أثر زمن الاختبار على الأداء التفاضلي للجنس في فقرات الاختبار الوطني لضبط نوعية التعليم في الرياضيات

إعداد

أ/ ميساء عبد الله الحسبان

جامعة اليرموك

أ.د/ يوسف محمد سوالمة

جامعة اليرموك

أثر زمن الاختبار على الأداء التفاضلي للجنس في فقرات الاختبار الوطني لضبط نوعية التعليم في الرياضيات

أ/ ميساء عبد الله الحسبان و أ.د/ يوسف محمد سوالمة *

الملخص:

هدفت الدراسة تعرف أثر زمن الاختبار على الأداء التفاضلي للفقرات وفقًا لمتغير الجنس في الاختبار الوطني لضبط نوعية التعليم في مبحث الرياضيات للصف الثامن الأساسي. وتكون الاختبار المستهدف من ٤٠ فقرة من نوع الاختيار من متعدد لكل منها أربعة بدائل. وتكونت عينة الدراسة من (١٤٢٠) طالب وطالبة من طلبة الصف الثامن الأساسي في محافظة الزرقاء، تم اختيارهم وفق أسلوب العينة العشوائية العنقودية مع مراعاة تمثيلها لمتغير الجنس، وتم تقسيمهم عشوائيا إلى مجموعتين: مجموعة ضابطة بزمن ساعة ونصف للاختبار، ومجموعة تجريبية بزمن ساعة واحدة للاختبار. وتم استخدام طريقة نسبة الأرجحية لمانتل هانزل MH-LOR للكشف عن الأداء التفاضلي للفقرات. وأسفرت النتائج في المجموعة التجريبية عن ظهور الأداء التفاضلي في ١٧ فقرة حيث كان الأداء التفاضلي منتظمًا في (١١) فقرة منها (٦) فقرات لصالح الذكور و(٥) فقرات لصالح الإناث، وغير منتظم في (٦) فقرات منها فقرتين لصالح الذكور و(٤) فقرات لصالح الإناث. وفي المجموعة الضابطة ظهر الأداء التفاضلي في (١٥) فقرة حيث كان الأداء منتظماً في (٥) فقرات منها (٤) فقرات لصالح الذكور وفقرة واحدة لصالح الإناث، وغير منتظم في (١٠) فقرات منها (٦) فقرات لصالح الذكور و(٤) فقرات لصالح الإناث. كما تشير النتائج إلى عدم وجود تأثير لزمن الاختبار (ساعة، ساعة ونصف) على الأداء التفاضلي لفقرات الاختبار الوطني لضبط نوعية التعليم.

الكلمات المفتاحية: الأداء التفاضلي للفقرات، الاختبار الوطني لضبط نوعية التعليم، زمن الاختبار، مانتل هانزل MH-LOR.

* أ/ **ميساء عبد الله الحسبان، أ.د/ يوسف محمد سوالمة:** جامعة اليرموك.

The Effect of Test Time on DIF in the National Mathematics Test for Controlling Instructional Quality According to Gender

Maysa Abdullah Alhusban
Yarmouk University\ Jordan

Prpf. Yousef Mohammad Sawalmeh
Yarmouk University\ Jordan

Abstract:

The study aimed to investigate the effect of test time on the differential item functioning in the National Mathematics Test for Instructional Quality control according to the gender using the Mantel-Haenszel MH-LOR method. The test consists of (40) items, each with four alternatives, distributed over five domains (numbers and operations on them, algebra, measurement, geometry, and statistics and probability). The study sample consisted of (1420) eighth-grade students from the Zarqa Governorate who were selected using a stratified random cluster sampling method from the first Zarqa Directorate, the second Zarqa Directorate, and the Russeifa Directorate, with consideration for gender representation. They were randomly divided into two groups: a control group with an hour and half test time, and an experimental group with one-hour test time. The Mantel-Haenszel MH-LOR odds ratio method was used to detect the differential functioning of the. The results in the experimental group revealed differential performance in 17 items, where performance was consistent in 11 items, with 6 favoring males and 5 favoring females, and inconsistent in 6 items, with 2 favoring males and 4 favoring females. In the control group, differential performance was observed in 15 items, with consistent performance in 5 items, 4 favoring males and 1 favoring females, and inconsistent performance in 10 items, with 6 favoring males and 4 favoring females. Furthermore, the results indicate no effect of test time (one hour, one and a half hours) on the differential performance of items in the national test for educational quality control.

Keywords: Test time, differential item performance, differential distractor performance, national test for controlling instructional quality.

المقدمة:

تعد الاختبارات وسيلة تقييم ذات أهمية بالغة في ميدان التربية والتعليم. وتم تطويرها واستخدامها لاستكشاف جوانب مختلفة من القدرات العقلية والسمات الشخصية للأفراد. وتفيد الاختبارات التحصيلية في العديد من القرارات التربوية، وفي تزويد المعلمين بالمعلومات اللازمة لتصميم المناهج الدراسية، وفي تقييم تأثير السياسات المختلفة على تعلم الطلاب. وتسهم الاختبارات في تحسين جودة التعليم من خلال توجيه الجهود التعليمية نحو تلبية احتياجات الطلبة بشكل أفضل، وتحسين استراتيجيات التدريس، وتوجيه الطلبة نحو المجالات التي تتناسب مع قدراتهم وميولهم الفردية مما يعزز الاستعداد الذهني والدافعية للدراسة والتحسن المستدام في جودة التعليم (علام، ٢٠١٩).

وقد تتأثر فعالية التقييم وعدالته بسرعة إجابة أسئلة الاختبار. فقد تؤدي سرعة الاختبار إلى الحد من دقة التقييم، لأن الطلاب الذين يستغرقون وقتًا أطول في الإجابة على الأسئلة قد يكونوا في الواقع أكثر معرفةً من الطلاب الذين يستغرقون وقتاً أقل ويجيبون بسرعة أكبر. وتُعَرَّف سرعة الاختبار على أنها وجود قيود زمنية على الاختبار تحول دون إكمال مجموعة المفحوصين لجزء غير قليل من أسئلة الاختبار (Bejar,1985).

وتعد قيود الوقت من أبرز العوامل التي تؤدي إلى زيادة سرعة المفحوص أثناء الاختبار، خاصة في سياق التقييمات عالية المخاطر. والتفسير الأكثر شيوعًا لذلك هو أن الكثير من المفحوصين يقضون وقتًا طويلاً للغاية في الاستجابة لمجموعة الفقرات الأولى في الاختبار، مما يترك وقتًا غير كافٍ لإجابة ما تبقى من فقرات الاختبار. ويتسبب ذلك في التسرع في إجابة الفقرات المتبقية دون استخدام الاستراتيجيات المثلى في الإجابة، مما يؤدي إلى التخمين و/أو ترك الفقرات دون إجابة (Suh, et al., 2012).

ويؤدي تحديد وقت طويل للاختبار إلى تفاوت في وقت الاستجابة؛ بعض الطلاب يجيب عن الأسئلة ويسلم إجاباته في وقت قليل من وقت الاختبار، وبعضهم الآخر يستغرق كل وقت الاختبار. هذا التفاوت في وقت الاستجابة قد يؤثر في أداء الطلاب. لكن هذا التأثير قليل نسبيا مقارنة بالتأثير الناتج عن التفاوت في الوقت المحدد مسبقا للاختبار (Pommerich, 2004). وقد كشف لاندرم وآخرون (Landrum et al., 2009) عن وجود ارتباط سالب بين الوقت اللازم لإكمال الاختبار والأداء عليه.

يُعد تحديد الوقت المناسب للاختبار أحد أكثر التحديات أهمية في تطوير الاختبار، ومن المرجح أن تحدد قرارات مطوري الاختبار في هذا المجال حدود العدالة والمنفعة والصدق في

عملية التقييم. فعدم توفير الوقت الكافي للمفحوصين قد يقلل درجاتهم على الاختبار عن القدرات الحقيقية لهم. وإذا تم منح الكثير من الوقت، فقد يقيس الاختبار سلوكيات غير التي وضع من أجلِها وقد يسمح بإدخال تأثيرات عملية غير مرغوب فيها تتعلق بزيادة التكاليف. ويساعد معالجة هذه المشكلات بشكل صحيح في وقت مبكر من عملية تطوير الاختبار في زيادة عدالة التقييمات للمفحوصين وجعلها أكثر صدقا فيما يتعلق بتحقيق أغراض الاختبار (Sireci, & Botha, 2020).

ويعتمد تحديد الزمن المناسب للاختبار على عواملَ عملية وسيكومترية متعددة، مثل الوقتِ المتاحِ للاختبار وساعاتِ العمل وساعاتِ الدراسة وتوافرِ القاعات والمراقبين وموارِدِ الكمبيوتر وتكاليفِ إدارةِ الاختبار (; Martone & Sireci, 2009; Sireci, 1998; Sireci & Faulkner-Bond, 2014). وهناك عواملٌ **سيكومترية** مهمة يجبُ مراعاتها عند تحديدِ القيودِ الزمنيةِ، وتساهمُ في تعظيمِ دقةِ وصحةِ التقييم مثل خاصية الصدق التي تتطلبُ أن تمثلَ الاختباراتُ بشكلٍ كافٍ البنيةَ (المحتوى) المطلوبة كما هو محددٌ في مواصفاتِ الاختبارَ (Crocker, 2003; Messick, 1989; Sireci,1998)، وخاصية الثبات التي تشير إلى المدى الذي يعطي الاختبار النتيجة نفسها في المحاولات المتكررة؛ إذ يجب ألا تختلف درجات المتقدمين للاختبار بشكل كبير عبر أشكال متكافئة من الاختبار أو عبر مناسبات اختبار مختلفة. وبالتالي، عند تحديد القيود الزمنية للاختبار، يواجه مصممو الاختبارات كلًا من القيود العملية المتعلقة بتكاليف وموارد الاختبار بالإضافة إلى الخصائص السيكومترية المتمثلة في الصدق والثبات (Wainer & Thissen, 1996).

