أثر طريقة ترتيب المموهات في فقرات الاختيار من متعدد في اختبار محكي المرجع على درجة القطع واللاتغير في القياس

جمال شحادة الشويكي[ii]
أ.د. يوسف سوالمة[i]

تاريخ الاستلام
2024/4/23
تاريخ القبول
2024/5/20

الملخص

هدفت الدراسة إلى الكشف عن أثر طريقة ترتيب المموهات في فقرات الاختيار من متعدد في اختبار محكي المرجع على درجة القطع للاختبار باستخدام طريقة ندلسكي، وعلى اللاتغير في القياس. وتم استخدام المنهج التجريبي لتحقيق هدفي الدراسة. وتكونت أداة الدراسة من اختبار محكي المرجع مكون من (20) فقرة في مبحث العلوم للصف الثامن الاساسي حيث تم عمل نموذجين متكافئين في المحتوى وموقع الإجابة الصحيحة ومختلفين في ترتيب المموهات (ترتيب عشوائي في النموذج الأول، وترتيب تنازلي من الأقوى إلى الأضعف في النموذج الثاني). وتكونت عينة الطلبة من (290) طالبا وطالبة تم توزيعهم عشوائيا على نموذجي الاختبار. وتكونت عينة المحكمين من (40) محكما ومحكمة تم توزيعهم عشوائيا بالتساوي على نموذجي الاختبار لتحديد درجة القطع بمعرفة صعوبة الفقرات وبدونها. وبينت نتائج الدراسة أن درجة القطع للاختبار محكي المرجع عند ترتيب المموهات من الأقوى إلى الأضعف أعلى منها عند ترتيب المموهات عشوائيا. وأشارت النتائج إلى تحقق خاصية اللاتغير في القياس عبرطريقتي ترتيب المموهات.

الكلمات المفتاحية: ترتيب المموهات، اختبار اختيار من متعدد، اختبار محكي المرجع، درجة القطع، اللاتغير في القياس.

[i] جامعة اليرموك
[ii] جامعة اليرموك

The Effect of Distractors Ordering Method in Multiple-Choice Criterion- Referenced Test on Cut-Off Score and Measurement Invariance

Abstract

The study aimed to reveal the effect of distractors ordering method in multiple-choice criterion-referenced test on the cut-off score for the test using the Nedelsky model, and on the measurement invariance. The experimental design was used to achieve the two objectives of the study. The study instrument consisted of a criterion-referenced test containing 20 items in the Science subject for the eighth-grade, where two test forms are equivalent in content and the location of the correct answer, but different in the distractors ordering (random order in the first form, and descending order from strongest to weakest in the second). The students' sample consisted of (290) male and female students who were randomly distributed to the two test forms. The sample of raters consisted of (40) male and female raters who were randomly distributed equally among the two test forms to determine the cut-off score with and without knowing the values of items' difficulty. The results of the study showed that the cut-off score for the criterion-referenced test when the distractors were arranged from strongest to weakest was higher than when it was ordered randomly. The results indicated that measurement invariance was achieved across the two methods of arranging the distractors.

Key words: Distractors ordering, Multiple-choice tests, Criterion-referenced test, Cut-off score, Measurement invariance.

المقدمة:

تلعبُ الاختبارات بشكل عام والاختباراتِ التحصيليةِ بشكلٍ خاصٍ، دورًا مهمًا في عمليتي التعليم والتعلم. ويقيسُ الاختبار التحصيلي الجيد التغيرات التي يحدثها التعلم في طريقةِ تفكيرِ الطالبِ، وميوله النفسية، وأسلوبهِ في حل المشكلات، وقدْرته على النقدْ البناء، واكتسابهِ مهارات وخبرات مفيدة (العبيدي والجبوري، 1981).

وتُعد اختبارات الاختيارِ من متعددٍ من أكثرِ الاختبارات المستخدمة في المؤسسات التعليمية، وهي مدار اهتمَام خبراء القياسِ. وتتكون فقرة الاختيار من متعدد من جذر وخيارات (بدائل) للإجابة، أحدها هو الإجابة الصحيحة، وتسمى البدائل الخاطئة بالمموهات (أو المشتتات). وتستغرق الفقرة المحكمة البناء وقتاً لإعدادها، وليس من السهل توفيرِ البدائلِ الجذابةِ (Gronlund, 1993).

ويرى سواميناثان وهامبلتون (Hambleton & Swaminathan, 1985) أنَّ مستوى أداء الطالب في الاختبار يعتمَدُ على خصائصه وخصائصِ فقراتِ الاختبارِ مثل الصعوبة والتمَييز وفعالية المموهات وغيرها. ويرى علام (2006) أن صياغة فقرات الاختيار من متعدد ترتبط بتجانس البدائل، وسلامة العرض اللغوي، والتوزيع العشوائي للبدائل الصحيحة، وهي من العوامل التي تؤثر على صعوبة فقرات الاختيار من متعدد.

وتولي الأدبياتُ التربوية أهميةً كبيرة لموقع الإجابة الصحيحة في اختبارِ الاختيارِ من متعددٍ، ويعدُ الاهتمَامُ بترتيب البدائل من الموضوعات التي تناولتها الكثير من الدراسات (أبو فودة، 2014؛ سليمان والصالح، 2016). أما الاهتمام بترتيب المموهات فهو أحد الأفكارِ الجديدة في عمليةِ تطوير الاختباراتِ وبناء الفقرات، ويكون التركيز في عملية الترتيب على فاعلية تلك المموهات. وقد أشار ليونز وآخرون (Lions et al.,2021) إلى أنَّ اقوى المموهات يكون موجودا وبشكل متكرر بين الخيارات الوسطى وأضعف المموهات وبشكل متكرر في الخيار الأخير، وأنَّ التوزيع غير المتوازن للمموهات قدْ يزيد من فرصة التخمين العشوائي. ويتوقع أن يؤثر موقع المموه القويُّ داخلَ فقرة الاختبار، سواء كان قريبا من الإجابة الصحيحة أو بعيدا عنها، بشكلٍ مباشرٍ على استجاباتِ الأفراد المفحوصين، وهو ما ينعكس إحصائيا على الخصائص السيكومترية لفقرات الاختبار (Johnstone &Ambusaidi,2001).

وبّين الكعابنة وعودة (2019) أنَّه لا يوجد أثر لطريقة اختيار المموهات على كل من مواقع الافراد على تدريج متصل السمة، وتقديرات معلمة التخمين، والاخطاء المعيارية لتقديرات معلمة الصعوبة. بينما هناك أثر لطريقة اختيار المموهات على كل من الأخطاء المعيارية لتقدْيرات معلمة القدْرة، والعلامة الخام، وتقدْيرات معلمة الصعوبة، وتقدْيرات معلمة التمَييز، والاخطاء المعيارية في تقدير كل من التمييز والتخمين. وفي إطار النظرية الكلاسيكية، هناك أثر على قيم معاملات الصعوبة وقيم معاملات التمَييز في ضوء اختلاف طريقة اختيار المموهات في فقرات اختبار الاختيار من متعدد.

وفي معرض تحليل تمييز الفقرات وفاعلية المموهات، أكد اساموه واوكانسي (Asamoah & Ocansey, 2019) أن التمَييز الموجب للفقرة ينتج عندما تكون نسبةالإجابات الصحيحة للفقرة في الفئة العليا أعلى منها في الفئة الدنيا. أما المموه الفاعل فهوالذي يتمَ اختياره من قبل أفراد الفئة الدنيا باحتمال أعلى منه في الفئة العليا. وهذا يعني أن فاعلية الفقرة ممثلة بمعامل تمييزها من فاعلية المموهات فيها، وأن فاعلية المموه الذي تمَييزه صفر مشكوك فيها، ولا بد من استبعاد المموه الذي يتم اختياره من قبل أفراد الفئة العليا باحتمال أكبر من الفئة الدنيا.

ويُعدُ الاختبارُ التحصيلي محكي المرجع أحد الاتجاهات الحديثة في عملية القياس النفسي والتربوي، ويتم تفسير أداء الفرد عليه من خلال مقارنته بمعيار محدد (درجة القطع)، وبالتالي تصنيفه إلى متقن وغير متقن لما يقيسه الاختبار. وظهرت الاختبارات محكية المرجع نتيجة قصور الاختبارات معيارية المرجع التي تركز على الفروق الفردية بين الطلاب دونَ مراعاةِ مدى إتقانِ الطالب للمعلوماتِ والمهاراتِ التي يتمَ اختبارها وقياسها (سليمان والصالح، 2016).

وهناكَ العديدُ من الطرقِ لتحديدِ درجةِ القطعِ، وتصنفُ تلكَ الطرقِ في ثلاث فئات: تعتمَدُ الفئة الأولى على التحكيم فقط، وتعتمَدُ الفئة الثانيةَ على التحكيم وتسترشد بالبيانات التجريبية، وتعتمَدُ الفئة الثالثة على البيانات التجريبية وتسترشد بالتحكيم. ويصنفها الاتجاه الحديث في القياس النفسي والتربوي إلى طرقِ تتمركز حول الاختبار وتعتمد بصفة رئيسية على أحكام الخبراء المتعلقة بالاختبار وفقراته، ومن أهمها طريقة ندلسكي وطريقة أنجوف وطريقة إبل، وطرق متمركزة حول المفحوص وتعتمَدُ بصفة رئيسية على البيانات التجريبية التي يتمَ الحصول عليها من تطبيق الاختبار الفعلي وفقراته، ومن أشهرها طريقة المجموعة المحكية، وطريقة المجموعات المتناقضة، وطريقة المجموعة الحدية (بني عطا، 2005).

