تصحيح معاملات صعوبة الفقرات لأثر التخمين في اسئلة الإختيار من متعدد:
صورة معدّلة لمعادلة جيلفورد

ا.د. احمد سليمان عودة prof. Ahmad Sulieman Audeh
جامعة جدارا /اربد- الأردن Jadara University /Irbid-Jordan
Prof_ahmad_audeh@yahoo.com

مقدمة

يتزايد الاهتمام بصناعة الأسئلة من نوع الاختيار من متعدد لكثرة المزايا التي تتمتع بها، وتزايد استخداماتها في الاختبارات واسعة النطاق (Large scale assessment) التي تطبق على أعداد كبيرة من المفحوصين، وقدرتها على تحقيق درجة عالية من صدق المحتوى التعليمي؛ ولكن يعاب عليها أنها عرضة للتخمين Guessing، الذي يعتبر مصدراً من مصادر الأخطاء(E) في نظرية القياس التقليدية القائمة على العلاقة بين الدرجات الحقيقية (T) والدرجات الظاهرية (X) حيث أن X = T + E، وتأثيره كذلك (اي التخمين c) على احتمال إجابة مفحوص قدرته (θ) على الفقرة االتي تعرّف بصعوبتها (b) وتمييزها (a) وفقا للمعادلة الشاملة لمعالم الفقرة الثلاثة(الصعوبة والتمييز والتخمين $P(\theta) = c + (1 - c)/[1 + e^{-1.7a(\theta - b)}]$ في نظرية استجابة الفقرة، واشير اليهما بإختصار وعلى التوالي نظرية الاختبار (Test theory) ونظرية الفقرة (Item theory). وبما أن التخمين هو نوع من الخطأ في القياس، فهو عامل مؤثر في دقة تقدير الدرجة الحقيقية للمفحوص من الدرجة الظاهرية ، طالما أن احتمال الإجابة عن الفقرة عند قدرة معينة P(θ) يعتمد على صعوبة الفقرة المتأثرة بالتخمين.

ووفقا لنظرية القياس القائمة على الاختبار فان معامل صعوبة الفقرة المقدر بنسبة الإجابات الصحيحة R/N (حيث أن N عدد المفحوصين، R عدد الذين أجابوا إجابة صحيحة عن الفقرة)، تفترض بأن كل الإجابات الصحيحة (R) ناتجة عن معرفة حقيقية، اي انّ $R = R_T$.، ولا يوجد للتخمين أي دور في تحديد الإجابات الصحيحة، بمعنى أن $R_g = 0$. وأكدت أدبيات القياس على عدم صحة هذا الافتراض، وأن مجرد استخدام فقرات (أسئلة) الاختيار من متعدد يعني بالمنطق الرياضي أن هذا الافتراض لا يتحقق (Violated)، وأن درجة عدم التحقق تزداد بنقصان عدد البدائل، وتزداد مقاومة الاختبار (Robustness) لأثر التخمين بزيادة عدد البدائل.

وبما أن عدد البدائل في الحدود القصوى قد يتراوح بين (8−2)، فإن فرصة التخمين تتراوح نظريا بين (125.−50.) ولذلك أكدت أدبيات القياس على أهمية الحد من تأثير التخمين على دقة القياس بأساليب إحصائية منسجمة منطقيا مع تعليمات الإختبار. فقد يطلب من المفحوص في تعليمات الاختبار أن لا يلجأ إلى التخمين إذا شعر بأنه لا يمتلك المعرفة للإجابة عن الفقرة، وأن يتركها دون إجابة، وكأنه يقول للمفحوص بأنك ستخسر إذا لم تلتزم (وهذا هو الشائع في الأوساط التربوية) بمعنى ان على المفحوص ان لايغامر او يخاطر في الإجابة. أو أن يطلب منه الإجابة عن جميع الفقرات، وكأنه يقول للمفحوص عليك الاستفادة من أي فرصة للتخمين طالما أنك لا تخسر شيئاً من درجتك مفترضاً (أي الفاحص) تساوي تأثر جميع المفحوصين إيجابياً بفرصة التخمين، الا أن تباين الخطأ

المنتظم الناجم عن التخمين إلى جانب الخطأ العشوائي يستحق الإشارة إليه صراحة في المعادلة الرئيسية للنظرية التقليدية في القياس ليصبح $X = T_t + T_g + e$ ، بمعنى أن جزءاً من الدرجة الظاهرية (X) حقيقياً فعلاً (T_t) وناتجاً من الاختلاف في قدرات المفحوصين على السمة المقاسة ، وجزءاً آخر ناتجاً عن الاختلاف في القدرة على التخمين الذي يجمع بين التخمين العشوائي (Blind Random) والتخمين الذكي (Intelligent) القائم على المعرفة الجزئية (Partial knowledge)، والمخاطرة (Risk taking)، مع الإفتراض الضمني بأن جميع المفحوصين يستفيدون من فرصة التخمين، وفي نفس الوقت يختلفون في قدرتهم على التخمين، كما يختلفون في درجة التزامهم بالتعليمات عندما يطلب منهم أن لايخمنوا؛ بمعنى أنهم يختلفون في درجة تقديرهم لأثر المخاطرة في الإجابات انطلاقاً من مبدأ الفراق الفردية في السمات الشخصية ذات الصلة بموقف الاختبار وتقديراتهم لمضامين هذه التعليمات، وتوقعاتهم لآثار الالتزام أو عدم الالتزام بها.

ومن زاوية أخرى، يمكن ان ينظر الى التخمين على انه عامل (Factor) من العوامل المحددة للعلامة X، ويضيف تباينا حقيقياً ($s^2_{T_g}$) يفترض بأنه لا يرتبط بالتباين في القدرة الحقيقية الخاصة بما وضع الاختبار لقياسه ($S^2_{T_t}$)، وقد يترتب عليه بشكل جوهري تخفيض صدق الاختبار الذي يعتمد إحصائياً على نسبة التباين الحقيقي المرتبط بالسمة المقاسة من التباين الكلي ($S^2_{T_t} / S^2_x$)، بمعنى أن نسبة التباين الناتجة من عامل التفاعل بين التخمين والمخاطرة ($S^2_{T_g} / S^2_x$) تساهم في نسبة التباين المحددة للصدق (S^2_T / S^2_x). وعندها قد ينعكس ذلك جوهرياً أيضاً على تشبعات (Loadings) فقرات الاختبار بهذا العامل؛ وبالتالي عدم تحقق افتراض أحادية البعد (Unidimentionality) التي تقوم عليه النماذج الرئيسية في النظرية الحديثة في القياس (نظرية الفقرة) الأحادية والثنائية والثلاثية المعلمة، التي تفترض جميعها(اي النماذج) أن دقة كل المعالم التي تقدنها البرامج الإحصائية القائمة على هذه النماذج مبنية على مدى تحقق افتراض أحادية البعد.