وعند تحديد القيود الزمنية المعقولة للاختبار، يأخذ باني الاختبار في الاعتبار عدة عوامل مختلفة، من ضمنها عدد الفقرات في الاختبار (طول الاختبار)، ومستوى الصعوبة الإجمالي للاختبار، ومستوى قدرة الأفراد المستهدفين بالاختبار. وعدم أخذ هذه العوامل في الاعتبار أثناء عملية تحديد زمن الاختبار، قد يؤدي إلى التشكيك في صحة النتائج خاصة في ظل عدم كفاية الوقت المحدد واضطرار المفحوص إلى الإجابة بسرعة، ويشكل عامل السرعة في هذه الحالة عاملا دخيلًا مشوشًا على تأثير القدرة في الأداء الكلي على الاختبار (De Boeck et al., 2011). وقد بين العمري (١٩٨٥) أن زيادة وقت الاختبار عن الوقت المناسب لا تؤدي إلى زيادة دالة إحصائيا في الوسط الحسابي لعلامات المفحوصين عليه، بينما تقصير وقت الاختبار يؤدي إلى نقص دال إحصائيا في الوسط الحسابي عند مختلف فئات الأفراد في مستوى القلق ودرجة المخاطرة. وتوصل ودلي وآخرون (Wadley et al., 2013) إلى أنه لا يوجد أثر

لتمديد وقت الاختبار على الوقت المستغرق في إجابة فقراته ولا على الأداء عليه من قبل الأفراد من ذوي فرط الحركة ونقص الانتباه.

وتحديد مدة غير كافية للاختبار قد يؤثر سلبًا على صدق الاختبار، ويحول دون قياس الاختبار ما صمم لقياسه. وغالبا ما تُعزى سرعة الاختبار إلى تأثير عدم كفاية وقت الاختبار إلى الحد الذي لا يجد عنده معظم المفحوصين الوقت الكافي للتفكير والإجابة عن كل فقرة في الاختبار، وعندها يقال إن الاختبار سريع "Speeded"، وكذلك يسمى الاختبار سريعا إلى مدى معين، إذا كان المفحوصون يأخذون علامات أقل من العلامات التي يأخذونها إذا كان لديهم الوقت الكافي لإنهاء الاختبار، أو أتبين أنهم بحاجة إلى وقت إضافي غير محدد لإنهائه (Bridgeman et al, ٢٠٠٤). وقد بين مانديناك وآخرون (Mandinach et al., 2005) إلى أنه لا يوجد أثر لتمديد وقت الاختبار على الوقت المستغرق في إجابة فقراته ولا على الأداء عليه من قبل الأفراد من ذوي فرط الحركة ونقص الانتباه.

وتحديد مدة غير كافية للاختبار قد يؤثر سلبًا على صدق الاختبار، ويحول دون قياس الاختبار ما صمم لقياسه. وغالبا ما تُعزى سرعة الاختبار إلى تأثير عدم كفاية وقت الاختبار إلى الحد الذي لا يجد عنده معظم المفحوصين الوقت الكافي للتفكير والإجابة عن كل فقرة في الاختبار، وعندها يقال إن الاختبار سريع "Speeded"، وكذلك يسمى الاختبار سريعا إلى مدى معين، إذا كان المفحوصون يأخذون علامات أقل من العلامات التي يأخذونها إذا كان لديهم الوقت الكافي لإنهاء الاختبار، أو أتبين أنهم بحاجة إلى وقت إضافي غير محدد لإنهائه (Bridgeman et al, ٢٠٠٤). وقد بيّن مانديناك وآخرون إلى أن تمديد وقت الاختبار يفيد الافراد ذوي القدرة المتدنية ومع ذلك فإنه لا يفيد من لا يملك المعرفة والمهارة. أما تولنتو–ميلر وآخرون (Talento-Miller et al., 2013) فقد أشاروا إلى أن سرعة الأداء على اختبارات اللغة الإنجليزية التكيفية لا تختلف باختلاف اللغة الام عند ضبط مستوى القدرة.

وهناك ازدياد مطرد لاستخدام الاختبارات محددة الوقت، وتميل هذه الاختبارات إلى أن تكون اختبارات قوة تُطبق ضمن حدود زمنية. وقد وجد الباحثون أن اختبارات السرعة والقوة في المهام المعقدة لا تقيس البنية نفسها. فالسرعة في اختبار غير محدد التوقيت غير مرتبطة بدرجة الاختبار. وقد اهتمت بعض الأبحاث النفسية التي تركز على الدرجات في الاختبارات بتطوير نماذج تأخذ وقت الاستجابة بالاعتبار. وتختلف تلك النماذج فيما بينها من حيث توزيع وقت الاستجابة، والعلاقة المفترضة بين مستوى القدرة التي تقيسها الفقرة وسرعة الاستجابة لها (Wang & Sireci, 2013).

وقد اهتم علماء القياس النفسي وعلماء النفس المعرفي بالعلاقة بين السرعة والدقة، فقد أكد علماء النفس الإدراكيون على وجود علاقة عكسية بين السرعة والدقة. ويرى علماء القياس أن وقت الاختبار يلعب دورا اساسيًا في فرص إكمال الاختبار. وقد ثبت أن تحيز السرعة يؤدي إلى تحيز تقديرات معلمات النموذج في سياق أدبيات نظرية الاستجابة للفقرة IRT (Oshima, 1994) ويؤثر أيضًا على صدق الاختبار. وعلى الرغم من أن السرعة تُعزى في الغالب إلى عدم كفاية الوقت المخصص للاختبار، إلا أن هناك عوامل أخرى قد تمنع المفحوص من إنهاء الاختبار في حدود زمنية محددة (Wise et al, 2015).

وتعد سرعة الاختبار مصدر قلق لدي باني الاختبار حيث قد تتعرض الخصائص السيكومترية للخطر (Crocker & Algina, 1986). وبشكل خاص عندما لا تكون السرعة جزءًا من بنية الاختبار موضع الاهتمام (Lord et al., 1968). وقد أشار أتالي (Attali,2005) إلى أن تسريع الاستجابة في الاختبار يزيد من فرص التخمين العشوائي في الفقرات الأخيرة في الاختبار مما يقلل من اتساقها مع الفقرات الأولى في الاختبار وبالتالي يُضعف من ثبات الاختبار. وقد بين ايرلام وآخرون (Erlam et al.,2016) أن تحديد زمن للاختبار له تأثير كبير على طول وجودة المحتوى في الكتابة التفاعلية ولكنه لا يؤثر على دقة الكتابة أو درجة تعقيدها، أي أن الزمن قد لا يشكل تهديدا لصدق الاختبار.

ويشير مصطلح وقت الاستجابة (Response Time) في الاختبار إلى الوقت الذي يقضيه المتقدم للاختبار على الفقرة في الاختبار (Schnipke & Scrams, ٢٠٠٢). والسرعة هي مصدر محتمل لتباين الخطأ الذي قد يصبح تهديدًا خطيرًا لصدق الاختبار، وللتفسير المناسب للنتائج (Lu & Sireci ,2007). وتزيد الخصائص الشخصية للمفحوص المتمثلة بالقلق والمخاطرة واللغة وضعف القدرة والدافعية من فرص تعرضه لضغط الوقت في الاختبار (Goegebeur et al., 2008). والمفحوص لا يمتلك حرية اختيار قدراته ومستوى سرعة إجابته للفقرة، لكن هناك علاقة تبادلية بين السرعة والقدرة، ويتم تمثيل فكرة العلاقة التبادلية باعتبار أن سرعته في الإجابة تحدد الوقت الذي يستغرقه في إجابة الفقرة وأن معلمة القدرة له تحدد صحة الاستجابة (van der Linden, 2006). وقد أكد كلين وآخرون (Klein et al., 2009) على أن العلاقة التبادلية بين السرعة والدقة هي ظاهرة فردية، وأن الشخص الذي يخضع للاختبار يتحكم في السرعة أثناء الاختبار.

نظرًا لأن معظم الاختبارات الموحدة يتم إجراؤها في حدود زمنية محددة، فإنه يزداد الانتباه إلى سرعة الاختبار بغض النظر عن شكله. ويعزى ذلك إلى أن سرعة الاختبار غير المنضبطة

بين نماذج الاختبار قد تؤدي إلى وجود نسبة عالية من المتقدمين للاختبار لا يستطيعون إكماله لعدم كفاية الوقت، وقد يؤثر ذلك على صدق الاختبار (Lu & Sireci, 2007).

وقد يكون لقيود الوقت تأثير كبير على صحة الدرجات حيث يعتمد القياس الدقيق على افتراض أن درجات الاختبار تمثل الكفاءة الحقيقية للمفحوص فيما يتعلق بالبنى المقصودة. فزمن الاختبار الذي لا يسمح للمفحوصين بإظهار كفاءتهم الحقيقية، يمكن أن يكون له عواقب سلبية من خلال إدخال مصدر للتباين ليس له صلة ببنية الاختبار. ويمكن أن يؤدي الوقت غير المناسب للاختبار إلى نتائج مشكوك فيها أو حتى غير عادلة (متحيزة) إذا كانت الدرجة التي تؤثر بها السرعة على النتائج مختلفة عما هو متوقع بناءً على بنية الاختبار (Lu & Sireci, 2007).