وفي الدراسة الحالية سيتم استخدام طريقة ندلسكي، ويطلبُ في هذه الطريقة من كل محكم تحديد الخيارات (البدائل) في الفقرة التي يمكن استبعادها من قبل الطالب الذي يمتلك الحد الأدنى المقبول من الكفاية في مجال الاختبار لكونها غير صحيحة. وبذلك فإن احتمَال الإجابة الصحيحة للفقرة هو مقلوب عدد الخيارات (البدائل) المتبقية. على سبيل المثال، بالنسبة لفقرة تتكون من أربعة بدائل، ووجد المحكم أنَّ الطالبَ الحاصلَ على الحدِ الأدنى المقبول من الكفاية يمكنه استبعاد أحد البدائل، فإنَّ احتمَال أنَّ يجيبَ الطالب بشكل صحيح على الفقرة هو (0.33)، واحتمَال أنْ يجيبَ الطالب على الفقرة بشكل صحيح اذْا ما تمَ استبعاد بديلين هو (0.5). في هذه الحالة، تكون درجة القطع لكل محكم هي مجموع تقدْيراته لجميع فقرات الاختبار، ودرجة القطع للاختبار بأكمله هي الوسط الحسابي لمجموع تقدْيرات جميع المحكمين (Rrilly e al.,1984).

وقد بين الشريم والسوالمة (2006) أن طريقة ندلسكي تفيد في تصنيف الطلبة بصورة دقيقة وأن نتائجها لا تختلف بصورة جوهرية عن طريقة أنجوف.

ويعد صدق نتائج الاختبار من أهم الخصائص السيكومترية، ويتعلق الصدق بقضية تفسير تلك النتائج بدلالة السمة المستهدفة بعملية القياس. وعندما يتحقق الصدق للاختبار، يكون التباين في العلامات بدلالة الخاصية موضع الاهتمام بالقياس وليس بدلالة متغيرات أخرى لا علاقة لها بتلك الخاصية. لذلك زاد الاهتمام في العقود القليلة الماضية بالأداء التفاضلي للفقرات واللاتغير في القياس لارتباطهما بالصدق. وتعد خاصـية اللاتغيـر في معالم الأفراد والفقرات من أهم السمات المميزة لنماذْج نظرية الاستجابة للفقرة التي تتجلى أهميتها في حل العديد من المشكلات التي واجهت ْنظريةَ القياس الكلاسيكية مثل تلك المتعلقة بمعادلة الاختبارات والاختبارات المكيفة (الشهري، 2017). واستقلال القياسِ هو نقطة الفصل بين النظرية التقليدية ونظرية الاستجابة للفقرة، وهذا ما يسميه لورد (Lord,1980) باللاتغير، فاذْا تحققتْ مثل هذه الخاصية فيمكن تنفيذ بعض التطبيقات، مثل معادلة الاختبارات، وبناء بنوك الأسئلة، والكشف عن تحيز الفقرات والقياس التكيفي (بني عطا والشريفين، 2012). وقد تم فحص اللاتغير في القياس في كثير من الدراسات باستخدام التحليل العاملي التوكيدي والتحليل العاملي متعدد المجموعات لتوفير أدلة على صدق البناء العاملي للاختبار (السعيد والشريفين، 2021) (السمكي، 2020؛ Tsaousis,2015).

وتكشف مراجعة الأدب التربوي الكثير من الدراسات حول كل من طرقِتحديد درجات القطع في الاختبارات محكية المرجع والمقارنة فيما بينها واللاتغير في القياس لكثير من الاختبارات والمقاييس دون ربط ذلك بترتيب المموهات في الفقرات. وتلقي الدراسة الحالية الضوء على أثر ترتيب المموهات في اختبار محكي من نوع الاختيار من متعددفي العلوم للصف الثامن الاساسي على كل من درجة القطع باستخدام طريقة ندلسكي واللاتغير في القياس.

مشكلة الدراسة وسؤاليها:

تعد الاختبارات الموضوعية وخاصة الاختيار من متعدد من أكثر الاختبارات استخداماً في المؤسسات التربوية لقياس التحصيل نظراً لشمولها المحتوى بشكلٍ جيدٍ، وسهولة التطبيق، وموضوعية التصحيح وتمَتعها بدرجة عالية من الثبات والصدق، حيث إنَّ تطبيقها يعتمَدُ على عدة عوامل كطبيعة المحتوى المراد قياسه وعدد الطلبة والهدف من القياس ومهارة كتابة الاسئلة.

ومع ازدياد التحول في كثير من التقييمات من الاختبارات الورقية إلى الاختبارات الالكترونية، وازدياد الحاجة إلى نماذج متعددة من الاختبار، تزداد الحاجة إلى خلط البدائل في فقرات الاختيار من متعدد وتغيير ترتيبها لاعتبارات تتعلق بتطبيق الاختبار وأمنه. ويمكن تصنيف الصور المختلفة في خلط البدائل في فئتين: في الفئة الأولى يتم المحافظة على موقع الإجابة الصحيحة، وفي الفئة الثانية لا يتم المحافظة على موقع الإجابة الصحيحة. وتقتصر صور الخلط في الفئة الأولى على طرق ترتيب المموهات. حيث يمكن ترتيب المموهات عشوائيا أو ترتيبها من المموه الأضعف إلى الأقوى أو من الأقوى إلى الأضعف أو وضع المموه القوي بجوار الجواب الصحيح أو وضع المموه القوي بعيدا عن الجواب الصحيح. ويتوقع أن يكون لطريقة ترتيب المموهات تأثيرا على خصائص الاختبار المحكي المرجع وعلى درجة القطع له. وقدْ ركزت معظم الدراسات على موقع البديل الصحيح والقليل منها ركز على ترتيب المموهات. وفي حدود علم الباحثان لا توجد دراسات حول أثر تغيير ترتيب المموهات على كل من اللاتغير في القياس لاختبار محكي المرجع ودرجة القطع له.

لذلك جاءتْ هذهِ الدراسةِ لتلقي الضوء على أثر طريقة ترتيب المموهات في فقرات الاختيار من متعدد في اختبار محكي المرجع في العلوم للصف الثامن الأساسي على درجة القطع له واللاتغير في القياس.

وبالتحديد تسعى الدراسة إلى إجابة السؤالين التاليين:

1-هل تختلفُ درجة القطع لاختبار محكي المرجع في العلوم للصف الثامن الاساسي باستخدام طريقة ندلسكي باختلاف طريقة ترتيب المموهات (ترتيب عشوائي مقابل ترتيب من الأقوى إلى الأضعف)؟

2-هل يتحققُ اللاتغير في القياس لاختبار محكي المرجع في العلوم للصف الثامن الاساسي عبر طريقتي ترتيب المموهات؟

أهداف الدراسة

تهدفُ الدراسة الحالية إلى الكشف عن أثر ترتيب المموهات على تحديد درجة القطع للاختبار المحكي المرجع، والتعرف على أثر ترتيب المموهات على اللاتغير في القياس.

أهمية الدراسة

تكمنُ الأهمية النظرية لهذه الدراسة في انه سيستفاد من نتائجها من قبل الباحثين المتخصصين والاخصائيين التربويين في التعرف على أثر ترتيب المموهات على تحديد درجة القطع المناسبة للتمَييز بين الطلبة المتمَكنين وغير المتمَكنين في اختبار محكي المرجع وفي الكشف عن تحقق خاصية اللاتغير في القياس باختلاف ترتيب المموهات. ويتوقع أن توفر هذه الدراسة معلومات تساعد الجهات المختصة في الميدانِ التربوي في كتابة الفقرات وتطوير الاختبارات التحصيلية.
وتتحقق الأهمية التطبيقية للدراسة من خلال:
1-توفير أداة قياس تمَت مراجعتها وتحكيمها من قبل متخصصين ومحكمين يمكن الاعتمَاد عليها في قياس تحصيل الطالبات والطلاب.
2-تحديد درجة القطع في الاختبارات التحصيلية محكية المرجع اذْ يعول عليها في الحكم على ما يستطيع وما لا يستطيع ان يتقنه الطالب في محتوى اختبار محكي المرجع في العلوم للصف الثامن الأساسي، ومن ثم اتخاذُ القرارات التربوية المناسبة حول مدى الحاجة لتطوير أساليب التدريس والتقويم.

التعريفات المفاهيمية والاجرائية

المموه القوي: ويقصد به البديل غير الصحيح الذي اختاره العدد الأكبر من المفحوصين ومعظمهم من الفئة الدنيا.
ترتيب المموهات: المموه هو البديل الخاطئ في فقرة الاختيار من متعدد، وترتيب المموهات يعني الطريقة التي يتم من خلالها تحديد مواقع تلك المموهات. وفي الدراسة الحالية حيث يوجد أربعة بدائل للفقرة، تم في كل فقرة تحديد موقع البديل الصحيح وتم توزيع المموهات على المواقع الثلاث المتبقية بطريقة عشوائية في النموذج الأول للاختبار ومن المموه الأقوى إلى المموه الأضعف في النموذج الثاني للاختبار.
درجة القطع للاختبار محكي المرجع :يعرفها هامبيلتون (Hambleton, 1982) بأنَّها درجة على متصل السمة المقاسة، والتي تستخدم لتصنيف المفحوصين إلى فئتين تعكس مستويات الكفاءة بالنسبة لمجموعة من الأهداف أو المهارات التي يقيسها الاختبار، بحيث يمكن من خلالها الحكم على مستوى أداء المفحوص بأنه متمَكن أو غير متمَكن.
اللاتغير في القياس: أنْ يقيسَ الاختبار البناء نفسه وبنفس الطريقة لجميع المجموعات الفرعية المختلفة ولنفس نطاق القياس من حيث البنية الأساسية التي يجري تقييمها (Reise et al., 1993).

