولا، ثم استعراض علاقة بقية الطرق ببعضها البعض ثانيا .إلا أنه سيتم مقارنة الطرق التي تحتاج لعينات كبيرة أولا، ثم الطرق الأخرى .

طرق السمات الكامنة

يوضح جدول رقم ١ أن أعلى ارتباط لنموذج المعْلمين بطرق السمات الكامنة الأخرى كان بطريقة المعْلمين المعتمدة على المساحة المؤشرة signed area ، حيث بلغ ٠,٨٧ ، يليه الارتباط بطريقة راوش ٠,٨٦ ، ثم بطريقة الثلاثة معْالم ٠,٦٥ ، فالمساحة المؤشرة لنموذج راوش ٠,٦٤ .أما أدنى ارتباط لطريقة المعْلمين فقد كان بالمساحة المطلقة لنموذج المعْلمين ٠,٠٢ ، وبالمساحة المؤشرة المغلقة لنموذج الثلاثة معالم closed signed area ٠,١٧ ، وكذلك المساحة المطلقة المغلقة لنموذج الثلاثة معْالم ٠,٣٥ .

جدول رقم ٢ . نسب الاتفاق بين طريقة المعلمين ، طريقة أنغوف ، طرق معامل التمييز ، وبقية طرق تحيز بنود الاختبار .

	نموذج الثلاثة	نموذج المعلمين			راوش		أنغوف		كاي تربيع		معامل
	معالم DIF-3	DIF-2	SA-2	UA-2	DIF-1	SA-1	أنغوف المعدله	أنغوف	كميلي	منتل-هنزل	التمييز
نموذج المعلمين DIF-2	٨٣%	-	٩٢%	٧٥%	٨٣%	٩٢%	-	-	٧٥%	٧٥%	-
أنغوف	٤٧%	٧٥%	٧٣%	٦٠%	٥٣%	٦٠%	-	-	٤٧%	-	-
أنغوف المعدله	٦٩%	٧٥%	٦٩%	٦٢%	٦٢%	٦٩%	٧٧%	-	٧٧%	-	-
معامل التمييز	٣٠%	٤٢%	٢٧%	٤٤%	٣٨%	٣٣%	٢٧%	٢٣%	٣٣%	٢٧%	-
معامل التمييز المعدل	٧٥%	٧٥%	٦٧%	٥٨%	٨٣%	٩٢%	٥٨%	٦٧%	٦٧%	٦٧%	٣٣%
منتل-هنزل	٦٧%	٧٥%	٧١%	٢٣%	٧٥%	٧٥%	٤٧%	٧٧%	٨٥%	-	-

طرق التحيز الأخرى

بلغ أعلى ارتباط للطرق الأخرى بنموذج المعلمين ٠,٨١ ، وأقل ارتباط ٠,٤٢. وقد كان كميلي كاي تربيع ومنتل-هنزل أعلى الطرق ارتباطا بنموذج المعلمين، حيث بلغت قيمة الارتباط ٠,٨١ لكل منهما، يليهما طريقة أنغوف المعدلة ٠,٧٧ ، ثم طريقة التمييز المعدلة ٠,٦٠ ، فطريقة التمييز ٠,٥٣ ، وأخيرا طريقة أنغوف ٠,٤٢ . وتعد هذه النتائج إجابة عن التساؤل الأول والمتعلق بمدى العلاقة بين نموذج المعلمين وبقية الطرق الأخرى.

ارتباط الطرق الأخرى ببعضها

أظهرت النتائج أن أعلى معامل ارتباط لطريقة الثلاثة معالم DIF-3 كان مع طريقة المساحة المؤشرة لنموذج راوش ٠,٦٤ ،ثم مع طريقة التمييز المعدلة ٠,٦١ ، ثم مع طريقة كميلي وطريقة منتل-هنزل ٠,٥٣ -لكل منهما - ثم مع طريقة أنغوف المعدلة ٠,٥٢ ، بالإضافة إلى طريقة معامل التمييز ٠,٥١ .أما ارتباط طريقة أنغوف بطريقة الثلاثة معالم فقد كان ضعيفا، ٠,٣١ .

أما طريقة المساحة المغلقة المؤشرة CSA-3 . فقد تراوحت ارتباطاتها ببقية الطرق بين -٠,٤٠ و ٠,٣٣ ، حيث كان ارتباطها بالمساحة المطلقة لطريقة المعلمين -٠,٤٠ وكان أعلى ارتباط لها مع طريقة كميلي كاي تربيع ، بيد أن كل الارتباطات ضعيفة بشكل عام.

