بناء اختبار لقياس القدرة على التَّفكير المجرّد باستخدام نظريَّة الاستجابة للفقرة

إيمان أبو وردة و نضال الشريفين*

Doi: //10.47015/18.4.1

تاريخ تسلم البحث: 2020/10/22 **تاريخ قبوله: 2020/12/24**

ملخص: هدفت الدراسة إلى بناء اختبار لقياس القدرة على التفكير المجرّد باستخدام نظرية استجابة الفقرة، ولتحقيق هدف الدراسة، تم بناء الصورة الأولية للاختبار من 62 فقرة من نوع الاختيار من متعدد. وقد طبق الاختبار على عينة مكونة من 1015 طالبًا وطالبة، وتم التحقق من افتراض أحادية البعد والاستقلال الموضعي في البيانات، وبيّنت النتائج مطابقة 40 فقرة لافتراضات النموذج ثلاثي المعلّمات، كأحد نماذج نظرية الاستجابة للفقرة، وتضمن الاختبار بصورته النهائية 40 فقرة بخصائص سيكومترية مناسبة؛ إذ بلغت قيمة معامل الثبات النظري والتجريبي للمقياس 0.91 و0.89 على التوالي، كما تمتع الاختبار بدلالات متعددة للصدق. وأظهرت نتائج التحليل تمتع معالم الفقرات (الصعوبة، والتمييز، والتخمين) بمؤشرات جيدة ومقبولة في القياس النفسي. وأظهرت النتائج وجود لا تغير تكويني (شكلي)، ولا تغير متري، ولا تغير عددي في القياس باختلاف الجنس، وذلك بالإستناد إلى التحليل العاملي التوكيدي متعدد المجموعات، مما يؤكد صدق الاختبار. وجاءت نتائجة مؤكدة لما تم التوصل إليه في نظرية الاستجابة للفقرة.

(الكلمات المفتاحية: التفكير المجرد، نظرية الاستجابة للفقرة، البناء العاملي التوكيدي)

مقدمة: يُعد القياس النفسي جوهر علم النفس؛ إذْ يُعنى بقياس الظواهر والوظائف النفسية والعقلية المختلفة كالانتباة والإدراك والتفكير وغيرها الكثير من الظواهر، فيهتم بشرح العلاقة بين السمات الملاحظة وغير الملاحظة. وقد وجد علماء النفس المعرفيون فائدة كبيرة في القياس النفسي كحقل وأداة لقياس السلوك غير المُلاحظ. وبذلك أصبحت عملية القياس النفسي نوعًا من الاختبار يخضع له الفرد وتكون نتيجته تقويمًا لقدراته العقلية أو تقويمًا لأوجه شخصيته المختلفة. ولأن القياس والتقويم يُمثلان أساسًا قويًا في عملية التطوير والتجديد، فهما يُعدان بمثابة جهاز الكشف عن مواطن القوة والضعف في عملية التّعلّم والتّعليم، وإدخال التحسينات عليها وتطويرها وتجديدها. فهُما من المجالات المهمة في الحقل التربوي والنفسي التي لا يمكن الاستغناء عنها؛ وذلك لأهميتهما في اتّخاذ القرارات في مختلف الميادين، مما ينعكس على تقدم العلوم المعرفية وتطورها، الأمر الذي من شأنه أن يسهم في مواكبة التطورات السريعة في مختلف ميادين الحياة.

ويحظى التفكير باهتمام واسع من قبل العلماء والباحثين في الميادين المختلفة على اختلاف تخصصاتهم؛ ذلك أَنَّ التفكير يعكس تفرد الإنسان عن بقية المخلوقات، ويُمثل أرقى أشكال النشاط العقلي، كما أنَّه مؤشّر هام لتميّز الحضارات وتقدّم الدول التي تجتهد في استثمار طاقاتها لتطوير هذه القدرة لدى أفراد مجتمعها، الأمر الذي يجعل الإنسان قادرًا على تحسين ظروف حياته في المجالات المختلفة (Abu Jadu & Nofal, 2010). ويمكن وصف التفكير في حد ذاته بأنه تطبيق تسلسلي صريح لمجموعة من القواعد على مشكلة رسمية (Gilhooly, 2004).

Constructin of a Test to Measure the Abstract Reasoning Ability Using Item Response Theory

Eman Abuwardih and Nedal Al-Shraifin, *Yarmouk University, Jordan.*

Abstract: This study aimed at constructing a test to measure the ability of abstract reasoning by using item response theory. To achieve the objectives of the study, a preliminary multiple-choice version of 62 items was constructed. The scale was administered to a randomly selected sample of 1015 students. The un-idimensionlaity and local independence of the test were established by many different indices. The results indicated that 40 items fit the assumptions of the three-parameter logistic model with a theoretical and an empirical reliability coefficient of 0.91 and 0.89, respectively. The results of the analysis showed that the item parameters of (difficulty, discrimination and guessing) have good and acceptable indicators in psychometrics. The scale also has several validity indicators. The results showed that there are configural invariance, metric invariance and scalar invariance in the test measuring abstract reasoning ability according to gender, based on multi-group confirmatory factor analysis, This confirms the validity of the test and the results reached in the response item theory.

(**Keywords**: Abstract Reasoning, Item Response Theory, Confirmatory Factor Analysis)

إنّ اختبارات التفكير بشكل عام هي معيار شائع لرسم الخرائط أو الفرز أو الترتيب للاختيار بين الأفراد (Hunt, 2011) . لِهذا تُعدّ تربية التّفكير من المهمات العظيمة التي ينبغي للمؤسسات أنْ تُعنى بها، بدءاً بالأسرة وانتهاءً بالمؤسسات التربوية المتعددة. وللمدرسة دور كبير في تنمية التفكير؛ إذ إنّ التحصيل الدراسيّ على الأغلب هو المؤشر الوحيد المستخدم، والمعيار الأساس لمعرفة مدى تقدم الطالب أو تأَخُره في المقررات المدرسية، لهذا يجب أنْ يهتم التربويّون بالقدرة على التفكير المجرّد (Al-Otaibi, 2001) كمتنبئ آخر في تقدّم الطالب وتعلمه.