الدراسات السابقة

اختلفت الدراسات السابقة في منهجيات التعامل مع اساليب التصحيح لأثر التخمين بشكل عام. فقد اشارت دراسة مارتن واروين وجورج (Martin, Erwin & George, 1972) الى ضعف الأساس المنطقي لهذا التصحيح القائم على التباين في سرعة الطلبة التي تدفعهم الى الاجابة العشوائية عن الاسئلة الواقعة في نهاية الامتحان. واشارت لاحقا دراسة روبرت (Robert, 1988) الى أن العقاب لأثر التخمين يمكن أن يكون مناسبا لاختبارات السرعة المتطرفة او اختبارات القوة المتطرفة، ولكنها لا تناسب معظم الامتحانات الصفّية التي تجمع بين القوة والسرعة؛ ولذلك فان المعادلة التقليدية في التصحيح لأثر الخمين لا تراعي المعرفة الجزئية، كما أن معرفة الطلبة بأن هناك عقاب لأثر التخمين سيزيد من وقت الاجابة، ويؤثر بالتالي على صدق الاختباروفقراته. أما فلدت (Feldt, 1993) فقد اشار الى ان أدبيات القياس تقدم ارشادات ووصفات لمطوري الاختبارات ومنها اختيار معاملات صعوبة الفقرات بمتوسط = 0.50 للحصول على ثبات أعلى.وقد اشار في نتائج دراستة عن أثر توزيع صعوبة الفقرات على ثبات الاختيار الى ان معاملات الصعوبة في

حالة الاسئلة التي تصحح (1،0) لا بد وأن تكون أعلى من 0.50، ووجد ان الثبات كان أعلى عندما تراوحت معاملات الصعوبة بين 0.57 ، 0.67 المقابلة لفرصة تخمين تراوحت بين 0.50 ، 0.15 وأشار الى التأثير العكسي على ثبات الاختيار عندما تزداد فرصة التخمين الى 0.25 مع معاملات صعوبة تصل الى 0.74 . كما أشار لو ولو وهنج ويسب (Lau, Lau, Hong & Usop, 2011) الى أن استخدام الطريقة التقليدية للتصحيح لأثر التخمين تؤدي بشكل عام الى التحفظ في الاجابة الى الحد الذي يحرم الطلبة من توظيف المعرفة الجزئية، وقد أشاروا في دراستهم ووفق تصميم معين استخدموا فيه الحذف الجزئي للبدائل أن مراعاة المعرفة الجزئية في حساب الدرجة يؤدي الى تقليل خطأ القياس.

وفي هذا السياق ايضا قارن ثوماس ونيل والكس وآن (Thomas, Neal, Alex & Anne, 2006) الدرجات على اختبار فقراته على شكل اجابة قصيرة (بدون تخمين) مع الدرجات على الإختبار نفسه بوجود بدائل مع اوزان احتمالية 20% للتخمين الصحيح ويعطى (1) واحتمال 80% للتخمين غير الصحيح ويعطى (1/4) بحيث يكون مجموع ما يحصل المفحوص بالتخمين يساوي صفر [−0.20*1+1/4*0.80]،حيث اشارت النتائج الى ان التصحيح لأثر التخمين يزيد صدق الاختبار عندما تكون الفقرات وبدائل الاجابة غير قابلة لتوظيف المعرفة الجزئية، الذي اعتبراه غير واقعي. وبهذا الصدد فقد اشار لوسي وتريسي وجيمز ومارك Lucy, Tracey, James & Mark (2009) الى الإستماع الى الطلبة وردود افعالهم عند التعامل مع التصحيح لأثر التخمين في الوقت الذي يتزايد فيه استخدام اختبارات الاختيار من متعدد، حيث اشارت نتائج الدراسة الى أن الطلبة لا يفضلون استخدام التصحيح لأثر التخمين إلا في حالة الامتحان الذي يسمح فيه بفتح الكتاب (open–book) وأنهم يعاقبون على أدائهم عندما تصحح درجاتهم لأثر التخمين، وان الدرجات كانت أقل عندما طلب منهم ان لا يخمنوا، وفسر ذلك بأن هذا العقاب لأثر التخمين بالأسلوب التقليدي يؤدي الى أن تكون قيم معاملات صعوبة الاسئلة أقل مما هي عليه فعلاً.

قامت دراسات اخرى في تعاملها مع التخمين في الإطار نفسه على نظرية الفقرة (item theory)(او ما يطلق عليها نظرية استجابة الفقرة) مقابل دراسات اخرى قائمة على نظرية الإختبار (test theory). فقد قارنت دراسة كلوز وكريستيان (Klaus,. & Christian, 2007) نتائج التحليل لإجابات الطلبة على نماذج اختبار تختلف فقط في عدد بدائل فقراته بنتائج التحليل لنموذج مرجعي لايعطى فيه بدائل(اجابة قصيرة) ، وأشارت النتائج وفق نموذج راش ان صعوبة الفقرة بخمسة بدائل أواكثر تعادل صعوبتها الفقرة بدون بدائل (أي لا يكون هناك فرصة للتخمين)، بمعنى ان فرصة التخمين تصبح ضعيفة جداً بخمسة بدائل أو أكثر، وأن التغير في صعوبة الفقرة لا يكون جوهرياً. وبحثت دراسة شانج ولن ولن (Chang, Lin, & Lin, 2007) في اثر مراعاة المعرفة الجزئية القائمة على اسلوب الحذف المتدرج (subset elimination) والاختيار المتدرج (subset selection) على خصائص الاختبار والفقرات وفق النموذج المتدرج القائم على المعرفة الجزئية (partial credit step function) لنظرية الفقرة مقارنة بالأسلوب الذي يعتمد عدد الاجابات التي يتم اختيارها دون اي مراعاة للمعرفة الجزئية. وأكدت نتائج الدراسة على اهمية حوسبة الاختبارات واستخدام برامج تسهل