محددات الدراسة

تقتصر هذه الدراسة على اختبار محكي المرجع من إعداد الباحثين وفق الإجراءات المتعارف عليها في بناء الاختبارات المحكية في مادة العلوم للصف الثامن الأساسي، وبنموذجين توزع في الأول المموهات في الفقرة عشوائيا وترتب في الثاني من الأكثر فاعلية إلى الأقل فاعلية. كما تقتصر على عينة من طلاب وطالبات الصف الثامن الاساسي في مدارس مديرية التربية والتعليم للواء الاغوار الشمالية.

الدراسات السابقة

هدفت دراسة الشريم والسوالمة (2006) إلى تحديد درجة القطع لاختبار محكي المرجع في الرياضيات باستخدام نموذجي "أنجوف" و"ندلسكي" بمعرفة صعوبة الفقرات وعدم معرفتها. وقدْ تكوّن الاختبارُ من ثلاثين فقرة من نوع الاختيار من متعدد لكل منها أربعة بدائل. وتكوّنتْ عينة الدراسة من ثمانين محكماً ومحكمة. وقدْ تمَ تقسيمُ المحكمين بطريقة المزاوجة العشوائية إلى أربع مجموعات متساوية حددت كل منها درجة القطع للاختبار مرتين وفق الاسلوب المعين لها. وأشارت نتائج الدراسة إلى أن درجة القطع للاختبار تتراوح بين (.62) و(.68) باستخدام نموذج أنجوف وبين (.49) و(.57) باستخدام نموذج ندلسكي، وإنَّه لا يوجد فرق دال إحصائياً عند مستوى الدلالة الإحصائية (α=.05) بين معاملي ثبات النموذجين. كما بينت النتائج إنَّه لا يوجد فرق دال إحصائياً بين معاملي الارتباط لصعوبة الفقرات وتقدْيرات المحكمين لها في النموذجين عندما لا يزود المحكمين بصعوبة الفقرات. بينما يوجد فرق دال إحصائياً بين معاملي الارتباط عند معرفة المحكمين لصعوبة الفقرات. وأظهرت النتائج أن تصنيفات الطلبة جميعها كانت دقيقة.

وقامَ عباس (2020) بدراسة طرق تحديد درجات القطع لاختبارات الذكاء والقدْرات العقلية من خلال تطبيق ثلاث طرق لتحديد تلك الدرجة، الاولى تحكيمية وهي طريقة جيجر، والثانية تسترشد بالمحكمين وتعتمَد على البيانات وهي طريقة انجوف التجريبية المعدلة، والثالثة تعتمَدُ على البيانات وتسترشد بالمحكمين وهي طريقة المجموعة الحدية. وتعد عملية تحديد درجة القطع Cutting score في اختبارات الذكاء والقدْرات العقلية محكية المرجع محور ارتكاز بناء تلك الاختبارات لأنَّها تزود الباحث بتقدْيرات كمية لتحديد مدى تمَكن الفرد من المهارات والمعارف والحكم على ادائه ومن ثمَ اتخاذْ القرارات المناسبة.

وهدفتْ الدراسة التي أجراها أبو فودة (2014) إلى بيان أثر إعادة ترتيب بدائل الاستجابة في صعوبة فقرة الاختيار من متعدد، ولتحقيق ذلك تمَ بناء اختبار في مبحث الرياضيات لطلبة الصف التاسع الأساسي، يتكون من (20) فقرة، ثم جرى إعداد صورتين للاختبار تختلفان في ترتيب بدائل الاستجابة. ثمّ جرى تطبيق صورتي الاختبار على عينة من (600) طالبٍ وطالبة في تربية محافظة جرش، بواقعِ (300) طالبٍ وطالبةٍ في كل صورة، وذلك خلال عام 2013/2012م. وحسبت قيم معاملات صعوبة الفقرات في صورتي الاختبار، وجرى إيجاد الفرق بين قيم معاملي الصعوبة في الفقرات المتناظرة في الصورتين، وكذلك جرى إيجاد الفرق المطلق بينَ مواقع الاستجابة الصحيحة فيهما. وأظهرت النتائج أنَّ الفرق لم يكن دالاً إحصائيًا بين معاملات صعوبة الفقرات، وإنَّه لم يكن هناك نمط ثابت فيما يتعلق بإعادة ترتيب بدائل الاستجابة.

وقدم اسامويه واوكانسي (Asamoah, & Ocansey, 2019) تحليلاً مفصلاً لتمَييزِ الفقراتِ وفاعلية المشتتات (المموهات) في اختبارٍ تحصيلي من 30 فقرة في الرياضيات الأساسية لطلبة المرحلة الثانوية. وتبين أن 27 فقرةٍ من أصل 30 فقرةٍ كان تمَييزها إيجابيا، وفقرتين كان تمَييزها سلبيًا، وفقرة لديها تمَييز صفري. وتوصلتْ الدراسة إلى انه عندما يكون التمَييز للفقرات موجبا فإن نسبة اعلى من الطلابِ في الفئةِ العليا حصلَ على المعرفةِ أكثر من تلك التي حصل عليها الطلاب الضعاف، وأنَّ الفقرة ذات التمَييز صفر تكون إما سهلة أو صعبة للجميع. في تحليل المموهات كان المموه الفاعل هو الذي يتمَ اختياره بعدد أكبر من طلاب الفئة الدنيا والفعالية للموه الذي تمَييزه صفر مشكوك فيها بينما لم يكن مرغوباً في المموه الذي تمَ اختياره بعدد أكبر من

طلاب الفئة العليا. وفي الدراسة التي أجراها ليونز وآخرون (Lions et al.,2021) بعنوان موضع المشتتات في اختبار الاختيار من متعدد: الاقوى يسبق الأضعف، تبينَ أنَّ اقوى المشتتات تكون موجودة وبشكل متكرر بين الخيارات الوسطى وأضعف المشتتات وبشكل متكرر في الخيار الأخير. كما تبين أنَّ التوزيع غير المتوازن للمشتتات قدْ يزيد من فرصة التخمين العشوائي.

وقام الشهري (2017) بعمل دراسة بهدف فحص اللاتغير في تقديرات معالِم الفقرات والأفراد في ضوء النموذج اللوجستي ثلاثي المعلم عند تطبيق نموذجي اختبار اختيار من متعددٍ يتفقانِ في قياسِ الهدف التدريسي ويختلفان في قواعد صياغة الاختيار من متعدد، وقدْ بلغ حجم عينة الدراسة الصالحة للتحليل الإحصائي (1391) طالباً، ولتحقيقِ أهداف الدراسة صمم الباحث أداة الدراسة (نموذجي اختبار) من نوع الاختيار من متعدد من أربعة بدائل ضمن منهج علوم الصف السادس الابتدائي للفصل الدراسي الأول من العام 1436-1437هـ وتمَ تحميل البيانات باستخدامِ البرنامج الإحصائي XCalibre4.2.2. وتوصلت الدراسة إلى أنه لم يتحقق افتراض اللاتغير في قدْرات الأفراد عند تقدمهم لفقرات مختلفة في الصعوبة، حيث كانت الفروق بين متوسطات معالم القدرة دالة، ولم يتحقق افتراض اللاتغير في تقدْير (معلمة الصعوبة، التمَييز) باختلاف معالم القدرة، وتحققِ افتراضِ اللاتغير في تقدير معلمة (التخمين) باختلاف معالم القدرة.

وهدفتْ الدراسة التي قامتْ بها أبو مسرة والسوالمة (2021) إلى استخدامِ طريقة أشجار راش في التقسيم المتكرر للكشف عن الأداء التفاضلي للفقرة في اختبار الرياضيات الوطني لضبط نوعية التعليم للصف الرابع الأساسي في الأردن، تبعا للمتغيرات التي توفرها عملية التقييم، وهي: الجنس وموقع المدرسة والسلطة المشرفة على التعليم وطريقة تقدْيم الاختبار. إضافة إلى معرفة مدى تأثر الطريقة باختلاف حجم العينة. تكون مجتمَع الدراسة من استجابات (31136) طالبا وطالبة في الصف الرابع على (25) فقرة من نوع الاختيار من متعدد، تطابقت مع نموذج راش. وبتطبيق الطريقة على استجابات عينة عشوائية قوامها (1500) من المفحوصين، بينتْ النتائج أنَّ هناك أداءً تفاضلياً وفقًا لمتغيرِ الجنس فقط؛ اذْ تمَ الكشف عن (8) فقرات ذات أداء تفاضلي، أي ما نسبته (32%) من الفقرات. وكان الأداء التفاضلي لصالح الذكور في (5) فقرات، ولصالح الإناث في (3) فقرات. وأظهرتْ النتائج حساسية طريقة أشجارِ راش لحجم العينة؛ اذْ زادت فرص وجود أداء تفاضلي للفقرات بازدياد حجم العينة.