وزاد الاهتمام بمشكلة التحيز في الاختبارات العقلية نتيجة لحركة حقوق الإنسان، حيث عملت على تحقيق العدالة والمساواة بين الأفراد في فرص التعليم والعمل بجعل الاختبارات خالية من التحيز ما أمكن(Conoley, 2003) . وذكر كروكر وألجينا (Crocker & Algina, 1986) أن التحيز في فقرات الاختبار يعكس مصادر غير مقصودة للتباين تؤثر بصورة مختلفة على مجموعة من الأفراد مقارنة بمجموعة أخرى، وأضاف كول وموس (Cole & Moss, 1989) أنه يمكن أن يؤدي التحيز في درجات الاختبار إلى عدم تمثيل القدرات الحقيقية للطلاب، مما يؤدي إلى نتائج غير دقيقة وغير عادلة، وقد تؤثر سلبًا على قرارات تحديد المسار والتصنيف والتنبؤ القائمة على الدرجات.

ولاستكشاف هذا التحيز، يتم استخدام تحليل الأداء التفاضلي للفقرة (Differential Item Functioning DIF) لدراسة الاختلاف في احتمالية إجابة الأفراد من خلفيات مختلفة ولكن بنفس مستوى القدرة على فقرة اختبارية بشكل صحيح (Penfield & Camilli, 2006 ;Camilli et al.,1994). وقد يشير الأداء التفاضلي للفقرات DIF إلى التحيز وبالتالي عدم الإنصاف المحتمل للقياس. لذلك، يجب استخدام تحليل الأداء التفاضلي للفقرات DIF بشكل روتيني في تطوير الاختبارات التعليمية والنفسية والتحقق من صحتها (Martinková et al., 2017). وقد يكون الأداء التفاضلي منتظما عندما تعمل الفقرة بشكل ثابت لصالح مجموعة معينة عبر جميع مستويات القدرة وغير منتظم عندما تعمل الفقرة لصالح مجموعة معينة عند جزء من متصل القدرة بينما تعمل لصالح مجموعة أخرى عند الجزء الآخر من متصل القدرة.

طرق الكشف عن الأداء التفاضلي للفقرات (DIF):

هناك عدة طرق للكشف عن الأداء التفاضلي للفقرة (DIF) بالاعتماد على تحليل بيانات الأداء على الاختبار بناءً على خصائص الطلاب الديموغرافية أو النفسية أو الاجتماعية.

وتشمل الطرق الشائعة للكشف عن الأداء التفاضلي للفقرة (DIF) الانحدار اللوجستي Logistic Regression (LR) ويُستخدم هذا الأسلوب لمقارنة احتمالية نجاح الطلاب في فقرة ما بناءً على خصائصهم الديموغرافية أو النفسية أو الاجتماعية، ونظرية استجابة الفقرة Item Rresponse theory (IRT)ويُستخدم هذا الأسلوب لمقارنة منحنيات استجابة الطلاب للفقرة بناءً على خصائصهم الديموغرافية أو النفسية أو الاجتماعية، وتتطلب هذه الطرق عينات كبيرة لتكون فعالة (Hidalgo & Lopez-Pina, 2004). وطريقة مانتل هانزل Mantel Haenszel (MH)، ونسبة الأرجحية Odds Ratio (OR) ونسبة الأرجحية لمانتل هانزل Mantel Haenszel Log Odds Ratio (MHLOR) Penfield & Camilli, (2006).

وتُستخدم إحصائية نسبة الأرجحية لمانتل هانزل (MH-LOR) لكشف الأداء التفاضلي للفقرات (DIF) في اختبارات الاختيار من متعدد. في هذه الطريقة، يتم تقسيم الممتحنين إلى مجموعتين: مجموعة مرجعية ومجموعة مستهدفة. يتم تحديد المجموعة المرجعية بناءً على بعض المتغيرات الخارجية، مثل الجنس أو العرق أو الخلفية التعليمية. ويتم تحديد المجموعة المستهدفة بناءً على المقارنة التي يتم إجراؤها، مثل المقارنة بين الجنسين أو الأعراق أو الخلفيات التعليمية. أي يتم حسابها من خلال تقسيم المتقدمين إلى مجموعتين: مجموعة مرجعية ومجموعة مستهدفة. ومن ثم تسجيل استجابات كل فقرة (صحيحة/خاطئة) لكلتا المجموعتين. بالنسبة لكل فقرة، يتم حساب نسبة الأرجحية لمانتل هانزل MH-LOR على أنها نسبة احتمال الإجابة الصحيحة في المجموعة المرجعية إلى احتمال الإجابة الصحيحة في المجموعة المستهدفة (Mantel, & Haenszel, 2001).

وتتميز هذه الطريقة بأنها سهلة الفهم والتفسير ولا يلزم وجود نماذج إحصائية معقدة مثل نظرية استجابة الفقرة(IRT). كما أنها متينة لانتهاكات بعض افتراضات نظرية الاستجابة للفقرة. ويؤخذ عليها أنها أقل قوة من طرق نظرية الاستجابة للفقرة IRT في كشف أنواع معينة من الأداء التفاضلي للفقرة DIF. كما أنها تتطلب أحجام عينات كبيرة للحصول على نتائج دقيقة. تستخدم هذه الطريقة لتقييم الأداء التفاضلي للفقرات (DIF) في حالة الفقرات ثنائية التصحيح وتعتمد على جدول التوافق من خلال مصفوفة مربعة ثنائية تصنف الافراد لكل فقرة وعند كل علامة اختبار.

وقد حاز موضوع الكشف عن الأداء التفاضلي لفقرات الاختبار الوطني لضبط نوعية التعليم على اهتمام الباحثين، بوصفه اختبارًا وطنيًا يستهدف فئات هامة من الطلبة ويتنبأ

بمهارات الطلبة في مباحث مختلفة (البرصان، ٢٠١٣؛ سلمان وظاظا، ٢٠٢٢؛ العمري وشطناوي، ٢٠١٦). لذلك سيتم استخدامه لتحقيق أهداف الدراسة الحالية.

وتجدر الإشارة إلى أنه، بحسب علم الباحثين، لا توجد أي دراسة عربية او اجنبية اهتمت بدراسة أثر زمن الاختبار على الأداء التفاضلي للفقرات. وتتميز هذه الدراسة بأنها ستقوم بدراسة تأثير وقت الاختبار على الأداء في الاختبار ومؤشرات صدقه بدلالة الأداء التفاضلي لفقراته مما يساعد في تطوير الاختبار وتحسين موثوقية نتائجه.

مشكلة البحث:

نظرًا لندرة الدراسات التي تطرقت لدراسة موضوع أثر السرعة في الإجابة على الأداء التفاضلي للفقرات، جاءت هذه الدراسة لتلقي الضوء على أثر زمن الاختبار على الأداء التفاضلي للفقرات وفقًا لمتغير الجنس حيث لم يسبق التطرق له. فقد اهتمت غالبية الدراسات التي تطرقت إلى زمن الاختبار إلى تأثيره على الأداء في الاختبارات المختلفة ولفئات الأفراد المختلفة وتأثير ذلك على دقة القياس وعلاقته بسرعة الإجابة. ويعد الأداء التفاضلي للفقرات من المؤشرات الحديثة على عدالة الاختبارات والتي ينصح بدراستها لغرض تطوير الاختبارات وضمان عدالتها. والدراسات السابقة تناولت الكشف عن الأداء التفاضلي لاختبارات معينة وبطرق مختلفة كما تناولت عملية المقارنة بين طريقتين أو أكثر، ولكن لا توجد أي دراسة عربية أو أجنبية_ في حدود علم الباحثين_ تطرقت إلى الكشف عن تأثير وقت الاختبار على الأداء التفاضلي لفقراته وفقا لمتغير الجنس. وتجدر الإشارة إلى أن مسألة تحديد الزمن المناسب للاختبار أصبحت من القضايا المهمة خاصة مع التحول إلى الاختبارات الالكترونية في كثير من الاختبارات واسعة النطاق. وبالتحديد تحاول الدراسة الإجابة عن الاسئلة الآتية:

١- هل تظهر فقرات الاختبار الوطني لضبط نوعية التعليم في مبحث الرياضيات للصف الثامن الأساسي أداءً تفاضليا للجنس في المجموعة التجريبية حيث تم تقليل وقت الاختبار؟
٢- هل تظهر فقرات الاختبار الوطني لضبط نوعية التعليم في مبحث الرياضيات للصف الثامن الأساسي أداءً تفاضليا للجنس في المجموعة الضابطة حيث لم يتم تقليل وقت الاختبار؟
٣- هل يختلف سلوك الفقرات في الأداء التفاضلي باختلاف المجموعة حيث يختلف وقت الاختبار عند مستوى الدلالة الإحصائية (α=0.05)؟

هدف الدراسة:

هدفت الدراسة إلى بحث أثر زمن الاختبار على الأداء التفاضلي للفقرات وفقًا لمتغير الجنس في الاختبار الوطني لضبط نوعية التعليم في مبحث الرياضيات للصف الثامن الأساسي.

أهمية الدراسة:

تبرز أهمية هذه الدراسة في تحديد أثر وقت الاختبار على الأداء التفاضلي للفقرات للجنس. ويتوقع أن توفر الدراسة معلومات مفيدة لتطوير الاختبار مستقبلا من حيث انتقاء الفقرات وتحديد الوقت المناسب لإجابتها بحيث تحقق درجة أعلى من العدالة بين الأفراد .