عملية التصحيح بالأساليب التي تراعي المعرفة الجزئية. كما اكدت دراسة اسبنوسا وجارديزابال (Espinosa & Gardeazabal, 2010) على اهمية التعمق في النظر الى تصرف الطلبة عند اجاباتهم عن اسئلة الاختيار من متعدد والتي تتفاعل فيها المعرفة مع المخاطرة، والى ضرورة التمييز بين نظرية القرار (Decision theory) مقابل نظرية القياس،عند التعامل مع الجوانب النظرية والتطبيقية للإختبارات. فقد اشارت الى ان سلوك الطالب يميل الى التحفظ وعدم المخاطرة ، كلما توقع عقاباً عالياً بسبب التخمين. وبالتالي زيادة عدد الفقرات المتروكة دون اجابة، كما اشارا الى ان الإفتراضات التي تقوم عليها نظرية القياس القائمة على الفقرة لا تتحقق طالما أن هناك عامل مخاطرة، ولذلك تحدث عن نموذج توفيقي يجمع بين نظرية القرار ونظرية القياس يأخذ بالاعتبار الفقرات المتروكة بسبب ضيق وقت الامتحان بالنسبة للمفحوص مقابل الفقرات التي قامت اجابتها على معرفة جزئية أو المخاطرة التامة في الاجابة، وربط هذه العوامل المؤثرة في الإجابة على توزيع وقت الاجابة (Response time)

تعكس نتائج الدراسات السابقة بصورة مباشرة اوغير مباشرة جدلية التعامل مع التخمين في اسئلة الإختيار من متعدد في ضوء عدة عوامل وممارسات في الإطارين النظري والتطبيقي، وكيفية التعامل مع التخمين والمخاطرة والمعرفة الجزئية ، واختلاف تأثيرها على خصائص الإختبار وفقراته باختلاف التعليمات وعدد بدائل الإجابة ، وقد بينت هذه الدراسات بصورة غير مباشرة انه لايوجد وصفة سحرية لضبط أثر التخمين ، ولا الى معرفة حجم التأثيرعلى هذه الخصائص ودقة القرارات ، ولكنها عكست اهمية التصحيح لأثر التخمين بأساليب أكثر عدالة من الأسلوب التقليدي القائم على العقاب، ومن الطبيعي ان تكون صعوبة الفقرات في اختبارات الإختيار من متعدد محور هذا التصحيح.وهو الغرض الأساسي في هذه الدراسة.

مشكلة الدراسة:

أدرك المنظّرون في القياس مبكراً أهمية الحد من التلوث الإحصائي (Contamination) في التقديرات لدرجات المفحوصين وخصائص الفقرات بمعادلات إحصائية وفقاً لتعليمات الإختبار تحت عناوين مختلفة أبرزها التصحيح لأثر التخمين، إلا أن مساهمات الباحثين للحد من هذا التلوث في إطار النظرية الحديثة القياس في القياس مازالت متواضعة، وقد كانت أبرز هذه المساهمات من شو (Chiu, 2010)، وهو على اي حال ليس موضوع البحث هنا، لأن النظرية التقليدية مازالت مسيطرة ، وأنه ينظر لنظرية الفقرة بانها مكملة وليست منافسة لها(Osterlind,2006, p263)

وبالمقابل ، يأتي هذا البحث في إطار الجهود العلمية للباحثين المهتمين في القياس النفسي والتربوي بشكل عام والاختبارات من نوع الاختيار من متعدد بشكل خاص المرتكزة على النظرية التقليدية، وتحديدا تعديل معادلة تصحيح معامل صعوبة الفقرة لأثر التخمين التي بدأها جيلفورد (Guilford, 1936)، في ضوء تعدد معادلات تصحيح درجات المفحوصين لأثر التخمين التي تأخذ بالإعتبار الإختلاف بين التخمين العشوائي والتخمين الذكي والتخمين القائم على المعرفة الجزئية وتفاعل التخمين مع المخاطرة. وهو تعديل مقترح

لمعادلة جيلفورد لتصحيح صعوبة الفقرة لآثر التخمين، انطلاقا من تعظيم عدالة هذه المعادلة في ضوء مدى مقاومتها للافتراضات الضمنية التي قامت عليها اوانطلقت منها ، ومدى انسجامها مع جدلية التصحيح لأثر التخمين في ضوء تعليمات الاختبار من جهة، والأسس المنطقية التي ارتكز عليها التعدد في المعادلات المقترحة للتصحيح لأثر التخمين من جهة اخرى.

تعليمات الاختبار ومعادلات التخمين

تؤكد ادبيات القياس التربوي على اهمية الإلتزام بقواعد وشروط تطبيق الاختبارات العقلية بشكل عام والتحصيلية بشكل خاص وتصحيحها وتفسير درجاتها وفقا لهذه الشروط، ومن ضمنها التعليمات المتعلقة بالتخمين التي تعتبر عقداً بين الفاحص والمفحوص يترتب عليها اجراءات متعلقة بالتصحيح لأثر التخمين، كأن يطلب من المفحوص:

- الإجابة عن جميع الفقرات، وهي في هذه الحالة لا تخول الفاحص اتخاذ أي إجراء لمعاقبة المفحوصين لأثر التخمين، وأن صعوبة الفقرة الثنائية التدريج (0,1) تقدر نسبة المفحوصين الذين أجابوا إجابة صحيحة عن الفقرة (R/N) على افتراض أن N عدد المفحوصين.
- أن لا يخمن: وهي تعني ضمناً أن على المفحوص أن لا يخاطر في الإجابة، ويتوقع منه ان يحد من درجة مخاطرته، مع الاعتراف بالفروق بين المفحوصين في درجة المخاطرة أو التجاوب مع التعليمات، أو على المفحوص أن يكون مهيئاً لتعديل درجته الظاهرية (X)، علماً بأن التعديل قد يحتمل أكثر من خيار يمكن تلخيصها على النحو الآتي:

(1) العقاب التخمين وفقاً للمعادلة التقليدية، وذلك بتخفيض العلامة الظاهرية (X) بمقدار ما يمكن أن يحققه المفحوص من علامة إضافية عن طريق التخمين (عدد الفقرات تتناسب مع الإجابات التي أجاب عنها إجابة خاطئة (W))، الذي يفترض بأنه يتم بصورة عشوائية تامة، ولا تعترف بالتخمين الذكي أو استخدام المعرفة الجزئية، ولكنها تأخذ بالاعتبار درجة المخاطرة التي تتناسب مع عدد الفقرات المتروكة دون إجابة (O)، وتحسب الدرجة المصححة (X) بتخفيض (X) العدد المتوقع رياضياً للإجابات الصحيحة حسب الفرصة المهيأة للإجابة الصحيحة من البدائل الخاطئة (L−1)، على الافتراض أن للفقرة أربعة بدائل واحدة منها صحيحة، أي أن ($X^* = x - \frac{W}{L-1}$)، وقد اعتبر لتل (Little,1962) أن هذه المعادلة متشددة أو قدم معادلة أقل تشدداً، وذلك بتقليل التخفيض في الدرجة إلى النصف لتصبح ($X^* = x - \frac{W}{2(L-1)}$)، مفترضاً أن ما يقوم به المفحوص في الإجابة لا يمكن أن يكون عشوائياً تاماً أو تخميناً ذكياً تاماً، وافتراض أن التخمين يتم مناصفة بين النوعين.

(2) المكافأة لعدم المخاطرة، والتي تتناسب مع عدد الفقرات المتروكة، وهي تعطي المفحوص درجة إضافية تتناسب مع الاحتمال الرياضي للإجابة الصحيحة على الفقرة فيما لو أجاب عنها إجابة عشوائية لتصبح ($X^* = X + \frac{O}{L}$)، والتي يعاب عيها بأنها تبالغ (Overestimate) في تقدير قدرة المفحوص، بعكس طريقة العقاب التي يعاب عليها بأنها تبخس (Underestimate)

قدرة المفحوص؛ ولذلك جاء التفكير بموازنة التحيز (Balancing bias) بمعادلة تجميع بين الاستدلال المنطقي والاستدلال الإحصائي ($X^* = X - \frac{W}{L-1} + \frac{O}{L}$)، والسؤال المطروح هنا يتمثل بانعكاس هذه الخيارات على تصحيح معاملات صعوبة الفقرات وليس فقط على درجات المفحوصين.

معادلة جيلفورد لتعديل صعوبة الفقرة

لم يكن الهدف من هذا العرض التعريف بتصحيح الدرجات لأثر التخمين أو المخاطرة لأنها متوفرة في أدبيات القياس التربوي، ولكن الهدف هو الربط بين تصحيح الدرجات وتصحيح صعوبة الفقرات التي بدأها جيلفورد (Guilford, 1936)، حيث تعرّف صعوبة الفقرة إحصائياً بأنها نسبة المفحوصين الذين أجابوا عنها إجابة صحيحة (R/N)، وهو تعريف تقليدي يقدّر صعوبة الفقرة بأعلى مما هي عليه حقيقة رقمياً حسب التعريف الإجرائي للصعوبة، ولذلك قدم جيلفورد تصحيحاً لمعامل الصعوبة (d_c) قائما على أنه نسبة المفحوصين الذين أجابوا إجابة صحيحة بدون تخمين (S/N) معتبراً أن ($S = R - \frac{W}{L-1}$)، وانتهت عمليات الاشتقاق بالتعويض للتوصل إلى أن معامل الصعوبة المصحح لأثر التخمين ($d_c = \frac{S}{N} = \frac{Ld-1}{L-1}$)، وهي معادلة تصحيح قائمة على العقاب لأثر التخمين دون التمييز بين الفقرات المحذوفة (O)، والفقرات التي أجاب عنها إجابة خاطئة (W)، مفترضاً أن لدى جميع المفحوصين نفس الدرجة من المخاطرة، وكأنه لا علاقة بين التصحيح وتعليمات الاختبار، ولذلك فهي ممارسة غير منطقية من ناحية، والافتراضات القائمة عليها لا تتحقق من ناحية اخرى، ولا تتجاوب مع التحذيرات الخاصة بانتهاك (Violation) الافتراضات التي تقوم معادلة تصحيح الدرجات بالعقاب، ولذلك فهي(اي معادلة التصحيح بطريقة جيلفورد) تتجاوز افتراض عدم وجود تباين في المخاطرة بين المفحوصين، وتتجاوز حقيقة التخمين الذكي والمعرفة الجزئية، حيث أن (W) في معادلة جيلفورد غير تلك التي ترد عادة في معادلة التصحيح لأثر التخمين بالعقاب انسجاما مع التعليمات التي تنص على أن لا يخمن أو لا يخاطر، فهو يفترض (اي جيلفورد) أن جميع المفحوصين لا يتقيدون بالتعليمات كما لو أنه طلب منهم الإجابة عن جميع الفقرات.

التعديل المقترح لمعادلة جيلفورد

ينطلق التعديل المقترح من أن:

1. التعليمات تنص على أنه يطلب من المفحوص أن لا يخمن أو لا يخاطر في الإجابة العشوائية، ويتحمل بالتالي مسؤولية عدم الالتزام بهذه التعليمات.
2. عدد الإجابات الخاطئة (w) في التعديل المقترح تختلف عنها عند جيلفورد، فهي عند جيلفورد تساوي (N−R)، إلا أنها في التعديل المقترح (N−R−O).

3. معامل الصعوبة المصحح d_c^* قائم على نسبة الإجابات الصحيحة المعدّلة لأثر التخمين التي توازن بين المكافأة $\frac{1-d}{L}$ والعقاب، $\frac{Ld-1}{L-1}$ لأن التصحيح الذي يجمع بين المكافأة والعقاب يتوقع ان يحقق التوازن النسبي في التحيز بتقدير معامل صعوبة الفقرة وفقا للمعادلة المقترحة . $d_c^* = \frac{Ld-1}{L-1} + \frac{1-d}{L}$ ويبين الجدول التالي مقارنة ثنائية لقيم معاملات الصعوبة المعدلة بالمعادلتين d_c ، d_c^* المقابلة لمعاملات الصعوبة على متصل الصعوبة المحسوب من نسبة الإجابات الصحيحة (d).