التعقيب على الدراسات السابقة

تناولتْ بعض الدراسات السابقة تأثيرَ طرقِ تحديد درجات القطع في اختبارات محكية المرجع والمقارنة بينها ودراسة الخصائص السيكومترية لها دون ربط ذلك بترتيب المموهات (البدائل) للفقرات. بينما اهتمَتْ دراسات أخرى بأثر طرق اختيار المموهات في فقرات اختبار الاختيار من متعدد في خصائصها السيكومترية واخرى تناولتْ تأثير عدد بدائل الفقرات على الخصائص السيكومترية للاختبارِ ومعالم الفقرات وقدرات الأفراد، والبعض الآخر اهتمَ بتأثير موقع البديل الصحيح على دقة تقديرات معالم الفقرات والقدْرة وأخرى قدمتْ تحليلاً مفصلاً لتمَييز الفقرات وفاعلية المموهات دونَ دراسة تأثير ذلك على تحديد درجة القطع. واهتمَتْ دراسات بخاصية اللاتغير في القياس من خلال التحقق من البناء العاملي للاختبار والبعض الآخر اهتمَ بفحص اللاتغير في تقدْيرات معالم الفقرات والأفراد بحسب الجنس والموقع دون معرفة مدى تأثر خاصية اللاتغير في القياس باختلافِ ترتيبِ المموهاتِ.

بينما تفردتْ هذهِ الدراسةِ في القاء الضوء على أهمية تأثير ترتيب مموهات الاستجابة (البدائل) في اختبار محكي من نوع اختيار من متعدد في العلوم للصف الثامن الاساسي على تحديد درجة القطع وعلى تأثير هذا الترتيب للبدائل على خاصية اللاتغير في القياس. حيث سيتمُ تسليطُ الضوء أيضا على أثر ترتيب المموهات على الخصائص السيكومترية للفقرات (الصعوبة، التمَييز). حيث لم يهتمَ مطورو الاختبارات بترتيبِ المشتتات (المموهات) بحسب قوتها أثناء تجميع الاختبارات بقدْر اهتمَامهم بخلق المشتتات. ولم يعثر الباحث على دراسات تهتمَ بترتيب المموهات وأثرها على درجة القطع وكذلك لم تظهرْ دراسات تهتمُ بأثر ترتيب المموهات من الأقوى للأضعف على خاصية اللاتغير في القياس.

المنهجية والاجراءات

منهج الدراسة

استخدم الباحثان المنهجَ التجريبي في هذه الدراسةِ، وفقَ التصميم البعدي لمجموعتين عشوائيتين.

مجتمَع الدراسة

تكون مجتمَع الدراسة من جميع طلابِوطالباتِ الصفِ الثامنِ الاساسي في جميع مدارس تربية لواء الاغوار الشمالية للعام الدراسي 2022–2023م. ويبلغ عددهم وفقاً لإحصائيات مديرية التربية والتعليم للواءِ الاغوارِ الشماليةِ (1742) طالبا وطالبة (887 طالبة في 22 مدرسة، و855 طالباً في 19 مدرسة).

عينة الدراسة

تمَ اختيارُ عينة الدراسة بالطريقة العشوائية العنقودية من مجتمَع الدراسة، حيث كانت وحدة الاختيار هي المدرسة، وقد اختيرت (5) مدارس عشوائيا للذكور، و(5) مدارس عشوائيا للإناث. وقدْ اختيرتْ من كل مدرسة شعبة واحدة عشوائيا من بين شعب الصف الثامن الأساسي. وقد تكونت العينة من 290 طالب وطالبة (148 إناث، 142 ذكور). كما تمَ اختيار عينة مقصودة تتكون من (40) معلما ومعلمة ومشرفا ومشرفة تربويين (24 ذكور، 16 إناث) لمبحث العلوم كمحكمين بهدفِ تحديد درجة القطع على الاختبار بحسب طريقة ندلسكي، وتمَ تقسيمهم عشوائيا إلى مجموعتين متساويتين في كل منها (20) محكما. حيث قامت المجموعة الأولى بتقدير درجة القطع للنموذج الأول من الاختبار في جولتين واحدة قبل معرفة درجة الصعوبة والأخرى بعد معرفة درجة الصعوبة، وقامت المجموعة الثانية بتقدير درجة القطع للنموذج الثاني في جولتين كذلك.

أداة الدراسة

تكونت أداة الدراسة من اختبار مبحث العلوم لطلبة الصف الثامن الأساسي في الفصل الدراسي الثاني من العام الدراسي 2023/2022، ويتكون الاختبار في صورته الأولية من 40 فقرة اختيار من متعدد بأربعة بدائل. وقد تم بناء الاختبار وكتابة فقراته وفق خطوات بناء الاختبارات محكية المرجع (Haladynaet al. ,2002). وتم عرضه على (10) محكمين من المشرفين

التربويين والمعلمين الخبراء في تدريس العلوم للصفِ الثامن الاساسي للتحقق من تمَثيل الفقرات للمجال السلوكي وانسجامها مع نتاجات التعلم التي تقيسها. وطبق الاختبار على عينة استطلاعية (48 طالب وطالبة) من مجتمع الدراسة، ومن خارج العينة. واستخدمت معادلة كورد ريتشاردسون 20 لتقدير الثبات وبلغت قيمة الثبات (.75)، كما تم حساب معاملات الصعوبة والتمييز للفقرات. وتراوحت معاملات الصعوبة بين 0.19 و0.88، وتراوحت قيم معاملات التمَييز بين 0.15- و0.56. وقد تم حذف 9 فقرات لكونها لم تف بالمعايير المقبولة للصعوبة والتمييز. ويبين الجدول (1) معاملات الصعوبة والتمَييز للفقرات المتبقية وفعالية بدائلها.

جدول 1: معامل الصعوبة ومعامل التمَييز وفعالية المموهات لفقرات الاختبار

الرقم	معامل الصعوبة	معامل التمَييز	فعالية المموه (البديل)			
			أ	ب	ج	د
1	0.88	0.13	-0.08	+0.17*	-0.08	0.00
2	0.75	0.24	-0.04	-0.13	-0.08	+0.25*
3	0.77	0.22	-0.17	+0.04	+0.21*	-0.08
4	0.79	0.26	-0.08	+0.17*	0.00	-0.08
5	0.52	0.36	+0.38*	-0.25	-0.04	-0.08
6	0.33	0.47	+0.42*	-0.17	-0.29	+0.04
7	0.71	0.16	+0.04	0.00	-0.04	0.00*
8	0.75	0.35	-0.17	-0.13	+0.42*	-0.13
9	0.79	0.16	+0.17*	-0.08	-0.04	-0.04
10	0.71	0.15	-0.13	+0.25*	-0.04	-0.08
11	0.77	0.25	+0.21*	0.00	-0.17	-0.04
13	0.73	0.58	-0.13	+0.42*	-0.17	-0.13
14	0.81	0.32	-0.13	+0.29*	-0.08	-0.08
15	0.67	0.44	-0.33	-0.04	+0.33*	+0.04
18	0.54	0.20	+0.17*	-0.21	+0.08	-0.04
19	0.73	0.48	+0.38*	-0.13	-0.17	-0.08
22	0.65	0.34	-0.17	-0.00	-0.13	+0.29*
23	0.81	0.14	0.04	-0.21*	0.08	0.08
26	0.90	0.22	-0.08	+0.13*	-0.04	0.00
28	0.63	0.31	-0.25	0.00	+0.33*	-0.08
29	0.54	0.24	+0.17*	-0.13	0.00	-0.04
30	0.71	0.47	+0.33*	0.00	-0.25	-0.08
31	0.42	0.29	-0.13	-0.21	+0.33*	0.00
32	0.54	0.37	-0.17	+0.42*	-0.04	-0.21
33	0.69	0.47	-0.08	-0.13	+0.46*	-0.25
34	0.48	0.24	+0.04	-0.29	+.21*	+0.04
35	0.69	0.22	-0.17	-0.29*	0.13	0.00
36	0.79	0.17	-0.04	-0.04	+0.08*	-0.08
37	0.42	0.30	+.08	+0.17	+0.08	-0.33*
38	0.75	0.47	+0.33*	-0.17	0.00	-0.17
39	0.35	0.20	-0.08	+0.21*	0.00	-0.13

إجراءات الدراسة:

1. تمَ اعداد نموذجين من الاختبار يتكون كل منهما من الفقرات نفسها وبنفس الترتيب للفقرات، وبنفس موقع الإجابة الصحيحة، لكنها تختلف في ترتيب المموهات، حيث يكون ترتيب المموهات عشوائيا في النموذج الأول ويكون الترتيب من المموه الأقوى إلى الأضعف في النموذج الثاني. وقد تم استبعاد الفقرات التي فيها على الأقل مموهين لهما نفس الفعالية تقريبا وعددها (11) فقرة لجعل الترتيب للمموهات أكثر تمايزا. وبذلك أصبح كل نموذج يتكون من 20 فقرة.
2. تمَ تطبيق نموذجي الاختبار على طلبة كل شعبة بالتناوب حيث طبق النموذج الأول على الفرد الأول والنموذج الثاني على الفرد الثاني وهكذا بالنسبة لبقية طلبة الشعبة، وتصحيحهما واستخراج النتائج وإدراجها في ملف اكسل تمَهيدا للتحليل.
3. تمَ تقسيمُ المحكمين عشوائيا إلـى مجمـوعتين، كـلُ مجموعـةٍ تتكـونُ مـن (20) محكمـاً مـن معلمي ومشرفي موادِ العلومِ، وقامت المجموعة الأولى بتقدْير درجة القطع للنموذج الأول في جولتين قبل معرفة صعوبة الفقـرات وبعـد معرفـة صـعوبة الفقـرات، وفـق طريقـة ندلسـكي، وقامت المجموعة الثانية بتقدير درجة القطع للنموذج الثاني في جـولتين قبـل معرفـة صـعوبة الفقرات وبعد معرفة صعوبة الفقرات.وقدْ سبق ذلك تدريب أفراد المجموعتين على كيفية تنفيذ طريقة ندلسكي، وتوضيح الأساس المنطقي لها في تحديد الحد الأدنى لمستوى الأداء المطلوب في نموذجي الاختبار.