التعريفات الاصطلاحية والإجرائية:

- **تقليل وقت الاختبار**: تحديد زمن الاختبار بساعة واحدة فقط، أي بمعدل دقيقة ونصف لكل فقرة اختيار من متعدد في الاختبار، وهذا أقل من الوقت المحدد للاختبار من قبل وزارة التربية والتعليم.
- **عدم تقليل وقت الاختبار**: الالتزام بالوقت المحدد من قبل وزارة التربية والتعليم ومدته ساعة ونصف، أي بمعدل دقيقتين وربع لكل فقرة اختيار من متعدد في الاختبار.
- **سرعة الإجابة**: الوقت الذي يستغرقه المفحوص في الإجابة عن فقرات الاختبار (Schnipke & Scrams, ٢٠٠٢).
- **الأداء التفاضلي للفقرة**: هو مؤشر احصائي للتعبير عن الفروق في احتمالية إجابة الفقرة بشكل صحيح من قبل مجموعتين أو أكثر عند مستوى القدرة نفسه (Holland & Wainer, 2012).
- **الاختبار الوطني لضبط نوعية التعليم**: اختبار سنوي تعده وزارة التربية والتعليم ممثلة بمديرية الاختبارات في إدارة الامتحانات والاختبارات؛ لقياس مدى امتلاك الطلبة في المرحلة الدراسية المستهدفة، لمهارات التعلم الأساسية ومستويات أداء الطلبة بدلالة مؤشرات الأداء، ومدى امتلاكهم لمهارات اقتصاد المعرفة في المباحث التي اختبروا فيها (وزارة التربية والتعليم، ٢٠٢٢).

محددات الدراسة:

عند تعميم نتائج هذه الدراسة لا بد من أن يؤخذ بالاعتبار أنها اقتصرت على:

- الاختبار الوطني في ضبط نوعية التعليم في الرياضيات للصف الثامن الأساسي في الأردن للعام ٢٠٢٢.

- وقتين للاختبار هما الوقت المحدد من الوزارة وهو ساعة ونصف والثاني مدته ساعة واحدة.
- عينة من طلبة الصف الثامن في مدارس محافظة الزرقاء/ الأردن للعام الدراسي ٢٠٢٢/٢٠٢٣.

الطريقة والإجراءات:

- منهج الدراسة:

استخدمت الدراسة المنهج التجريبي الحقيقي وفق التصميم البعدي لمجموعتين عشوائيتين. مجموعة ضابطة أعطيت ساعة ونصف للتقدم للاختبار أي بمعدل دقيقتين وربع لكل فقرة اختيار من متعدد، ومجموعة تجريبية أعطيت ساعة واحدة للتقدم للاختبار أي بمعدل دقيقة ونصف لكل فقرة اختيار من متعدد.

- مجتمع الدراسة:

تكون مجتمع الدراسة من جميع طلبة الصف الثامن في المدارس الحكومية في محافظة الزرقاء للعام الدراسي (٢٠٢٢/٢٠٢٣)، والبالغ عددهم (١٨٧٤٨) طالب وطالبة موزعين على (٦٤) مدرسة في مديرية الزرقاء الأولى، و(٦٧) مدرسة في مديرية الزرقاء الثانية، و(٤٦) مدرسة في مديرية الرصيفة.

- عينة الدراسة:

تم اختيار عينة الدراسة بالطريقة العشوائية العنقودية من مجتمع الدراسة مع مراعاة تمثيلها لمديريات التربية في مجتمع الدراسة وتمثيلها لمتغير الجنس، وكانت وحدة الاختيار في المرحلة الأولى المدرسة وفي المرحلة الثانية الشعبة. وقد تم اختيار ٥ مدارس عشوائيا من مديرية الزرقاء الأولى ومنها تم اختيار ٢٠ شعبة عشوائيا من الصف الثامن الأساسي، كما تم اختيار ٦ مدارس عشوائيا من مديرية الزرقاء الثانية ومنها تم اختيار ١٠ شعب عشوائيا من الصف الثامن الأساسي، كما تم اختيار ٣ مدارس عشوائيا من مديرية الرصيفة ومنها تم اختيار ١٢ شعبة عشوائيا. وقد تم استبعاد استجابات ١٨٠ طالبا وطالبة بسبب نمطية الإجابات. وتكونت العينة بصورتها النهائية من ١٤٢٠ طالب وطالبة منهم (٦٨٩) طالب وطالبة بمديرية الزرقاء الأولى و(٢٦٣) طالب وطالبة بمديرية الزرقاء الثانية، و(٤٦٨) طالب وطالبة بمديرية الرصيفة.

أداة الدراسة:

تمثلت أداة الدراسة بالاختبار الوطني لضبط نوعية التعليم في المملكة الأردنية الهاشمية للصف الثامن الأساسي بمبحث الرياضيات، وتكون الاختبار من (٤٠) فقرة من نوع الاختيار

من متعدد، لكل فقرة أربع بدائل واحدة منها صحيحة، وتغطي خمسة محاور (الأعداد والعمليات عليها، الجبر، القياس، الهندسة، الإحصاء والاحتمالات). وقد تم تحديد وقتين لتطبيق الاختبار الأول مدته ساعة ونصف كما حددته وزارة التربية والتعليم والثاني مدته ساعة واحدة فقط.

إجراءات الدراسة:

تم تنفيذ إجراءات الدراسة بدءا بمراجعة الدراسات والأبحاث والمراجع المتعلقة بموضوع وقت الاختبار والسرعة في الإجابة والأداء التفاضلي للفقرات، ثم تم الحصول على الاختبار الوطني لضبط نوعية التعليم من إدارة الامتحانات والاختبارات في وزارة التربية والتعليم، ثم الحصول على كتاب تسهيل مهمة. ثم تحديد مجتمع الدراسة وعينة الدراسة. وقد تم تقسيم الشعب عشوائيا وبالتساوي في كل مديرية إلى مجموعتين: مجموعة ضابطة أعطيت ساعة ونصف لإجابة فقرات الاختبار، ومجموعة تجريبية أعطيت ساعة واحدة لإجابة فقرات الاختبار. ثم تم تطبيق الاختبار على أفراد المجموعتين من قبل الباحث الأول وبمساعدة معلمين ومعلمات في المدارس المشمولة في عينة الدراسة. وتم جمع أوراق الإجابة وتم تفريغ الإجابات في برمجية SPSS وتم التحقق منها واستبعاد الإجابات النمطية، وتم تصحيح الإجابات حيث تم إعطاء الإجابة الصحيحة العلامة ١ والإجابة الخاطئة العلامة، وتم إجراء التحليلات الإحصائية باستخدام برمجية Differential Distractor Functioning Software DDFS.

المعالجة الإحصائية:

١. لإجابة السؤال الأول والثاني تم إجراء تحليل الأداء التفاضلي للفقرات باستخدام طريقة نسبة الأرجحية لمانتل– هانزل MH-LOR من خلال البرمجية (DDFS)، وتم تفسير النتائج بناء على المؤشرات التالية:

- مؤشر Z (LOR): يحدد وجود أداء تفاضلي للفقرة أم لا، ويكون هناك أداء تفاضلي إذا كانت قيمة ($Z > 2$) أو ($Z < -2$).
- LOR (Log-odds ratio) كمؤشر لتحديد اتجاه وحجم الأداء التفاضلي:
  - يكون DIF لصالح المجموعة المرجعية إذا كانت ($LOR > 0$).
  - يكون DIF لصالح المجموعة المستهدفة إذا كانت ($LOR < 0$).
  - يكون DIF كبيراً إذا كانت ($|LOR| > 0.64$).
  - يكون DIF متوسطاً إذا كانت ($0.43 \leq |LOR| \leq 0.64$).
  - يكون DIF صغيرًا إذا كانت ($|LOR| < 0.43$) .

- مؤشر BD (The Breslow- Day chi-square test of trend in odds ratio) لشكل الأداء التفاضلي هل هو منتظم أو غير منتظم وذلك من خلال ما يلي:
  - يكون الأداء التفاضلي منتظمًا إذا كان (BD<٣.٨٤).
  - يكون الأداء التفاضلي غير منتظم إذا كان (BD>٣.٨٤).

٢. لإجابة السؤال الثالث تم استخدام اختبار فاي لمعرفة مدى توافق سلوك الأداء التفاضلي في المجموعة التجريبية مع نظيره في المجموعة الضابطة.

نتائج الدراسة ومناقشتها:

النتائج المتعلقة بالإجابة عن السؤال الأول: هل تظهر فقرات الاختبار الوطني لضبط نوعية التعليم في مبحث الرياضيات للصف الثامن الاساسي اداء تفاضليا للجنس في المجموعة التجريبية حيث وقت الامتحان ساعة واحدة؟

لإجابة هذا السؤال تم استخدم نسبة الأرجحية لمانتل –هانزل MH-LOR، من خلال البرمجية الإحصائية (DDFS)، ويبين الجدول (١) نتائج الأداء التفاضلي لفقرات الاختبار الوطني لضبط نوعية التعليم في مبحث الرياضيات للصف الثامن الأساسي في المجموعة التجريبية حيث وقت الاختبار ساعة واحدة وفقا للجنس.

