تكافؤ/ثبات القياس في البحوث النفسية والتربوية مقارنة بين التحليل العاملي التوكيدي متعدد المجموعات ونظرية الاستجابة للمفردة

أ. د. محمد حبشي حسين محمد
أستاذ علم النفس التربوي
كلية التربية جامعة الإسكندرية

ملخص :

تهدف العديد من البحوث النفسية والتربوية إلى إيجاد الفروق بين مجموعتين أو أكثر من الأفراد في متغير واحد أو في عدد من المتغيرات النفسية والتربوية، أو إيجاد الفروق بين أداء مجموعة واحدة من الأفراد عبر فترات زمنية مختلفة على مقياس أو أكثر. في تلك المواقف تظهر أهمية تكافؤ/ثبات القياس Measurement Invariance\Equivalence الذي يهدف إلى التحقق من أن الأداة تعمل بنفس الطريقة عبر المجموعات المختلفة وأن الفروق المحتملة بين المجموعات في تلك المتغيرات الكامنة التي تقيسها الأدوات ترجع بالفعل إلى فروق في مقدار ما تمتلكه كل مجموعة من السمة وليس إلى اختلاف في البنية العاملية أو في الأوزان النسبية لكل عبارة من مجموعة إلى أخرى. وقد هدفت العديد من الدراسات الأجنبية (e.g., Jabrayilov et al. 2017) إلى التحقق من شرط تكافؤ القياس كمتطلب قبلي للمقارنة بين الذكور والإناث(e.g., Elosua & Hermosilla, 2013)، أو بين ثقافات مختلفة (e.g., Ren et al. 2015)، أو مقارنة استجابات نفس المجموعة من الأفراد عبر فترات زمنية مختلفة (e.g., Moreira et al.2018). وقد استخدمت بعض تلك الدراسات التحليل العاملي التوكيدي متعدد المجموعات (e.g An et al., 2017)، في حين استخدم البعض الأخر نماذج نظرية الاستجابة للمفردة وبصفة خاصة الأداء الفارق أو المتمايز للمفردة (e.g., Sahhin, 2018)، وقارن عدد قليل جداً من الدراسات بين نتائج الطريقتين (e.g., Nelson, 2006). إلا إنه وفي حدود علم الباحث لا توجد دراسة عربية هدفت إلى التحقق من تكافؤ القياس باستخدام أي من المنهجين، لهذا فقد هدفت الدراسة الحالية إلى تقديم الإجراءات المتبعة في كل طريقة وذلك للتحقق من تكافؤ قياس قائمة حل المشكلة Problem Solving Inventory التي اعدها كل من Heppner and Petersen (1982) والتي تم تطبيقها على عينة مكونة من ٤٣١ طالبة من طالبات كلية التربية جامعة الإسكندرية المتزوجات (١٣٢) وغير المتزوجات (٢٩٩)، وقد أظهرت النتائج تقارب النتائج من الطريقتين.

تكافؤ/ثبات القياس في البحوث النفسية والتربوية مقارنة بين التحليل العاملي التوكيدي متعدد المجموعات ونظرية الاستجابة للمفردة

أ. د. محمد حبشي حسين محمد
أستاذ علم النفس التربوي
كلية التربية جامعة الإسكندرية

مقدمة

تستخدم مقاييس التقرير الذاتي Self-Report Measures بكثرة في البحوث النفسية والتربوية، بهدف تقييم الجوانب المختلفة للسلوك البشري. تتكون تلك المقاييس من عدد من العبارات التي تم إعدادها لتقييم ظاهرة كامنة بهدف متابعة الأفراد عبر فترات زمنية مختلفة كما يحدث في الدراسات الطولية، أو بهدف مقارنة المجموعات بعضها ببعض. لكي نضمن مصداقية تلك المقارنة يجب أن تقيس الأداة نفس التكوين الفرض أو المفهوم بنفس البناء عبر المجموعات. في تلك الحالة، يقال أن تلك الأداة متكافئة القياس. إذا أمكن إثبات تكافؤ القياس فإن المشاركون في المجموعات المختلفة يفسروا عبارات المقياس بنفس الطريقة. بمجرد أن اثبتنا تكافؤ القياس، فإن الدراسات المستقبلية يمكنها استخدام تلك الأداة لمقارنة محددات وعواقب درجات العامل الكامن الذي تقيسه الأداة. على الجانب الأخر إذا لم يتحقق تكافؤ القياس، فإن هذا معناه أن المجموعات أو الأفراد من وقت إلى أخر يستجيبوا بطريقة مختلفة لعبارات المقياس ونتيجة لذلك فإن متوسطات العامل لا يمكن مقارنتها بطريقة منطقية.

فقد أشار كلاً من Putnick and Bornstein (2016) إلى أن دراسة ما ربما تقيس نفس المعرفة أو السلوك (الشكل) عبر المجموعات أو الأوقات، لكن تلك المعرفة أو السلوك ربما يكون لها معنى مختلف (وظيفة) للمجموعات المختلفة أو في التوقيتات المختلفة. فقد تطبق المجموعات المختلفة معاني مختلفة لنفس المعلومة أو السلوك. لهذا، تعتمد المقارنة الصحيحة والمناسبة لتكوين فرضي بين المجموعات أو عبر الأوقات المختلفة في المقام الأول على التأكد من تكافؤ معنى المفهوم أو التكوين الفرضي. فعلى سبيل المثال، عند مقارنة الاكتئاب بين الذكور والإناث، فإن عبارات مثل البكاء المتكرر، أو زيادة الوزن أو مشاعر الوحدة النفسية لها معاني مختلفة بين الجنسين. فإن استخدام تلك العبارات في مقياس للاكتئاب للمقارنة بين الذكور والإناث في الاكتئاب يعد عمل مضلل، حيث إن تلك المؤشرات الثلاثة لها معاني مختلفة بين المجموعتين. في هذا المثال، فأن الرجال ربما

يعد كل من نمذجة المعادلة البنائية ونظرية الاستجابة للمفردة مدخلين لفنيات النمذجة اللذان يستخدمان بصورة واسعة لتقويم تكافؤ القياس في البحوث التربوية والنفسية. تربط نمذجة المعادلة البنائية استجابات المفردة المشاهدة بالسمات الكامنة غير المشاهدة، والبيانات المشاهدة يمكن أن تكون أما متصلة أو فئوية. يشار عادة إلى نمذجة المعادلة البنائية على أنها نمذجة بناء التباين المشترك Covariance Structure Modeling لأن تقديرات معالم النموذج يتم الحصول عليها بواسطة جعل الفروق بين مصفوفة التباين المشترك المشاهدة ومصفوفة التباين المشترك المشتقة من النموذج أصغر ما يمكن. تشمل عائلة نمذجة المعادلة البنائية أنواع مختلفة من فنيات النمذجة، مثل تحليل المسار، والتحليل العاملي التوكيدي، تحليل النموذج البنائي الكامل. يعد التحليل العاملي التوكيدي منهج النمذجة الشائعة لفحص ما إذا كانت العلاقات بين السمات الكامنة ومؤشرتها متسقة مع النموذج المقترح اعتماداً على النظريات والبحوث السابقة.

على الجانب الأخر، تستخدم نظرية الاستجابة للمفردة احتمالية أنساق الاستجابة لتقدير مستويات السمات الكامنة للمستجيبين. تستخدم نظرية الاستجابة للمفردة بصورة واسعة في تدريج إحصاءات المفردات ودرجات الممتحنين على المقاييس النفسية والتربوية. يمكن تقسيم نظرية الاستجابة للمفردة إلى نوعين اعتماداً على أبعاد التكوينات الافتراضية المستهدفة: نظرية الاستجابة للمفردة أحادية البعد التي تتطلب سمة كامنة وحيدة، ونظرية الاستجابة للمفردة متعددة الأبعاد حيث يوجد سمات متعددة. العبارات في كل نوع يمكن أن تكون أما ثنائية أو متعددة أو خليط من عدد مختلف من فئات الاستجابة عبر المفردات.

ناقشت دراسات سابقة التشابه والاختلاف بين نظرية الاستجابة للمفردة والتحليل العاملي التوكيدي على المستويين النظري والتطبيقي (Tate, 2003, Wirth & Edwards, 2007). السمة الكامنة، يطلق عليها أيضاً القدرة، في نموذج نظرية الاستجابة للمفردة أحادية البعد تقابل العامل الكامن في التحليل العاملي التوكيدي أحادي العامل. في حالة المقاييس ذات المقاييس الفرعية المتعددة، يمكن استخدام نظرية الاستجابة للمفردة متعددة الأبعاد (محمد حبشي حسين، ٢٠١٨) التي فيها تقابل السمات الكامنة المتعددة العوامل في نموذج التحليل العاملي التوكيدي متعدد العوامل. يمكن تحويل التشبعات المعيارية والعتبات الفارقة thresholds في نموذج التحليل العاملي التوكيدي المعياري إلى معاملات التمييز ومستويات صعوبة المفردات في نموذج نظرية الاستجابة للمفردة ثنائي البارامتر للمفردات الثنائية (Finch, 2010; Swygert, Mcleod, & Thissen, 2001).

مشكلة الدراسة

يعد تكافؤ/ثبات القياس أحد الطرق الإحصائية التي تمكن الباحثين من التحقق بطريقة إحصائية عما إذا كانت أداة قياس ما تعكس بصورة دقيقة فروق في تكوين فرضي ما. حيث أن تكافؤ القياس يعد بمثابة متطلب قبلي لإجراء مقارنات بين مجموعات. في الحقيقة، فإن الاستنتاجات المشتقة من تحليلات مقارنة ربما تكون متحيزة أو غير صادقة إذا كانت القياسات ليست لها نفس المعنى عبر المجموعات، بمعنى أن الاستنتاجات المتعلقة بفروق المتوسطات بين المجموعات على مقياس ما تكون غير صادقة إذا كان المفهوم الكامن الذي يفترض أنه يقاس يكون مختلف عبر المجموعات (Vandenberg & Lance, 2000).

بالرغم من تاريخ الاهتمام باختبار تكافؤ/ثبات القياس يمتد إلى أكثر من ٥٠ سنة (e.g., Meredith, 1964)، لكن الفنيات الإحصائية لاختبار التكافؤ أصبحت أكثر اتاحة لمجتمع الباحثين فقط حديثاً. في الحقيقة، فإن تكافؤ القياس أصبح بصورة سريعة أكثر دقة في البحوث النفسية والنمائية. بحلول القرن الواحد والعشرين، وجه المتخصصون في مجال العلوم النفسية والتربوية اهتمامهم بصورة متزايدة بدلالة ومغزى تكافؤ القياس، بصفة خاصة داخل إطار نمذجة المعادلة البنائية (Vandenberg, 2002). في البحثين الرائدين اللذان قام بهما Widaman and Reise (1997)، Vandenberg and Lance (2000) حيث اجريا مراجعة لأدبيات تكافؤ القياس، ورسما مدخل يشبه السلم لاختبار تكافؤ القياس، وقدما للباحثين توجيهات خطوة بخطوة لتنفيذ اختبارات التكافؤ. فقد لاحظ Vandenberg (2002) بعد الانتهاء مباشرة من نشر المقالتين السابقتين أن هناك زيادة كبيرة جداً في عدد البحوث التي اهتمت بالتحقق من تكافؤ المقاييس، وأن تلك الزيادة استمرت بصورة أشد في العقد التالي، إلا أن تلك الزيادة لم يصاحبها نصائح متسقة أو كافية، أو تفسير أو ممارسة جيدة أو فهم (Putnick & Bornstein, 2016). إذا كان هذا هو حال البحوث الأجنبية، ففي حدود علم الباحث أن هذا الاهتمام بالتحقق من تكافؤ القياس لم ينتقل إلى البحوث العربية، وذلك لعدم معرفة الكثير من الباحثين في وطنا العربي بهذا المفهوم، وعدم معرفة البعض الأخر بالخطوات التي يجب اتباعها لاختبار تحقق تكافؤ القياس. لهذا فإن هذا البحث الحالي هو هدف تعليمي في المقام الأول وذلك بتوجيه انظار الباحثين إلى أهمية هذا المفهوم وكيف يمكن اختباره باستخدام المنهجين الأكثر انتشاراً وهما التحليل العاملي التوكيدي ونماذج نظرية الاستجابة للمفردة، وتقديم مثال تطبيقي لتنفيذ تلك الخطوات، ولهذا فإن البحث الحالي يحاول الإجابة عن الأسئلة التالية:

١. ما الإجراءات المتبعة للتحقق من تكافؤ القياس باستخدام التحليل العاملي التوكيدي؟

٢. ما الإجراءات المتبعة للتحقق من تكافؤ القياس باستخدام نماذج نظرية الاستجابة للمفردة؟

٣. ما أجه الشبه والاختلاف بين المدخلين في اختبار تكافؤ القياس؟

٤. ما أهم النقاط التي يجب ذكرها في تقرير البحث في حالة استخدام كل مدخل من المدخلين السابقين؟

أهداف الدراسة

يهدف البحث الحالي إلى ما يلي:

١. توجيه انظار الباحثين إلى مفهوم تكافؤ/ثبات القياس واهمية هذا المفهوم في البحوث النفسية والتربوية التي تهتم بالمقارنة سواء بين الأفراد أو الثقافات أو بين فترات زمنية مختلفة لنفس الأفراد.