المعالجة الإحصائية

تم استخدامَ برنامجِ Excell وبرنامجِ SPSS لإيجاد خصائص الفقرات من حيث الصعوبة والتمييز وفعالية المموهات. وتم استخدامَ طريقة ندلسكي لتقدير درجات القطع لكلا النموذجين من قبل المحكمين في المجموعتين واستخدام الاختبار t-test للمقارنة بين الأوساط الحسابية للنموذجين للإجابة عن السؤال الأول في الدراسة. وقد تم استخدامَ برنامج AMOS 22 لإجراء التحليل العامليَ التوكيدي متعدد المجموعات للإجابة عن السؤال الثاني وفقاً لمؤشرات الجدارة في النموذج الخطي الهيكلي:(SEM)

1-مربع كاي/درجات الحرية.

2- مؤشرات المطابقة المطلقة ومنها:

*مؤشر المطابقة المقارن Comparative Fit Index(CFA).

*مؤشر وسط مربع الخطأ التقريبي Root Mean Square Error Of (RMSEA) Approximation.

*مؤشر الجذر المعياري لوسط البواقي Root Mean Square Residual(RMSR).

نتائج الدراسة

لقد هدفت الدراسة إلى الكشف عن أثر طريقة ترتيب المموهات في فقرات الاختيار من متعدد على تحديد درجة القطع لاختبار محكي المرجع في مادة العلوم للصف الثامن، والتحقق من خاصية اللاتغير في القياس عبر طريقتي ترتيب المموهات. وفيما يلي عرضٌ للنتائج بحسب ترتيب سؤالي الدراسة:

1- **النتائج الخاصة بالسؤال الأول** "هل تختلفُ درجة القطع لاختبار محكي المرجع في العلوم للصف الثامن الاساسي باستخدام طريقة ندلسكي باختلاف طريقة ترتيب المموهات (ترتيب عشوائي مقابل ترتيب من الأقوى إلى الأضعف)؟"

بين جدول 2 الأوساط الحسابية لتقديرات المحكمين لدرجات القطع للاختبار المكون من (20) فقرة للنموذج الأول في جولتي التقدير قبل وبعد معرفة المحكمين لصعوبة الفقرات (بيانات الدراسة)، حيث أن الوسط الحسابي لتقديرات المحكمين هو بمثابة درجة القطع لفقرة الاختبار، وأن الوسط الكلي لجميع الفقرات هو بمثابة درجة القطع للاختبار.

جدول 2: الأوساط الحسابية لتقديرات المحكمين لفقرات الاختبار للنموذج الأول قبل وبعد معرفة صعوبة الفقرات

الفقرة	الوسط الحسابي للتقديرات قبل معرفة صعوبة الفقرات	الوسط الحسابي للتقديرات بعد معرفة صعوبة الفقرات
1	0.26	0.43
2	0.30	0.42
3	0.27	0.43
4	0.26	0.42
5	0.29	0.43
6	0.35	0.44
7	0.38	0.42
8	0.36	0.45
9	0.26	0.40
10	0.29	0.43
11	0.35	0.42
12	0.26	0.45
13	0.37	0.43
14	0.37	0.41
15	0.25	0.42
16	0.31	0.43
17	0.36	0.44
18	0.32	0.42
19	0.42	0.43
20	0.36	0.40
الوسط الحسابي الكلي	0.319	0.426

يتبين من جدول 2 أن قيم أوساط التقدير للمحكمين لدرجة القطع لفقرات النموذج الأول قبل معرفة صعوبة الفقرات كانت في المدى (0.25–0.42)، وأن الوسط الكلي لجميع الفقرات يساوي (0.319)، وهذه القيمة تعد درجة قطع للاختبار باستخدام نموذج ندلسكي بدون معرفة معاملات

صعوبة الفقرات، وعند ضرب هذه النسبة بعدد فقرات الاختبار (20 فقرة) فإنها تساوي (6)، أي على الطالب في الصف الثامن الاساسي الإجابة على 6 فقرات من فقرات اختبار العلوم حتى يعد متمكنا من الوحدات الدراسية.

كما يتبين من جدول 2 أن قيم أوساط التقدير للمحكمين لدرجة القطع لفقرات النموذج الأول بعد معرفة صعوبة الفقرات كانت في المدى (0.40–0.45)، وأن الوسط الكلي لجميع الفقرات يساوي (0.426)، وعند ضرب هذه النسبة بعدد فقرات الاختبار (20 فقرة) فإنها تساوي (9)، أي على الطالب في الصف الثامن الاساسي الإجابة على (9) فقرات من فقرات اختبار العلوم حتى يعد متمكنا من الوحدات الدراسية.

ولاختبار الدلالة الإحصائية للفرق بين الوسطين تم استخدام اختبار t للعينات المرتبطة. ويبين الجدول 3 نتائج اختبار t.

جدول 3: نتائج اختبار (t) للعينات المترابطة

النموذج الأول	المتوسط الحسابي	الانحراف المعياري	قيمة اختبار t المحسوبة	درجة الحرية	الدلالة الإحصائية
قبل	0.320	0.050	9.399	19	0.000
بعد	0.427	0.014			

يبين جدول 3 أن الفرق دال إحصائيا عند مستوى الدلالة 0.01، أي أنه يوجد فرق دال إحصائيا بين متوسطي تقديرات المحكمين للنموذج الأول.

ويبين الجدول 4 الأوساط الحسابية لتقديرات المحكمين لدرجات القطع للاختبار المكون من (20) فقرة للنموذج الثاني، في جولتي التقدير قبل وبعد معرفة المحكمين لصعوبة الفقرات (بيانات الدراسة).

جدول 4: الأوساط الحسابية لتقديرات المحكمين لفقرات الاختبار للنموذج الثاني قبل وبعد معرفة صعوبة الفقرات

الفقرة	الوسط الحسابي للتقديرات قبل معرفة صعوبة الفقرات	الوسط الحسابي للتقديرات بعد معرفة صعوبة الفقرات
1	0.48	0.53
2	0.31	0.46
3	0.42	0.54
4	0.44	0.52
5	0.39	0.47
6	0.32	0.53
7	0.40	0.47
8	0.42	0.53
9	0.37	0.43
10	0.34	0.50
11	0.42	0.52
12	0.40	0.50
13	0.40	0.46
14	0.41	0.48

15	0.42	0.51
16	0.39	0.54
17	0.38	0.49
18	0.41	0.50
19	0.36	0.58
20	0.33	0.58
الوسط الحسابي الكلي	0.39	0.51

يتبين من الجدول 4 أن قيم أوساط التقدير للمحكمين لدرجة القطع للنموذج الثاني قبل معرفة صعوبة الفقرات كانت في المدى (0.31–0.48)، وأن الوسط الحسابي الكلي يساوي (0.39)، وهذه القيمة تعد درجة قطع للاختبار باستخدام نموذج ندلسكي بدون معرفة معاملات صعوبة الفقرات، وعند ضرب هذه النسبة بعدد فقرات الاختبار (20 فقرة) فإنها تساوي (8)، أي على الطالب في الصف الثامن الاساسي الإجابة على 8 فقرات من فقرات اختبار العلوم حتى يعد متمكنا من الوحدات الدراسية.

كما يتبين من الجدول 4 أن قيم أوساط التقدير للمحكمين لدرجة القطع لنفس النموذج الثاني بعد معرفة صعوبة الفقرات كانت في المدى (0.43–0.58)، وأن الوسط الحسابي الكلي يساوي (0.506)، وعند ضرب هذه النسبة بعدد فقرات الاختبار (20 فقرة) فإنها تساوي (10)، أي على الطالب في الصف الثامن الاساسي الإجابة على (10) فقرات من فقرات اختبار العلوم حتى يعد متمكنا من الوحدات الدراسية.

ولاختبار دلالة الفرق بين الوسطين، تم استخدام اختبار t للعينات المرتبطة، ويبين الجدول 5 نتائج اختبار t.

جدول 5: نتائج اختبار (t) للعينات المترابطة

النموذج الثاني	المتوسط الحسابي	الانحراف المعياري	قيمة اختبار t المحسوبة	درجة الحرية	الدلالة الإحصائية
قبل	0.390	0.042	9.473	19	0.000
بعد	0.506	0.037			

يبين جدول 5 أن الفرق دال إحصائيا عند مستوى الدلالة 0.01، أي أنه يوجد فرق دال إحصائيا بين متوسطي تقديرات المحكمين للنموذج الثاني.

ولاختبار دلالة الفرق بين وسطي النموذجين قبل معرفة صعوبة الفقرات، تم استخدام اختبار(t-test) للعينات المستقلة، ويبين الجدول 6 نتائج اختبار t.