جدول (١) نتائج الأداء التفاضلي لفقرات الاختبار الوطني
لضبط نوعية التعليم في مبحث الرياضيات للصف الثامن الأساسي
في المجموعة التجريبية حيث وقت الاختبار ساعة واحدة وفقا للجنس

الفقرة	LOR	الفئة المستفيدة	المقدار	SE(LOR)	Z(LOR)	BD	نوع الاداء التفاضلي
١	- 0.424	الإناث	متوسط	٠.١٧٧٣	.3914-2	٥.٨٧٦	غير منتظم
٢	−٠.٣٣٧٤	الإناث	صغير	٠.١٧٤٤	−١.٩٣٤٦	٦.٦٤٦	غير منتظم
٣	−٠.٨٤٩٧	الإناث	كبير	٠.٢٣٤٥	−٣.٦٢٣٥	١.٢٨٢	منتظم
٤	0.2258	الذكور	صغير	٠.١٨١١	١.٢٤٦٨	٠.١٥٨	منتظم
٥	−٠.٣٤٢٩	الإناث	صغير	٠.١٨٧٦	−١.٨٢٧٨	٩.٢٠١	غير منتظم
٦	٠.١٥٠٧	الذكور	صغير	٠.١٨٥٥	٠.٨١٢٤	٠.١٨٣	منتظم
٧	٠.١٠٨٨	الذكور	صغير	٠.١٩٧٣	٠.٥٥١٤	٠.٠٢	منتظم
٨	٠.٠٨٩	الذكور	صغير	٠.١٩٥	٠.٤٥٦٤	٢٠.٥٣٧	غير منتظم
٩	−٠.٠٤٣٢	الإناث	صغير	٠.١٩٧٨	−٠.٢١٨٤	٠.٠٠٨	منتظم
١٠	٠.٠٣٣	الذكور	صغير	٠.١٩٧٣	٠.١٦٧٣	٠.٤٦٨	منتظم
١١	٠.٤٢٣٢	الذكور	صغير	٠.١٧٥٩	٢.٤٠٥٩	٠.٦١٦	منتظم
١٢	٠.٥٦٥١	الذكور	متوسط	٠.١٨٩٤	٢.٩٨٣٦	١.٠٤٤	منتظم
١٣	١.٣٧٩٩	الذكور	كبير	٠.٢١٥٧	٦.٣٩٧٣	٧.١٣١	غير منتظم
١٤	٠.٤٨٢٥	الذكور	متوسط	٠.٢٣١٣	٢.٠٨٦	٣.١٤٩	منتظم
١٥	٠.٠٧٩٢	الذكور	صغير	٠.١٩١٦	٠.٤١٣٤	١١.٦٨٤	غير منتظم

الفقرة	LOR	الفئة المستفيدة	المقدار	SE(LOR)	Z(LOR)	BD	نوع الاداء التفاضلي
١٦	٠.٠٥٧٤	الذكور	صغير	٠.١٩	٠.٣٠٢١	٨.٨٢٤	غير منتظم
١٧	٠.٠٠٢	الذكور	صغير	٠.٢٣٥٤	٠.٠٠٨٥	٠.٠٠٤	منتظم
١٨	٠.٠٦٤١	الذكور	صغير	٠.٢٠٨	٠.٣٠٨٢	٣.٢٧٦	منتظم
١٩	٠.٣٥٠٨	الذكور	صغير	٠.٢٠٧٥	١.٦٩٠٦	٤.١٨٩	غير منتظم
٢٠	-٠.١٤٢٧	الإناث	صغير	٠.١٨٨٥	-٠.٧٥٧	١.٠٦٤	منتظم
٢١	-٠.٣٦٥٥	الإناث	صغير	٠.٢٢٣١	-١.٦٣٨٣	٤.٧٨٨	غير منتظم
٢٢	-٠.٠٢٠٧	الإناث	صغير	٠.٢١٦٤	-٠.٠٩٥٧	١.٩٢٧	منتظم
٢٣	-٠.٧١٤٧	الإناث	كبير	٠.٢٠٠١	-٣.٥٧١٧	١.٩٢	منتظم
٢٤	٠.٠٣٧	الذكور	صغير	٠.٢٠٢٤	٠.١٨٢٨	٨.٢٥٤	غير منتظم
٢٥	-١.٣٠٢٦	الإناث	كبير	٠.٢٦٣٦	-٤.٩٤١٦	١٦.٤٠٨	غير منتظم
٢٦	٢.١٠٩٧	الذكور	كبير	٠.٢٤٠٤	٨.٧٧٥٨	١.٩١٨	منتظم
٢٧	-٠.١٠٦٢	الإناث	صغير	٠.١٩١٧	-٠.٥٥٤	٠.٠٠٢	منتظم
٢٨	-٠.٦٣٢٣	الإناث	كبير	٠.٢٣٠١	-٢.٧٤٧٩	٥.٩٩٤	غير منتظم
٢٩	-٠.٣٦٧٦	الإناث	صغير	٠.٢٠٩٧	-١.٧٥٣	١٦.٠٠٨	غير منتظم
٣٠	-٠.٨٥٨٥	الإناث	كبير	٠.٢١٦٨	-٣.٩٥٩٩	٠.٧٤٢	منتظم
٣١	-٠.٠٦١٩	الإناث	صغير	٠.٢١٧٢	-٠.٢٨٥	١.٢٩٥	منتظم
٣٢	-١.٥٢٧٢	الإناث	كبير	٠.٢٤٢٢	-٦.٣٠٥٥	٣٢.٥٤٩	غير منتظم
٣٣	٠.٠٠٧	الذكور	صغير	٠.١٨١٢	٠.٠٣٨٦	٣.٧٣٩	منتظم
٣٤	-١.٦٥٣٥	الإناث	كبير	٠.٣٤٨٥	-٤.٧٤٤٦	٠.٦٢٤	منتظم
٣٥	0.0102	الذكور	صغير	٠.١٩٧٢	٠.٠٥١٧	١.٩٤	منتظم
٣٦	١.٤٣٦٨	الذكور	كبير	٠.٢٤٨٨	٥.٧٧٤٩	٣.١٦	منتظم
٣٧	-٠.٢٤٧٥	الإناث	صغير	٠.١٩٦٢	-١.٢٦١٥	٠.٣٨٨	منتظم
٣٨	١.٥٨١٣	ذكور	كبير	٠.٢٢٤٣	٢.٥٩١٦	١.١٢١	منتظم
٣٩	١.٣٣٣١	الذكور	كبير	٠.٢٣٢٢	٥.٧٤١٢	٧.٢٠٣	غير منتظم
٤٠	-٠.٥٢٣٧	الإناث	متوسط	٠.١٨٤٨	-٢.٨٣٣٩	٠.٠٠٢	منتظم

يتبين من الجدول (١) أن هناك ١٧ فقرة ذات أداء تفاضلي تبعاً لمتغير الجنس في المجموعة التجريبية حيث وقت الاختبار ساعة واحدة، وكان الأداء التفاضلي منتظمًا في ١١ فقرة وغير منتظم في ٦ فقرات. وقد تضمنت الفقرات ذات الأداء التفاضلي المنتظم ٦ فقرات لصالح الذكور وبحجم كبير في الفقرات (٣٨،٣٦،٢٦) وبحجم متوسط في الفقرتين (١٢،١٤) وبحجم صغير في الفقرة (١١)، و(٥) فقرات لصالح الإناث وبحجم كبير في الفقرات (٣،٢٣،٣٠،٣٤) وبحجم متوسط في الفقرة (٤٠). وقد تضمنت الفقرات ذات الأداء التفاضلي غير المنتظم فقرتين لصالح الذكور وبحجم كبير في الفقرتين (٣٩،١٣)، و٤ فقرات لصالح الإناث وبحجم كبير في الفقرات(٣٢،٢٨،٢٥) وبحجم متوسط في الفقرة (١) . ويبين الجدول (٢) توزيع الفقرات ذات الأداء التفاضلي بحسب مجالات مبحث الرياضيات للصف الثامن الأساسي في وحجم الأداء التفاضلي.

جدول (٢) توزيع الفقرات ذات الأداء التفاضلي بحسب مجالات مبحث الرياضيات للصف الثامن الأساسي في وحجم الأداء التفاضلي

المجال	الفقرات ذات الأداء التفاضلي		
	كبير	متوسط	صغير
الأعداد والعمليات عليها	٣	١	–
الجبر	١٣	١٤،١٢	١١
القياس	٢٨،٢٦،٢٥،٢٣	–	–
الهندسة	٣٤،٣٢،٣٠	–	–
الإحصاء والاحتمالات	٣٩،٣٨،٣٦	٤٠	–

كما تم حساب الاوساط الحسابية والانحرافات المعيارية لأداء الطلبة في مبحث الرياضيات للصف الثامن الأساسي بعد حذف الفقرات التي تظهر الأداء التفاضلي. ويبين الجدول (٣) الوسط الحسابي والانحراف المعياري لأداء طلبة الصف الثامن في مبحث الرياضيات حسب الجنس بعد حذف الفقرات ذات الأداء التفاضلي ودلالة الفرق بين الوسطين.

جدول (٣) الأوساط الحسابية والانحرافات المعيارية لأداء الطلبة في مبحث الرياضيات للصف الثامن الأساسي بعد حذف الفقرات التي تظهر الأداء التفاضلي وفقًا لمتغير الجنس في ظل تقليل وقت الاختبار

الجنس	العدد	الوسط الحسابي	الانحراف المعياري	قيمة t	مستوى الدلالة
ذكور	227	١٣.٦	٣.٦٩	- ٤.٣٣٤	.٠٠٠
اناث	460	١٤.٩	٤.٠٤		

يلاحظ من الجدول (٣) أن هناك فرق في الوسط الحسابي لأداء طلبة الصف الثامن الأساسي بمبحث الرياضيات بين الذكور والإناث بعد حذف الفقرات ذات الاداء التفاضلي للجنس، والفرق بين الوسطين دال احصائياً عند مستوى الدلالة ($\alpha = ٠.٠٥$)، وهذا يدل فعليًا على أن الفرق في الأداء على الاختبار بين الذكور والإناث يمثل فرقًا حقيقيًا.