٢. عرض لخطوات اختبار تكافؤ القياس باستخدام التحليل العاملي التوكيدي، ونظرية الاستجابة للمفردة، مع الإشارة إلى جوانب التشابه والاختلاف بين المنهجين.

٣. تقديم مثال تطبيقي على خطوات تكافؤ القياس باستخدام قائمة حل المشكلة

أهمية الدراسة

ترجع أهمية البحث الحالي إلى أهمية وحداثة الموضوع الذي يتناوله وهو تكافؤ/ثبات القياس، كما ترجع إلى ندرة البحوث العربية التي اهتمت بهذا المفهوم، وفي عدد وكم البحوث التي تهتم بعملية المقارنة، حيث أن هذا النوع من البحوث هو الأكثر انتشاراً بين أنواع البحوث الأخرى، فقد أظهرت المراجعة التي قام بها الباحث لقاعدة بيانات Proquest أن نسبة البحوث التي تهتم بالمقارنة هي النسبة الأكبر، كما أن التحقق من تكافؤ القياس لا يقتصر فقط على البحوث الوصفية، ولكن يمتد إلى الدراسات التجريبية، سواء تلك التي تستخدم منهج المجموعة الواحدة، أو المجموعتين، كما أن تكافؤ القياس يكون له دور هام في الدراسات عبر الثقافية، حيث أن التحقق من تكافؤ قياس الأدوات المستخدمة في عملية المقارنة قبل استخدام الدرجات الخام عملية سوف تسهم في تحسين جودة ونتائج البحوث وتجعلها أكثر دقة. كما تقدم الدراسة مقياس مقنن للقدرة المدركة على حل المشكلة يمكن استخدامه للمقارنة بين المتزوجات وغير المتزوجات من طالبات الجامعة.

مصطلحات الدراسة

تكافؤ/ثبات القياس: يقصد بتكافؤ القياس أن نفس التكوين الفرضي يقاس بطريقة مماثلة عبر المجموعات. السؤال الرئيسي الكامن خلف اختبارات تكافؤ القياس عبر المجموعات هو ما إذا كانت معالم (أو بارامتر) تحليلية عاملية معينة مثل التشبعات والثوابت وتباينات الخطأ وتباينات العوامل والتباين المشترك للعوامل ومتوسطات العوامل يمكن افتراض أنها متكافئة عبر المجموعات. يتحقق تكافؤ القياس عندما تقس أداة ما نفس التكوين الفرضي (مثل الاكتئاب) بنفس الطريقة عبر المجموعات المختلفة من المستجيبين؛ وفي هذه الحالة يفترض أن درجات المقياس تحمل نفس المعنى عبر المجموعات (Terluin et al., 2016)

وفي الدراسة الحالية يشير تكافؤ أو ثبات القياس إلى درجة المطابقة النموذج الذي يعبر عن العلاقة بين الدرجات المشتقة من أدوات القياس والتكوينات الافتراضية التي يتم قياسها عبر عدة مجموعات (Hayes, Smith, & Hunsley, 2011)

التحليل العاملي التوكيدي Confirmatory Factor Analysis إجراء إحصائي متعدد المتغيرات يستخدم للتحقق من جودة تمثيل عدد من المتغيرات المشاهد لعدد من التكوينات الفرضية. يعد التحليل العاملي التوكيدي أسلوب إحصائي مرن وقوي، وهو يركز على نمذجة العلاقة بين المتغيرات الظاهرة (المشاهدة) والتكوينات الافتراضية المستترة.

يعرف التحليل العاملي التوكيدي متعدد المجموعات Multiple Group Confirmatory Factor Analysis بأنه أسلوب إحصائي متعدد المتغيرات يستخدم لمقارنة البنية العاملية التي تمثل العلاقة بين عدد من المتغيرات المشاهدة والتكوينات الافتراضية التي تقيسها عبر المجموعات المختلفة أو عبر فترات زمنية مختلفة لنفس المجموعة.

تشير نظرية الاستجابة للمفردة Item Response theory التي يطلق عليها ايضاً اسم نظرية السمات الكامنة Latent Trait Theory إلى فئة من النماذج الرياضية التي تحاول أن تفسر العلاقة بين السمات الكامنة (الخصائص أو الصفات غير المشاهدة) ومظاهرها (أي الأداء أو الاستجابات أو النواتج المشاهدة). نظرية الاستجابة للمفردة هي مدخل يستخدم لبناء وتقويم وتصحيح أدوات القياس. تصف نماذج نظرية الاستجابة للمفردة العلاقة بين استجابات الفرد لعبارات أداة القياس ومكانه أو موقعه على التكوين الفرضي (مثل القدرة الحسابية، أو شدة الاكتئاب، أو مستوى الإجهاد) الكامن (غير القابل للمشاهدة)

الإطار النظري للدراسة

أولا: تكافؤ القياس

يعد JÖreskog (1971) أول من كتب حول تكافؤ البناء العاملي. وقد تم تقديم هذا المفهوم بواسطة Byrne, Shavelson and Muthén (1989)، ومنذ ذلك التاريخ أنطلق اختبار تكافؤ/ثبات القياس. يتحقق تكافؤ القياس عندما تتطابق الدرجات المشاهدة والتكوينات الفرضية الكامنة عبر المجموعات (Drasgrow & Kanfer, 1985). وفي أحد المراجعات التي قام بها كل من Vandenberg & Lance (2000) توصلا إلى أن زيادة في أعداد البحوث المهتمة باختبار تكافؤ القياس، كما أنهما قدما دليل ارشادي خطوة بخطوة لإجراء تكافؤ القياس (يمكن للقارئ المهتم الرجوع إلى تلك المقالة للحصول على مراجعة دقيقة لتاريخ تكافؤ القياس والمعادلات الرياضية المستخدمة). وقد لاحظ Vandenberg (2002) بعد ذلك بفترة قصيرة استمرار الاهتمام بقضية تكافؤ القياس، إلا أنه لم يصاحب هذا الاهتمام نصائح متسقة وكافية أو تحسن في الفهم أو التطبيق.

تكافؤ/ثبات القياس من منظور التحليل العاملي التوكيدي

يعد التحليل العاملي التوكيدي أسلوب إحصائي يستخدم للتحقق من البنية العاملية لعدد من المتغيرات المشاهدة.، ويستخدم هذا الأسلوب لاختبار الفروض المتعلقة بالعلاقة بين المتغيرات المشاهدة وتكويناتها الافتراضية الكامنة. أي أن التحليل العاملي التوكيدي إجراء إحصائي متعدد المتغيرات يستخدم للتحقق من جودة تمثيل عدد من المتغيرات المشاهد لعدد من التكوينات الفرضية. يعد التحليل العاملي التوكيدي أسلوب إحصائي مرن وقوي، وهو يركز على نمذجة العلاقة بين المتغيرات الظاهرة (المشاهدة) والتكوينات الافتراضية المستترة.

يعد التحليل العاملي التوكيدي متعدد المجموعات Multiple Group Confirmatory Factor Analysis المنهج الأكثر استخداماً لتقييم تكافؤ/ثبات القياس داخل إطار نمذجة المعادلة البنائية. في الأصل تم تطوير مدخل التحليل العاملي التوكيدي متعدد المجموعات ليناسب البيانات المتصلة لكن تم توسيعه ليشمل البيانات الفئوية (Millsap & Yun-Tein, 2004). في التحليل العاملي متعدد المجموعات، يتم مقارنة نماذج متداخلة بالنسبة إلى مطابقة بيانات النموذج العام، لكي يتم فحص ما إذا كانت المعالم المقيدة المضافة في النموذج الأقل تعقيداً تكون متكافئة عبر المجموعات. إذا كان نموذج القياس في كل مجموعة يشارك نفس النسق من العناصر المساوية للصفر وغير المساوية للصفر في مصفوفة تشبعات المفردات والعوامل، فإن التكافؤ الشكلي

تحقق. يتم في تلك الخطوة اختبار مطابقة بيانات النموذج ودقة بناء الأبعاد. إذا تكافأت مصفوفة تشبعات المفردات والعوامل عبر المجموعات، فإن هذا معناه أن التكافؤ/الثبات المتري قد تحقق. أي أن هذه الخطوة تتحقق من أن عدد العوامل والمفردات التي تشبعت على كل عامل كانت ثابتة عبر المجموعات. يتطلب ثبات التدريج أن القواطع Intercepts للمفردات تكون متطابقة عب المجموعات، الذي يعد متطلب سابق لمقارنة متوسطات السمات الكامنة أو المعالم البنائية عبر المجموعات.

قام Vandenberg and Lace (2000) بمراجعة دقيقة لأدبيات تكافؤ/ثبات القياس ولخص عدد من اختبارات النماذج المتداخلة الموصى بها لاكتشاف عدم تحقق تكافؤ/ثبات القياس. يجري الباحثون عادة اختبار عام لتساوي مصفوفة التباين المشترك؛ إذا اظهر الاختبار العام أنه لا توجد فروق في مصفوفات التباين المشترك بين فئات البيانات، في هذه الحالة يستنتج الباحث أن شروط تكافؤ/ثبات القياس تحققت لتلك البيانات. لكن، شكك بعض الباحثين في فائدة هذا الاختبار النوعي على أساس أن هذا الاختبار يمكن أن يشير إلى أن تكافؤ القياس متحقق بصورة منطقية عندما تجد الاختبارات الأكثر نوعية أو الأكثر تحديداً خلاف ذلك. بصرف النظر عما إذا كان الاختبار العام لتساوي مصفوفة التباين المشترك يشير إلى عدم تحقق تكافؤ/ثبات القياس، فإن سلاسل من اختبارات فرق مربع كاي للنماذج المتداخلة يمكن إجراءها لتحديد المصادر المحتملة للفروق.