أما طريقة المساحة المطلقة المغلقة لنموذج الثلاثة معالم CUA-3، فقد تراوحت ارتباطاتها ببقية الطرق بين -٠,١٩ و ٠,٥٤ وكان أعلى ارتباط لها بطريقة منتل-هنزل ٠,٥٤ ، ثم بطريقة كميلي كاي تربيع ٠,٥٣ ، وكان أدنى ارتباط لها بطريقة المساحة المطلقة لنموذج المعلمين CUA-2، وبطريقة أنغوف المعدلة .

كذلك تراوحت ارتباطات طريقة المساحة المؤشرة لنموذج المعلمين SA-2 ببقية الطرق بين ٠,٣٨ و ٠,٨٤ ، حيث كان أعلى ارتباط لها ٠,٨٤ مع طريقة كميلي كاي تربيع ، يليه ارتباطها مع طريقة منتل-هنزل . أما أدنى ارتباط لها ، فقد كان بطريقة معامل التمييز ٠,٣٨ ، وبطريقة أنغوف ٠,٣٩ .

وقد كانت معظم ارتباطات طريقة المساحة المطلقة لنموذج المعممين UA-2 ببقية الطرق سالبة ، إذ لم يكن هناك ارتباط موجب إلا مع طريقة الثلاثة معالم ٠,٢١ ، وطريقة أنغوف المعدلة ٠,٢٠ ، كما كان ارتباطها بطريقة المعلمين قريبة جدا من الصفر ٠,٠٢ .

وقد تراوحت ارتباطات نموذج راوش ببقية الطرق بين -٠,٠٨ و ٠,٨٤ بوسيط مقداره ٠,٦٦ ، حيث كان أعلى ارتباط لها بكل من كميلي كاي تربيع ومنتل-هنزل ٠,٨٤ - لكل منهما - يليه ارتباطها مع المساحة المؤشرة بالطريقة نفسها SA-1 ٠,٨٣ ، ثم ارتباطها بطريقة التمييز المعدلة ٠,٧٧ . أما أقل ارتباطات هذه الطريقة فقد كانت مع تلك الطرق التي تعتمد على المساحة المؤشرة لكل من نموذج المعلمين ونموذج الثلاثة معالم.

وتراوحت ارتباطات طريقة المساحة المؤشرة لنموذج راوش ببقية الطرق بين ٠,٥٩ و ٠,٧٩ ، حيث كان ارتباطها بطريقتي كميلي ومنتل-هنزل كاي تربيع ٠,٧٩ - لكل منهما - وبطريقة معامل التمييز المعدل ٠,٦٦ ، و أنغوف المعدلة ٠,٦٤ ، وبمعامل التمييز ٠,٦٠ ، وبطريقة أنغوف ٠,٥٩ .

أما طريقة أنغوف المعدلة ، فقد تراوحت ارتباطاتها ببقية الطرق بين ٠,٢٧ و٠,٧ ، حيث كان أعلى ارتباط لها مع طريقة كميلي ٠,٧٠ ، يليه ارتباطها مع طريقة منتل-هنزل ٠,٦٩ .وقد كان ارتباطها بطريقة معامل التمييز وبطريقة أنغوف منخفضا .

كما كان ارتباط طريقة أنغوف بطريقتي معامل التمييز وبطريقتي كاي تربيع مرتفعا ، إذ تراوحت الارتباطات بين ٠,٦٩ و ٠,٧٦ .كذلك كان ارتباط كميلي كاي تربيع عاليا جدا بمنتل-هنزل كاي تربيع ٠,٩٩ ، وجيدة بطرق التمييز ، حيث كان الارتباط مع

طريقة التمييز ٠,٦٠ ، ومع طريقة التمييز المعدلة ٠,٦٨ .وكانت ارتباطات طريقة منتل-هنزل بطريقتي التمييز مماثلة لارتباطات كميلي كاي تربيع بتلك الطرق .أما ارتباط طريقة التمييز بطريقة التمييز المعدلة فقد كان ٠,٨١ وهو ارتباط جيد. وتعد هذه النتائج إجابة عن سؤال البحث رقم (٢) والخاص بعلاقة بقية الطرق ببعضها البعض .