وقد تُعزى هذه النتيجة إلى عوامل تتعلق بالأفراد مثل الاختلافات في العوامل البيولوجية والاتجاهات نحو الرياضيات والجدية في التقدم للاختبار. كما قد تُعزى هذه النتيجة إلى الاختلافات بين الذكور والإناث في سرعة معالجة المعلومات التي تلزم لإجابة أسئلة الاختيار من متعدد، مما يؤثر على قدرتهم على إكمال جميع الأسئلة في وقت الاختبار المحدد. كما قد يختلف الذكور والإناث في استراتيجيات حل المشكلات التي يستخدمونها. فقد يميل البعض إلى القراءة المتأنية لكل سؤال وفهمه قبل البدء بالإجابة، بينما قد يميل الآخرون إلى الإجابة بسرعة على أكبر عدد ممكن من الأسئلة ثم العودة إلى الأسئلة الصعبة لاحقًا، ومع قلة الوقت، قد يتّخذ الطلاب قرارات خاطئة أو يتّسرعون في الإجابة على الأسئلة، مما يؤثر على دقة إجاباتهم. كما

قد يعاني البعض من قلق الاختبار بشكل أكبر من الآخرين مع نقصان الوقت، مما قد يُؤثّر على تركيز الطلبة وقدرتهم على التفكير بشكل واضح وحلّ الأسئلة بدقة. كما أن مهارات إدارة الوقت قد تختلف بين الطلبة اللذين يُجيدون إدارة وقتهم والطلبة اللذين يُواجهون صعوبة في إدارة وقتهم.

وتتفق هذه النتائج مع نتائج دراسة كل من كيم وآخرون (Kim, et al, ١٩٩٥)، ودراسة جيري وآخرون (Giray, et al,.1999)، ودراسة الطراونة (٢٠١١)، ودراسة البرصان (٢٠١٣)، ودراسة الفريحات وسواقد (٢٠١٤)، ودراسة العمري وشطناوي (٢٠١٨)، ودراسة العجلوني والسوالمة (٢٠١٦)، ودراسة أبو مسرة والسوالمة (٢٠٢٠) بوجود أداء تفاضلي في فقرات اختبارات مبحث الرياضيات.

وتتفق هذه النتائج مع دراسة غارنر وانغلهارد (Garner & Englhard, ١٩٩٩) في أن الذكور أفضل من الإناث في الهندسة والقياس وتتعارض معها في أن الطالبات أفضل من الطلاب في الجبر. وقد يعزى هذا التعارض إلى اختلاف في مجتمع الدراسة وفي مستوى الاختبار وطبيعته. وجدير بالذكر أن اختبار الرياضيات المستخدم في الدراسة الحالية يصنف ضمن اختبارات الأداء العادي وليس ضمن اختبارات أقصى أداء لكون العلامة ليست مهمة للطالب ولا يترتب عليها أي عواقب على تقييمه. وعلاوة على ذلك فقد كانت الأوساط الحسابية للذكور والإناث في الدراسة الحالية أعلى من نظيراتها في التطبيق الوزاري، ويعود ذلك إلى أن السياسة المتبعة في تطبيق هذه الاختبارات لدى الوزارة لا تتيح للطالب معرفة موضوع الاختبار (رياضيات أم علوم أم لغة عربية أم لغة إنجليزية) الذي سيكون من نصيبه إلا ساعة الاختبار، وعليه لا يكون هناك أي استعداد مسبق للاختبار.

النتائج المتعلقة بالإجابة عن السؤال الثاني: هل تظهر فقرات الاختبار الوطني لضبط نوعية التعليم في مبحث الرياضيات للصف الثامن الأساسي أداءً تفاضليا للجنس في المجموعة الضابطة حيث وقت الاختبار ساعة ونصف؟

ومن أجل الكشف عن الأداء التفاضلي لفقرات الاختبار تبعًا لمتغير الجنس في المجموعة الضابطة حيث لم يتم تقليل وقت الاختبار تم استخدم نسبة الأرجحية لمانتل –هانزل، من خلال البرمجية الاحصائية (DDFS). ويبين الجدول (٤) نتائج الأداء التفاضلي لفقرات الاختبار الوطني لضبط نوعية التعليم في مبحث الرياضيات للصف الثامن الاساسي تبعاً لمتغير الجنس في المجموعة الضابطة حيث لم يتم تقليل وقت الاختبار.

جدول (٤)

نتائج تحليل الأداء التفاضلي لفقرات الاختبار الوطني لضبط نوعية التعليم في مبحث الرياضيات للصف الثامن الاساسي تبعاً لمتغير الجنس في ظل عدم تقليل وقت الاختبار باستخدام طريقة نسبة الأرجحية لمانتل–هانزلMH-LOR

الفقرة	LOR	الفئة المستفيدة	المقدار	SE(LOR)	Z(LOR)	BD	نوع الاداء التفاضلي
١	-٠.٠٧٩٦	الإناث	صغير	٠.١٧٥٤	-٠.٤٥٣٨	١.٥٢١	منتظم
٢	-٠.٤٣٤٩	الإناث	متوسط	٠.١٧٨٤	-٢.٤٣٧٨	٣.٤٢٧	منتظم
٣	٠.١٧٢٨	ذكور	صغير	٠.٢٦٧٦	٠.٦٤٥٧	١.٢١٣	منتظم
٤	-0.0186	الإناث	صغير	٠.١٨٤٢	-٠.١٠١	٠.٠٤٨	منتظم
٥	٠.٩٧٧٤	الذكور	كبير	٠.١٩٦٨	٤.٩٦٦٥	٨.٣٩٦	غير منتظم
٦	-٠.١٥٢٩	الإناث	صغير	٠.١٩٥٦	-٠.٧٨١٧	٠.٣١٥	منتظم
٧	٠.٠٩٤٣	الذكور	صغير	٠.١٩٦٩	٠.٤٧٨٩	٠.٠٩٤	منتظم
٨	٠.١٦٩٧	الذكور	صغير	٠.٢١٤٨	٠.٧٩	٣.٥٠٧	منتظم
٩	-٠.١٥٣٢	الإناث	صغير	٠.٢١٠٥	-٠.٧٢٧٨	١.٢٩٧	منتظم
١٠	٠.٤٩١٣	الذكور	متوسط	٠.١٩٩	٢.٤٦٨٨	٤.٧٠٨	غير منتظم
١١	-٠.٢٦٢١	الإناث	صغير	٠.١٧٤٤	-١.٥٠٢٩	٠.٠٠١	منتظم
١٢	-٠.٣٥٥٣	الإناث	صغير	٠.١٨٨٩	-١.٨٨٠٩	٣.٩٣٩	غير منتظم
١٣	-١.٩١٦١	الإناث	كبير	٠.٢٢٥٢	-٨.٥٠٨٤	٣١.١٢٢	غير منتظم
١٤	-٠.٠٧٦٩	الإناث	صغير	٠.٢٤٨٨	-٠.٣٠٩١	٠.٢٠٣	منتظم
١٥	٠.٢٤٠٢	الذكور	صغير	٠.١٩٣٣	١.٢٤٢٦	١٨.٠٠١	غير منتظم
١٦	٠.١٨٨	الذكور	صغير	٠.١٨٥٢	١.٠١٥١	٦.٠٢٣	غير منتظم
١٧	-٠.٠٩٥٢	الإناث	صغير	٠.٢٤٥٣	-٠.٣٨٨١	٠.٠٠٠	منتظم
١٨	-٠.٢٤٤٧	اناث	صغير	٠.٢٠٣٢	-١.٢٠٤٢	٠.٤١٨	منتظم
١٩	٠.٠٠٥٨	الذكور	صغير	٠.٢٠٣٩	٠.٢٨٤٥	٠.٢٢٤	منتظم
٢٠	-٠.١٩٦٨	الإناث	صغير	٠.٢٠٣٨	-٠.٩٦٥٧	٣.٤٩٩	منتظم
٢١	-٠.٠٩٣٩	الإناث	كبير	٠.٢٥٣٣	-٠.٣٧٠٧	١.٤	منتظم
٢٢	-٠.٤٣٦٤	الإناث	متوسط	٠.٢٢٣١	-١.٩٥٦١	٠.٠٧٥	منتظم
٢٣	٠.٤٢٨٥	الذكور	متوسط	٠.١٨٠٧	٢.٣٧١٣	٠.٦٢٤	منتظم
٢٤	٠.٢٢٩٣	ذكور	صغير	٠.٢٠٩٩	١.٠٩٢٤	٠.٤٨٨	منتظم
٢٥	١.٣٥٤٦	الذكور	كبير	٠.٢٣٩٥	٥.٦٥٥٩	٨٧.١٥٢	غير منتظم
٢٦	-١.٨٩٣٩	الإناث	كبير	٠.٢١٠٥	-٨.٩٩٧١	٥.٩٤٩	غير منتظم
٢٧	-٠.١٤٠٤	الإناث	صغير	٠.٢٠٠٣	-٠.٧٠٠٩	١.١٨٢	منتظم
٢٨	٠.٧٩٩	ذكور	كبير	٠.٢٢٩٣	٣.٤٨٤٥	٢.٥٧٥	منتظم
٢٩	-٠.٠٢٠٥	الإناث	صغير	٠.٢٣٩٢	-٠.٠٨٥٧	١٠.٢٥٦	غير منتظم
٣٠	٠.٧٢٥٨	الذكور	كبير	٠.٢١٥٢	٣.٣٧٢٧	٠.٢١٣	منتظم
٣١	-٠.٠٧٩٦	الإناث	صغير	٠.٢٠٤٩	-٠.٣٨٨٥	٠.٣٥٤	منتظم
٣٢	٢.٥٠١٢	الذكور	كبير	٠.٢٩١١	٨.٥٩٢٢	٣٥.٢٨٩	غير منتظم