يحدد تكافؤ/ثبات القياس من منظور التحليل العاملي التوكيدي متعدد المجموعات من خلال مستويات تكافؤ/ثبات متمايزة، مع قيود تساوي إضافية على معالم النموذج عبر المجموعات تضاف للمستوى الأعلى. على سبيل المثال، يفترض التكافؤ/الثبات الشكلي Configural Invariance فقط أن بنية العوامل بصورة عامة تكون واحدة عبر المجموعات. بمعنى أن، عدد العوامل ونسق العوامل تكون واحدة عبر المجموعات. الثبات المتري (الضعيف) Metric Invariance يتطلب أن تكافؤ تشبعات العوامل عبر المجموعات. يتطلب ثبات التدريج (القوي) Scalar Invariance تكافؤ أو ثبات كل من تشبعات العوامل والعتبات الفارقة Thresholds عبر المجموعات. يتطلب تكافؤ/ثبات القياس الصارم Strict Measurement Invariance أن تكون تباينات بواقي المفردات ثابتة عبر المجموعات بالإضافة على تشبعات العوامل والعتبات الفارقة.

وصف Marsh, Muthén, Asparouhov, Lüdtke, Robitzsch, Morin, and

Trautwein (2009) تصنيف مكون من ثلاث عشر نماذج متداخلة بصورة جزئية تتباين بين نموذج للتكافؤ/الثبات الشكلي إلى نموذج للتكافؤ/الثبات التام حيث قيود التساوي وضعت على تشبعات العوامل، وقواطع المفردات، وتباين بواقي المفردات، بالإضافة إلى متوسطات وتباينات والتباينات المشتركة للعوامل الكامنة. وقد أشار Milfont and Fischer (2010) إلى أنه في حالة نماذج القياس يوجد أربع نماذج شائعة: التكافؤ الشكلي، والمتري والتدريج والخطأ. وسوف نتناول تلك النماذج في الجزء التالي.

التكافؤ/الثبات الشكلي Configural Invariance

الخطوة الأولى، والأقل تقييد، في سلم تكافؤ القياس هي التكافؤ الشكلي أو ثبات شكل النموذج. صممت تلك الخطوة لاختبار ما إذا كانت المفاهيم الكامنة تحمل نفس النسق من التشبعات الحرة والثابتة (على سبيل المثال، تلك التي تم تقديرها بواسطة النموذج والأخرى التي تم تثبتها لتساوي صفر) عبر المجموعات. يقصد بالتكافؤ على المستوى الشكلي أن التنظيم القاعدي أو الأساسي للمفاهيم الكامنة تم اثباتها في المجموعات محل المقارنة. عدم التكافؤ الشكلي يعني أن نسق التشبعات للمفردات على العوامل الكامنة يختلف في المجموعات محل المقارنة، أي إنه في أحد المجموعات يوجد على الأقل مفردة واحدة تشبعت على عامل مختلف، أو تشبعت بصورة مركبة على عاملين أو أكثر.

وجود عدم تكافؤ شكلي يتضمن بديلين: (١) إعادة تعريف التكوين الفرضي (مثل حذف بعض المفردات وإعادة اختبار النموذج) أو (٢) افتراض أن المفهوم غير متكافؤ وعدم الاستمرار في تكافؤ القياس واختبار الفروق بين المجموعات (Putnick & Bornstein, 2016). إعادة تعريف التكوين الفرضي في أي خطوة لاختبار التكافؤ (على سبيل المثال، بحذف مفردات أو إجراء أي تعديلات بعدية للنموذجpost-hoc alternatives) استراتيجية محكومة بواسطة البياناتdata-driven، أكثر من النظريةtheory-driven، بمعنى أنها ذات طبيعة استكشافية، ويجب أن تفسر النتائج في ضوء ذلك. للتأكد من مصداقية تعديلات النموذج الاستكشافية، يتم تكرار الدراسة الميدانية باستخدام عينات مماثلة أمر في غاية الأهمية، ويطلق على تلك العلمية عبر المصداقية Cross Validation.

التكافؤ/الثبات المتري Metric Invariance

إذا تحقق التكافؤ الشكلي، الخطوة التالية اختبار التكافؤ المتري، أو تكافؤ تشبعات المفردات على العوامل. يعنى التكافؤ المتري أن كل مفردة تسهم في المفاهيم الكامنة بدرجة مماثلة عبر

المجموعات. يختبر التكافؤ المتري بواسطة جعل تشبعات العوامل متكافئة عبر المجموعات. يقارن النموذج المقيد العوامل بالنموذج المتكافئ من ناحية الشكل لتحديد المطابقة. إذا كان مطابقة النموذج أسوأ بصورة دالة في نموذج التكافؤ المتري مقارنةً بنموذج تكافؤ الشكل، فإن هذا يشير إلى أن تشبع واحد على الأقل غير متكافأ عبر المجموعات، وأن التكافؤ المتري غير متحقق. إذا كان مطابقة النموذج العام ليست سيئة بصورة دالة في نموذج التكافؤ المتري مقارنة بنموذج التكافؤ الشكلي، فإن هذا دليل على أن افتراض تساوي التشبعات عبر المجموعات لم يؤثر بصورة دالة على جودة النموذج، ومن ثم فإن هذا يثبت أن النموذج متكافأ من ناحية المترية.

اكتشاف عدم التكافؤ المتري يطرح ثلاثة اختيارات: (١) فحص مصدر عدم التكافؤ عن طريق تحرير (في مدخل الرجوع للخلف Backward approach) أو إضافة (في مدخل التقدم للأمام Forward approach يمكن الرجوع إلى Jung & Yoon, 2016)قيود على تشبعات العوامل وإعادة اختبار النموذج إلى أن يتحقق التكافؤ أو الثبات الجزئي؛ (٢) حذف المفردات ذات التشبعات غير المتكافئة وإعادة اختبار النموذجين الشكلي والمتري؛ (٣) افتراض أن التكوين الفرضي غير متكافئ وعدم مواصلة تكافؤ القياس واختبار الفروق بين المجموعات.

تكافؤ التدريج Scalar Invariance

نحتاج إلى تكافؤ التدريج أو المقدار الثابت intercept لمقارنة متوسطات السمات الكامنة، يشير تحقق تكافؤ التدريج إلى أن الدرجات المشاهدة ترتبط بالدرجات الكامنة، بمعنى أن من لهم نفس الدرجة على السمة الكامنة يحصلون على نفس الدرجة المشاهدة بصرف النظر عن المجموعة التي ينتمون إليها. يتم التحقق من تكافؤ قياس التدريج من خلال تثبيت المقادير الثابتة للمفردات بحيث تصبح واحدة عبر المجموعات (Milfont and Fischer, 2010). إذا أظهرت النتائج أن مطابقة النموذج بصفة عامة أسوأ بصورة دالة في نموذج تكافؤ التدريج مقارنةً بنموذج التكافؤ المتري، فإن هذا يعد بمثابة مؤشر على أن يوجد على الأقل ثابت مفردة واحدة يختلف عبر المجموعات، وأن البيانات لا تؤيد تكافؤ التدريج. على الجانب الأخر إذا أظهرت النتائج أن المطابقة العامة للنموذج ليس أسوأ بصورة دالة إحصائية في نموذج تكافؤ التدريج مقارنةً بنموذج التكافؤ المتري، فإن هذا مؤشر على أن تثبيت قواطع المفردات عبر المجموعات لا يؤثر بصورة دالة إحصائياً على مطابقة النموذج، وأن البيانات تؤيد تكافؤ التدريج.

إذا لم يتحقق تكافؤ التدريج فإن الباحث يجد نفسه أمام ثلاث اختيارات: (١) فحص مصدر عدم

التكافؤ من خلال التخلص المتتابع (في مدخل الرجوع للخلف) أو الإضافة (في مدخل التقدم للأمام) من قيود قاطع المفردة وإعادة اختبار النموذج حتى نحصل على نموذج متكافئ بصورة جزئية، (٢) حذف المفردات ذات القواطع غير المتكافئة وإعادة اختبار نماذج التكافؤ الشكلي والمتري والتدريج، أو (٣) افتراض أن التكوين الفرضي غير متكافئ وعدم مواصلة اختبار تكافؤ القياس والفروق بين المجموعات (Putnick and Bornstein, 2016)

تكافؤ البواقي Residual Invariance

إذا أظهرت النتائج تحقق تكافؤ التدريج، فإن الخطوة الأخيرة تتمثل في تكافؤ القياس تتمثل في اختبار تكافؤ بواقي المفردات المتكافئة على المستوى المتري والتدريج. يقصد بتكافؤ البواقي أن مجموع التباين النوعي (تباين المفردة غير المشترك مع العامل) وتباين الخطأ (خطأ القياس) يكون متماثل عبر المجموعات. بالرغم أن تكافؤ البواقي يعد مكون متطلب للوصول إلى التكافؤ التام Full Invariance، إلا أن اختبار تكافؤ البواقي ليس متطلب قبلي لاختبار فروق المتوسطات لأن البواقي ليست جزء من العامل الكامن، لهذا فإن تكافؤ بواقي المفردات ليس له صلة بتفسير فروق متوسطات السمات الكامنة (Vandenberg & Lance, 2000). على هذا الأساس يحذف كثير من الباحثين تلك الخطوة. إلا أننا ندرجها هنا نظراً لأن تكافؤ البواقي مازال يذكر في تقرير بحوث تكافؤ القياس (Putnick and Bornstein, 2016)

يتم اختبار تكافؤ البواقي من خلال تقييد بواقي المفردات لكي تكون متكافئة عبر المجموعات. هذا القيد يدخل على المفردات التي أظهرت تكافؤ متري وتدريج، كما هو الحال في تكافؤ التدريج فإن أي مفردة تحمل تشبعات مختلفة و/أو قواطع مختلفة يسمح لها أن تختلف عبر المجموعات (أي لا تقيد) في نموذج تكافؤ البواقي. يقارن النموذج المقيد في بواقي المفردات بنموذج تكافؤ التدريج لتحديد المطابقة. إذا كانت مطابقة النموذج بصورة عامة أسوأ بصورة دالة في نموذج تكافؤ البواقي مقارنةً بنموذج تكافؤ التدريج فإن هذا مؤشر على أنه يوجد على الأقل بواقي مفردة واحدة مختلف عبر المجموعات، وأن تكافؤ البواقي لم يتحقق. على الجانب الأخر، إذا كانت المطابقة العامة ليست أسوأ بصورة دالة في نموذج تكافؤ البواقي مقارنة بنموذج تكافؤ التدريج فإن هذا يدل على تحقق تكافؤ البواقي.

إذا لم يتحقق تكافؤ البواقي، يمكن للباحث أن (١) يفحص مصدر على تكافؤ البواقي من خلال التخلص التدريجي (في مدخل الرجوع للخلف) أو إضافة (في مدخل التقدم للامام) قيود بواقي المفردات وإعادة اختبار النموذج إلى أن نصل إلى نموذج متكافئ بصورة جزئية، أو (٢) قبول

عدم تكافؤ البواقي واستكمال اختبارات فروق المتوسطة أو العلاقات الفارقة في العوامل الكامنة عبر المجموعات. (Putnick and Bornstein, 2016)

مطابقة نماذج تكافؤ القياس

في التحليل العاملي التوكيدي متعدد المجموعات، يستخدم اختبار الفرق في قيمة مربع كاي ومؤشرات المطابقة الوصفية مثل مؤشر جودة المطابقة Comparative Fit Index (CFI)، مؤشر تكر ولويس Tucker and Lewis Index (TLI)، جذر متوسط مربعات خطأ التقريب Root Mean Square Error of Approximation (RMSEA)، جذر متوسط البواقي المعياري Standardized Root Mean Residual (SRMR) (Chen, 2007; Thompson & Green, 2013). يؤشر فرق مربع كاي الدال إلى أن يوجد بارامتر مقيد واحد على الأقل في النموذج الأكثر تقيداً يخلف بصورة دالة عبر المجموعات. على الجانب الأخر، يشير اختبار مربع كاي غير الدال إلى أن مطابقة بيانات-النموذج للنموذج الأكثر تقيداً ليس أسوأ بصورة دالة من النموذج الأقل تقيداً، وعليه فإن المعالم المقيدة تكون مستقلة عبر المجموعات وأن مزيد من خصائص تكافؤ القياس يمكن دراساتها. تستخدم مؤشرات التعديلات Modification Indices للبحث عن المعالم غير الثابتة أو المتكافئة، وتقت الإجراءات التدريجية لاختبار تكافؤ القياس باستخدام التحليل العاملي التوكيدي بصورة جيدة (Millsap & Yun-Tein, 2004; Thompson & Green, 2013).