وقد ظهرت لاختبارات التفكير المجرّد عدة نظريات تفسره وتُحدد مكوناته؛ إذْ ترجع اختبارات قياس التفكير المجرّد إلى الأبحاث التي قام بها سبيرمان (Sperman) الذي خلص إلى وجود عامل عام وعوامل خاصة، وعرّف سبيرمان العامل العام بالقدرة الفطرية على إدراك وإيجاد العلاقات المشتركة (Newton & Bristoll, 2009).

* جامعة اليرموك، الأردن.

© حقوق الطبع محفوظة لجامعة اليرموك، إربد، الأردن، 2022.

إنّ التفكير المجرّد هو أحد أنواع التفكير المُركّب التي انبثقت من نظرية بياجيه للنمو المعرفي. وفي الوقت نفسه، تشكل مرحلة التفكير المجرد المرحلة الرابعة والأخيرة من مراحل النمو العقلي للفرد، ويُطلق على مرحلة التفكير المجرّد أحيانًا مرحلة التفكير الشكليّ أو العمليات الرمزية أو العمليات الرسمية، وفيها تحدث لِبنى الطفل تغيرات نوعية، مما يجعله ينتقل من التركيز على المحتوى إلى التركيز على شكل الفكرة نفسها. ويستطيع المراهق وضع الفروض واختبارها، ويستطيع التعامل مع المشكلات، ويطور استراتيجيات حلّها، ويفكر على نحو مجرّد، ويصل إلى النتائج المنطقية دون الرجوع إلى الأشياء الماديّة أو الخبرات المباشرة (Abdullah & Azzawi, 2019).

وبناءً على ذلك، فإنّ استخدام اختبار لقياس قدرة التّفكير المجرّد على نطاق واسع ربما يساعد في التنبؤ بقدرة الأفراد على حل المشكلات المرتبطة بمستقبلهم والكشف عن قدرتهم على اتّخاذ القرارات المهمة في حياتهمْ، كما يُساعدهم على تطوير قدرتهم على التفكير، وكذلك التخطيط للبرامج التعليمية وتطوير المناهج الدراسيّة بحيث تكون مناسبة للمستويات العمرية، ومراعية لعمليات النمو المعرفيّ للطّلبة، ومراعية للتسلسل في تنمية قدرات الأفراد وتفكيرهم. كذلك يستخدم اختبار قياس قدرة التفكير المجرّد على نطاق واسع لأهداف أكاديميّة وتوجيهيّة وإرشادية، ربما لأنه يُساهم في توجيه الأفراد إلى تَخَصصات وأعمال ومهن تتناسب مع قدراتهم؛ إذ يمكن تطبيقه للتمييز بين الطلبة ذوي القدرات العالية وذوي القدرات المتدنية، لمساعدتهم في وضع برامج وخطط مناسبه لهم، بحيث تُنمي جوانب القوة لديهم، وتُشخص وتُعالج جوانب الضعف. وحتى ينجح المتعلمون في تحقيق الهدف الأسمى لوجودهم في بيئة المدرسة، كان لا بُد من معرفة وتقصّي المعلومات عن المُتعلم، وعن نُموه العقلي والمعرفي والقدرات التي يمتلكها، وهذا لا يأتي الا بتوفير أداة تتصف بالصدق والثبات. ولذلك كانت الحاجة مُلحّة لبناء اختبار لقياس قدرة التفكير المجرّد. وهناك مستويان مختلفان لمفاهيم التفكير المجرّد، هما: مفهوم الأدوات والاستعلامات التي تعتمد على التجريد الإدراكي؛ لأن هذا النوع من التجريد يعتمد على العلاقات الموجودة بين المدركات الحسية والخبرة المستخلصة من العلاقات المُتضمنة فيها، ومفهوم الزمان الذي يعتمد على العلاقات بين الرموز البحتة وعلى زيادة في المعلومات والتفكير المجرّد (Stanger, cited in Slavin, 1996).

وقد أكد إنهلدر وبياجيه (Inhelder & Piaget, cited in Abdullah & Azzawi, 2019) على عدد من المهارات التي يجب أَنْ تتوافر لدى أصحاب التفكير المجرّد، مثل: الاستدلال الافتراضي، والاستنتاجي، والمنطقي، واقتراح الحلول، وحل المشكلات، وأيضًا الاستدلال التناسبي، والاستدلال التركيبي (Combinatorial reasoning)، وتحديد وضبط المتغيرات (Controlling variables)، والتعليل الارتباطي (Correlational reasoning)، والاستدلال الاحتمالي (Probabilistic reasoning). وقد أشار برايون (Bryon, 2008) إلى اختلاف أنواع الأسئلة في اختبارات التفكير المجرد، لكنها إجمالاً تقيس المبادئ الأساسية نفسها الآتية: التدوير، حيث يتم تدوير الشكل بزوايا مختلفة، والتغيير، حيث يتغير الشكل إلى شكل آخر ثم يتغير مرة أخرى، والاتساق، حيث يتم إجراء تغيير على الشكل ثم يتم تطبيقه باستمرار على الصور التالية، والاستبدال، حيث تستبدل بالشكل أشكال أخرى، والانتباه إلى التفاصيل، حيث يجب النظر بعنايةٍ في المخططات والانتباة إلى أدق التفاصيل. ويُعد مقياس التفكير المجرّد غير اللفظي مشابهًا لمصفوفات رافن المتقدّمة واختبار كاتل المتحرر من الثقافة، وهو اختبار يُستخدم لقياس التفكير غير اللفظي المستقل عن المعرفة اللفظية أو المكتسبة. ويتكون مقياس التفكير المجرّد من عدة أنواع مختلفة من الفقرات، تشمل المقارنات الشكلية والتسلسلات والأنماط والتجمعات، ويتم تقديم الفقرات في شكل سلسلة من الأشكال مع جزء مفقود. ثم يُطلب من المشارك اختيار الخيار الصحيح الأكثر منطقية من بين العديد من الخيارات المعروضة (Bautista, 2013; Hamel & Schmittmann, 2006).