جدول 6: نتائج اختبار tللعينات المستقلة للنموذجين الأول والثاني قبل معرفة صعوبة الفقرات

النموذج	المتوسط الحسابي	الانحراف المعياري	قيمة اختبار t المحسوبة	درجة الحرية	الدلالة الإحصائية
الأول	0.32	0.05	-4.82	38	0.000
الثاني	0.39	0.04			

يبين جدول 6 أن الفرق بين وسطي النموذجين قبل معرفة الصعوبة دال إحصائيا عند مستوى الدلالة 0.01، أي أنه يوجد فرق دال إحصائي بين متوسطي تقديرات المحكمين لصالح النموذج الثاني.

ولاختبار دلالة الفرق بين وسطي النموذجين بعد معرفة صعوبة الفقرات، تم استخدام اختبار(T-Test) للبيانات المستقلة، ويبين الجدول 7 نتائج اختبار t.

جدول 7: نتائج اختبار t للعينات المستقلة للنموذجين الأول والثاني بعد معرفة صعوبة الفقرات

النموذج	المتوسط الحسابي	الانحراف المعياري	قيمة اختبار (ت) المحسوبة	درجة الحرية	الدلالة الإحصائية
الأول	0.43	0.01	-8.921	38	0.000
الثاني	0.51	0.04			

يبين جدول 7 أن الفرق بين وسطي النموذجين بعد معرفة الصعوبة دال إحصائيا عند مستوى الدلالة 0.01، أي أنه يوجد فرق دال إحصائيا بين متوسطي تقديرات المحكمين لصالح النموذج الثاني.

نتائج السؤال الثاني "هل يتحققُ اللاتغير في القياس لاختبار محكي المرجع في العلوم للصف الثامن الاساسي عبر طريقتي ترتيب المموهات؟**":**

للإجابة عن السؤال الثاني؛ تم أولا إجراء التحليل العاملي الاستكشافي من أجل تحديد النموذج المعتمد في التحليل العاملي التوكيدي متعدد المجموعات (MGCFA). وقد أفرز التحليل ثلاثة عوامل تتشبع عليها (16) فقرة؛ وبفحص محتوى الفقرات تبين أن الأبعاد الثلاثة تمثل الاحياء، والفيزياء، وعلوم الأرض. ولفحص تكافؤ البناء العاملي المُقترح للاختبار المحكي عبر طريقتي ترتيب المموهات فقد تم اختبار التكافؤ التكويني/الشكلي (Configural Invariance) والتكافؤ المتري (Metric Invariance) والتكافؤ العددي (Scalar Invariance) وفق متغير ترتيب المموهات من خلال برنامج التحليل الاحصائي (AMOS).

أولًا: التكافؤ التكوينيConfigural Invariance

تم أولًا فحص مطابقة النموذج المُقترح بشكل منفصل لكل مجموعة. ويبين جدول 8 نتائج مطابقة النموذج المقترح لكل من طريقتي ترتيب المموهات بشكل منفصل.

جدول 8: نتائج مطابقة البيانات لـ طريقتي ترتيب المموهات للنموذج العاملي المقترح

ترتيب المموهات	χ^2	DF	χ^2/DF	CFI	RMSEA	SRMR	الدلالة الإحصائية χ^2
عشوائي	122.623	99	1.239	.946	.041	.064	.054
ترتيب تنازلي	109.349	99	1.105	.968	.027	.058	.224

يوضح جدول 8 نتائج مؤشرات المطابقة لاختبار مبحث العلوم للصف الثامن الاساسي؛ حيث أن قيم (Chi/DF) المحسوبة أقل من 3 (Marsh & Hocevar,1985)، وأن قيمة (CFI) Comparative fit index مؤشر المواءمة أكبر من 0.90 (Hair,et al.,1995)، وأن قيمة مؤشر جذر متوسط مربع الخطأ التقريبي Root Mean Square Error (RMSEA) of Approximation أقل من 0.08 (Browne & Cudeck ,1993)، وقيمة مؤشر (SRMR) أقل من 0.09 (Hu & Bentler, 1999)؛ وهذا يدلل على مطابقة النموذج للبيانات في كل من طريقتي ترتيب المموهات. وتبين من قيم مؤشرات التعديل وجود ارتباطات بين الأخطاء في المجموعتين. وبعد ذلك تم استخدام التحليل العاملي التوكيدي متعدد المجموعات (MGCFA) مع البيانات الكاملة لاختبار التكافؤ التكويني/الشكلي (ConfiguralInvariance) بدون فرض قيود، ويبين جدول9 نتائج التحليل لاختبار اللاتغير التكويني.

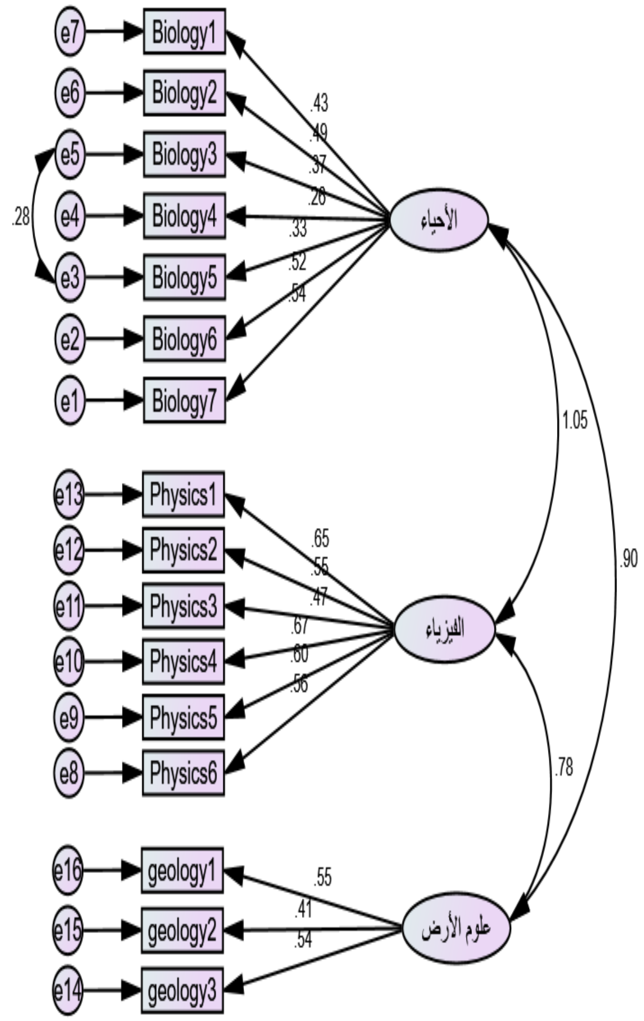
جدول 9: نتائج مطابقة بيانات اختبار العلوم للصف الثامن الأساسي

المؤشر الاحصائي	χ^2	DF	χ^2/DF	CFI	RMSEA	SRMR	الدلالة الإحصائية χ^2
طريقتي ترتيب المموهات	231.970*	198	1.172	.955	.025	.064	.049

*دالة إحصائياً عند مستوى الدلالة الإحصائية (α=0.05)

يوضح جدول 9 وجود دلالة إحصائية لقيمة (χ^2)؛ وقد بلغت قيمة χ^2/DF المحسوبة (1.172)، وهي أقل من 3، وتزيد قيمة (CFI) على 0.90، وأما مؤشر جذر متوسط مربع الخطأ التقريبي(RMSEA) فكانت قيمته أقل من 0.08، وقيمة مؤشر SRMR أقل من 0.09 (Chen, 2007)؛ ويدلل ذلك على تحقق خاصية اللاتغير التكويني بشكل جيد وفق متغير طريقتي ترتيب المموهات. ويبين الشكل 1 نموذج التكافؤ التكويني.

شكل 1: نموذج التكافؤ التكويني لاختبار مبحث العلوم للصف الثامن الأساسي



ثانياً: اللاتغير المتري

للتحقق من اللاتغير المتري تم حذف أوزان الانحدار (regression weight) التي يفترضها البرنامج للفقرات (Bilology1,Physics1, geology1)؛ إذ تم عمل وحدة قياس للمتغير الكامن، وذلك بإعطاء قيمة أولية لمعامل الانحدار لأفضل فقرة تشبعاً على العامل الكامن تساوي

واحد. وذلك ليتمكن البرنامج من تقدير معالم النموذج، ولأن اللاتغير المتري يفرض قيد تساوي التشبع للفقرات على المتغيرات الكامنة عبر المجموعات المختلفة. وتم فحص تحقق اللاتغير المتري في فقرات اختبار مبحث العلوم للصف الثامن الأساسي. ويبين جدول 10 نتائج اختبار نموذج اللاتغير المتري.

جدول 10: نتائج اختبار نموذج اللاتغير المتري (فرض قيود) لـ اختبار العلوم للصف الثامن الأساسي

المؤشر الاحصائي / النموذج	χ^2	DF	χ^2/DF	CFI	RMSEA	SRMR	الدلالة الإحصائية χ^2
اللاتغير التكويني (دون قيود)	231.970*	198	1.172	.955	.025	.064	.049
اللاتغير المتري	242.985	214	1.135	.962	.022	.066	.085

*دالة إحصائياً عند مستوى الدلالة الإحصائية ($\alpha=0.05$)

يتبين من جدول 10 وجود دلالة إحصائية لقيمة (χ^2) في النموذج التكويني، وعدم وجود دلالة إحصائية لقيمة (χ^2) في النموذج المتري، وكانت قيمة (χ^2/DF) في نموذج اللاتغير المتري أقل من 3 وفق متغير طريقتي ترتيب المموهات، وللتحقق من توفر خاصية اللاتغير المتري، تم حساب الفرق بين قيم المؤشرات (ΔRMSEA،ΔSRMR،ΔCFI) بين نموذج اللاتغير التكويني ونموذج اللاتغير المتري، في ظل فرض قيود على تساوي التشبع للفقرات على العوامل في طريقتي ترتيب المموهات.