الفقرة	LOR	الفئة المستفيدة	المقدار	SE(LOR)	Z(LOR)	BD	نوع الاداء التفاضلي
٣٣	-٠.٠٨٣	الإناث	صغير	٠.١٧٦١	-٠.٤٢١٣	٣.٣٣٣	منتظم
٣٤	١.٣٩٥٤	الذكور	كبير	٠.٢٧٨٩	٥.٠٠٣٢	٧.٨٠٩	غير منتظم
٣٥	-0.1741	الإناث	صغير	٠.٢٠٣	-٠.٨٥٧٦	٦.٤١١	غير منتظم
٣٦	-١.٢٩٩٩	الإناث	كبير	٠.٢١٧٨	-٥.٩٦٨٣	١٣.١٦٤	غير منتظم
٣٧	٠.٤٧٥٨	الذكور	متوسط	٠.٢١	٢.٢٦٥٧	١.٥٥١	منتظم
٣٨	-٠.٣١٩٤	الإناث	صغير	٠.٢١٥٩	-١.٤٧٩٤	٠.٢٨٥	منتظم
٣٩	-١.٥٤٥٩	اناث	كبير	٠.٢١٧٣	-٧.١١٤١	٩.٠٠٦	غير منتظم
٤٠	٠.٧٩٩٧	الذكور	كبير	٠.١٧٦٥	٤.٥٣٠٩	٥.٢١٩	غير منتظم

يتضح من الجدول (٤) أن هناك ١٥ فقرة ذات أداء تفاضلي تبعاً لمتغير الجنس في المجموعة الضابطة حيث وقت الاختبار ساعة ونصف، وكان الأداء التفاضلي منتظمًا في ٥ فقرات وغير منتظم في ١٠ فقرات. وقد تضمنت الفقرات ذات الأداء التفاضلي المنتظم ٤ فقرات لصالح الذكور وبحجم كبير في الفقرتين (٢٨، ٣٠) وبحجم متوسط في الفقرتين (٢٣،٣٧)، وفقرة واحدة (٢) لصالح الإناث وبحجم متوسط. وقد تضمنت الفقرات ذات الأداء التفاضلي غير المنتظم (٦) فقرات لصالح الذكور وبحجم كبير في الفقرات (٥، ٢٥، ٣٢، ٣٤، ٤٠) وبحجم متوسط في الفقرة (١٠)، و٤ فقرات لصالح الإناث وبحجم كبير في الفقرات (١٣، ٢٦، ٣٦، ٣٩). ويبين الجدول (٥) توزيع الفقرات ذات الأداء التفاضلي بحسب مجالات مبحث الرياضيات للصف الثامن الأساسي وحجم الأداء التفاضلي. ويلاحظ من الجدول أن كلاً من مجال القياس ومجال الإحصاء والاحتمالات قد استحوذ على أربع فقرات ثم يليه مجال الهندسة بواقع ثلاث فقرات، يليه كلاً من مجال الأعداد والعمليات عليها ومجال الجبر بواقع فقرتين.

جدول (٥)

توزيع الفقرات ذات الأداء التفاضلي بحسب مجالات مبحث الرياضيات للصف الثامن الأساسي وحجم الأداء التفاضلي

المجال	الفقرات ذات الأداء التفاضلي		
	كبير	متوسط	صغير
الأعداد والعمليات عليها	٥	٢	-
الجبر	١٣	١٠	-
القياس	٢٥،٢٦،٢٨	٢٣	-
الهندسة	٣٠، ٣٢،٣٤	-	-
الإحصاء والاحتمالات	٣٩، ٤٠، ٣٦	٣٧	-

كما تم حساب الاوساط الحسابية والانحرافات المعيارية لأداء الطلبة في مبحث الرياضيات للصف الثامن الأساسي في المجموعة الضابطة بعد حذف الفقرات التي تظهر الأداء

التفاضلي وفقًا لمتغير الجنس. ويبين الجدول (٦) الوسط الحسابي والانحراف المعياري لأداء طلبة الصف الثامن في مبحث الرياضيات حسب الجنس بعد حذف الفقرات ذات الأداء التفاضلي ودلالة الفرق بين الوسطين.

جدول (٦)

الأوساط الحسابية والانحرافات المعيارية لأداء الطلبة في مبحث الرياضيات للصف الثامن الاساسي بعد حذف الفقرات ذات الأداء التفاضلي للجنس في المجموعة الضابطة

الجنس	العدد	الوسط الحسابي	الانحراف المعياري	قيمة t	مستوى الدلالة
ذكور	٤٦٩	١٦.٣	٣.٩٢	– ٦.٢٠٥	.٠٠٠0
اناث	٢٦٤	١٨.٢	٤.١٤		

يلاحظ من الجدول (٦) ان هناك فرق ظاهري في الوسط الحسابي لأداء طلبة الصف الثامن الأساسي بمبحث الرياضيات وفقًا لمتغير جنس الطالب، والفرق بين الوسطين دال احصائياً عند مستوى الدلالة ($\alpha = ٠.٠٥$) بعد حذف الفقرات ذات الأداء التفاضلي، ويعني ذلك أن الفروق حقيقية بين الجنسين ولصالح الإناث.

ويمكن عزو هذه النتيجة إلى العوامل البيولوجية والاستعداد والاتجاهات نحو الرياضيات والجدية في الاستجابة لفقرات الاختبار، فقد تؤدي الاختلافات البيولوجية بين الذكور والإناث دورًا في اختلاف القدرات المعرفية بين الجنسين، مما قد يُؤثّر على أداء الطلبة في بعض مهام الرياضيات. كما تؤدي التوقعات الثقافية دورًا هامًا في تشكيل دوافع واتجاهات الطلبة نحو مجالات دراسية معينة، مما قد يُؤثّر على اختياراتهم الدراسية وأدائهم الأكاديمي. كما قد تختلف فرص التعلم والتطوير المتوفرة لكل من الذكور والإناث في مجال الرياضيات، مما قد يُؤثّر على أدائهم في الاختبارات.

واشارت بعض الدراسات إلى وجود اختلافات طفيفة في مهارات المعالجة المكانية بين الجنسين كما قد يكون لدى الذكور مهارات أفضل في تدوير الأشكال وتصورها، بينما قد تتمتع الإناث بمهارات أفضل في اللغة والذاكرة اللفظية. وقد تلعب هذه الاختلافات دورًا في أداء بعض أنواع المسائل الرياضية (Aladaileh, 1996).

كما قد تختلف مهارات التعلم بين الجنسين، حيث قد يُفضل الذكور التعلم من خلال حل المشكلات، بينما قد تُفضل الإناث التعلم من خلال التعاون والتفاعل كما قد تؤدي هذه الاختلافات إلى تفضيل الطلبة لأنواع معينة من الأنشطة التعليمية والتي قد تُؤثر على أدائهم في الاختبارات.

النتائج المتعلقة بسؤال الدراسة الثالث: هل يختلف سلوك الفقرات من حيث الأداء التفاضلي باختلاف المجموعة حيث يختلف وقت الاختبار عند مستوى الدلالة الإحصائية (α=0.05)؟

وللإجابة عن هذا السؤال تم حساب اختبار فاي Phi لمعرفة مدى توافق سلوك الأداء التفاضلي في المجموعة التجريبية مع نظيره في المجموعة الضابطة. ويبين جدول ٧ توزيع فقرات الاختبار في المجموعة التجريبية والضابطة بحسب سلوك الأداء التفاضلي.

جدول (٧)

توزيع فقرات الاختبار الوطني لضبط نوعية التعليم في المجموعة التجريبية والضابطة بحسب سلوك الأداء التفاضلي

			المجموعة التجريبية		
			لا يوجد أداء تفاضلي	يوجد أداء تفاضلي	المجموع
المجموعة الضابطة	لا يوجد أداء تفاضلي	العدد	19	6	25
	يوجد أداء تفاضلي	العدد	4	11	15
	المجموع	العدد	23	17	40
قيمة اختبار PHI ٠.٤٨٣			الدلالة الإحصائية ٠.٠٠٢		

يتبين من الجدول (٧) أن (٤٧.٥%) من الفقرات ليس لديها أداء تفاضلي في كلا المجموعتين (الضابطة، التجريبية) وأن (٢٧.٥%) لديها أداء تفاضلي في كلا المجموعتين. أي أن نسبة الاتفاق بين سلوك الأداء التفاضلى للفقرات فى المجموعتين ٧٥% ونسبة عدم الاتفاق (٢٥%). وقد بلغت قيمة اختبار فاي Phi (٠.٤٨٣) وبلغت قيمة الدلالة الإحصائية لها (٠.002). أي أن هنالك توافق ذو دلالة معنوية فى سلوك الأداء التفاضلى للفقرات فى المجموعة التجريبية وسلوكها فى المجموعة الضابطة مما يدلل على ان تأثير الزمن على الأداء التفاضلى للفقرات هو تأثير غير جوهري. وهذه النتائج تدعم فكرة أن زمن الاختبار لا يؤثر فى جانب مهم من خصائص الاختبار وهو الأداء التفاضلي لفقراته. وبناء على ذلك فإنه من الممكن استخدامه في ظل تقييد زمن تطبيقه.

التوصيات:

في ضوء النتائج التي تم التوصل لها توصي الدراسة بما يلي:

١. استمرار استخدام اختبار ضبط نوعية التعليم في الرياضيات لقياس مهارات الطلاب، نظرًا لموثوقيته وثباته ضد تأثير الزمن.

٢. إجراء دراسات معمقة لفهم العوامل التي تساهم في الاختلافات في أداء طلاب الصف الثامن الأساسي في مبحث الرياضيات بين الذكور والإناث، مع مراعاة العوامل الاجتماعية والنفسية والتعليمية.
٣. دراسة أثر زمن الاختبار على الأداء التفاضلي للفقرات لمباحث أخرى غير الرياضيات .
٤. تطوير اختبارات عادلة ودقيقة واستخدام أساليب تصميم اختبارات حديثة لضمان عدالة الاختبارات ودقتها في قياس قدرات الطلبة.

المراجع

أبو مسرة، آيات (٢٠٢٠). *استخدام طريقة أشجار راش في التقسيم المتكرر للكشف عن الأداء التفاضلي للفقرة في اختبار الرياضيات الوطني لضبط نوعية التعليم (رسالة دكتوراه غير منشورة)*. جامعة اليرموك، إربد.