ويُعد تفكير الراشدين من وجهة النظر النمائية (Developmental) نِتاجًا مُعقدًا لتاريخ طويل من الارتقاء يبدأ من فترة مبكرة جدًا من الحياة. وتتغير عمليات التفكير استجابة للمطالب التي تُفرض من خلال تفاعل طويل ومتنوع مع البيئة المادية والاجتماعية. والأمر الحاسم هو أن التفكير المجرّد يرتقي؛ بمعنى أنه يتغير بشكل منهجي كُلما تقدم الإنسان في العمر من الطفولة حتى الرشد (Solso, MacLin & MacLin, 2005). ويؤكد وادزورث (Wadsworth, cited in Abu Roman, 1991) أنه تنمو لدى المراهق أكثر العمليات المعرفية تطورًا، وتظهر لديه القدرة على التفكير المجرد، والتفكير بغير المحسوس، والتفكير فيما وراء الحاضر أو المعلومات، فيصبح الفرد أكثر قدرة على التعامل مع المثيرات الاكثر بُعدًا في الزمان والمكان، فتنمو لديه القدرة على بناء النظريات والأنظمة بمعناها الواسع. وقد أشارت دراسة ألتر وإيجان (Alter & Egan, 1997) إلى أنه بالرّغم من أنّ الفرد يدخل في مرحلة التفكير الرسميّ (Formal Operations) من عمر إحدى عشرة سنة، فإنّ هناك إخفاقًا لخريجي الجامعات في التعامل بكفاءة مع المشكلات التي تتطلّب مهارات التفكير المجرّد، حيث تتميز هذه المرحلة بقدرة الفرد على تبرير الأفكار المجردة الافتراضية والعكسيّة، ويتمكن من صياغة واختبار الفرضيات المتعددة المهتمة بربط السبب مع النتيجة، ويستطيع أيضًا في هذه المرحلة فصل وضبط المتغيرات. وقد وجد هازان وزازكيز (Hazzan & Zazkis, 2005) أنّ أخطاء الطلاب في حلّ المشكلات غالبًا ما تكون لقلة وصولهم لمستويات التجريد، وحدّد ثلاثة تفسيرات مختلفة لمستويات التجريد: مستوى التجريد وجودة العلاقات بين موضوع التّفكير والشّخص المفكّر، ومستوى التجريد باعتباره انعكاسًا لازدواجية العملية المستهدفة، ومستوى التجريد كدرجة لتعقيد مفهوم التفكير.

وقد بُذلت جهود حثيثة منذ عقود مضت لتطوير أدوات قياس للتفكير والعمليات العقلية، وأسفرت جهود الباحثين عن بناء أدوات لقياس التفكير المجرّد (الرسمي). فقد أجرى بيرني (Burney, 1974) دراسة هدفت إلى بناء اختبار لقياس التفكير المجرّد وتقنينه والكشف عن مرحلة البنية العقلية المعرفية التي وصل إليها أفراد عينة الدراسة. وتكونت عينة الدراسة من (128) طالبًا وطالبة، من المدارس الإعدادية وما فوق. اعتمد الباحث في دراسته على اختبار الاستدلالات المنطقية الذي يتكون من 21 مسألة غطت مجموعة من المفاهيم. وبيّنت النتائج عدم وجود فروق ذات دلالة إحصائية في البنية العقلية المعرفية بين الذكور والإناث.

وتوالت بعدها العديد من الدراسات، وكان من اهمها دراسة بوتيستا (Bautista, 2013) التي هدفت لتقييم الخصائص السيكومترية لمقياس التفكير المجرد. وتكونت عينة الدراسة من 6624 طالبًا و6653 طالبة من طلبة الصفوف التاسع والعاشر والحادي عشر والثاني عشر. قام الباحث بمقارنة نموذج راش بالنموذج ثنائي وثلاثي المعلمات، وتم التأكد مما إذا كان نموذج راش لقياس التفكير المجرّد ثابتًا عبر مستوى الصف والجنس. وأشارت النتائج إلى أنه لا توجد فروق ذات دلالة للتفاعل بين الصف والجنس على مقياس التفكير المجرد، وأشارت النتائج إلى وجود فقرة واحدة لا تُطابق النموذج، وأشارت أيضًا إلى أنّ الثبات كان متدنيًا عند المستويات العُليا للقدرة بالاعتماد على دالة معلومات الاختبار. وحقق النموذج ثنائي المعْلمات أعلى معلومات عند متوسطي القدرة. وهدفت دراسة باكير وبيسر (Bakir & Bicer, 2015) إلى تحديد مستويات التفكير الشكلي (المجرد) والتطور المعرفي لدى معلمي العلوم قبل الخدمة. اشتملت عينة الدراسة على (421) معلمًا ومعلمة ممن التحقوا ببرنامج تدريس العلوم، وتم تقييم المجموعة باستخدام اختبار التفكير المنطقي (GALT)، وأشارت النتائج إلى عدم وجود فروق دالة إحصائيًا في متوسطات درجات اختبار التفكير المنطقي لدى معلمي العلوم تعزى للجنس. وأجرى التميمي (Al-Tammi, 2015) دراسة حول تطوير اختبار القدرة على الاستدلال المجرّد لبول نيوتن على طلبة المرحلة الإعدادية باستعمال نموذج أُحادي المعلمة وفقًا لنظرية الاستجابة للفقرة. طبق الاختبار على عينة بلغت 500 طالب وطالبة من الصفوف (الرابع والخامس والسادس). وحُللت الفقرات باستخدام نموذج راش، وأشارت النتائج إلى تحقق أُحادية البُعد، وتمت مطابقة الفقرات للنموذج اعتمادًا على قيمة مربع كآي، واستبعدت 7 فقرات من الاختبار الأصلي البالغ عدد فقراته 68 فقرة ليصبح الاختبار بصورته النّهائية مكونًا من 61 فقرة، وبلغ معامل الثبات للاختبار 0.882.