جدول 1: معاملات الصعوبة المقدرة من نسبة الإجابات الصحيحة (d) المشار اليها بالحرف G مقابل المعاملات المعدلة بمعادلة جيلفورد d_c ، والمعدلة بالمعادلة المقترحة d_c^* المشار اليها بالحرف A محسوبة عند عدد من البدائل(2−8) .

d	G2	A2	G3	A3	G4	A4	G5	A5	G6	A6	G7	A7	G8	A8
.99	.980	.985	.985	.988	.987	.989	.988	.990	.988	.990	.988	.990	.989	.990
.98	.960	.970	.970	.977	.973	.978	.975	.979	.976	.979	.977	.980	.977	.980
.97	.940	.955	.955	.965	.960	.968	.963	.969	.964	.969	.965	.969	.966	.969
.96	.920	.940	.940	.953	.947	.957	.950	.958	.952	.959	.953	.959	.954	.959
.95	.900	.925	.925	.942	.933	.946	.938	.948	.940	.948	.942	.949	.943	.949
.94	.880	.910	.910	.930	.920	.935	.925	.937	.928	.938	.930	.939	.931	.939
.93	.860	.895	.895	.918	.907	.924	.913	.927	.916	.928	.918	.928	.920	.929
.92	.840	.880	.880	.907	.893	.913	.900	.916	.904	.917	.907	.918	.909	.919
.91	.820	.865	.865	.895	.880	.903	.888	.906	.892	.907	.895	.908	.897	.908
.90	.800	.850	.850	.883	.867	.892	.875	.895	.880	.897	.883	.898	.886	.898
.89	.780	.835	.835	.872	.853	.881	.863	.885	.868	.886	.872	.887	.874	.888
.88	.760	.820	.820	.860	.840	.870	.850	.874	.856	.876	.860	.877	.863	.878
.87	.740	.805	.805	.848	.827	.859	.838	.864	.844	.866	.848	.867	.851	.868
.86	.720	.790	.790	.837	.813	.848	.825	.853	.832	.855	.837	.857	.840	.858
.85	.700	.775	.775	.825	.800	.838	.813	.843	.820	.845	.825	.846	.829	.847
.84	.680	.760	.760	.813	.787	.827	.800	.832	.808	.835	.813	.836	.817	.837
.83	.660	.745	.745	.802	.773	.816	.788	.822	.796	.824	.802	.826	.806	.827
.82	.640	.730	.730	.790	.760	.805	.775	.811	.784	.814	.790	.816	.794	.817
.81	.620	.715	.715	.778	.747	.794	.763	.801	.772	.804	.778	.805	.783	.807
.80	.600	.700	.700	.767	.733	.783	.750	.790	.760	.793	.767	.795	.771	.796
.79	.580	.685	.685	.755	.720	.773	.738	.780	.748	.783	.755	.785	.760	.786
.78	.560	.670	.670	.743	.707	.762	.725	.769	.736	.773	.743	.775	.749	.776
.77	.540	.655	.655	.732	.693	.751	.713	.759	.724	.762	.732	.765	.737	.766
.76	.520	.640	.640	.720	.680	.740	.700	.748	.712	.752	.720	.754	.726	.756
.75	.500	.625	.625	.708	.667	.729	.688	.738	.700	.742	.708	.744	.714	.746
.74	.480	.610	.610	.697	.653	.718	.675	.727	.688	.731	.697	.734	.703	.735
.73	.460	.595	.595	.685	.640	.708	.663	.717	.676	.721	.685	.724	.691	.725
.72	.440	.580	.580	.673	.627	.697	.650	.706	.664	.711	.673	.713	.680	.715
.71	.420	.565	.565	.662	.613	.686	.638	.696	.652	.700	.662	.703	.669	.705