ويبين جدول 11 نتائج هذه القيم (ΔRMSEA،ΔSRMR، ΔCFI)

جدول 11: نتائج اختبار نموذج اللاتغير المتري (فرض قيود) لـ اختبار العلوم للصف الثامن الاساسي

المؤشر الاحصائي	$\Delta\chi^2$	ΔDf	ΔCFI	ΔRMSEA	ΔSRMR	الدلالة الإحصائية $\Delta\chi^2$
اللاتغير المتري	11.014	16	.007	.003	.002	.809

يبين جدول 11 أن قيم الفرق في المؤشرات (ΔRMSEA،ΔSRMR،ΔCFI) بين نموذج اللاتغير التكويني (غير المقيد)، ونموذج اللاتغير المتري (المقيد) تدل على تحقق كامل (بدون تحرير لأي من قيم التشبع للفقرات على العوامل) لخاصية اللاتغير المتري في اختبار علوم الصف الثامن الأساسي، بالإضافة إلى أن قيمة الاختبار الاحصائي (الفرق في قيمة كاي تربيع) بين نموذج اللاتغير التكويني، ونموذج اللاتغير المتري غير دالة إحصائياً. وهذا دليل على تحقق خاصية اللاتغير المتري. أي أن فرض قيد تساوي قيم التشبع للفقرات على العوامل حسب متغير طريقتي ترتيب المموهات لا تؤثر بصورة دالة احصائياً في مطابقة البيانات.

ثالثاً: اللاتغير العددي

يتحقق اللاتغير العددي في فقرات اختبار العلوم للصف الثامن الأساسي حسب متغير طريقتي ترتيب المموهات، بفرض قيد تساوي قيم التشبع للفقرات على المتغيرات الكامنة عبر متغير

ترتيب المموهات، وفرض قيد تساوي التقاطعات (الارتباطات) بين المتغيرات الكامنة بناءً على متغير طريقتي ترتيب المموهات على النموذج المقترح، ويبين الجدول 12 نتائج اختبار نموذج اللاتغير العددي.

جدول12: نتائج اختبار نموذج اللاتغير العددي (فرض قيود) لـ اختبار العلوم للصف الثامن الأساسي

المؤشر الاحصائي / النموذج	χ^2	DF	χ^2/DF	CFI	RMSEA	SRMR	الدلالة الإحصائية χ^2
اللاتغير المتري	242.985	214	1.135	.962	.022	.066	.085
اللاتغير العددي	245.070	217	1.129	.963	.021	.067	.093

يتبين من جدول 12عدم وجود دلالة إحصائية لقيمة (χ^2) في النموذج المتري، والنموذج العددي، وكانت قيمة (χ^2/DF) في كل من نموذج اللاتغير المتري، ونموذج اللاتغير العددي أقل من 3 عبر متغير طريقتي ترتيب المموهات، وللتحقق من توفر خاصية اللاتغير العددي، تم حساب الفرق بين قيم المؤشرات (ΔRMSEA،ΔSRMR،ΔCFI) بين نموذج اللاتغير المتري ونموذج اللاتغير العددي، في ظل فرض قيود على قيم التشبع للفقرات على العوامل والتقاطعات بينها عبر طريقتي الترتيب للمموهات. ويبين الجدول13 نتائج هذه القيم (ΔCFI ،ΔSRMR، ΔRMSEA).

جدول 13: نتائج اختبار نموذج اللاتغير العددي لـ اختبار العلوم للصف الثامن الأساسي

المؤشر الاحصائي	$\Delta\chi^2$	ΔDf	ΔCFI	ΔRMSEA	ΔSRMR	الدلالة الإحصائية $\Delta\chi^2$
اللاتغير العددي	2.085	3	.001	.001	.001	.555

يبين جدول 13 أن قيم الفرق في المؤشرات (ΔCFI,ΔSRMR,ΔRMSEA) بين نموذج اللاتغير المتري، ونموذج اللاتغير العددي) تدل على تحقق كامل (بدون تحرير لأي من قيم التشبع للفقرات على العوامل) لخاصية اللاتغير العددي في اختبار علوم الصف الثامن الأساسي ككل وفق متغير طريقتي ترتيب المموهات، بالإضافة إلى أن قيمة الاختبار الاحصائي (الفرق في قيمة كاي تربيع) بين نموذج اللاتغير المتري، ونموذج اللاتغير العددي غير دالة إحصائياً. وهذا دليل على تحقق خاصية اللاتغير العددي؛ أي أن فرض قيد تساوي قيم التشبع للفقرات على العوامل والتقاطعات بين العوامل حسب متغير ترتيب المموهات لا تؤثر بصورة دالة احصائياً في مطابقة البيانات.

مناقشة النتائج والتوصيات

نتائج السؤال الأول: هل تختلفُ درجة القطع للاختبار باختلاف طريقة ترتيب المموهات؟

تشير النتائج إلى اختلاف في درجة القطع باختلاف طريقة ترتيب المموهات سواء بمعرفة صعوبة الفقرات أو بدون معرفتها، حيث كانت درجة القطع في النموذج الذي يكون فيه ترتيب المموهات من الأقوى إلى الأضعف أعلى من درجة القطع في النموذج الذي يكون فيه ترتيب المموهات عشوائيا. وتفسر هذه النتيجة أنه عند ترتيب المموهات من الأقوى إلى الأضعف تزداد فرصة إدراك المحكم للخطأ في المموه الأقوى وعليه تزداد فرص استبعاده، وهذا بالتأكيد يزيد من احتمال إجابة الطالب الذي يمتلك الحد الأدنى المقبول من الكفاية. وهذا يعمل على رفع درجة القطع. ويلاحظ أن درجات القطع كانت منخفضة نسبيا للنموذجين. وقد يعزى الإنخفاض إلى طبيعة طريقة ندلسكي المستخدمة. وهذا يتفق مع ما أشار له شانغ (Chang, 1999) حيث يعزو انخفاض درجة القطع وفق نموذج ندلسكي للأسباب التالية: السبب الأول أن طريقة ندلسكي تضع قيودا على المحكمين فيما يتعلق بعدد القيم الاحتمالية التي تنتج من تقديراتهم، حيث أنها تؤدي إلى قيم منفصلة والفروق بينها غير متساوية، فإذا قدر المحكم أن الطالب قادر على استبعاد بديلا واحدا من أربعة بدائل فإنه يعطي الفقرة درجة قطع قيمتها (0.33)، أما إذا استبعد بديلين فإنه يعطيها درجة قطع قيمتها (0.5)، وإذا استبعد ثلاثة بدائل يعطيها درجة قطع قيمتها (1.00)، أما إذا لم يستطع استبعاد أي بديل فإنه يعطيها درجة قطع قيمتها (0.25)، ومن خلال النظر إلى هذه القيم نجد أن معظمها لا يتجاوز القيمة (0.5)، أي أن النموذج لا يسمح للمحكمين بإعطاء قيم مرتفعة، كما أن المحكمين نادرا ما يعطوا احتمال واحد صحيح، ولهذا السبب جاءت درجات القطع لنموذجي الاختبار منخفضة، وهنالك سبب آخر وهو صعوبة تحديد البدائل التي يرى المحكمون أن الطلبة الذين يمتلكون الحد الأدنى من الكفاية قادرون على استبعادها كونها خاطئة، وهذا ما لاحظه الباحث من بعض المحكمين عند استخدامهم للنموذج. كما يمكن أن يعزى إلى اعتبارات تتعلق بإدراك المحكم للطالب الذي يمتلك الحد الأدنى المقبول من الكفاية والاعتقاد السائد لديه بأنه ضعيف.

ومن ناحية أخرى يلاحظ أن درجة القطع للنموذجين ارتفعت بعد تزويد المحكمين بصعوبة الفقرات. وبالنظر إلى صعوبة الفقرات يلاحظ أن الغالبية العظمى من الفقرات كانت سهلة حيث كانت درجات الصعوبة لغالبية الفقرات فوق 0.60، ومن الطبيعي أن يقوم المحكم بمراجعة تقديراته بعد معرفته بأن الفقرات سهلة. ويتوقع في الفقرات السهلة أن تزداد الخيارات التي يتوقع أن تستبعد من الطالب الذي يمتلك الحد الأدنى المقبول من الكفاية. ومن المنطقي أن تكون درجة القطع في الاختبار السهل أعلى منها في الاختبار الصعب بالاعتماد على الإدراك نفسه لمفهوم الفرد ذي الحد الأدنى من الكفاية.

وقد ارتفعت تقديرات المحكمين لدرجات القطع وفق نموذجي الاختبار بمعرفة صعوبة الفقرات، مما يدلل على أهمية توفير مؤشرات عن صعوبة الفقرات وهذا يؤكد أهمية تجريب الاختبار وفقراته لمعرفة المؤشرات الإحصائية كالصعوبة والتمييز وفاعلية المموهات وتوفيرها للخبراء الذين يقومون بتقدير درجات القطع حيث لا تستند الأحكام على الانطباعات الشخصية فقط بل تدعم وتعزز بالبيانات التجريبية.