المؤشرات التي تم استخدامها في الدراسة الحالية لتقييم المطابقة العامة للنموذج: مربع كاي المعياري، نسبة مربع كاي إلى درجة الحرية، جذر متوسط مربعات خطأ التقريب، وجذر متوسط مربعات البواقي المعياري Standardized Root Mean Square Residual (SRME). نسبة مربع كاي إلى درجة الحرية ٣:١ أو أقل تشير إلى مطابقة جيدة (Carmines & Mclver, 1981). قيم جذر متوسط مربعات خطأ التقريب التي تقع في مدى ٠.٠٠٦ إلى ٠.٠٠٨ تشير إلى تطابق مقبول في حين تشير القيم التي تقع في مدى ٠.٠٠٨ إلى ٠.١٠ إلى تطابق معتدل، والقيم أكبر من ٠.١٠ تشير إلى تطابق سيء (Browne & Cudeck, 1993; Cheung & Rensvold, 2002)، كما استخدم الفرق في قيم كاي تربيع، ومؤشر المطابقة المقارن، ومؤشر عبر المصادقة Cross Validation index كمؤشرات مطابقة تزايدي لحساب التحسن عبر النماذج المتنافسة. يشير اختبار مربع كاي الدال إلى أن النموذج ذو القيمة الأصغر لمربع كاي يحمل مطابقة أفضل من الناحية الإحصائية. إلا أن هذا

الاختبار يحمل نفس عيوب اختبار مربع كاي من حيث أنه يتأثر بحجم العينة. أي إنه في حالة العينات كبيرة الحجم، فإن الفروق البسيطة تعطي نتائج اختبار دالة. لهذا فإن اختبار مربع كاي يستخدم فقط كمؤشر على التحسينات الدالة. مؤشر المطابقة المقارن القرب من ٠.٩٥ يشير إلى مطابقة مقبولة. القيم الأصغر لمؤشر عبر المصادقة المتوقع يعكس النموذج ذو المطابقة الأفضل (Milfont & Fischer, 2010).

نظراً لتأثر اختبار فرق مربع كاي بحجم العينة، فقد اقترح Cheung and Rensvold (2002) محك أكثر قابلية للتطبيق: مقدار التغير في قيمة مؤشرة المطابقة المقارن لتحديد ما إذا كانت النماذج المتداخلة متكافئة من الناحية التطبيقية. في الدراسة الحالية، عندما زاد الفرق في مؤشر جودة المطابقة المقارن عن ٠.٠١ بين النموذجين المتداخلين، فإن النموذج الأكثر تقيداً يرفض حيث أن إضافة قيود على النموذج أحدثت مطابقة أسوأ من الناحية التطبيقية. لكن عندما لا يتحقق هذا المحك وبعض المعالم (مثل تشبعات العوامل أو القواطع) لم يتم تحددها على أنها متساوية بين المجموعتين، يمكن في هذه الحالة النظر بعين الاعتبار إلى نموذج متكافئ القياس بصورة جزئية.

تكافؤ/ثبات القياس من منظور نظرية الاستجابة للمفردة

بالرغم أن التحليل العاملي التوكيدي متعدد المجموعات هو الطريقة الأكثر استخداماً للتحقق من تكافؤ القياس في البحوث النفسية والتربوية، إلا أن مناهج نظرية الاستجابة للمفردة استخدمت أيضاً لنفس الغرض، وفي بعض الحالات قدمت معلومات مختلفة وأكثر فائدة للتحقق من تكافؤ القياس (Raju, Laffitte, Byrne, 2002). تشير نظرية الاستجابة للمفردة التي تعرف أيضا بنظرية الاستجابة الكامنة إلى فئة من النماذج الرياضية التي تحاول أن تفسر العلاقة بين السمات الكامنة (الخصائص أو الصفات غير المشاهدة) ومظاهرها (أي الأداء أو الاستجابات أو النواتج المشاهدة). تؤسس نظرية الاستجابة للمفردة رابطة بين خصائص المفردات على مقياس ما والتي يجيب عليها الأفراد والسمة الكامنة موضوع القياس. تقدم نظرية الاستجابة للمفردة أساس للطرق الإحصائية التي تستخدم في سياقات مثل إعداد الاختبارات، تحليل المفردات، وبنوك الأسئلة والاختبارات المتوافقة بالكمبيوتر.

تستخدم نظرية الاستجابة للمفردة لتحقيق واحدة أو أكثر من الأهداف التالية: (أ) تقويم الخصائص السيكومترية لأدوات القياس؛ (ب) اختبار تكافؤ القياس في الاستجابات لمقاييس تم تطبيقها عبر مجتمعات مختلفة؛ (ج) ربط أداتين أو أكثر يقيسان مجالات متشابهة على تدريج

مشترك؛ (د) إعداد اختبارات متوافقة تقدر مكانة الفرد على تكوين فرضي ما بأقل عدد ممكن من الأسئلة.

يفترض إطار نظرية الاستجابة المفردة نموذج علاقة منحنية وليست خطية لوصف العلاقة بين استجابات المفردة المشاهدة ومستوى السمة الكامنة التي يرمز لها بالرمز θ. تتحدد الطبيعة الدقيقة لهذا النموذج بواسطة فئة من معالم المفردة التي تكون فريدة لكل مفردة. يعتمد تقدير السمة الكامنة لفرد معين على استجابات المفردات المشاهدة بمعلومية معالم المفردات. يوجد بارامتران عادة ما يتم تقديرهم لكل مفردة. التمييز أو البارامتر الذي يمثل ميل خط مسار المفردة (يطلق عليه المنحنى المميز للمفردة أو الدالة المميزة للمفردة) الذي يحدد العلاقة بين السمة الكامنة والدرجة المشاهدة). النوع الثاني من معالم أو بارامتر المفردة هو موقع المفردة أو مستوى الصعوبة (يمكن الرجوع إلى محمد حبشي حسين، ٢٠١٨ لمزيد من المعلومات حول نظرية الاستجابة للمفردة). في كثير من الأحوال، سوف يستخدم الباحثون مقاييس متعددة الاستجابة أكثر من التصحيح الثنائي البسيط. بالرغم من أن هناك عدة نماذج في نظرية الاستجابة للمفردة لتلك النوع من المفردات المتعددة الاستجابة، فإن واحدة من تلك النماذج الأكثر استخداماً هو نموذج الاستجابة المتدرجة Graded Response Model (Samejima, 1969). في حالة نموذج الاستجابة المتدرجة، فإن العلاقة بين احتمال أن شخص ما ذو قدرة ما يختار بديل استجابة مفردة معينة يمكن التعبير عنه بصورة بيانية باستخدام دالة استجابة فئة المفردة (لمزيد من التفاصيل حول ذلك يمكن الرجوع إلى صلاح الدين محمود علام، ٢٠٠٥).

يماثل معامل التمييز في نظرية الاستجابة للمفردة ويرتبط بصورة رياضية بتشبعات العوامل في منهج التحليل العاملي التوكيدي (McDonald, 1989). في المقابل لا يوجد مكافئ واضح لبارامتر معامل الصعوبة في التحليل العاملي التوكيدي، بالرغم أن قواطع أو ثوابت intercepts المفردات في التحليل العاملي التوكيدي التي تعرف القيمة المشاهدة للمفردة عندما تكون قيمة التكوين الفرضي الكامن مساوية للصفر هو أكثر تشابه من الناحية المفاهيمية لبارامتر معامل الصعوبة. لكن، بينما يوجد ثابت واحد فقط يمكن تقديره لكل مفردة في التحليل العاملي التوكيدي، تقدر نظرية الاستجابة للمفردة المعتمدة على نموذج التقدير الجزئي عدة معالم للصعوبة لكل مفردة. اختبارات نظرية الاستجابة للمفردة لتكافؤ القياس تختلف اعتماداً على النمط النوعي للمنهجية المستخدمة. أكثر الطرق استخداماً هي اختبار نسبة الأرجحية Likelihood Ratio Test . ينفذ اختبار الارجحية العظمى على مستوى المفردة، ومثل مناهج التحليل العاملي التوكيدي، يستخدم تقدير الارجحية العظمى maximum likelihood

estimation لتقدير معالم المفردة التي تنتج قيمة لمطابقة النموذج تعرف على إنها دالة المطابقة fit function. في اختبار الارجحية العظمى، تلك القيمة لدالة المطابقة تعد مؤشر على جودة مطابقة النموذج المعطى للبيانات كنتيجة لإجراءات تقدير الارجحية العظمى المستخدمة في تقدير معالم المفردة (Camilli & Shepard, 1994). يتضمن اختبار الارجحية العظمى مقارنة مطابقة نموذجين النموذج المدمج compact model والنموذج الزائد المقارن. أولاً، يقيم النموذج القاعدي baseline model حيث تقدر معالم جميع المفردات تحت قيد أن معالم المفردة للمفردات المتشابهة (مثل المفردة الأولى للمجموعة الأولى والمفردة الأولى للمجموعة الثانية) تكون متساوية عبر المواقف (Thissen, Steinberg, & Wainer, 1988). يقدم هذا النموذج المدمج قيمة ارجحية قاعدية لمطابقة معالم المفردة للنموذج. يلي ذلك، تختبر كل مفردة، واحدة في كل مرة، للأداء المتمايز للمفردة. لاختبار الأداء المتمايز لكل مفردة، بصورة منفصلة

في نظرية الاستجابة للمفردة، يعرف تحليل تكافؤ/ثبات القياس عادة على إنه اكتشاف الأداء المتمايز للمفردة Differential Item Functioning (DIF). يوجد أداء متمايز للمفردة عندما تكون العلاقة بين الدرجة على المفردة والسمة الكامنة ليست متطابقة عبر المجتمعات الفرعية (Drasgow, 1984). الأداء المتمايز التراكمي لجميع المفردات على مقياس تمثل الأداء المتمايز للاختبار differential test functioning حيث العلاقة بين الدرجة على الاختبار والسمة الكامنة غير متطابقة عبر المجتمعات الفرعية. يوجد مرجعات متعددة لإجراءات/إحصاءات الأداء المتمايز للمفردة (e.g., Penfield & Camilli, 2007). لاختبار الأداء المتمايز للمفردة لكل مفردة، يتم إجراء تحليل منفصل لكل مفردة حيث تقيد جميع تقديرات معالم المفردات المتماثلة لتصبح متساوية عبر المواقف عبر المواقف (مثل الفترات الزمنية أو مجموعات الأفراد)، باستثناء معالم المفردة موضع الاختبار للأداء المتمايز. يقدم هذا النموذج الزائد قيمة ارجحية تقترن بتقدير معالم المفردة بصورة منفصلة لكل مجموعة. تلك القيمة للارجيحية يمكن مقارنتها بقيمة الارجحية للنموذج المدمج حيث المعالم لجميع المفردات المتماثلة تكون مقيدة لتكون متساوية عبر المجموعات. تعطي المعادلة التالية صيغة اختبار نسبة الأرجحية:

$$LR_i = \frac{L_C}{L_{Ai}} \qquad (1.1)$$

حيث يشير L_C إلى دالة الارجحية للنموذج المدمج (الذي يحتوي على عدد أقل من البارامتر) ويرمز L_{Ai} إلى دالة الارجحية للنموذج الزائد حيث يسمح للمفردة محل الاهتمام