مدى الاتفاق بين طرق التحيز

يوضح جدول رقم ٢ نسبة الاتفاق بين طريقة المعلمين و بقية الطرق ، إذ تراوحت نسبة الاتفاق في البنود التي اعتبرت متحيزة بطريقة المعلمين وبقية الطرق بين ٤٢٪ و ٩٢٪ .حيث كانت أعلى نسب الاتفاق بين طريقة المعلمين و طرق المساحة المؤشرة لنموذج المعلمين ولنموذج راوش ٩٢٪ لكل منها ، يليها نسبة اتفاق هذه الطريقة مع نموذج راوش ٨٣٪ ، ثم مع طرق كاي تربيع و طرق أنغوف وطريقة التمييز المعدلة وطريقة المساحة المطلقة لنموذج المعلمين ٧٥٪ لكل من هذه الطرق .أما نسبة اتفاق طريقة المعلمين بطريقة التمييز ، فقد كانت منخفضة إذ بلغت ٤٢٪ .وتعد هذه النتائج إجابة عن تساؤل البحث رقم (٢) والخاص بمدى اتفاق طريقة المعلمين مع بقية طرق تحيز البنود.

كذلك يوضح جدول رقم ٢ مدى الاتفاق بين طريقة منتل-هنزل وبقية الطرق ، إذ تراوحت نسبة الاتفاق بين ٢٣٪ و ٩٠٪ ، بوسيط مقداره ٧٥٪ ، حيث كانت أعلى نسب الاتفاق مع طريقة كميلي كاي تربيع ٩٠٪ ، يليها نسبة الاتفاق مع أنغوف المعدلة ٧٧٪ ، ثم مع كل من طريقة المعلمين وطريقة المعلم الواحد وطريقة المساحة المؤشرة لنموذج المعلم الواحد ٧٥٪ لكل منها ،ثم بطريقة المساحة المؤشرة لنموذج المعلمين ٧١٪ ، فطريقة الثلاثة معالم وطريقة معامل التمييز المعدلة ٦٧٪ لكل منها .أما طرق المساحة المطلقة لنموذج المعلمين ومعامل التمييز وأنغوف ، فقد كان اتفاقها مع طريقة منتل-هنزل ضعيفا ٢٣٪ ، ٢٧٪ ، ٤٧٪ على التوالي. وتعد هذه النتائج إجابة عن السؤال الرابع من أسئلة البحث والخاص بمدى الاتفاق بين طريقة منتل-هنزل وبقية الطرق.

كما يوضح جدول رقم ٢ ، أيضا ، مدى الاتفاق بين طريقة أنغوف المعدلة وبقية الطرق في عدد البنود المتحيزة ، حيث تراوحت نسب الاتفاق بين ٢٣٪ و ٧٧٪ .فقد كان الاتفاق عاليا بين طريقة أنغوف المعدلة وكل من طرق كاي تربيع (كميلي ومنتل- هنزل) ، وطريقة أنغوف ، إذ بلغت نسبة الاتفاق مع هذه الطرق ٧٧٪ ، يليها نسبة اتفاق هذه الطريقة مع نموذج الثلاثة معالم وطريقتي المساحة المؤشرة لنموذجي المعلمين والمعلم الواحد ٦٩٪ لكل منها .كذلك فإن نسبة اتفاق بقية الطرق مع هذه الطريقة يعد جيدا عدا معامل التمييز الذي كانت نسبة اتفاقه معها ٢٣٪ .وتعتبر هذه النتائج إجابة عن تساؤل البحث الخامس والمتعلق بمدى الاتفاق بين طريقة أنغوف المعدلة وبقية الطرق.

ويوضح جدول رقم ٢ ، أيضا ، نسبة اتفاق طريقة معامل التمييز المعدلة مع بقية الطرق ، إذ تراوحت نسبة الاتفاق بين ٣٣٪ و ٩٢٪ ، بوسيط مقداره ٦٧٪. فقد كان اتفاق هذه الطريقة مع طريقة المساحة المؤشرة لنموذج راوش عاليا جدا ٩٢٪ ، يليه نسبة الاتفاق مع نموذج راوش ٨٣٪ ، ثم طريقة الثلاثة معالم ٧٥٪ ، فطريقتي كاي تربيع وأنغوف المعدلة ٦٧٪ لكل منها .وقد كان اتفاق هذه الطريقة مع المساحة المطلقة لنموذج المعلمين ومع طريقة أنغوف جيدا حيث بلغ ٥٨٪ .أما اتفاق طريقة التمييز المعدلة مع طريقة التمييز الأساسية فلم يكن جيدا ، حيث بلغ ٣٣٪ . وتعد هذه النتائج إجابة عن تساؤل البحث السادس والخاص بمدى الاتفاق بين طريقة معامل التمييز المعدلة وبقية الطرق.