نتائج السؤال الثاني: "هل يتحقق اللاتغير في القياس للاختبار المحكي عبر طريقتي ترتيب المموهات؟"

للاجابة عن هذا السؤال تم اعتماد التحليل العاملي التوكيدي متعدد المجموعات (MGCFA)، وقد أفرز التحليل ثلاثة عوامل تتشبع عليها (16) فقرة؛ وبفحص محتوى الفقرات تبين أن الأبعاد الثلاثة تمثل الاحياء، والفيزياء، وعلوم الأرض. ولفحص تكافؤ البناء العاملي المُقترح للاختبار المحكي عبر طريقتي ترتيب المموهات فقد تم اختبار التكافؤ التكويني/الشكلي (Configural Invariance) والتكافؤ المتري (Invariance Metric) والتكافؤ العددي (Scalar Invariance) وفق متغير ترتيب المموهات من خلال برنامج التحليل الاحصائي (AMOS).

حيث كشفت نتائج هذا السؤال عن تحقق كامل لخاصية اللاتغير في القياس للاختبار المحكي في كل من طريقتي ترتيب المموهات، إذ كشفت النتائج عن تحقق خاصية اللاتغير التكويني عبر متغير ترتيب المموهات، وذلك دليل على عدم وجود اختلاف في البنية العاملية المفترضة للاختبار المحكي باختلاف طريقتي ترتيب المموهات، ويعزى ذلك إلى الصدق البنائي للاختبار المحكي، وتمثيل الفقرات للمحتوى، والدقة بإجراءات تطبيق الاختبار المحكي.

وكشفت النتائج أيضاً تحقق كامل لخاصية اللاتغير المتري للاختبار المحكي عبر متغير ترتيب المموهات، وتحقق كامل لخاصية اللاتغير العددي على مستوى الاختبار، وهذا يعني أن الفرق بين الأوساط الحسابية في الأداء على النموذجين للاختبار المحكي (إن وجد) لا يعزى لاختلاف في خصائص الاختبار باختلاف ترتيب المموهات. وهذا يدلل على أن اختلاف ترتيب المموهات لا يؤثر في خصائص الاختبار، وعليه فإنه يمكن في الاختبارات الالكترونية خلط البدائل وتغيير مواقعها دون أن يكون لذلك أثر سلبي على الاختبار.

التوصيات

في ضوء نتائج الدراسة يوصي الباحثان بما يأتي:

-إجراء دراسات أخرى للكشف عن أثر ترتيب المموهات في اختبار اختيارٍ من متعددٍ محكيّ المرجع باستخدام طرق أخرى مختلفة عن الطريقة المستخدمة في هذه الدراسة.

-إجراء دراسات مشابهه لتسليط الضوء على طرق إحصائية جديدة في الكشف عن صدق الاختبارات.

-ضرورة تدعيم الأحكام الشخصية ببيانات إحصائية حول الفقرات في عملية تحديد درجات القطع ومستويات الأداء في الاختبارات المختلفة.

-إعطاء مطوري بنوك الأسئلة حرية ترتيب المموهات في فقرات الاختيار من متعدد بالطريقة التي يرتاحون لها حيث لن يغير ذلك فيما تقيسه تلك الفقرات.

-يمكن لمطوري نماذج الاختبار المتكافئة عمل تلك النماذج من خلال تغيير ترتيب المموهات في كل فقرة لكون اختلاف ترتيب المموهات لا يؤثر في خصائص الاختبار.

المراجع

أولاً: المراجع العربية:

أبو فودة، باسل خميس سالم. (2014). أثر إعادة ترتيب بدائل الاستجابة في صعوبة فقرة الاختيار من متعدد. مجلة الدراسات العربية في التربية وعلم النفس، 53، 265-287.

أبو مسرة، آيات خالد فالح، والسوالمة، يوسف محمد (2021) .استخدام طريقة أشجار راش في التقسيم المتكرر للكشف عن الأداء التفاضلي للفقرة في اختبار الرياضيات الوطني لضبط نوعية التعليم. المجلة الاردنية في العلوم التربوية، 17(1)،119-134.

بني عطا، زايد. (2005). استخدام نموذج ذي الحدين لفحص تقدْيرات المحكمين لدرجة القطع لاختبار بدلالة عدد بدائل الفقرة. رسالة دكتوراة غير منشورة. جامعة اليرموك، الأردن.

بني عطا، زايد، والشريفين، نضال. (2012). أثر اختلاف شكل توزيع القدْرة على معالم الفقرة ودالة المعلومات للاختبار. المجلة الأردنية في العلوم التربوية، 8(2)، 151- 166.

السمكي، لطفي. (2020). الصدق والثبات في القياس لاختبار تمَس "TIMSS 2011" في العلوم للصف الثامن عبر موقع المدرسة والجهة المشرفة في الأردن. مجلة العلوم التربوية والنفسية، 4(2)، 45-67.

الشهري، بندربن سالم. (2017). فحص اللاتغير في تقدْيرات معالم الفقرات والأفراد في ضوء النموذج اللوجستي الثلاثي المعلم . مجلة كلية التربية، 33(1)،411-444.

الشريم، احمد، والسوالمة، يوسف. (2006). تحديد درجة القطع لاختبار محكي المرجع في الرياضيات باستخدام نموذجي "انجوف" و "ندلسكي": دراسة مقارنة بمعرفة صعوبة الفقرات وعدم معرفتها .المجلة الأردنية في العلوم التربوية، 2(1)، 1-10.

السعيد، أريج والشريفين، نضال. (2021). الكشف عن اللاتغير في الاختبار التقييمي لمهارات الرياضيات لطلبة الصف الثالث الأساسي عبر مديريات التربية في الأردن المجلة الدولية للدراسات التربوية والنفسية، 9(1)، 113-128.

سليمان، شاهر والصالح، علي. (2016). بناء اختبار تحصيلي محكي المرجع في القياس والتقويم باستخدام النموذج اللوغاريتمَي ثلاثي المعلم لنظرية الاستجابة للمفردة الاختبارية. مجلة التربية، 168(1)، 175-201.

العبيدي، غانم والجبوري، حنان. (1981). اساسيات القياس والتقويم في التربية والتعليم. دار العلوم للطباعة والنشر.

الكعابنة، نايل، وعودة، أحمد (2019) أثر طريقة اختيار المموهات لفقرات اختبار الاختيار من متعدد في خصائصها السيكومترية وفق النظرية الكلاسيكية والنظرية الحديثة في القياس. المجلة العلمية لجامعة الملك فيصل (العلوم الانسانية والادارية)، 26(1)،85-110.

عباس، حيدر جليل. (2020). طرق تحديد درجات القطع لاختبارات الذكاء والقدْرات العقلية .مجلة أبحاث الذكاء، 14(29)،144-199.

علام، صلاح الدين محمود. (2006). الاختبارات والمقاييس النفسية والتربوية. دار الفكر العربي، 241-243.

ثانياً: المراجع الاجنبية

Asamoah, D., & Ocansey, M. (2019). Item discrimination and distractor Analysis: A technical report on thirty multiple choice core mathematics achievement test items. International Journal of Research and Scientific Innovation, 6 (9), 2321, 2705.

Browne, M., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen and J. S. Long (Eds.), Testing structural equation models. (pp. 136-162).

Chang, L. (1999). Judgmental item analysis of the Nedelsky and Angoff standard-setting methods. Applied Measurement in Education, 12(2), 151-165.

Chen, F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. Structural equation modeling: a multidisciplinary journal, 14(3), 464-504.

Gronlund, N. (1993). How to make achievement tests and assessments? (5th Ed.). Boston: Allyn & Bacon.

Haladyna, T., Downing, S., & Rodriguez, M. (2002). A review of multiple-choice item-writing guidelines for classroom assessment. Applied measurement in education, 15(3), 309-333.

Hambleton, R. (1982). Test Score Validity and Standards-setting Mmethods. In Berk, R. A. (ed.), Criterion-referenced Measurement: the State of Art, 2ed. London: The John Hopkins Press Ltd.

Hambleton, R., & Swaminathan, H. (1985). Test equating. Item ResponseTheory: Principles and Applications, 197-223.

Hair, F., Anderson, E., Tatham, L., & Black, C. (1995). Multivariate data analysis with Reading (4thED). Prentice-Hall.

Hu, L., & Bentler, M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Structural equation modeling: a multidisciplinary journal, 6(1), 1-55.

Johnstone,A. H., &Ambusaidi, A. (2001).Fixed-response questions with a difference. Chemistry Education Research and Practice, 2(3), 313-327.

Lions, S., Monsalve, C., Dartnell, P., Godoy, M. I., Córdova, N., Jiménez, D., ... &Lemarié, J. (2021). The position of distractors in multiple-choice test items: The strongest precede the weakest. In Frontiers in Education (p. 391). Frontiers.

Lord, F. (1980). Applications of item response theory to practical testing problems (1st ed.). Routledge.

Marsh, H., & Hocevar, D. (1985). Application of confirmatory factor analysis to the study of self-concept: First-and higher order factor models and their invariance across groups. Psychological bulletin, 97(3), 562.

Reise, S., Widaman, K., & Pugh, H. (1993). Confirmatory factor analysis and item response theory: two approaches for exploring measurement invariance. Psychological bulletin, 114(3), 552.

Rrilly, R., Zink, D., & Israelski, W. (1984). Comparisonof direct and indirect methods for setting minimum passing scores. Applied Psychological Measurement, 8, 421-429.

Tsaousis,I.(2015). Factorial invariance and latent mean differences of scores on SAAT across Gender. National Center for Assessment in Higher Education (NCA) at Riyadh, Saudi Arabia.