أن تختلف عبر المواقف.

$$\chi^2(M) = -2\ln(LR) = -2\ln(L_C) + 2\ln(L_{Ai}) \quad (1.2)$$

تحويل اللوغاريتم الطبيعي لتلك الدالة يمكن اخذه والناتج احصاءة اختبار تتوزع على صورة مربع كاي في حالة تحقق الفرض الصفري، حيث M تساوي الفرق في عدد معالم المفردة المقدرة في النموذج المدمج مقابل النموذج الزائد (أي الفرق في درجة الحرية بين النموذجين). هذا يعد مؤشر على سوء المطابقة حيث تشير النتائج الدالة إلى أن النموذج المدمج يطابق بصورة أكثر سوءً من النموذج الزائد. لكي نستخدم اختبار نسبة الارجحية، يتم حساب قيمة مربع كاي لكل مفردة في الاختبار. تلك المفردات ذو قيم مربع كاي الدالة يقال أنها تظهر أداء متمايز للمفردة (أي استخدام تقديرات معالم مفردة مختلفة يحسن المطابقة العامة للنموذج). بالرغم أن اختبار نسبة الارجحية يجرى عادة بواسطة عدد من التنفيذيات المجهدة لبرنامج MULILOG (Thissen, 1991)، فقد لاحظ الباحث الحالي أن Thissen (2001) أدرج مدخل نسبة الارجحية بطريقة ابسط خلال برنامجه الحديث IRTLRDIF. هذا التعديل للمدخل الذي نقدمه هنا اسهل بكثير للباحثين من الطريقة التقليدية. بالإضافة إلى أن فحص Cohen, Kim, and Wollack (1996) قدرة اختبار نسبة الارجحية لاكتشاف الأداء المتمايز للمفردة تحت تنوع من المواقف باستخدام بيانات تم توليدها بنظام المحاكاة Simulation. بالرغم أن البيانات التي تم توليدها ثنائية أكثر من أنها متعددة في طبيعتها، فإنهم وجدوا أن معدلات الخطأ من النوع الأول لمعظم النماذج التي تم فحصها تقع داخل المستويات المقبولة. بصفة عامة، فقد استنتجوا أن المؤشر يسلك بطريقة ممتازة بدرجة معقولة.

التشابه والاختلاف بين المدخلين

تكافؤ القياس في التحليل العاملي التوكيدي والأداء المتمايز أو الفارق للمفردة في نظرية الاستجابة للمفردة يحملان تشابهات واختلافات. يعبر الأداء الفارق للمفردة على إنه عدم تحقق لتكافؤ القياس، الذي يركز على فحص ما إذا كان المفحوصين الذين لهم مستوى متساوي من القدرة لهم احتمالات مختلفة للإجابة على المفردة بطريقة صحيحة. إذا كان المفردة تعمل بصورة مختلفة عبر المجموعات، فإن المفردة يضع عليها علامة على إنها مفردة تؤدي بصورة متمايزة مما يشير إلى احتمال تحيز المفردة. تفحص الممارسة الشائعة لطرق اكتشاف الأداء المتمايز أو الفارق للمفردة اعتماداً على نظرية الاستجابة للمفردة ما إذا كانت مفردة ما تعمل بطريقة متماثلة عبر المجموعات، الأمر الذي يشابه التكافؤ المتري وتكافؤ التدريج في التحليل

العاملي التوكيدي. بالرغم من بعض الدراسات تستخدم مصطلحي تكافؤ القياس في التحليل العاملي التوكيدي والأداء المتمايز للمفردة في نظرية الاستجابة للمفردة كمرادفين (Suh & Cho, 2014, Tay, Meade, & Cao, 2015)، إلا إنه يوجد فروق في الاستخدام التطبيقي حيث أن التحليل العاملي التوكيدي ونظرية الاستجابة للمفردة ليسا بالضبط نفس تكنيك النمذجة. بالإضافة إلى أن أغراض استخدام التحليل العاملي التوكيدي ونظرية الاستجابة للمفردة لاختبار التكافؤ القياس مختلفان من الناحية التاريخية.

الهدف في حالة تحليل تكافؤ القياس باستخدام التحليل العاملي التوكيدي غالباً ما يكون هو أثبات أن الدرجة المشاهدة للمجموعات (أي الدرجات الخام للمقياس ككل) قابلة للمقارنة بحث يمكن التوصل إلى تفسيرات جوهرية اعتماداً على الفروق المشاهدة بين المجموعات. ينظر إلى إجراءات تكافؤ القياس باستخدام التحليل العاملي التوكيدي على أنها متطلب قبلي للفروق بين متوسطات المجموعات محل المقارنة على المتغير محل الدراسة. هناك محاولات لاستخدام تكافؤ القياس باستخدام التحليل العاملي التوكيدي لأثبات أن تكافؤ تفسير المتغير الكامن عبر المجموعات (e.g., Mencl, Tay, Schwoerer, & Drasgow, 2012). إلا إن الاهتمام والتركيز في حالة التحليل العاملي التوكيدي عادة ما يكون على مستوى بنية التكوين الفرضي أكثر من الأداء المفردات الفردية.

على عكس تكافؤ القياس باستخدام التحليل العاملي التوكيدي، طورت إجراءات تكافؤ القياس باستخدام نظرية الاستجابة للمفردة لتقييم تكافؤ القياس للمفردات أو بصورة أكثر دقة عدم تكافؤ القياس للمفردات، ومن ثم الأداء الفارق أو المتمايز للمفردة. وهذا بسبب أن نظرية الاستجابة للمفردة طورت في سياق القياس للتركيز على تحديد مفردات الاختبار النوعية التي ربما تظهر تحيز ضد مجموعة معينة (هامة جداً لبنوك الأسئلة). هذا ليس معناه أن تكافؤ درجة المقياس أو الاختبار ليس مهمة، حيث يستطيع الباحث أن يحدد مدى التأثير المتراكم للأداء المتمايز للمفردات على الفروق في المتوسطات المشاهدة بين المجموعات (أي الأداء المتمايز للاختبار) حيث منحى الاستجابة للاختبار يكون مختلف عبر المجموعات. بسبب الاهتمام التطبيقي لبناء الاختبار في نظرية الاستجابة للمفردة، يوجد أيضا اهتمام أقل بالفروق في بنية التكوين الفرضي بين المجموعات عندما لا يوجد تكافؤ في القياس. بالأحرى، يستخدم الأداء المتمايز للمفردة لتحديد القضايا المحتملة في محتوى وصياغة المفردة. كما أن التأكيدات المختلفة بين نظرية الاستجابة للمفردة والتحليل العاملي التوكيدي، على الترتيب، قادت إلى اختلافات في المفاهيم/الإجراءات بالرغم من وجود تشابهات ضمنية. أحد الفروق البارزة في الخطوات

المعتادة لتحليل تكافؤ القياس، تكافؤ القياس في التحليل العاملي التوكيدي يختبر بصفة عامة بطريقة تدريجية، تقدير أنماط التكافؤ من شكلي إلى متري إلى تدريجي إلى متفرد (Vandenberg & Lance, 2000). في كل خطوة، توضع قيود أكثر على نموذج ما، وإذا وجد أن المطابقة أصبحت أسوأ بصورة دالة، فإن هذا مؤشر على أن نمط التكافؤ محل الاهتمام لم يتحقق (Cheung & Rensvold, 2002). في هذا الإجراء يتم التأكيد على أنماط التكافؤ ويتم ربطها بمعاني جوهرية. على سبيل المثال، افتقار المقياس للتكافؤ الشكلي سوف يشير إلى أن المفهوم محل الاهتمام يحمل تعريفات مختلفة بصورة جوهرية عبر المجموعات. في حالة نظرية الاستجابة للمفردة، تقدير أحادية البعد يقابل الخطوة الأولى للتكافؤ الشكلي في التحليل العاملي التوكيدي. من غير الشائع (بالرغم أنه محتمل نتيجة لتكافؤ المعالم لبعض نماذج نظرية الاستجابة للمفردة) أن يتم توظيف نفس خطوات التحليل العاملي التوكيدي لأن الهدف هو تحديد ما إذا كانت مفردات معينة لها أداء فارق أو متمايز.

لاختبار تكافؤ القياس باستخدام نظرية الاستجابة للمفردة، يتم في الغالب إجراء تحليل على مستوى المفردات. يحدث هذا غالباً من خلال مقارنة نماذج نظرية الاستجابة للمفردة باستخدام أما مدخل الخط الأساسي الحر free-baseline أو الخط الأساسي المقيد تماماً fully constrained baseline approach (Stark, Chernyshenko, & Drasgow, 2006). يلاحظ أن كلا المدخلين الحر والمقيد تماماً يمكن تطبيقهما أيضاً في إطار تكافؤ القياس باستخدام التحليل العاملي التوكيدي، إلا أن هذا نادراً في إطار تكافؤ القياس باستخدام التحليل العاملي التوكيدي.

الفرق الجوهري الثاني هو أنه بالرغم من أنه يمكن باستخدام نظرية الاستجابة للمفردة أن نختبر الفروق بين الصور المختلفة للأداء المتمايز على مستوى المفردة، فإن التركيز كان على وجود أداء متمايز للمفردة أكثر من الطبيعة النوعية للأداء المتمايز للمفردة. ليس هذا معناه أن شكل أو صورة الأداء المتمايز للمفردة ليس له أهمية. بل أن الهدف عادة ما يكون هو تحديد ما إذا كان الأداء المتمايز للمفردة يؤدي إلى أداء متمايز للاختبار الذي له تأثير جوهري على مقارنة درجة الاختبار (e.g., stark et al., 2004). ربما ينتج الأداء المتمايز غير المنتظم للمفردة Nonuniform DIF عدم تأثير على الدرجات المشاهدة حيث ربما يتم تعويض الدرجات المنخفضة للمجموعة البورية (عند مقارنتها بالمجموعة المرجعية) على سمة كامنة ما بالدرجات المرتفعة على مستوى سمة أخرى. أكثر من ذلك، في عدد من إجراءات نظرية الاستجابة للمفردة، يوجد بصفة عامة قوة أقل لاكتشاف الأداء المتمايز غير المنتظم للمفردة (من الأداء

المتمايز المنتظم للمفردة (e.g., Finch & French, 2007, Tay, Vermunt, & Wang, 2013).هذا يكون جنب إلى جنب بالاهتمام بالجوهري المتضمن في صورة عدم التكافؤ في التحليل العاملي التوكيدي لتكافؤ القياس (مثل التكافؤ المتري أو التكافؤ التدريجي).

الفرق الثالث هو المتطلب القبلي لأحادية البعد لمعظم نماذج نظرية الاستجابة للمفردة التي تستخدم حالياً، حيث تتطلب تلك النماذج أن يتكون المقياس من بعد واحد، فمن المستحيل أن نقيم تكافؤ القياس للعلاقات بين تكوينات افتراضية بنفس الطريقة التي تتم في التحليل العاملي التوكيدي. بالرغم من أنه تم تطوير نماذج حديثة متعددة الأبعاد لنظرية الاستجابة للمفردة (محمد حبشي حسين، ٢٠١٨) إلا أن تلك النماذج لم يتم استخدامها بصورة واسعة في البحوث العربية والأجنبية.