مناقشة النتائج

تراوحت معاملات ارتباط الطرق المختلفة لنماذج السمات الكامنة مع نموذج المعلمين بين ٠,٠٢ و ٠,٨٧ .فقد أظهرت النتائج الحالية أن هناك تشابها بين طريقة المعلمين وبعض الطرق الأخرى المعتمدة على نظرية السمات الكامنة ، إذ كان أعلى تشابه لهذه الطريقة مع طريقة المساحة المؤشرة لنموذج المعلمين .وتبدو هذه النتائج منطقية ، إذ أن

مساحة المعلمين تعتمد على المنحنيات المشتقة من معلمي الصعوبة والتمييز للطريقة نفسها (نموذج المعلمين) .كما كان هناك تشابه بين هذه الطريقة والطرق المعتمدة على نموذج راوش، وبينها وبين طريقة المعالم الثلاثة .ويبدو من نتائج هذه الدراسة أن طرق المساحة المؤشرة لنموذج المعلمين ونموذج المعلم الواحد لها الفعالية نفسها التي تتمتع بها طريقة المعلمين أو قريبة منها ؛ إلا أن الطرق المعتمدة على المساحة المطلقة، لا تتمشى مع مفهوم التحيز الذي تعتمد عليه طريقة المعلمين، إذ نجد أن طرق المساحة المطلقة المعتمدة على نموذج المعلمين أو تلك المعتمدة على نموذج الثلاثة معالم لا ترتبط ارتباطا عاليا بنموذج المعلمين . كذلك فإن تشابه طريقة المساحة المؤشرة لنموذج الثلاثة معالم ونموذج المعلمين لم يكن جيدا.

أما عن نسب الاتفاق، فقد كانت أعلى نسبة للاتفاق في البنود المتحيزة بين طريقة المعلمين وطرق المساحة المؤشرة لنموذج المعلمين ونموذج المعلم الواحد، يليها اتفاق طريقة المعلمين وكل من طريقة راوش وطريقة الثلاثة معالم، ثم طريقة المعلمين المطلقة. ويبدو من خلال استعراض نتائج الدراسة الحالية أنها تتفق مع نتائج Shepard وآخرين، حيث إن مقدار التشابه يعتمد على الطريقة الأساسية للحصول على تقدير للمعالم، وأن نسبة اتفاق طرق المساحة المطلقة مع بقية الطرق غير جيدة [٩، ص٣٦٠] .

وحيث إن عدد معاملات الارتباط كبير، فإنه تم استخدام التحليل العاملي لتبسيط العلاقة بين الطرق المختلفة ليمكن تفسيرها بيسر وسهولة .وقد أوضح جدول رقم ٣ أن الطرق التي تأخذ في الاعتبار معلمي الصعوبة والتمييز بما فيها طرق المساحة المؤشرة لنماذج المعلمين وراوش تشبعت على العامل الأول. وجدير بالملاحظة أن طريقة أنغوف المعدلة تشبعت تشبعا عاليا على هذا العامل، وأصبحت ضمن مجموعة الطرق المهمة لقياس التحيز .ويبدو أن دهشتنا تتلاشى عندما نعلم أن أساس التعديل لهذه الطريقة يأخذ

في الاعتبار دور معامل التمييز في مفهوم التحيز، ويمكن تسمية هذا العامل بعامل "طرق السمات الكامنة" نظرا لتشبع نماذج السمات الكامنة العالي على هذا العامل.

جدول رقم ٣ . العوامل المستخلصة من مصفوفة ارتباط طرق قياس التحيز ببعضها البعض بطريقة التدوير المتعامد بطريقة Varimax .

الطريقة / العامل	العامل الأول	العامل الثاني	العامل الثالث	العامل الرابع
نموذج الثلاثة معالم:				
DIF-3	٠.٨٢	٠.٠٢	-٠.٠٨	٠.١٧
CSA-3	-٠.١٤	٠.١٠	٠.٩٢	-٠.٠٧
CUA-3	٠.١٩	٠.٠٥	٠.٨٣	-٠.٠١
نموذج المعلمين:				
DIF-2	٠.٨٢	٠.٤٧	٠.٠٠٤	-٠.١٣
SA-2	٠.٨٧	٠.١٧	٠.٢٨	-٠.١٥
UA-2	-٠.٠٢	-٠.٠٧	-٠.٠٧	٠.٩٩
نموذج راوش :				
DIF-1	٠.٧٦	٠.٥٦	-٠.٠١	-٠.٠٦
SA-1	٠.٧٥	٠.٤٩	٠.١٥	٠.٣٤
أنغوف	٠.٢٤	٠.٨٤	٠.١٨	-٠.١٥
أنغوف المعدلة	٠.٨١	٠.٣٦	-٠.١٣	-٠.٠٢
كاي تربيع :				
كمبلي	٠.٧٧	٠.٥٦	٠.١٠	-٠.٠٩
منتل-هنزل	٠.٧٧	٠.٥٦	٠.١٠	-٠.٠٩
معامل التميز	٠.٣٩	٠.٨٧	-٠.٠١	-٠.٠٥
معامل التميز لمعدل	٠.٢٧	٠.٩٤	٠.٠٥	٠.٠٣