البرصان، إسماعيل. (٢٠١٣). الأداء التفاضلي لفقرات الاختبار الوطني الأردني لضبط نوعية التعليم لمادة الرياضيات للصف العاشر وفق متغير الجنس. *دراسات تربوية ونفسية. مجلة كلية التربية بالزقازيق*، ٢٨(٧٩)، ٢٢٩-٢٧٠.

الخرشة، طه (٢٠٢٣). الأداء التفاضلى لفقرات اختبار القدرات الرياضية (TOMA-3) تبعا لمتغير الجنس وأثره على دقة تقدير القدرة للأفراد ودالة المعلومات للاختبار وفق نظرية الاستجابة للفقرة. مجلة بحوث التربية النوعية، ٧٧،٤٠٣-٤٢٣.

سلمان، مفلح وظاظا، حيدر. (٢٠٢٢). الأداء التفاضلي لفقرات الاختبار الوطني لضبط نوعية التعليم في مباحث العلوم والرياضيات واللغتين العربية والإنجليزية للصف الثامن الأساسي في الأردن. *جرش للبحوث والدراسات*، ٢٣(٢) ، ٢٦٦٩ - ٢٦٩٦.

الطراونة، محمود ، و البكور، نائل. (٢٠١١). *تحري التحيز في فقرات الاختبار الوطني لطلبة الصف الثامن الأساسي في مادتي الرياضيات والعلوم لعام ٢٠١٠ م (رسالة ماجستير غير منشورة)*. جامعة مؤتة، مؤتة.

العجلوني، جهاد (٢٠١٦). فحص الأداء التفاضلى لمموهات فقرات اختبار الرياضيات فى التقييم الوطنى الأردنى لاقتصاد المعرفة وفقًا لجنس الطالب وموقع المدرسة. *مجلة العلوم الإنسانية*، جامعة النجاح للأبحاث، ٣٠ (١١)، ٢٢٦٦-٢٢٨٦.

علام، صلاح الدين. (٢٠١٩). *القياس والتقويم في العملية التدريسية*. دار المسيرة للنشر والتوزيع.

العمري، حسان وشطناوى، عبد الله. (٢٠١٦). الأداء التفاضلي لفقرات الاختبار الوطني لضبط نوعية التعليم في الرياضيات للصف العاشر تبعا لمتغير الجنس. *مجلة جامعة النجاح للأبحاث – العلوم الإنسانية*، ٣٠(٨)، .1554 - 1529

العمري، حمزة. (١٩٨٥). *تأثير طول وقت الاختبار التحصيلى على أداء المفحوصين عليه وعلى خصائصه السيكومترية، عند مفحوصين مختلفين فى درجات مخاطرتهم ومستوى قلقهم العام* (رسالة ماجستير غير منشورة). جامعة اليرموك، إربد .

فريحات، نجود (٢٠١٤). *الأداء التفاضلي لفقرات الاختبار الوطني لضبط نوعية التعليم في العلوم والرياضيات للصف العاشر الأساسي تبعاً لطريقة تقديم الاختبار (رسالة دكتوراه غير منشورة)*. جامعة اليرموك، إربد.

وزارة التربية والتعليم. (٢٠٢٢). *نتائج الاختبار الوطني لضبط نوعية التعليم للعام الدراسي ٢٠٢١-٢٠٢٢ للصف الثامن الأساسي*. عمان: وزارة التربية والتعلم.

Aladaileh, S. (1996). *The influence of sex and educational level in confidence in learning mathematics and its relationship to academic achievement in mathematics for sixth and tenth grades in the government schools of Karak Governerate*. Master Thesis, Mutah University, Al-Karak, Jordan.

Attali, Y. (2005). Reliability of speeded number-right multiple-choice tests. *Applied Psychological Measurement*, 29, 357-368.

Bejar, I. I. (1985). *Test speededness under number-right scoring: An analysis of the Test of English as a Foreign Language*. ETS Research Report Series, 1985(1), i-57.

Bridgeman, B., Trapani, C., & Curley, E. (2004). Impact of fewer questions per section on SAT I scores. *Journal of Educational Measurement*, 41, 291–310.

Camilli, G., Shepard, L., & Shepard, L. (1994). *Methods for identifying biased test items* (Vol. 4). Sage.

Cole, N. & Moss, P. (1989). Bias in test use. In R. L. Linn (Ed.), *Educational measurement* (pp. 201–219). Macmillan Publishing Co, Inc; American Council on Education

Conoley, C. A. (2003). *Differential item functioning in the Peabody Picture Vocabulary Test-: Partial correlation versus expert judgment*. Texas A&M University.

Crocker, L. (2003). Teaching for the test: Validity, fairness, and moral action. *Educational Measurement: Issues and Practice, 22*(3), 5-11.

Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. Holt, Rinehart and Winston.

De Boeck, P., Cho, S., & Wilson, M. (2011). Explanatory secondary dimension modeling of latent differential item functioning. *Applied Psychological Measurement, 35*(8), 583-603.

Erlam, R., & Wu, J. (2016). The effect of timing on the quantity and quality ot test-taker' writing. *New Zealand Studies in Applied Linguistics*, *22*(2), 21–34.

Garner, M. & Engelhard Jr, G. (1999). Gender differences in performance on multiple- choice and constructed response mathematics items. *Applied Measurement in Education, 12*(1), 29-51.

Goegebeur, Y., De Boeck, P., Wollack, J. & Cohen, A. (2008). A speeded item response model with gradual process change. *Psychometrika*, *73*, 65-87.

Hidalgo, M., & López-Pina, J. (2004). Differential item functioning detection and effect size: A comparison between logistic regression and Mantel-Haenszel procedures. Educational and Psychological Measurement, 64(6), 903-915

Holland, P. W., & Thayer, D. T. (1986). Differential item functioning and the Mantel-Haenszel procedure. *ETS Research Report Series, 1986*(2), i-24.

Holland, P. W., & Wainer, H. (2012). *Differential item functioning*. Routledge.

Klein, R., Fox, J., & van der Linden, W. (2009). A multivariate multilevel approach to the modeling of accuracy and speed of test takers. *Psychometrika*, *74*, 21-48.

Landrum, R., Carlson, H., & Manwaring, W. (2009). The relationship between time to complete a test and test performance. *Psychology Learning & Teaching*, 8(2), 53–56.

Lord, F., Novick, M., & Birnbaum, A. (1968). *Statistical theories of mental test scores.* Addison-Wesley.

Lu, Y., & Sireci, S. (2007). Validity issues in test speededness. *Educational Measurement: Issues and Practice*, *26*(4), 29-37.

Mandinach, E., Bridgeman, B., Cahalan-Laitusis, C., & Trapani, C. (2005). The impact of extended time on SAT® test performance. ETS Research Report Series, 2005(2), 1-35.

Mantel, J. G., & Haenszel, W. (2001). *Detecting differential item functioning using item response theory*. Cambridge, UK: Cambridge University Press.

Martinková, P., Drabinová, A., Liaw, Y.., Sanders, E., McFarland, J., & Price, R. (2017). Checking equity: Why differential item functioning analysis should be a routine part of developing conceptual assessments. *CBE—Life Sciences Education*, *16*(2), rm2.

Martone, A., & Sireci, S. G. (2009). Evaluating alignment between curriculum, assessments, and instruction, *Review of Educational Research*, 4, 1332–1361.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), Educational measurement (3rd ed., pp. 13–104). New York, NY: American Council on Education and Macmillan.

Oshima, T. (1994). The effect of speededness on parameter estimation in item response theory. *Journal of Educational measurement*, *31*(3), 200-219.

Penfield, R. D., & Camilli, G. (2006). 5 differential item functioning and item bias. *Handbook of Statistics*, 125–167. https://doi.org/10.1016/s0169-7161(06)26005-x

Pommerich, M. (2004). Developing computerized versions of paper-and-pencil tests: Mode effects for passage-based tests. *The Journal of Technology, Learning and Assessment*, *2*(6).

Schnipke, D., & Scrams, D. (2002). Exploring issues of examinee behavior: Insights gained from response-time analyses. *Computer-based testing: Building the foundation for future assessments*, *34*, 237-266.

Sireci, S. & Botha, S. (2020). Timing considerations in test development and administration. In Integrating timing considerations to improve testing practices (pp. 32-46). Routledge.

Sireci, S. G., & Faulkner-Bond, M. (2014). Validity evidence based on test content. Psicothema, 26, 100–107. doi: 10.7334/psicothema2013.256.

Suh, Y., Cho, S., & Wollack, J. (2012). A comparison of item calibration procedures in the presence of test speededness. *Journal of Educational Measurement, 49*(3), 285-311.

Talento-Miller, E., Guo, F., & Han, K. (2013). Examining test speededness by native language. *International Journal of Testing*, 13(2), 89-104.

van der Linden, W. (2006). A lognormal model for response times on test items. *Journal of Educational and Behavioral Statistics, 31*(2), 181-204.

Wadley, M., & Liljequist, L. (2013). The effect of extended test time for students with attention-deficit hyperactivity disorder. *Journal of Postsecondary Education and Disability, 26*(3), 263-271.

Wainer, H., & Thissen, D. (1996). How is reliability related to the quality of test scores? What is the effect of local dependence on reliability?. *Educational Measurement: Issues and Practice*, 15(1), 22–29.

Wang, X., & Sireci, S. G. (2013). *Investigating the relationship between item response time and cognitive level.* Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Wise, S., Kingsbury, G., & Webb, N. (2015). Evaluating content alignment in computerized adaptive testing. *Educational Measurement: Issues and Practice, 34*(4), 41-48.