الفرق الأخير هو أنه في تحليل تكافؤ القياس باستخدام التحليل العاملي التوكيدي تستخدم غالباً مؤشرات العاملة لنمذجة المعادلة البنائية لمطابقة النموذج للبيانات. تشمل تلك المؤشرات مؤشر المطابقة المقارنة Comparative Fit Index (CFI) ومؤشر المطابقة ير المعياري Non-normed fit index أو ما يطلق عليه مؤشر تكر ولويس Tucker and Lewis Index (TLI)، وجذر متوسط مربعات خطأ التقريب Root Mean Square Error of Approximation (RMSEA) (e.g., Cheung & Rensvold, 2002). على الجانب الأخر، تستخدم بصورة واسعة مؤشرات مستوى المفردة لاكتشاف الأداء المتمايز للمفردة في تحليل تكافؤ القياس باستخدام نظرية الاستجابة للمفردة. تلك المؤشرات تكون أكثر اتساعاً من اختبار تكافؤ معالم المفردة (e.g., Lord, 1980)، وتشمل أشياء كثيرة من بينها: المساحة بين منحنيات الاستجابة للمفردة، والانحدار اللوجستي، وفروق الدرجة المشاهدة (Holland & Thayer, 1988; Narayanan & Swaminathan, 1994)

مثال تطبيقي

نتناول في هذا الجزء مثال تطبيقي على الإجراءات المتبعة عن التحقق من تكافؤ القياس باستخدام الطريقتين: التحليل العاملي التوكيدي متعدد المجموعات، ونظرية الاستجابة للمفردة.

المشاركون في الدراسة

شارك في الدراسة الحالية ٤٣١ طالبة من طالبات كلية التربية المقيدون بالفرقة الرابعة تخصص طفولة في العام الجامعي ٢٠١٨/٢٠١٩م، وقسمت العينة إلى فئتين: المتزوجات وعددهن (١٣٢) وغير المتزوجات وعددهن ٢٩٩، وقد بلغ متوسط العمر الزمني للطالبات

المتزوجات المشاركات في البحث ٢١,٤٦ وانحراف معياري ٠,٩١، في حين المتوسط والانحراف المعياري للطالبات غير المتزوجات ٢١,٥٥، ٠,٧٧ومدى عمر زمني للعينة ككل هو(٢٠-٢٦) عام.

أداة الدراسة

استخدمت الدراسة الحالية قائمة حل المشكلة Problem Solving Inventory (PSI) من إعداد Heppner (1988)، تستخدم القائمة لتقييم إدراكات الفرد لسلوكيته واتجاهاته الخاصة بحل المشكلة. صممت القائمة لقياس ادراك الفرد لقدراته على حل المشكلة وليس لقياس المهارات الفعلية لحل المشكلة. تتكون القائمة من ثلاث اختبارات فرعية: الثقة في حل المشكلةproblem solving confidence، أسلوب الإقدام-التجنب approach-avoidance style، والضبط الشخصيPersonal Control . بالإضافة إلى الدرجات على الأبعاد الفرعية توجد درجة كلية ينظر إليها على أنها درجة عامة وحيدة لحل المشكلة.

يعرف البعد الأول وهو الثقة في حل المشكلة على إنه الثقة بالنفس أثناء خبرة الاندماج في حل أنشطة حل المشكلة؛ يقترن المقياس الفرعي للإقدام-التجنب بالنزعة العامة للفرد للإقدام أو تجنب مواقف حل المشكلة. يقترن المقياس الفرعي للتحكم الشخصي بكون الفرد متحكم في سلوكه وانفعاله أثناء الاندماج في مواقف حل المشكلة. يتم الحصول على الدرجة الكلية من خلال جمع كل الدرجات لجميع الاستجابات. تتكون القائمة من ٣٥ مفردة، تستبعد منهم ثلاث عبارات لا تدخل في حساب الأبعاد، جميع العبارات تعتمد على أسلوب التقرير الذاتي باستخدام أسلوب ليكرت ذو الست نقاط تتدرج من ١ (موافق بشدة) إلى ٦ (غير موافق بشدة). عندما تجمع الدرجات من المقاييس الفرعية الثلاثة فإننا نعطي درجة تتراوح بين ٣٢ إلى ١٩٢ تعكس قدرة الفرد المدركة على حل المشكلة؛ تشير الدرجات الأقل إلى سلوكيات واتجاهات تقترن بحل المشكلة الناجح. يقاس البعد الأول الثقة في حل المشكلة الابتكارية بإحدى عشر مفردة، ويقاس البعد الثاني أسلوب الأقدام-التجنب بست عشر عبارة، ويقاس البعد الثالث الضبط الشخصي بخمس عبارات. يستغرق تطبيق الأداة من ١٠-١٥ دقيقة. القائمة صممت بصورة أساسية للتطبيق على الراشدين، مؤشرات ثبات القائمة عالية حيث بلغت حوالي ٠.٨٥ لكل اختبار فرعي، ٠,٨٩ للقائمة ككل Heppner (1988). أظهرت القائمة صدق تلازمي وتكوين فرضي قوي عبر الدراسات المختلفة التي استخدمت القائمة ووجدت ارتباطات دالة بين العوامل والدرجة الكلية Heppner (1988).

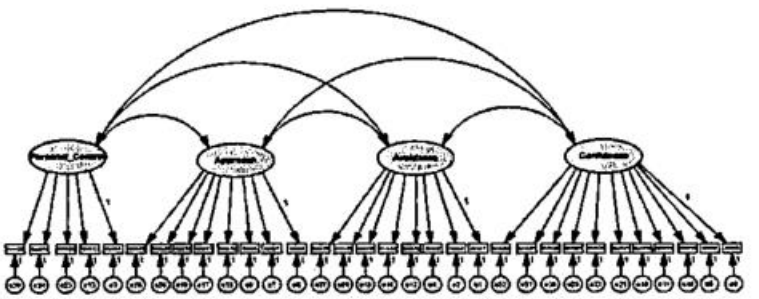
وقد استخدمت القائمة مع عينات مختلفة، ففي دراسة قام بها Terzioglu (2006) على عينة من مديرات التمريض عددها ٧١، وقد أظهرتْ أنّ مديري التمريض الذين ذكروا أنهم تلقوا تعليم حول الإدارة وقرأوا مجلات علمية منشورة وحضروا لقاءات علمية أدركوا انفسهم على أنهم أفضل في مهارات حل المشكلة. كما أظهرت الدراسة على إنه كلما طالت فترة العمل كمدير كلما زادت ادراكات المدراء لقدراتهم على حل المشكلة ولكن بصورة غير دالة إحصائياً.

وفي دراسة حديثة قام بها Ji et al. (2018) للتحقق من الخصائص السيكومترية للقائمة على عينة من مقدمي الرعاية الصحية للأفراد الذين يعانون من فقدان الذاكرة، وقد أظهرت الدراسة أن قيم معاملات الثبات باستخدام كرونباك ألفا تتراوح بين ٠,٨٢ إلى ٠.٩٠ للأبعاد الثلاثة وكانت ٠,٩٢ للمقياس ككل. كما أظهرت الدراسة أن هناك ارتباط متوسط موجب بين القائمة وقائمة بيك للاكتئاب، واظهرت النتيجة أيضاً وجود علاقة سالبة متوسطة بين الدرجة على قائمة حل المشكلة والدرجة على مقياس الكفاءة الذاتية في إدارة الأمراض المزمنة. في حدود علم الباحث لا توجد ترجمة عربية منشورة للقائمة، لهذا فقد قام الباحث بترجمة القائمة إلى اللغة العربية ويوجد في مرفق (١) القائمة بعد عملية الترجمة.

نتائج تكافؤ القياس في إطار التحليل العاملي التوكيدي متعدد المجموعات

تم إجراء أربع تحليلات عاملية توكيدية متعددة المجموعات وذلك لتقييم التكافؤ الشكلي والمتري والتدرجي والبواقي لقائمة حل المشكلة بين المتزوجات وغير المتزوجات من طالبات الجامعة. وقد تم تقدير تلك النماذج باستخدام الارجحية العظمى ذات التقدير القوي للخطأ Maximum Likelihood Robust (MLR) باستخدام برنامج AMOS 25. تلك الطريقة في التقدير تعني أن قيمة الأخطاء المعيارية واختبار مربع كاي تكون قوية بحيث أنها تظل تعطي قيم دقيقة في حالة عدم الاعتدالية وعدم استقلالية المشاهدات (Muthén & Muthén, 2012). وقد قام الباحث بإجراء تحليل عاملي توكيدي منفصل للمتزوجات وغير المتزوجات، وبعد ذلك تم إجراء تحليل عاملي متعدد المجموعات لتقييم التكافؤ الشكلي والمتري والتدريج والبواقي لقائمة حل المشكلة.

الخطوة الأولى للتحقق من تكافؤ القياس في إطار التحليل العاملي التوكيدي متعدد المجموعات هو إيجاد نموذج يتطابق من ناحية الشكل بين المجموعتين (المتزوجات وغير المتزوجات من الطالبات)، وقد استخدم الباحث برنامج AMOS 25 لاختبار تكافؤ القياس في إطار التحليل العاملي التوكيدي متعدد المجموعات.



شكل (١) نموذج قائمة حل المشكلة لطالبات الجامعة المتزوجات وغير المتزوجات.

جدول (١)

مؤشرات جودة مطابقة نموذج قائمة حل المشكلة الابتكارية بين المتزوجات وغير المتزوجات

الفئة	كا^2 (دح) χ^2	مؤشر جودة المقارنة CFI	نسبة مربع كاي لدرجة الحرية	جذر متوسط مربعات الخطأ التقريبي RMSEA	جذر متوسطات البواقي المعياري RMSR
المتزوجات	٦٨٩.٧(٤٥٨)	٠.٩٠٢	١.٥٠٦	٠.٠٧٤	٠.٠٤٦
غير المتزوجات	٧١٣.٤(٤٥٨)	٠.٩٣٥	١.٥٥٨	٠.٠٥٣	٠.٠٥١

يوضح شكل (١) النموذج الذي حقق أفضل جودة مطابقة للمجموعتين، ويوضح جدول (١) مؤشرات جودة مطابقة النموذج لدى عينة الطالبات المتزوجات وغير المتزوجات. كما يتضح من جدول (١) فإن نسبة قيمة مربع كاي إلى درجة الحرية أقل من ٣، كما أن مؤشر جودة المطابقة المقارن أكبر من ٠.٩٠، وايضاً جذر متوسط مربعات الخطأ التقريبي وجذر متوسط البواقي المعيار أقل من ٠.٠٨ مما يدل على أن النموذج يتطابق بدرجة مقبولة داخل المجموعتين، مما يدل على تحقق التكافؤ الشكلي للنموذج المقترح. ولهذا يمكننا الانتقال إلى الخطوة الثانية للتحقق من التكافؤ المتري، حيث نختبر تكافؤ تشبعات المفردات على العوامل التي تنتمي إليها. يوضح جدول (٢) نتائج اختبار التكافؤ المتري، والتدريج والبواقي.

جدول (١)

اختبارات تكافؤ قياس قائمة حل المشكلة بين المتزوجات وغير المتزوجات

التكافؤ	كا^2 (دح) χ^2	مؤشر جودة المقارنة CFI	جذر متوسط مربعات الخطأ التقريبي RMSEA	الفرق في قيمة كا^2 (دح) $\Delta\chi^2$	الفرق في قيمة مؤشر جودة المقارنة ΔCFI
الشكلي	١٤٠٤.٦٦ (٩١٦)	٠,٩٢٧	٠.٠٤٣	———	———
المتري	١٤٣٩.٠١٢ (٩٤٤)	٠,٩٢٣	٠.٠٤٢	(٢٨)٣٤,٣٥	٠.٠٠٤
التدريج	١٤٧٥,٨٥ (٩٧٦)	٠,٩٢١	٠.٠٤٢	(٣٢)٣٦,٨٤	٠,٠٠٢
البواقي	١٥٢١,٧٥ (١٠١٨)	٠,٩٠٩	٠.٠٤١	(٤٢)٤٥,٩٠	٠,٠٠٢

يتضح من جدول (٢) تحقق تكافؤ القياس على المستوى المتري والتدريج والبواقي، فقد كانت الفروق في قيمة مربع كاي غير دالة إحصائياً، كما أن الفروق في قيم مؤشر جودة المطابقة أقل من ٠.٠١، بل أن قيمة جذر متوسط مربعات أخطاء التقريب انخفضت مما يدل على تحسن جودة مطابقة النموذج بإضافة مزيد من القيود.