أما العامل الثاني فهو خليط من طرق تركز على الصعوبة وأخرى تركز على التمييز، وثالثة تركز على الصعوبة والتمييز معا، وقد يكون هذا التقارب نتيجة للعلاقة الوثيقة بين الصعوبة والتمييز .لذا فإنه قد يعكس تشابه طرق التمييز بتلك التي تعتمد على

الصعوبة .الأهم من هذا إن التشبعات على هذا العامل كانت للطرق التي لا تعتمد على طرق السمات الكامنة ، لذا يمكن تسميته بعامل " الطرق الإحصائية البسيطة." أما العامل الثالث ، فيبدو أنه يقتصر على طرق المساحة المعتمدة على نماذج السمات الكامنة وبالذات نموذج الثلاثة معالم ، وعليه فإنه يمكن تسميته بعامل "طرق المساحة لنموذج الثلاثة معالم." أما العامل الرابع ، فيبدو أنه لا يعكس توجها عاما بل قد يعكس ما يعرف بـ statistical artifact (ناتج إحصائي راجع لطبيعة المعادلة ولا يعبر عن معنى معين).

وعندما نود تبسيط العوامل الواردة في مصفوفة العوامل ، فإنه يمكن القول إن ما يميز طرق تحيز البنود هو : (أ) كونها طرق تعتمد على نظرية السمات الكامنة أم على أساليب مبسطة ؛ و (ب) كونها أحادية المعلم أم ثنائيته .

ونظرا لأن طرق التحيز المعتمدة على نظرية السمات الكامنة تتطلب عينات كبيرة وبرامج إحصائية معقدة ، فإنه سيتم التركيز على صلاحية الطرق التي لا تتطلب عينات كبيرة أو برامج إحصائية معقدة .وعليه فإن نتائج الدراسة الحالية أظهرت أن طريقتي كميلي ومنتل-هنزل كاي تربيع تعدان أفضل الطرق نظرا لارتباطهما العالي بنموذج المعلمين وارتباطهما الجيد بنموذج المعلم الواحد ، ونظرا لنسبة اتفاقهما مع طريقة المعلمين. ويبدو أن هذه النتائج متفقة مع نتائج الأبحاث السابقة ، سواء الإمبريقية منها أم التمثلية. إذ نجد أن Nugester [٢١]، و Instasuwan [٢٢]، و Ironson and Subkoviak [٤]، وShepard et al. [٩]، Rudner [١٥]، و Merz and Grossen [٢٠]، وكذلك Burrill [١٩]، يعدون طريقة كاي تربيع من بين الطرق المفضلة لدراسة التحيز .فقد عدت Instasuwan طريقة كاي تربيع أكثر معقولية حين تكون العينة محدودة [٢٢].كما توصلت Ironson وSubkoviak [٤] إلى نتائج مماثلة .أما Shepard وآخرون ، فقد أوضحوا أن "طريقة كاي تربيع قد تكون البديل العملي لنموذج الثلاثة معالم " [٩ ، ص٢٢٠] .

وتؤكد هذه النتائج أهمية طريقة كميلي كاي تربيع وإمكانية استخدامها في دراسة تحيز بنود الاختبار .ويعزز هذا التوجه قدرة الطريقة على التخلص من الأخطاء الموجبة (أن يكون البند متحيزا بناءا على هذه الطريقة في حين أنه غير متحيز بناء على طريقة المعلمين) ، إذ أن هذه الطريقة تعد من أقل الطرق من حيث الأخطاء الموجبة .فقد كانت نسبة الأخطاء الموجبة لكميلي كاي تربيع ٨٪ ، في حين كانت الأخطاء الموجبة لطريقة منتل-هنزل ١٧٪ .لذا فإنه ، على الرغم من تماثل الارتباط ونسب الاتفاق بين طريقتي كاي تربيع وطريقة المعلمين ، إلا أن الأخطاء الموجبة لطريقة كميلي أقل من تلك التي أظهرتها طريقة منتل-هنزل .لذا فإنه حين يكون معيار اختيار الطريقة هو قدرتها على التخلص من الأخطاء السالبة والموجبة معا ، فإن طريقة كميلي كاي تربيع يعتبر أفضل من ربيبتها (طريقة منتل-هنزل) .