.70	.400	.550	.550	.650	.600	.675	.625	.685	.640	.690	.650	.693	.657	.695
.69	.380	.535	.535	.638	.587	.664	.613	.675	.628	.680	.638	.683	.646	.684
.68	.360	.520	.520	.627	.573	.653	.600	.664	.616	.669	.627	.672	.634	.674
.67	.340	.505	.505	.615	.560	.643	.588	.654	.604	.659	.615	.662	.623	.664
.66	.320	.490	.490	.603	.547	.632	.575	.643	.592	.649	.603	.652	.611	.654
.65	.300	.475	.475	.592	.533	.621	.563	.633	.580	.638	.592	.642	.600	.644
.64	.280	.460	.460	.580	.520	.610	.550	.622	.568	.628	.580	.631	.589	.634
.63	.260	.445	.445	.568	.507	.599	.538	.612	.556	.618	.568	.621	.577	.623
.62	.240	.430	.430	.557	.493	.588	.525	.601	.544	.607	.557	.611	.566	.613
.61	.220	.415	.415	.545	.480	.578	.513	.591	.532	.597	.545	.601	.554	.603
.60	.200	.400	.400	.533	.467	.567	.500	.580	.520	.587	.533	.590	.543	.593
.59	.180	.385	.385	.522	.453	.556	.488	.570	.508	.576	.522	.580	.531	.583
.58	.160	.370	.370	.510	.440	.545	.475	.559	.496	.566	.510	.570	.520	.573
.57	.140	.355	.355	.498	.427	.534	.463	.549	.484	.556	.498	.560	.509	.562
.56	.120	.340	.340	.487	.413	.523	.450	.538	.472	.545	.487	.550	.497	.552
.55	.100	.325	.325	.475	.400	.513	.438	.528	.460	.535	.475	.539	.486	.542
.54	.080	.310	.310	.463	.387	.502	.425	.517	.448	.525	.463	.529	.474	.532
.53	.060	.295	.295	.452	.373	.491	.413	.507	.436	.514	.452	.519	.463	.522
.52	.040	.280	.280	.440	.360	.480	.400	.496	.424	.504	.440	.509	.451	.511
.51	.020	.265	.265	.428	.347	.469	.388	.486	.412	.494	.428	.498	.440	.501
.50	.000	.250	.250	.417	.333	.458	.375	.475	.400	.483	.417	.488	.429	.491
.49		.235	.235	.405	.320	.448	.363	.465	.388	.473	.405	.478	.417	.481
.48		.220	.220	.393	.307	.437	.350	.454	.376	.463	.393	.468	.406	.471
.47		.205	.205	.382	.293	.426	.338	.444	.364	.452	.382	.457	.394	.461
.46		.190	.190	.370	.280	.415	.325	.433	.352	.442	.370	.447	.383	.450
.45		.175	.175	.358	.267	.404	.313	.423	.340	.432	.358	.437	.371	.440
.44		.160	.160	.347	.253	.393	.300	.412	.328	.421	.347	.427	.360	.430
.43		.145	.145	.335	.240	.383	.288	.402	.316	.411	.335	.416	.349	.420
.42		.130	.130	.323	.227	.372	.275	.391	.304	.401	.323	.406	.337	.410
.41		.115	.115	.312	.213	.361	.262	.380	.292	.390	.312	.396	.326	.399
.40		.100	.100	.300	.200	.350	.250	.370	.280	.380	.300	.386	.314	.389
.39		.085	.085	.288	.187	.339	.237	.359	.268	.370	.288	.375	.303	.379
.38		.070	.070	.277	.173	.328	.225	.349	.256	.359	.277	.365	.291	.369
.37		.055	.055	.265	.160	.317	.212	.338	.244	.349	.265	.355	.280	.359
.36		.040	.040	.253	.147	.307	.200	.328	.232	.339	.253	.345	.269	.349
.35		.025	.025	.242	.133	.296	.187	.317	.220	.328	.242	.335	.257	.338
.34		.010	.010	.230	.120	.285	.175	.307	.208	.318	.230	.324	.246	.328
.33		.000	.000	.218	.107	.274	.162	.296	.196	.308	.218	.314	.234	.318
.32				.207	.093	.263	.150	.286	.184	.297	.207	.304	.223	.308
.31				.195	.080	.252	.137	.275	.172	.287	.195	.294	.211	.298
.30				.183	.067	.242	.125	.265	.160	.277	.183	.283	.200	.287
.29				.172	.053	.231	.112	.254	.148	.266	.172	.273	.189	.277
.28				.160	.040	.220	.100	.244	.136	.256	.160	.263	.177	.267
.27				.148	.027	.209	.087	.233	.124	.246	.148	.253	.166	.257

.26		.137	.013	.198	.075	.223	.112	.235	.137	.242	.154	.247
.25		.125	.000	.187	.062	.212	.100	.225	.125	.232	.143	.237
.24		.113		.177	.050	.202	.088	.215	.113	.222	.131	.226
.23		.102		.166	.037	.191	.076	.204	.102	.212	.120	.216
.22		.090		.155	.025	.181	.064	.194	.090	.201	.109	.206
.21		.078		.144	.012	.170	.052	.184	.078	.191	.097	.196
.20		.067		.133	.000	.160	.040	.173	.067	.181	.086	.186
.19		.055		.122		.149	.028	.163	.055	.171	.074	.176
.18		.043		.112		.139	.016	.153	.043	.160	.063	.165
.17		.032		.101		.128	.004	.142	.032	.150	.051	.155
.16		.020		.090		.118	.000	.132	.020	.140	.040	.145
.15		.008		.079		.107		.122	.008	.130	.029	.135
.14		.000		.068		.097		.111	.000	.120	.017	.125
.13				.057		.086		.101		.109	.006	.114
.12				.047		.076		.091		.099	.000	.104
.11				.036		.065		.080		.089		.094
.10				.025		.055		.070		.079		.084
.09				.014		.044		.060		.068		.074
.08				.003		.034		.049		.058		.064
.07				.000		.023		.039		.048		.053
.06						.013		.029		.038		.043
.05						.002		.018		.027		.033
.04						.000		.008		.017		.023
.03								.000		.007		.013
.02										.000		.002
.01												.000

الإستنتاجات والتوصيات

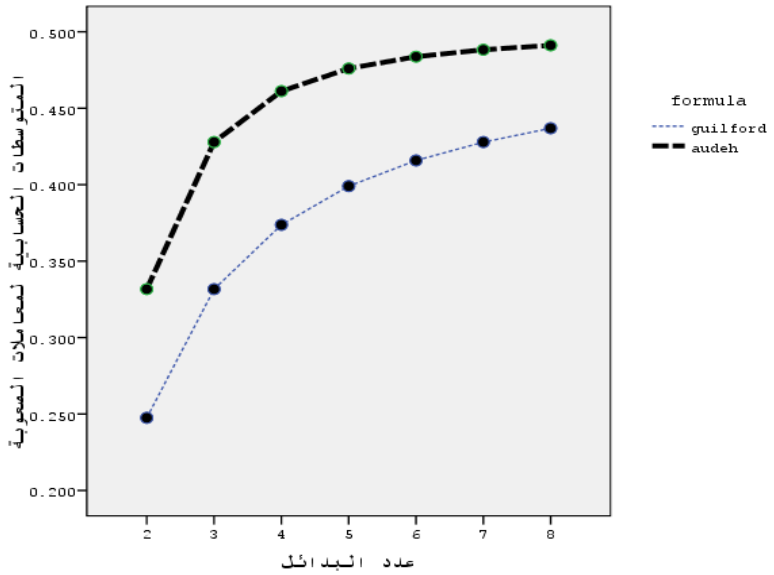
الملاحظ على النتائج من خلال البيانات في الجدول اعلاه مايلي:

اختلاف النهايات لمعاملات صعوبة الفقرات المعدلة بالمعادلتين (جيلفورد وعودة)، اي قيمة معامل الصعوبة قبل التعديل (d) المقابل للمعامل الذي يؤول الى الصفر في كل منهما(d_{cG})، (d^*_{cA}) كما هو مبين في الجدول المصغّر رقم (2).