وتزخر المكتبات بالعديد من أدوات القياس التي تقيس العديد من الخصائص النفسية، وكذلك القدرات، لكنها لم تعط ذلك القدر من الأهمية لقدرة التفكير المجرد. فتلاحظ قلة الدراسات العربية والأجنبية التي اهتمّت بذلك (Abdel Wahab, 2010)؛ فثمة حاجة مُلحة لبناء مقياس يتصف بخصائص القياس الموضوعي بحيث لا تتأثر خصائصه السيكومترية من صدق وثبات بخصائص مجموعة الأفراد التي تخضع له لقياس تلك القدرة. ونظرًا للتقدم المستمر في مجال القياس النفسي، وتزايد اهتمام الدول المتقدمة بضبط إجراءات وأساليب بناء المقاييس وتحليل فقراتها والتأكد من صدقها وثباتها، فقد دعت الضرورة إلى استخدام هذه النظريات وما ينبثق عنها من نماذج في محاولة جادة للتخلص من عيوب النظرية الكلاسيكية التي سيطرت على منهجيات القياس طوال القرن الماضي، وما زالت أغلب الدراسات والبحوث تعتمد عليها في بناء وتقنين المقاييس. لهذا هدف البحث الحالي إلى بناء اختبار لقياس القدرة على التفكير المجرّد باستخدام نظرية الاستجابة للفقرة.

إن الدراسات التي استخدمت مقاييس التفكير المجرد أو تلك التي سعت إلى بناء مقياس للتفكير المجرد اعتمدت في اختيار فقراتها على معايير منبثقة عن النظرية الكلاسيكية في القياس النفسي والتربوي، منها ما هو متعلق بإحصائيات الفقرة، ومنها ما هو متعلق بالخصائص السيكومترية، ولكن هذه النظرية تعرضت لجوانب نقد نظرًا لعجزها عن تفسير بعض القضايا في القياس النفسي، فظهرت نظرية الاستجابة للفقرة (Item Response Theory) كنظرية بديلة للنظرية الكلاسيكية، وقدمت حلولاً لمعظم المشكلات التي عجزت النظرية الكلاسيكية عن تفسيرها. لهذا تمّ الاعتماد على معايير نظرية الاستجابة للفقرة في بناء اختبار القدرة على التفكير المجرد، لاهتمامها بالربط بين استجابة المفحوص على الفقرة وقدرته، حيث تفترض هذه النظرية أنه يمكن التنبؤ بأداء الأفراد أو يمكن تفسير أدائهم على فقرة من فقرات الاختبار في ضوء خاصية مميزة لهذا الأداء تسمى السمة. ونظرًا لصعوبة ملاحظة السمات، فإنه يجب تقديرها، ولهذا يطلق عليها السمات الكامنة. وتقوم هذه النظرية على مجموعة من الافتراضات الأساسية: أحادية البعد (Uni-dimensionality)، والاستقلال الموضعي (Local Independence)، ومنحنى خصائص الفقرة (Item Characteristic Curve). كذلك تنبثق عنها مجموعة من النماذج التي استخدمت في بناء المقاييس والاختبارات والتي يمكن من خلالها الحصول على مؤشرات إحصائية للفقرة لا تعتمد على خصائص المفحوصين وتقديرات المفحوصين ولا تعتمد على صعوبة فقرات المقياس (Crock & Algina, 1986).

مشكلة الدراسة وأسئلتها

هدفت الدراسة إلى بناء اختبار لقياس القدرة على التفكير المجرّد وفق نظرية الاستجابة للفقرة؛ إذ تمثلت مشكلة هذه الدراسة في الحاجة إلى وجود أداة تتمتع بخصائص سيكومترية مقبولة، للتعرف إلى مستوى قدرة التفكير المجرّد لدى طلبة الصف العاشر الأساسي، بحيث تتوفر فيه خصائص القياس الموضوعي. ففي حدود اطلاع الباحثين، لم تتوفر في الأدب التربوي أداة مقننة لقياس القدرة على التفكير المجرد رغم أهميتها والحاجة الماسة إليها في تقييم مستوى التفكير المجرّد للأفراد. وحتى تكون الأداة دقيقة، لا

بد من بنائها وفق أسس وقواعد دقيقة. لذا قام الباحثان بالاعتماد على نظرية الاستجابة للفقرة، وتحديدًا على النموذج ثلاثي المعلمات في بناء هذه الأداة، مستفيدين من النظرية التقليدية في القياس التي عجزت عن تفسير بعض القضايا المهمة في القياس النفسي (Hambleton, Swaminathan & Rogers, 1991)، وبشكل محدد، فقد سعت الدراسة إلى الإجابة عن الأسئلة الآتية:

1. ما درجة مطابقة الاستجابات عن فقرات الاختبار للنموذج ثلاثي المعلمات لنظرية الاستجابة للفقرة؟
2. هل يختلف البناء العاملي المقترح لاختبار قياس قدرة التفكير المجرّد بصورته النِّهائية باختلاف جنس الطالب؟
3. ما دلالات الصدق والثبات (الخصائص السيكومترية) للاختبار؟
4. ما قيم تقديرات معالم الفقرات على الاختبار وفقًا للنموذج ثلاثي المعلمات في نظرية الاستجابة للفقرة؟

أهمية الدراسة

تنبثق أهمية هذه الدراسة من سعيها إلى الكشف عن قدرة التفكير المجرّد لدى طلبة الصف العاشر الأساسي، والتوّصل إلى أداة قياس موضوعية لقياس سمة التفكير المجرّد بشكل دقيق، كما يتوقع أن تكون للاختبار قيمة تربوية خاصة بإضافة أداة مقننة ودقيقة وذات خصائص سيكومترية مقبولة إلى المكتبة العربية في القياس النفسي. وستكون أداة قياس سهلة وسريعة التطبيق تستخدمها المؤسسات المعنيّة لقياس القدرة على التفكير المجرّد وفق نظرية الاستجابة للفقرة، حيث فقرات الاختبار متحررة من الأفراد وكذلك عينة الأفراد متحرره من عينة الفقرات. وفي ضوء نتائج القياس، يتم التخطيط لاستخدام هذا الاختبار كمتنبئ آخر بالاضافة إلى التحصيل الدراسي لغايات تصنيف الطلبة في التخصصات المختلفة والفروع الأكاديمية والمهنية، لا سيما التخصصات العلمية منها للمرحلة الثانوية. ويمكن التوسع في استخدامة في الجامعات بدلاً من الاعتماد على معدل الطالب في الثانوية العامة لقبوله في التخصصات المختلفة؛ لا سيما العلمية والهندسية. ويمكن الاعتماد عليه أيضًا في قبول طلبة الدراسات العليا في التخصصات المختلفة في الجامعات بدل الاعتماد على معدلات الطلبة كمتنبئ في قبولهم في البرامج المختلفة، وكذلك في قبول الأفراد في الوظائف الإدارية القيادية والحساسة؛ إذ يمكّن صناع القرار من تقييم الإمكانات الفكرية العامة ويعطيهم القدرة على تقييم المفاهيم الجديدة للافراد. وفي ضوء نتائج القياس أيضًا، يتم التخطيط لوضع برامج تدريبية لتنمية القدرة على التفكير المجرّد بما يتلاءم مع مستويات قدرات الطلبة، وإعدادهم للمستقبل بما ينعكس إيجابًا على حياتهم وعلى المجتمع بشكل عام.

محددات الدراسة

- اقتصرت الدراسة على عينة من طلبة الصف العاشر الأساسي في المدارس الحكومية، وبالتالي فان إمكانية تعميم النتائج تتحدد بمدى تمثيل هذه العينة لطلبة المدارس.
- المفاهيم والمصطلحات المستخدمة في هذه الدراسة محددة في التعريفات الإجرائية، وبالتالي فإن إمكانية تعميم النتائج تتحدد في ضوء هذه التعريفات.
- اقتصر اختبار قياس القدرة على التفكير المجرّد على الجانب غير اللفظي.
- اقتصر أسلوب التحليل على النموذج ثلاثي المعلمات كأحد نماذج نظرية الاستجابة للفقرة.

التعريفات الاصطلاحية والإجرائية

القدرة على التفكير المجرّد: هي القدرة على حلّ المشكلات باستخدام العمليات المنطقية (Piaget, 1970)، وتتمثل في القدرة أو الوظيفة المعرفية المتصلة بحل مشكلات معتمدة على تجريد خواص أو كيفيات أو مفاهيم أو مبادئ معينة من بين عدد من الجزئيات بينها درجة من التشابه أو الاختلاف (Solso, MacLin & MacLin, 2005). وتعتبر عملية عقلية تهدف إلى استخلاص المعاني المجردة للأشياء واستنباط النتائج والعلاقات بواسطة التفكير الافتراضي من خلال التعاميم والرموز، والقدرة على وضع الافتراضات، والتأكد من صحتها (Al-Atoum, 2004).

وتُقاس إجرائيًا من خلال الدرجة التي يحصل عليها الفرد على الاختبار الذي أُعد لهذه الدراسة وتكوّن من المجالات الآتية:

- **إدراك العلاقات والخصائص العامة:** هو تحليل الأشكال إلى أجزاء وإدراك القاعدة التي تربط هذه الأجزاء، وهذا يتطلب من المفحوص تدوير الاشكال. ويقاس بأسئلة المربع المفقود، وهو نمط من الأسئلة تظهر فيه مصفوفة ثلاثية تحتوي على أشكال ورسومات متسلسلة بشكل منطقي، ينقصها المربع التاسع، والمطلوب من المفحوص اختيار بديل واحد صحيح من بين أربعة بدائل بناءً على نمط او قاعدة أو عدة قواعد يكتشفها بنفسه.
- **المنطق الإدراكي:** يوضح كيف تتطور الفكرة من أجل معرفة ما سيحدث، ويهدف لقياس مدى جودة القرار المنطقي الذي يُتخذ بناءً على المعلومات المرئية. ويقاس من خلال أسئلة إكمال التسلسل، وهو نمط من الأسئلة تظهر فيه سلسلة من الصور أو الخطوط أو الرسوم البيانية التي تقدم، وتكون مهمة المفحوص رؤية كيف تسير معًا، ثم تحديد الشكل التالي في السلسلة بحيث تتبع الأشكال نمطًا منطقيًا.

- **الانتباة إلى التفاصيل**: ويُقاس من خلال أسئلة الشكل المخالف، وهو نمط من الأسئلة تظهر فيه سلسلة من الصور التي تحتوي جميعها على ميزة مشتركة باستثناء صورة واحدة، يتم عرضها على المفحوص ويجب عليه تحديد "الشكل المخالف" لبقية الصور.