أما طريقة أنغوف ، فلم يكن التشابه بينها وبين طريقة المعلمين عاليا ، فقد كان معامل الارتباط بينهما منخفضا .وعندما ينظر إلى ارتباط هذه الطزيقة بطريقة راوش وكيف أنه مرتفع (٠,٦٤) ، ونتائج الدراسات التمثلية المتعلقة بمعلم الصعوبة ، نجد أن نتائج هذه الدراسة تتفق وما أشار إليه Rudner وآخرون من أن "طريقة أنغوف غير حساسة للتحيز الناتج عن تمييز البند " [١٥ ، ص٢٢٧] .وهذا الضعف هو الذي جعل Shepard وآخرين [٧] يقترحون تعديلا على هذه الطريقة يقوم على اختزال قدرة البند على التمييز من هذه الطريقة.

وعند تقويم فعالية تعديل Shepard وآخرين ، وجدت الدراسة الحالية أن تشابه طريقة أنغوف المعدلة مع بقية الطرق ارتفع بدرجة عالية ، فقد أصبحت تضاهي طرق كاي تربيع من حيث التشابه مع نموذج المعلمين والثلاثة معالم ، كما زادت نسبة اتفاقها مع بقية الطرق .وتعد هذه النتائج مماثلة لما توصلت إليه دراسة Shepard وآخرين ، إذ أوضحت أن ارتباط طريقة أنغوف المعدلة "بالمحك (نماذج السمات الكامنة) كان مقاربا جدا لارتباط

كميلي كاي تربيع بنفس المحك" [٧ ، ص٩٨] .كما أن الأخطاء الموجبة لطريقة أنغوف المعدلة تعد قليلة جدا عندما تقارن بتلك المتعلقة بالطريقة الأساسية ، حيث بلغت نسبة الأخطاء الموجبة للطريقة المعدلة ٢٥٪ ، في حين بلغت هذه النسبة للطريقة الأساسية ٥٨٪. وهذا يعني أن تعديل Shepard وآخرين لطريقة أنغوف زاد من فعاليتها وجعلها في موقع تنافسي مع طرق كاي تربيع .

أما طريقة معامل التمييز المعدلة ، فقد أظهرت تشابها أكبر مع طرق السمات الكامنة وبقية الطرق مقارنة بالطريقة الأساسية لمعامل التمييز ، بيد أن معاملات الارتباط لم تكن ذات فارق كبير .ويبدو أن فعالية الطريقة المعدلة لمعامل التمييز زادت عندما حسبت نسبة الاتفاق بينها وبين بقية الطرق ، إذ بلغت نسبة اتفاق الطريقة المعدلة مع طريقة المساحة المؤشرة لنموذج راوش ٩٢٪ ، ولم تقل نسبة اتفاقها مع الطرق الأخرى عن ٥٨٪ ، بينما لم تتجاوز نسبة اتفاق الطريقة الأساسية مع بقية الطرق في أفضل أحواله ٤٤٪ . وهذا يعني أن الطريقة المعدلة لمعامل التمييز قادرة على التعرف على البنود التي اعتبرتها بقية الطرق متحيزة .وتعكس نسبة الاتفاق المنخفضة - ٣٣٪ - للطريقة المعدلة مع الطريقة الأساسية مدى الاختلاف بين الطريقة المعدلة وطريقة معامل التمييز الأساسية في القدرة على التعرف على البنود المتحيزة .وهذا يعني أن طريقة معامل التمييز المعدلة قادرة على التعرف على التحيز في البنود التي اعتبرتها الطريقة الأساسية غير متحيزة. بيد أنه من نافلة القول أن جدة هذه الطريقة لا تمكننا من اعتبارها من بين الطرق المنافسة لقياس التحيز ، حيث إن ذلك يتطلب بعض الدراسات عن مدى ثبات هذه الطريقة. لذا فإن الدراسات المستقبلية عن هذه الطريقة ينبغي أن تركز على مفهوم الثبات ، ومدى اتساق النتائج مع نتائج الدراسة الحالية.