جدول 2: معاملات الصعوبة قبل التعديل (d) المقابلة للمعاملات التي تؤول الى الصفر في كل من معادلتي التصحيح (d_{cG})، (d^*_{cA})

عدد البدائل(L)							
8	7	6	5	4	3	2	
.12	.14	.16	.20	.25	.33	.50	(d_{cG})
.01	.02	.03	.04	.07	.14	.33	(d^*_{cA})

كما يبن الشكل 1 الفروق التفصيلية بين متوسطات القيم المقدرة بالمعادلتين القائمتين على التصحيح لأثر التخمين مقارنة بقيم معاملات الصعوبة الأصلية عندعدد مختلف من بدائل الإجابة(2–8). مع ملاحظة ان المتوسط بدون اجراء اي تصحيح = 0.5 متبوعا بالجدول 3 الذي يبين هذه المتوسطات والإنحرافات المعيارية للقيم المقدرة على متصل الصعوبة (0.0– 1.0)، ثم نتائج فحص الفروق بين هذه المتوسطات باستخدام الإختبار الإحصائي(t) للبيانات المتجانسة في التباين حسب اختبار ليفين(Levene)المبينة في جدول 4.



الشكل 1: الفروق التفصيلية بين القيم المقدرة بالمعادلتين القائمتين على التصحيح لأثر التخمين عندعدد مختلف من

جدول **3** : **المتوسطات الحسابية والإنحرافات المعيارية لمعاملات الصعوبة المقدرة بمعادلتي جيلفورد وعوده عند عدد مختلف من البدائل**

	formula	Mean	Std. Deviation
d	guilford	.500	.287
	audeh	.500	.287
I2	guilford	.247	.321
	audeh	.332	.332
I3	guilford	.332	.332
	audeh	.428	.319
I4	guilford	.374	.330
	audeh	.461	.307
I5	guilford	.399	.326
	audeh	.476	.300
I6	guilford	.416	.322
	audeh	.484	.296
I7	guilford	.428	.319
	audeh	.488	.294
I8	guilford	.437	.316
	audeh	.491	.292

جدول 4:.نتائج فحص الفروق بين المتوسطات الحسابية والإنحرافات المعيارية لمعاملات الصعوبة المقدرة بمعادلتي جيلفورد وعوده عند عدد مختلف من البدائل.

	Levene's Test for Equality of Variances		t-test for Equality of Means	
	F	Sig.	t	Sig. (2-tailed)
I2	.503	.479	-1.812	.071
I3	.571	.451	-2.078	.039
I4	1.381	.241	-1.931	.055
I5	1.555	.214	-1.727	.086
I6	1.471	.227	-1.544	.124
I7	1.307	.254	-1.387	.167
I8	1.143	.286	-1.255	.211

ويلاحظ من الشكل 1 أن العلاقة بين عدد البدائل ومعاملات الصعوبة المقدرة بالمعادلتين علاقة وتيرية (monotonic)، ، وأن الفروق الرقمية واضحة على مستوى المتوسطات، وعلى مستوى الفقرات عند كل بديل من البدائل المبينة في جدول 1. ويلاحظ من جدول 5 الموضح للفروق بين متوسطات التقديرات بالمعادلتين الانخفاض الجوهري مقارنة بالمتوسط المرجعي عند معامل الصعوبة المتوسط في غياب التصحيح لأثر التخمين (d= 0.50) وخاصة عندما يكون عدد البدائل (2)، ويتناقص هذا الفرق مع تزايد عدد البدائل، ويكاد يتلاشى عند العدد (8) من البدائل وفق معادلة جيلفورد، كما يصبح الفرق بين التقديرات بالمعادلتين غير دال احصائياً عند هذا العدد.

أما دلالة هذه الفروق عن (d= 0.50) كما هي مبينة في جدول 5 فتشير الى ان معادلة جيلفورد المعدّلة أكثرعدالة منها قبل التعديل في مراعاة المعرفة الجزئية ، فمن المنطقي ان يصبح الفرق غير دال احصائيا بزيادة عدد البدائل.

جدول 5 : الفروق بين متوسطات التقديرات بالمعادلتين والمتوسط العام لمعاملات صعوبة الفقرات دون اي تعديل المشار اليه بالمتوسط المرجعي عند (d= 0.50)

	Original Guilford Correction Formula				Modified Guilford Correction Formula			
	mean	s.d.	t-value	sig.	mean	s.d.	t-value	sig.
Reference	.500	.287			.500	.287		
I2	.247	.321	-7.824	0.000	.332	.332	-5.037	0.000
I3	.332	.332	-5.037	0.000	.428	.319	-2.253	0.027
I4	.374	.330	-3.807	0.000	.461	.307	-1.259	0.211
I5	.399	.326	-3.080	0.003	.476	.300	-0.795	0.428
I6	.416	.322	-2.601	0.011	.484	.296	-0.547	0.586
I7	.428	.319	-2.253	0.027	.488	.294	-0.397	0.692
I8	.437	.316	-1.989	0.050	.491	.292	-0.301	0.764

وترجح نتائج المقارنة اهمية اعتماد الصورة المعدلة لمعادلة جيلفورد في تقدير معاملات صعوبة الفقرات التي توفر توازنا نسبيا بين عوامل التخمين العشوائي والمخاطرة والمعرفة الجزئية، ومن السهل على الباحثين ومطوري الاختبارات وبنوك الاسئلة استخدام الجدول المعدّ في هذه الدراسة لهذا الغرض. والتوصية أيضا باعتماد النهايات الطرفية المعدلة في تقدير معامل الصعوبة الذي اعتاد المنظرون في مجال القياس التربوي والإختبارات التحصلية بشكل خاص والقائم على الصيغة ($\frac{1+c}{2}$) حيث تقابل فرصة التخمين كما وردت في الجدول رقم(2)، وللتوضيح بمثال، فالمتوسط المقترح لمعاملات الصعوبة في حال الفقرات ذات البديلين ($\frac{1+0.33}{2}$) وليس ($\frac{1+0.50}{2}$) وهكذا بالنسبة للبدائل الأخرى الواردة في الجدول رقم(2). وترجح الأرقام في الجدول (وتحديدا عند مقارنة قيم الدلالات الإحصائية لقيمة t بالمعادلتين مع تزايد عدد البدائل) التوصية بعدم ضرورة التصحيح لأثر التخمين عندما يكون عدد بدائل فقرات الإختبار (4 فأكثر)، وهذا يتفق مع ماأشارت اليه دراسة كلوز وكريستيان(Klaus & Christian, 2007) بأن فرصة التخمين تصبح ضعيفة جداً بخمسة بدائل أو أكثر، وأن التغير في صعوبة الفقرة لا يكون جوهرياً، وهذا يرجح التعامل مع جداول التصحيح المقدمة في هذه الدراسة، وأن التصحيح لمعاملات الصعوبة بالمعادلة المقترحة أكثر مصداقية وواقعية (more authentic). ومن خلال الربط بين الأسس التي انطلقت منها الدراسة الحالية ونتائجها، والدراسات السابقة بشكل عام، ودراسة اسبنوسا وجارديزابال (Espinosa & Gardeazabal, 2010) بشكل خاص، يقترح الباحث على السيكومتريين اعادة النظر في التعريف الإجرائي لإختبار السرعة من جهة وافتراض التحرر من السرعة speediness من جهة اخرى ، وعلاقته بالإفتراضات الأخرى التي تقوم عليها نظرية الفقرة ونماذجها بشكل عام وافتراض أحادية البعد (unidimensionality) بشكل خاص.