نتائج تكافؤ القياس باستخدام نظرية الاستجابة للمفردة

للتحقق من تكافؤ القياس باستخدام نظرية الاستجابة للمفردة، ونظرا لأن قائمة حل المشكلة تستخدم أسلوب ليكرت السداسي، فكان لابد من استخدام برنامج للتحليل يتعامل مع هذا النوع من المفردات للكشف عن الأداء المتمايز لمفردات القائمة، لهذا فقد تم تحليل الأداء المتمايز للمفردة باستخدام الحزمة lordif في برنامج R (2011) Choi, Gibbons, & Crane. تستخدم الحزمة lordif إطار ا لانحدار اللوجستي الترتيبي ordinal logistic regression، وتستخدم نموذج الاستجابة المتدرجة Graded Response Model لتقدير السمة في إطار نظرية الاستجابة للمفردة (Crane, Gibbons, Jolley, & van Belle, 2006). وقد تم الاعتماد على محكين لاكتشاف الأداء المتمايز للمفردة الدال في الدراسة الحالية: (١) إحصاءه مربع معامل الارتباط المتعدد الزائف Pseudo R^2 >0.13 (Zumbo, 1999) و (٢) ١٠% تغيرات في قيمة بيتا (Crane et al., 2007). وباستخدام مدخل Cook et al. (2011) تم تقدير الأداء المتمايز للمفردة على الدرجات؛ وذلك بحساب معامل الارتباط بين درجات الأفراد المعدلة للأداء المتمايز للمفردات ودرجات الأفراد الأصلية لدراسة وجود أثر دال للأداء المتمايز للمفردات على الدرجات. الارتباط القوي يشير إلى أن التعديل في الدرجات لمراعاة الأداء المتمايز للمفردات يحمل فروق مهملة في درجات الأفراد. مما يعني أن معالم المفردة التي تم حسابها من خلال دمج المجموعات سويا يمكن استخدامها بدون الاهتمام بالتأثيرات المحتملة للأداء المتمايز للمفردات على تصحيح درجات الأفراد.

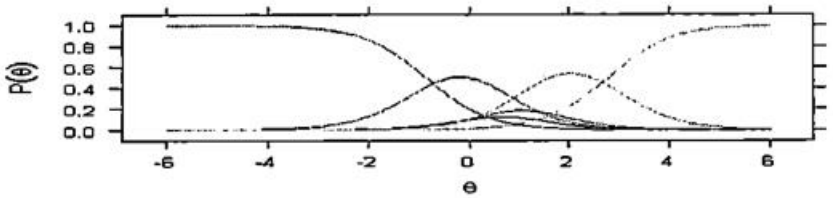
جدول (٣)

قيم معالم المفردة رقم ٢٤ لبعد الثقة في حل المشكلة للمتزوجات وغير المتزوجات.

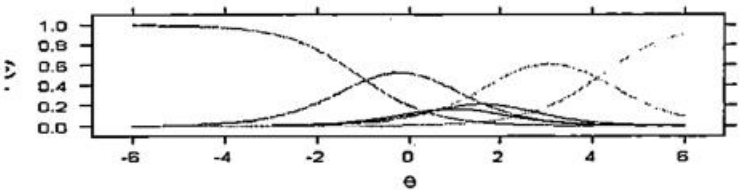
		العتبات الفارقة				
العينة	التمييز	الأولى	الثانية	الثالثة	الرابعة	الخامسة
غير المتزوجات	١.٦٥	٠.٨٨-	٠.٤٦	٠.٧٦	١.٢١	NA
المتزوجات	٠.٨٧	١.٧٢-	٠.٨٩	١.٦٣	٢.٦١	NA

وقد أظهرت النتائج اعتماداً على المحكين التغير في قيمة بيتا أو مربع معامل الارتباط المتعدد الزائف أن جميع مفردات عبارات البعد الأول "الثقة في حل المشكلة" لم تظهر أداء متمايز

باستثناء المفردة رقم ٢٤ التي تنص على " عندما أواجه بموقف جديدة، فأنني يكون عندي ثقة أنني استطيع معالجة المشاكل التي تظهر." فقد أظهرت النتائج، كما يوضحه جدول (٣) أن معامل تمييز تلك المفردة لدى الطالبات غير المتزوجات (١.٦٥٢) أعلى من نظريتها لدى المتزوجات (٠.٨٦٦). كما أظهرت النتائج أيضا أن قيم العتبات الفارقة بمعنى مستوى القدرة اللازم لكي ينتقل المستجيب من بديل أقل إلى بديل أعلى كانت بصورة مستقلة أعلى لدى غير المتزوجات مقارنةً بالمتزوجات، أي أن المتزوجات يحتجن إلى مستوى قدرة أعلى لكي تنتقل مثلا من بديل موافق إلى بديل موافق بشدة، ونظراً لأن قائمة حلا لمشكلة تتضمن ست بدائل فإنه يوجد خمس عتبات فارقة. يوضح الشكلين (٢، ٣) المنحنيات المميزة للمفردة رقم ٢٤ للمتزوجات، وغير المتزوجات على الترتيب.



شكل (٢) المنحنيات المميزة للمفردة رقم ٢٤ للمتزوجات



شكل (٣) المنحنيات المميزة للمفردة رقم ٢٤ لغير المتزوجات

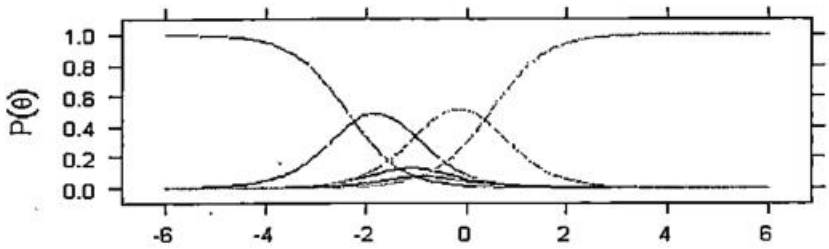
كذلك أظهرت النتائج أن المفردة رقم ٢٣ التي تنص على أن " بشرط توافر الوقت والجهد الكافي، اعتقد أنني استطيع حل معظم المشاكل التي تواجهني" ،في البعد الرابع الضبط الشخصي تؤدي بصورة متمايزة بين

جدول (٤)

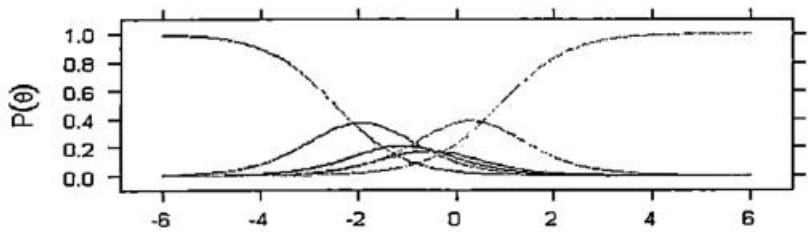
قيم معالم المفردة رقم ٢٣ لبعد الضبط الشخصي للمتزوجات وغير المتزوجات.

		العتبات الفارقة				
العينة	التمييز	الأولى	الثانية	الثالثة	الرابعة	الخامسة
غير المتزوجات	١.٣٤	١.٣٧-	٠.٢٦-	٠.٩٢	NA	NA
المتزوجات	١.٩	١.٥-	١.٠٨-	٠.١٥	NA	NA

المتزوجات وغير المتزوجات، كما يتضح من جدول (٤) أن معامل تمييز المفردة لعينة المتزوجات أعلى من عينة غير المتزوجات، وأن العتبات الفارقة للمتزوجات أقل من نظيرتها لدى عينة غير المتزوجات. ويوضح الشكلين (٤، ٥) المنحنيات المميزة للمفردة لدى المتزوجات وغير المتزوجات على الترتيب.



شكل (٤) المنحنيات المميزة للمفردة رقم ٢٣ للمتزوجات



شكل (٥) المنحنيات المميزة للمفردة رقم ٢٣ للمتزوجات

في حين أن لا توجد فروق في أداء المفردات في البعدين الثاني أسلوب التجنب، والثالث أسلوب الأقدام، أي أن المفردات في هاذين البعدين يعملان بنفس الطريقة لدى الطالبات المتزوجات وغير المتزوجات، كما أظهرت النتائج أن معامل الارتباط بين الدرجات الخام والدرجات المعادلة قوية جداً وتراوحت بين ٠.٩١ لبعد الثقة في حل المشكلة إلى ٠.٩٣ لبعد أسلوب الأقدام. مما يدل على أن المقياس صالح للاستخدام للمقارنة بين طالبات الجامعة المتزوجات وغير المتزوجات.

المناقشة

هدفت الدراسة الحالية إلى توجيه انظار الباحثين إلى قضية تكافؤ القياس تلك القضية التي لابد من مواجهتها في أي بحث يستخدم أداة للمقارنة بين مجموعتين أو أكثر، وقد هدفت الدراسة إلى عرض منهجين هم الأكثر من حيث الاستخدام في التصدي لتلك القضية: التحليل العاملي

التوكيدي متعدد المجموعات، والأداء المتمايز للمفردة في إطار نظرية الاستجابة للمفردة. وقد عرضت الدراسة للخطوات المتبعة في كل طريقة بصورة نظرية وبصورة تطبيقية، وقد استخدم الباحث قائمة حل المشكلة لعرض الجانب التطبيقي في الدراسة الحالية للتحقق من تكافؤ قياس القائمة بين المتزوجات وغير المتزوجات من طالبات الجامعة. وقد أظهرت النتائج أن الطريقتان لا تعطيان نفس النتائج، ففي حين أن إجراء التحليل العاملي التوكيدي متعدد المجموعات أظهر تكافؤ تام لجميع عبارات قائمة حل المشكلة بين المتزوجات وغير المتزوجات، فإن نظرية الاستجابة للمفردة أظهرت وجود مفردتين اظهرتا عدم تكافؤ في القياس. وقد اتفقت تلك النتيجة مع ما توصلت إليه دراسة Meade & Lautenschlager (2004) من أن التحليل العاملي التوكيدي متعدد المجموعات أقل كفاءة من نظرية الاستجابة للمفردة، وأن التحليل العاملي التوكيدي يهتم فقط بتشبعات المفردات على العوامل وأن تلك التشبعات تقابل معاملات تمييز المفردات في نظرية الاستجابة للمفردة، إلا أن التحليل العاملي التوكيدي لا يهتم بمعامل صعوبة المفردة، وأن المقدار الثابت أو القواطع لا تساوي معاملات الصعوبة. ونظراً إلى الفروق التي توجد بين الطريقتين فقد دمجت بعض الدراسات بين الطريقتين (e.g., Thombs, 2004)

المراجع

صلاح الدين محمود علام (٢٠٠٥). نماذج الاستجابة للمفردة الاختيارية أحادية البعد ومتعدد الأبعاد وتطبيقاتها في القياس النفسي والتربوي. القاهرة: دار الفكر العربي.

محمد حبشي حسين (٢٠١٨). الخصائص السيكومترية لمقياس اليقظة العقلية: مقارنة بين نظرية القياس التقليدية والنماذج الأحادية والمتعددة لنظرية الاستجابة للمفردة. المجلة المصرية للدراسات النفسية، ٢٨(٩٩)، ١٧-٧٦.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fiIt. In K. A. Bollen & J. S. Long (Eds.), (pp. 136–162). Newbury Park, CA: Sage.

Camilli, G., & Shepard, L. A. (1994). *Methods for identifying biased test items*. Thousand Oaks, CA: Sage.