خلاصـــــــة

تعد طرق السمات الكامنة أفضل الطرق لقياس تحيز بنود الاختبار من الناحية النظرية، إلا أن بعض الصعوبات العملية تقلل من إمكانية استخدامها لدراسة تحيز بنود الاختبارات (منها على سبيل المثال، صعوبة تقدير المعالم، وضرورة الحصول على عينات كبيرة، ومدى القدرة على التعامل مع البرامج الإحصائية التي تتطلبها هذه الطرق). أمام مثل هذه الصعوبات، فإن البديل العملي لهذه الطرق هو كميلي كاي تربيع، نظرا لتشابهه من حيث القاعدة النظرية مع طرق السمات الكامنة ولسهولة استخدامه .كما أن طريقة منتل-هنزل تعد بديلا منافسا لطريقة كميلي كاي تربيع على الرغم من زيادة الأخطاء الموجبة لهذه الطريقة مقارنة بطريقة كميلي. أما عندما لا يتمكن الباحث من استيفاء شروط مستويات القدرة لأي من طرق كاي تربيع، فإن طريقة أنغوف المعدلة تعد بديلا جيدا لهذه الطرق.

ونظرا لحداثة طريقة معامل التمييز المعدلة، فإنه على الرغم مما أظهرت من تشابه مع الطرق الأخرى ومن اختلاف مع طريقة معامل التمييز الأساسية، فإنه يفضل زيادة التقصي عن هذه الطريقة وتقدير مدى ثباتها قبل الاعتماد عليها في تقدير تحيز البنود .

المراجـــــع

[١] القاطعي، عبدالله. "تحيز بنود اختبار وكسلر لذكاء الأطفال المعدل (الصورة السعودية) حسب الجنس." *مجلة جامعة الملك سعود*، م٥، *العلوم التربوية والدراسات الإسلامية*، ٢(١٤١٣هـ)، ٣٧٣ – ٣٩٠.

[٢] Berk, R. ed. *Handbook of Methods for Detecting Test Bias*. Baltimore, MD: Johns Hopkins University Press, 1982.

[٣] Willson, V., R. Nolan , C. Reynolds and R. Kamphuas. "Race and Gender Effects on the Functioning of the Kaufman Assessment Battery for Children." *Journal of School Psychology*, 27 (1989), 289-96.

[٤] Ironson, G., and M. Subcoviak. "A Comparison of Several Methods of Assessing Item

Bias." *Journal of Educational Measurement*, 16 (1979), 209-25.

[٥] Angoff , W. and S. Ford. "Item-race Interaction on a Test of Scholastic Aptitude." *Journal of Educational Measurement*, 10 (1973), 95-105.

[٦] Angoff, W. "Use of Difficulty and Discrimination Indices for Detecting Item Bias." In: R. A. Berk, ed. *Handbook of Methods for Detecting Test Bias*. Baltimore, MD: Johns Hopkins University Press, 1982 , 96-116.

[٧] Shepard, L., G. Camilli, and D. William. "Validity of Approximation Techniques for Detecting Item Bias. "*Journal of Educational Measurement*, 22 , no. 2 (1985) , 77- 105.

[٨] Green, B., and J. Draper. *Exploratory Studies of Bias in Achievment Tests*. Monterey, CA: CTB/McGraw – Hill, 1972.

[٩] Shepard, L., G. Camilli. and M. Averil. "Comparison of Procedures for Detecting Test-item Bias with both Internal and External Ability Citeria." *Journal of Educational Statistics*, 6 (1981), 317-75.

[١٠] Scheuneman, J. "A New Method of Assessing Bias in Test Items." Paper Presented at the meeting of the American Educational Research Association , Washington, 1975.

[١١] Scheuneman, J. "A Method of Assessing Bias in Test Items." *Journal of Educational Measurement* , 16 (1979), 143-52.

[١٢] Holland, P., and D. Thayer. "Differential Item Performance and the Mantel-Haenszel Procedure." In: H. Wainer and H. Braun, ed. *Test Validity*. Hillsdale, NJ: Lawrence Erlbaum Association , 1988, 129-70 .

[١٣] Holland, P., and D. Thayer. "Differential Item Performance and the Mantel-Haenszel Procedure." Paper presented at the annual meeting of the American Research Association, San Francisco, 1986.

[١٤] Raju, N., R. Bode, and V. Larsen. "An Empirical Assessment of the Mantel-Haenszel Statistic for Studying Differential Item Performance." *Applied Measurement in Education*, 2 (1989), 1-13 .

[١٥] Rudner, L., P. Getson and D. Knight. "Biased Item Detection Techniques." *Journal of Educational Statistics*, 6 (1980) , 213-33.

[١٦] Lord, F. M. *Application of Item Response Theory to Practical Testing Problems*. Hillsdale, NJ: Erlbaum, 1980.

[١٧] Raju, N. "The Area between Two Item Characteristic Curves." *Psychometrika*, 53 (1988), 495-502.