References

Chang, S.-H., Lin, P. -C., & Lin, Z. C. (2007). Measures of partial knowledge and unexpected responses in multiple-choice tests. *Educational Technology & Society*, 10 (4), 95-109.

Chiu, T. (2010). *Correction for guessing in the framework of the 3pl item response theory*. A Dissertation submitted to the Graduate School-New Brunswick Rutgers, The State University of New Jersey.

Espinosa, M. P. & Gardeazabal, J. (2010). Optimal correction for guessing in multiple-choice tests. *Journal of Mathematical Psychology*, 54, 415 - 425.

Feldt, L. S. (1993).The relationship between the distribution of item difficulties and test reliability. *Applied Measurement In Education*, 6(1), 37-48.

Guilford, J. (1936). The determination of item difficulty when chance success is a factor. *Psychometrika*, 1(4), 259-264.

Klaus, D. K., & Christian, H. G. (2007). Item difficulty of multiple choice tests dependant on different item response formats – An experiment in fundamental research on psychological assessment. *Psychology Science*, 49 (4), 361-374.

Lau, P. N., Lau, S. H., Hong, K. S. & Usop, H. (2011). Guessing, partial knowledge, and misconceptions in multiple-choice tests. *Educational Technology & Society*, 14 (4), 99–110.

Little, E. (1962). Overestimation for Guessing in multiple choice test scoring. *Journal of Educational Research*,2,245-252.

Lucy, R. B., Tracey J. E., James, H. & Mark, T. (2009). Does correction for guessing reduce students' performance on multiple-choice examinations? Yes? No? Sometimes? *Assessment and Evaluation in Higher Education*, 34(1), 1-15.

Martin, G. W., Erwin, K. T., & George W. H. (1972). The formula score -a correction for chance or a chance correction*? Personnel Psychology*, 25, 75-93.
Osterlind, S. (2006). Modern measurement: theory, principles, and applications of mental appraisal. Columbus: Pearson.
Robert, B. F. (1988). Formula Scoring of multiple choice tests (correction for guessing). *Educational Measurement: Issues and Practice*, 7(2), 33-38.
Thomas, J., Neal, R., Alex, C., & Anne, C. (2006). Correcting for guessing increases validity in multiple-choice examinations in an oral and maxillofacial pathology course. *Journal of Dental Education*, 70 (4), 378-386.

Abstract

Revision of Guilford formula to Correct Item Difficulty for Guessing
in Multiple Choice Test Items

The original Guilford formula for estimation of multiple choice item difficulty was based on the penalty for guessing. This penalty was originally based on completely random or blind guessing, which means that it is purely based on mathematical estimation and on significantly violated assumptions. While authentic and fair estimation is expected to be based on mixed scoring formula which adds another correction factor to integrate measurement theory with decision theory based on partial knowledge and risk- taking behavior. New formula with two correction factors related to guessing, partial knowledge and risk-taking is presented in this paper. Further studies are suggested for reviewing the validation of the main assumptions of item theory models.(**Keywords**: item difficulty, correction for guessing, risk-taking, multiple choice item, partial knowledge, scoring formula)

الملخص

تصحيح معاملات صعوبة الفقرات لأثر التخمين في اسئلة الإختيار من متعدد:
صورة معدّلة لمعادلة جيلفورد

بنيت معادلة جيلفورد الأصلية لتصحيح معامل صعوبة الفقرة لأثر التخمين في اختبار الإختيار من متعدد على العقاب مع الإفتراض بأن التخمين يتم عشوائيا، الا أن التقدير لهذا ا لمعامل الصعوبة يقوم على اساس رياضي بحت لايتحقق فيه الإفتراض الذي قام عليه. وأن المؤشرات العملية والمنطقية ترجح اهمية تقديم صورة معدلة لهذه المعادلة لتكون اكثر عدالة ومصداقية في تقديرمعاملات الصعوبة لفقرات الإمتحانات العقلية بشكل عام، والتحصيلية بشكل خاص، ترتكزفيه هذه الصورة المعدلة على التوفيق والتكامل بين التخمين القائم على المعرفة الجزئية، والفروق الفردية لدرجة المخاطرة في ضوء التعليمات ذات الصلة بالتخمين وما يترتب عليها من اجراءات في التصحيح، وتحليل الدرجات تحليلا جماليا وتفصيليا عاى مستوى الفقرات. وقد قدمت الدراسة الحالية هذه الصورة المعدّلة مرفقة بجداول لإستخراج معاملات الصعوبة القائمة على التخمين العشوائي مقابل معاملات الصعوبة المقدرة بالمعادلتين الأصلية والمعدلة وترجح نتائج المقارنة لهذه التقديرات توافقا اعلى بين نتائج التقديرات لمعاملات الصعوبة المشتقة من الصورة المقترحة للمعادلة والأساس النظري الذي ترتكز عليه، وهي معادلة توفيقية تجمع بين نظرية القرار ونظرية القياس.(**الكلمات المفتاحية** :صعوبة الفقرة، التصحيح لأثر التخمين، المخاطرة في الإجابة، فقرات الإختيار من متعدد، المعرفة الجزئية، معادلة التصحيح).