- **الاستنتاج الإداركي:** هو تكوين أفكار أو مبادئ تربط الأشياء ببعضها البعض مع استبعاد البعض الآخر، ومعرفة ما هي صلة المعلومات ببعضها البعض. ويقيس مدى جودة الاستنتاج من المعلومات المرئية التي قُدمت. ويقاس من خلال أسئلة التماثل، وهو نمط من الأسئلة التي تبيّن كيفية ارتباط صورتين بعضهما ببعض، وما على المفحوص الا أنْ يرى كيف ترتبط الصورة التالية بأحد البدائل الأربعة المقدمة.

الطريقة والإجراءات

يتناول هذا الجزء من الدراسة وصفًا لأفراد الدراسة، وطريقة اختيار أفراد عينتها، والإجراءات التي اتبعت لبناء اختبار قياس القدرة على التفكير المجرّد وفق النموذج ثلاثي المعلمات في نظرية الاستجابة للفقرة، وذلك على النحو الآتي:

أفراد الدراسة

بلغ حجم عينة الدراسة (1015) طالبًا وطالبة من طلبة الصف العاشر الأساسي موزعين إلى (329) طالبًا و(686) طالبة من مدارس الذكور والإناث في مديرية تربية وتعليم محافظة عجلون، وقد تم اختيارها بالطريقة العشوائية الطبقية تبعًا لمتغير جنس المدرسة، ومن هذه المدارس تم اختيار شعب بالطريقة العشوائية العنقودية؛ لأن وحدة الاختيار هي المجموعة ممثلة بالشعبة.

إجراءات بناء الاختبار

الخطوة الأولى: تحديد الغرض من الاختبار: هدف الباحثان من دراستهما إلى بناء اختبار لقياس قدرة التفكير المجرّد وفق نظرية الاستجابة للفقرة، يساعد الطلبة على التخطيط لمستقبلهم، وعلى اختيار مجال تخصصهم ومجال دراستهم بما يتناسب مع قدراتهم ورغباتهم، وذلك من خلال الكشف عن الجوانب التي يجب أنْ تتوافر لدى الطلبة، وبيان الجوانب التي تحتاج إلى تطوير.

الخطوة الثانية: تحديد أبعاد المقياس: استفاد الباحثان من الإطار النظري الذي وضعه مصممو أدوات قياس التفكير بشكل عام والتفكير المجرّد بشكل خاص في تحديد مفهوم التفكير المجرد، ومكوناته، وأبعاده. ومن الدراسات التي تناولت موضوع التفكير المجرّد واستفاد منها الباحثان (Drager, 2014; Bakir & Bicer, 2015; Groves, 2015; Al-Azawi, 2016) في اشتقاق الأبعاد التي تكوّن بمجملها التفكير المجرد، وهي: إدراك العلاقات والخصائص العامة، والمنطق الإدراكي، والانتباه إلى التفاصيل، والاستنتاج الإدراكي.

الخطوة الثالثة: كتابة فقرات المقياس: بالاستفادة من الدراسات الأجنبية والعربية التي قامت ببناء مقاييس التفكير المجرّد أو التي استخدمت مقاييس التفكير المجرّد، مثل (Al-Tammi, 2015; Newton & Bristoll, 2009; Bautista, 2013; Helland, 2016; Chan, 2018; Nordstrom, Rendel & Tavares, 2015) والمواقع الإلكترونية الخاصة بالاختبارات النفسية وبالتفكير المجرّد (Wikijob, Psytech, Test PrepPack, Practice aptitude test)، تمكن الباحثان من كتابة 120 فقرة تغطي أبعاد المقياس الأربعة. وقد راعى الباحثان عند كتابة الفقرات أن تغطي الفقرات المجالات الرئيسة للتفكير المجرّد من نوع الاختيار من متعدد بأربعة بدائل، ولكل فقرة بديل واحد صحيح، وخُصصت لكل نمط اسئلة 30 فقرة.

الخطوة الرابعة: تحكيم الاختبار: تم عرض فقرات المقياس بصورته الأولية المكونة من (120) فقرة على هيئة تَّحكيم، تكونت من 16 مُحمكًا من أصحاب الخبرة والاختصاص من أساتذة جامعات ومشرفين مُتخصصين في علم النفس التربوي، والقياس والتقويم، وطُلب من كل مُحكّم بيان رأيه في الفقرات والمجالات من حيثُ: انسجام المجال مع التعريف، وملاءمة تعليمات الاختبار، وإمكانية حذف أي فقرة غير ملائمة للمجال الذي تنتمي إليه أو صعوبتها للمستوى العمري الذي تقدم إليه، ووضوح الفقرات، ومدى اتساق الفقرات وسلامتها اللغوية، وملاءمة بدائل الفقرة الواحدة، أو أية ملاحظة يرونها مناسبة. وفي ضوء ملاحظات المحكمين، تم تغيير وتعديل الكلمات غير الواضحة التي اتصفت بالغموض وعدم الوضوح، وتم تصويب الأخطاء اللغوية، وتمَّ الإبقاء على الفقرات التي اتفق عليها المحكّمون بنسبة 80% فما فوق، كذلك تم تبديل بعض الفقرات من فقرات ملونة إلى فقرات غير ملونة بناءً على ملاحظاتهم، وضبط متغير الصف باختيار طلبة الصف العاشر فقط لإجراء وتطبيق الاختبار عليهم. وبناءً على ذلك، تم حذف 15 فقرة ليصبح عدد فقرات المقياس 105 فقرات شُكلت الصورة الأولية المعدّلة للمقياس.