Carmines, E. G., and J. P. McIver. (1981). Analyzing models with unobserved variables. In *Social measurement: Current issues*, ed. G. W. Bohrnstedt and E. F. Borgatta, 65–115. Beverly Hills, CA: Sage.

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of

measurement invariance. *Structural Equation Modeling, 14*،464–504.

Cheung, G. W., & Rensvold, R. B. (2000). Assessing extreme and acquiescence response sets in cross-cultural research using structural equation modeling. *Journal of Cross-Cultural Psychology*, 31, 187–212.

Cheung G W., Rensvold R. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Struct Equation Modeling,9*: 233-255

Choi, S. W., Gibbons, L., & Crane, P. K. (2011). Lordif: An R package for detecting differential item functioning using iterative hybrid ordinal logistic regression/item response theory and Monte Carlo simulations. *Journal of Statistical Software, 39*(8), 1–30.

Cohen, A. S., Kim, S. H., & Wollack, J. A. (1996). An investigation of the likelihood ratio test for detection of differential item functioning. *Applied Psychological Measurement, 20*, 15-26.

Cook, K. F., Bombardier, C. H., Bamer, A. M., Choi, S. W., Kroenke, K., & Fann, J. R. (2011). Do somatic and cognitive symptoms of traumatic brain injury confound depression screening?. *Archives of Physical Medicine and Rehabilitation, 92*, 818–823.

Crane, P. K., Gibbons, L. E., Jolley, L., & van Belle, G. (2006). Differential item functioning analysis with ordinal logistic regression techniques: DIFdetect and difwithpar. *Medical Care, 44*(Suppl 3), S115–S123.

Crane, P. K., Gibbons, L., Ocepek-Weiklson, K., Cook, K., Cella, D., Narasimhalu, K., et al. (2007). A comparison of three sets of criteria for determining the presence of differential item functioning using ordinal logistic regression. *Quality of Life Research, 16*(Suppl 1), 69–84.

Drasgow, F. (1984). Scrutinizing psychological test: Measurement equivalence and equivalent relations with external variables are the central issues. *Psychological Bulletin, 95*, 134–135.

Drasgow, F., & Kanfer, R. (1985). Equivalence of psychological measurement in heterogeneous populations. *Journal of Applied Psychology, 70*, 662–680.

Finch, H. (2010). Item parameter estimation for the MIRT model: bias and

precision of confirmatory factor analysis-based models. *Applied Psychological Measurement, 34*(1), 10-26.

Finch, H. & French, B. F. (2007). Detection of Crossing Differential Item FunctioningA Comparison of Four Methods. *Educational and Psychological Measurement 67*(4):565-582.

Haynes, S., Smith, G. T., & Hunsely, J. (2011*). Scientific foundations of clinical assessment.* New York: Routledge.

Heppner, P. P. (1988). The problem solving inventory: Manual. Palo Alto, CA:Consulting Psychologists.

Heppner, P. P., and Peterson, C. H. (1982). The development and implications of a personal problem-solving inventory. *Journal of Counseling Psychology, 29*, 66-75.Holland, P. W., & Thayer, D. T. (1988). Differential item functioning and the MantelHaenszel procedure. In H. Wainer & H. I. Braun (Eds.), *Test validity* (pp. 129-145). Hillsdale, NJ: Erlbaum.

Ji M, Sereika SM, Rohay JM, Erlen JA. (2018). Psychometric Properties of the Problem Solving Inventory in Caregivers of Individuals With Memory Loss. Journal of Gerontological Nursing. 44: 25-32

Joreskog, K. G. (1971). Statistical analysis of sets of congeneric tests. *Psychometrika, 36*, 109-133.

Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Hillside, NJ: Erlbaum.

Marsh, H. W., Muthe´n, B., Asparouhov, T., Lu¨dtke, O., Robitzsch, A., Morin, A. J. S., & Trautwein, U. (2009). Exploratory structural equation modeling, integrating CFA and EFA: Application to students' evaluations of university teaching. *Structural Equation Modeling, 16*, 439– 476.

McDonald, R. P. (1989). An index of goodness-of-fit based on noncentrality. *Journal of Classification, 6*, 97–103.

Meade, A. W., & Lautenschlager, G. J. (2004). A Monte-Carlo Study of Confirmatory Factor Analytic Tests of Measurement Equivalence/Invariance. Structural Equation Modeling, 11(1), 60-72.

Mencl, J., Tay, L., Schwoerer, C. E., & Drasgow, F. (2012). Evaluating quantitative and qualitative types of change: An analysis of the malleability of general and specific self-efficacy constructs and

measures. *Journal of Leadership & Organizational Studies, 19*(3), 378-391.

Meredith W.(1964). Notes on factorial invariance. *Psychometrika. 29*:199–185.

Millsap, R. E., & Yun-Tein, J. (2004). Assessing factorial invariance in ordered-categorical measures. *Multivariate Behavioral Research, 39* (3), 479–515.

Milfont, T. L., & Fischer, R. (2010). Testing measurement invariance across groups: Applications in cross-cultural research.International. *Journal of Psychological Research, 3*(1), 111-130.

Muthén, L. K.., & Muthén, B. O. (2012). *Mplus User's Guide* (Seventh Edition). Los Angeles, CA: Muthén & Muthén.

Narayanan, P., & Swaminathan, H. (1996). Identification of items that show nonuniform DIF. *Applied Psychological Measurement, 20*, 257-274.

Penfield, R.D., & Camilli, G., Differential item functioning and item bias. In S. Sinharay & C.R. Rao (Eds.), *Handbook of Statistics*, Volume 26: Psychometrics (pp. 125-167). New York: Elsevier, 2007.

Putnick, D. L. and Bornstein, M. H. (2016). Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. *Developmental Review, 41*:71 – 90.

Raju, N. S., Laffitte, L. J., & Byrne, B. M. (2002). Measurement equivalence: a comparison of methods based on confirmatory factor analysis and item response theory. *The Journal of Applied Psychology, 87*, 517–29.

Samejima, F. (1969). Calibration of latent ability using a response pattern of graded scores. *Psychometrika Monograph Supplement*, No. 17

Stark, S., Chernyshenko, O. S., & Drasgow, F. (2004). Examining the effects of differential item/test functioning (DIF/DTF) on selection decisions: When are statistically significant effects practically important? *Journal of Applied Psychology*, 89, 497–508.

Stark, S., Chernyshenko, O. S., & Drasgow, F. (2006). Detecting differential item functioning with CFA and IRT: Toward a unified strategy. *Journal of Applied Psychology, 91*, 1292–1306.

Suh, Y., & Cho, S. J. (2014). Chi-square difference tests for detecting

functioning in a multidimensional IRT model: A Monte Carlo study. *Applied Psychological Measurement, 38*, 359–375.

Swygert, K., McLeod, L. D. and Thissen, D. (2001). Factor analysis for items scored in more than two categories. In *Test scoring*, Edited by: Thissen, D. and Wainer, H. 217–250. Hillsdale, NJ: Erlbaum.

Tate, R. (2003). A comparison of selected empirical methods for assessing the structure of responses to test items. *Applied Psychological Measurement, 27*, 159-203.

Tay, L., Meade, A. W., & Cao, M. (2015). An overview and practical guide to IRT measurement equivalence analysis. *Organizational Research Methods, 18*, 3–46.

Terzioglu F.(2006). The perceived problem-solving ability of nurse managers. *Journal of Nursing Management*: 14, 340–347.

Terluin, B., Smits, N., Brouwers, M., & Vet, H. (2016). The Four-Dimensional Symptom Questionnaire (4DSQ) in the general population: scale structure, reliability، measurement invariance and normative data: a cross-sectional survey. *Health and Quality of Life Outcomes*, 114-130.

Thissen, D. (1991). MULTILOG: Multiple, categorical item analysis and test scoring using item response theory. Chicago, IL: Scientific Software. [Computer Program.]

Thissen, D. (2001). IRTLRDIF v.2.0b: Software for the computation of the statistics involved in item response theory likelihood-ratio tests for differential item functioning. Retrieved February 10, 2018, from http://www.unc.edu/~dthissen/dl.html.

Thissen, D., Steinberg, L., & Wainer, H. (1988). Use of item response theory in the study of group differences in trace lines. In H. Wainer & H. I. Braun (Eds.), *Test validity* (pp. 147-169). Hillsdale, NJ: Erlbaum.

Thombs, D. B. (2004). Measurement invariance of the childhood trauma Application of Structural equation modeling and item response theory (unpublished doctoral dissertation). Proquest Dissertations and Theses.

Thompson, B. (2004). *Exploratory and Confirmatory Factor Analysis: Understanding Concepts and Applications*. Washington: American Psychological Association.

Thompson, M. S. and Green, S. B. (2013). Evaluating Between-Group Differences in Latent Variable Means. Pp. 163-218 in *Structural Equation Modeling: A Second Course, 2nd ed.*, edited by Gregory R. Hancock and Ralph O. Mueller. Charlotte, NC: Information Age Publishing.

Tay, L., Vermunt, J. K., & Wang, C. (2013). Assessing the item response theory with covariate (IRT-C) procedure for ascertaining differential item functioning. *International Journal of Testing, 13*(3), 201–222.

Vandenberg, R. J. (2002). Toward a further understanding of and improvement in measurement invariance methods and procedures. *Organizational Research Methods, 5*, 139–158.

Vandenberg, R., j. & Lance, C. (2000). A Review and Synthesis of the Measurement Invariance Literature: Suggestions, Practices, and Recommendations for Organizational Research. *Organizational Research Methods 3*(1):4-69

Wirth, R. J., & Edwards, M. C. (2007). Item factor analysis: Current approaches and future directions. *Psychological Methods, 12*, 58-79.

Zumbo, B. D. (1999). *A handbook on the theory and methods of differential item functioning (DIF): Logistic regression modeling as a unitary framework for binary and Likert type (ordinal) item scores*. Ottawa, ON: Directorate of Human Resources Research and Evaluation Department of National Defense

The measurement Invariance/Equivalence in Psychological and Educational Studies a comparison between Multiple Group Multiple-Group Confirmatory Factor Analysis and Item Response Theory

Prof. Mohamed Habashy Hussein

Professor of Educational Psychology

College of education- Alexandria University

A great number of psychological and educational studies aim to find the differences between two groups of more than two groups in one or more than educational and psychological variable or finding the differences among the same group across several time points on one or several scale. At this context, the importance of measurement equivalence/invariance emerges. measurement equivalence/invariance aims to verify that the scale works with the same manner across the different groups and the potential differences between groups on these latent variables as measured by the scales are due to real differences exist between groups and not a result of a difference in factorial structures of the scale or due to the differences in relative weights for each item from group to another. Several foreign studies (e.g., Jabrayilov et al. 2017) aimed to examines the measurement equivalence/invariance as s prerequisite for comparing males and females (e.g., Elosua & Hermosilla, 2013), or among multiple cultures (e.g., Ren et al. 2015), or comparing the response of the same group across different points of time (e.g., Moreira et al.2018). Some of these studies used multiple-group confirmatory factor analysis (e. g. An et al., 2017), whereas other studies utilized models of item response theory especially Differential Item Functioning (DIF). On the other hand, a few studies compared the two strategies (e.g., Nelson, 2006). Limit to my knowledge, there is no any Arabic study utilized one of the two strategy to investigate the measurement equivalence/invariance before using any measure to compare the different groups. Therefore the main goal of the current study to introduce these two strategies to the Arabian studyers, and gave a demonstration to the two strategies using problem solving inventory created by Heppner and Petersen (1982) after administering the inventory on a sample of married (132) and unmarried (299) female college students (431) from college of education in Alexandria university. The general conclusion from the results indicated that both strategies gave a similar result.