[١٨] Raju, N. "Determining the Significance of Estimated Signed and Unsigned Areas between Two Item Response Functions." *Applied Psychological Measurement*, 14 (1990), 197-207.

[١٩] Burrill, L. "Comparative Studies of Item Bias Methods." In: R. A. Berk , ed., *Handbook of Methods for Detecting Test Bias*. Baltimore, MD: Johns Hopkins University Press, 1982, 161-79 .

[٢٠] Merz, W., and N. Grossen. "An Empirical Investigation of Six Methods for Examining Test Item Bias." Paper presented at the annual meeting of the National Council on Measurement in Education, San Francisco, April 1977(ED 178 566) (1977).

[٢١] Nugester, R. "An Empirical Investigation of Three Models of Item Bias." *Dissertation Abstracts International* , 28 (1977) , 272A.

[٢٢] Instasuwan, P. "A Comparison of Three Approaches for Determining Item Bias in Coss-national Tests." Unpublished Doctoral Dissertation , University of Pittsburgh, 1979.

[٢٣] Raju, N., and J. Normand. "The Regression Bias Method: A Unified Approach for Detecting Item Bias and Selection Bias." *Educational and Psychological Measurement* , 45 (1985), 37-54.

[٢٤] Perlman, C., et al. "Investigating the Ability of Four Methods for Estimating Item Bias." Paper presented at the annual meeting of the National Council for Measurement in Education, New Orleans, April 1988 (ED 296-003).

[٢٥] DeMauro, G. "Effects of Representation of Gender Groups in the Examinee Population on the Mantel-Haenzel Procedure." Paper presented at the annual meeting of the American Educational Research Association, Boston , MA. ED 318-747, 1990.

[٢٦] النافع ، عبدالله ، وعبدالله القاطعي ، والجوهرة ، السليم. *إعداد اختبارات ومقاييس للتعرف على الموهوبين والكشف عنهم: (القسم أ) صدق وثبات اختبار وكسلر لذكاء الأطفال المعدل*. الرياض : مدينة الملك عبدالعزيز للعلوم والتقنية ، ١٤١١هـ ، ٥-٢٥.

[٢٧] Reckase, M. "Unifactor Latent Trait Models to Multifactor Tests: Results and Implications." *Journal of Educational Statistics*, 4 (1979), 207-30.

[٢٨] Mislevy, R. and R. Bock. BILOG-3: *Item Analysis and Test Scoring with Binary Logistic Models*. 2d ed. Mooresville: Scientific Software , Inc., 1990.

[٢٩] Loyed, B., and H. Hoover. "Vertical Equating Using the Rasch Model." *Journal of Educational Measurement*, 17 (1980), 179-93.

[٣٠] Baker, F., and A. Al-Karni. "A Comparison of Two Procedures for Computing IRT

Equating Coefficients." *Journal of Educational Measurement*, 28 (1991), 147-62.

[٣١] Kim, S., and A. Cohen. "IRTDIF: A Computer Program for IRT Differential Item Functioning Analysis." *Applied Psychological Measurement* , 16, 2 (1992) , 158.

[٣٢] Kim, S., and A. Cohen. "Effects of Linking Methods on Detection of DIF." *Journal of Educational Measurement*, 29 (1992), 51-66.

[٣٣] Ironson, G. A. "Use of Chi-square and Latent Trait Approaches for Detecting Item Bias." In: R. A. Berk, ed. *Handbook of Methods for Detecting Test Bias*. Baltimore , MD: Johns Hopkins University Press, 1982, 117-60.

Comparative Study of Some Item Bias Detection Methods

Abdullah A. Al-Qataee
Associate Professor, Department of Psychology, College of Education, King Saud University, Riyadh, Saudi Arabia

Abstract. Several methods of item bias detection indices were introduced in the last decades. The IRT methods were said to be more elegant compared to others. However, due to some practical difficulties, the search for an alternative method is necessary. The purpose of the present study was to compare the ability of several methods to detect item bias compared to the two parameter IRT models. Thus, DIF methods, the transformed item difficulty (TID) method, chi-square methods, and item discrimination indices were compared using data from the Information Subtest for the Normative Sample of the Saudi version of the WISC-R. Sex biases were considered. The results indicate that IRT methods are more elegant than other methods, thus confirming previous findings. Chi-square methods were found to be a good substitute for IRT methods. More specifically, Cammilli chi-square was most effective followed by MH, and Angoff modification of the TID. The new modification of the item discrimination (ID) index seemed to be more effective than the original one. However, further study of the new modification of ID is needed.