الخطوة الخامسة: التجريب الأولي للاختبار: تم تطبيق الصورة الأولية المعدّلة للمقياس، بواقع جلستين بلغ وقت كل جلسة منهما 45 دقيقة، وأخذ الطالب استراحة لمدة 15 دقيقة بين الجلستين، وبلغ عدد الطلبة 302 من الطلاب والطالبات، وانحصر الغرض من التجريب الأوّلي في: تحديد الزمن الذي يستغرقه الطلبة في الإجابة عن الاختبار، والكشف عن الفقرات الغامضه، ودراسة الخصائص السيكومترية للفقرات من أجل تحديد الفقرات المناسبة منها، والتعرف إلى مستوى صعوبة الفقرات وقدرتها التمييزية، ومعرفة مدى وضوح تعليمات الاختبار للطلبة. وبعد التجريب الأوّلي، حُدَّد الزمن المناسب للاختبار بـ 65 دقيقة، وتمت دراسة معاملات التمييز والصعوبة للفقرات، وحُذفت الفقرات ذات التمييز المنخفض (أقل من 0.20)، وتم الابقاء على الفقرات التي تراوحت معاملات الصعوبة

لها بين (0.22 و0.72) (Odeh, 2015). وتَكوّن الاختبار في صورته النِّهائية من 62 فقرة.

الخطوة السادسة: تطبيق الاختبار بصورتة النهائية: تمت عملية التطبيق للاختبار بصورتة النِهائية على عينة بلغت 1015 طالبًا وطالبة، وتم اتباع الإجراءت اللازمة لضمان تطبيق الاختبار بالصورة التي يجب أن يُطبق بها، وأُدخلت جميع إجابات الطلبة إلى ذاكرة الحاسوب، وتم استخراج علامات الطلبة بنهاية عظمى مقدارها 62 ونهاية صغرى مقدارها 0.

الخطوة السابعة: اختيار النموذج: تم اختيار النموذج ثلاثي المعْلمات (Three- parameter Logistic Model (PLM_3)؛ إذْ يعد النموذج اللوجستي ثلاثي المعْلمات المستخدم لأغراض تحليل بيانات هذه الدراسة أقل النماذج أحادية البعد تشددًا. فهو يسمح بأن تختلف فقرات الاختبار في صعوبتها وتمييزها، وهو ما يلاحظ عادة في بناء الاختبارات؛ إذْ من الصعب إيجاد مجموعة من الفقرات التي تميز بدرجة واحدة بين مستويات السمة أو القدرة التي يقيسها اختبار معين، وهذا هو الافتراض الذي اعتمده النموذج أحادي المعلمة (One-parameter Logistic Model)، كذلك يفترض تأثر الإجابات بعامل التخمين وهو المعلم الثالث الذي تميز به عن النموذج ثنائي المعْلمات (Two – parameter Logistic Model)، الذي أضافه لورد وأطلق عليه معلم الخط التقاربي الأدنى (Lower Asymptote Line) أو معْلمة التخمين (Guessing Parameter)، وهذه المعْلمة تحدد احتمال أن يجيب فرد يفترض أن مستوى قدرته متدنٍ جدًا ومع هذا يجيب إجابة صحيحة عن فقرات الاختبار عن طريق التخمين. والحقيقة أن إضافة هذه المعْلمة الثالثة هي بمثابة تعويض عن عدم مطابقة المنحنيات المميزة للفقرات عند النهاية السفلى لمتصل القدرة؛ نتيجة تأثر إجابات هؤلاء الأفراد بعامل التمييز(Lord, 1980; Hambleton, 1989). ويوصف النموذج اللوجستي ثلاثي المعْلمات بالمعادلة الرياضية الآتية:

$$p_i(\theta) = c_i + \frac{1-c_i}{1+e^{Dai(\theta-bi)}} (i = 1,2,....n\ )$$

حيث ($p_i(\theta)$: احتمال أن يجيب المفحوص ذو القدرة (θ) عن الفقرة (i) إجابة صحيحة،D: ثابت التدريج ويساوي تقريبًا 1.7، b_i: معْلمة الصعوبة للفقرة (i) وهي النقطة التي تقع على متصل القدرة عندما يكون ميل منحنى خصائص الفقرة أكبر ما يمكن، a_i: معْلمة تمييز الفقرة، وهي ميل منحنى خصائص الفقرة عند انعطاف المنحنى، c_i: خط التقارب السفلي لمنحنى خصائص الفقرة الذي يمثل احتمالية إجابة المفحوص بأقل قدرة إجابة صحيحة على الفقرة (معْلمة التخمين للفقرة).

المعالجة الإحصائية

للإجابة عن أسئلة الدراسة، تمت المعالجات الإحصائية للبيانات على النحو الآتي:

أولاً: التحقق من مطابقة البيانات لافتراضات نظرية الاستجابة للفقرة

للإجابة عن سؤال الدراسة الأول، تمّ أولاً التحقق من مطابقة البيانات للافتراضات الأساسية لنظرية الاستجابة للفقرة أُحادية البُعد باستخدام برمجية SPSS، والاستقلال الموضعي باستخدام برمجية IRTPro v4.2. وفيما يأتي توضيح لذلك:

تمّ اختبار افتراض أحادية البُعد للبيانات المتحققة على المقياس، واعتمد الباحثان على عدد من المؤشرات التي أشار إليها هتي (Hattie, 1985) للدلالة على أُحادية البُعد. ومن هذه المؤشرات:

- حساب معاملات الارتباط بين الاستجابات على الفقرات، والاستجابات على المقياس الكلي، بعد حذف الفقرة نفسها من المقياس (Correcte-item Total Correlation)، حيث تراوحت قيم معاملات الارتباط بين (0.23-0.72) بمتوسط حسابي بلغ (0.56). وكانت جميعها موجبة، وكانت جميعها ذات دلالة إحصائية عند مستوى الدلالة ($\alpha = 0.01$). وهذا يدل على أن الفقرات تتشارك في قياس بُعد واحد تعبر عنه العلامة الكلية التي تقيس التفكير المجرد. ويمكن اعتبار هذه النتيجة مؤشرًا على أُحادية البُعد للاختبار.
- تم استخدام أسلوب تحليل المكونات الأساسية (Principal Component Analysis)، وذلك لاستجابات الطلبة على فقرات اختبار التفكير المجرد. وقد تم إيجاد قيم الجذور الكامنة (Eigenvalues) ونسب التباين المفسّر (Explained Variance) لكل عامل من العوامل، وأُجري التدوير المائل (Oblique Rotation) للعوامل، وكانت قيمة الجذر الكامن لها أكبر من واحد، وكان عددها (4) عوامل فسرت (43.63%) من التباين الكلي. ويتضح ذلك من الجدول (1) الذي يبين قيم الجذور الكامنة، ونسبة التباين المفسر لكل عامل من العوامل، وكذلك نسبة التباين المفسّر التراكمي المقابلة لكل عامل من العوامل.

الجدول (1)

قيم الجذر الكامن ونسبة التباين المفسر والتباين المفسر التراكمي لكل عامل من العوامل.

	الجذور الكامنة								
	1	2	3	4	5	6	7	8	9
قيمة الجذر الكامن	20.19	3.3	2.08	1.49	0.99	0.95	0.85	0.77	0.77
نسبة التباين المُفسر	32.56	5.32	3.35	2.4	1.6	1.52	1.37	1.25	1.24
نسبة التباين المُفسر التراكمي	32.56	37.88	41.23	43.63	45.23	46.75	48.12	49.37	50.61

يتبين من الجدول (1) أن قيمة الجذر الكامن للعامل الأول 20.19، وهي قيمة مرتفعة إذا ما قورنت بقيم الجذور الكامنه لبقية العوامل، وتفسر 32.56% من تباين درجات الطلبة على الصورة الأولية لاختبار التفكير المجرّد قبل حذف الأفراد غير المطابقين والفقرات غير المطابقة للنموذج ثلاثي المعْلمات، وهذا يتماشى مع ما اقترحه ريكيس (Reckase, 1979) من أنه إذ أمكن أن يفسر العامل الأول 20% من التباين المفسر على الأقل، وكانت نسبة الجذر الكامن للعمل الأول إلى الجذر الكامن للعامل الثاني كبيرة وتزيد على 2. وهذا متحقق في هذه الدراسة، فهذا يعد مؤشرًا آخر على أحادية البعد لاختبار التفكير المجرّد المستخدم في هذه الدراسة وفق ما أشار إليه لورد (Lord, 1980). كذلك أظهرت نتائج عملية قسمة حاصل طرح الجذر الكامن للعامل الثاني من الجذر الكامن للعامل الأول على حاصل طرح الجذر الكامن للعامل الثالث من الجذر الكامن للعامل الثاني قيمة كبيرة تزيد على 14، مما يشير إلى تحقق افتراض أحادية البعد كمؤشر ثالث (Hattie, 1985). وللتحقق من افتراض الاستقلال الموضعي، تم استخدام برنامج IRTpro v4.2 لحساب قيمة X^2 للاستقلال الموضعي وتحويلها إلى قيم معيارية لكل زوج مُتشكل من أيّ فقرتين؛ فإذا كانت قيمة X^2 للاستقلال الموضعي أقل من 5، فمعنى ذلك أنّ زوج الفقرتين يُحقّق الاستقلال الموضعي، وإذا كانت قيمتة أكبر من أو تُساوي 5، فهذا يُعني أنّ زوج الفقرات لا يحقق الاستقلال الموضعي. ثم تم رصد التكرارات والنسب المئوية لحالة الاستقلال الموضعي لكل زوج من أزواج الفقرات، كما يظهرها الجدول (2) (Cai, Du Toit & Thissen, 2011).

الجدول (2)

التكرارات والنسب المئوية لحالة الاستقلال الموضعي لفقرات اختبار التفكير المجرد

حالة التبعية الموضعية	التكرار	النسبة المئوية
غير مستقلة	183	%9.68
مستقلة	1708	%90.32
الكلي	1891	%100

يلاحظ من الجدول (2) أَنّ الاستقلال الموضعي مُتحقّق لفقرات الاختبار، حيث بلغت النسبة المئوية للحالات التي تحققت لها الاستقلالية 90.32%؛ وهي بذلك تجاوزت نسبة 80% التي تعد النسبة الدنيا لتحقق فرض الاستقلال الموضعي كما حددها الأدب التربوي. كذلك تم التأكد من عدم تأثير عامل الوقت المخصص للإجابة على تمكن الطالب من الوصول إلى الإجابة الصحيحة على المقياس.

ثانيًا: مطابقة الأفراد والفقرات للنموذج ثلاثي المعْلمات في نظرية الاستجابة للفقرة

تم استخدام برمجية Bilog-Mg3 لمُطابقة الأفراد، وتم استخدام مؤشر مطابقة الأفراد L_z من خلال برمجية Pre-Fit وباستخدام برمجية R ولمطابقة الفقرات على النموذج اللوجستي ثلاثي المعْلمات، وذلك باستخدام اختبار مربع كآي عند مستوى الدلالة ($\alpha = 0.01$).

التحقق من البنية العاملية للمقياس بصورته النّهائية باختلاف جنس الطالب

للإجابة عن سؤال الدراسة الثاني، تم التحقق من تكافؤ القياس للصورة النهائية لاختبار التفكير المجرّد المكون من 40 فقرة، باستخدام التحليل العاملي التوكيدي متعدد المجموعات Multiple Groups Confirmatory Factor Analysis (MGCFA). وقد تم إجراء جميع التحليلات الإحصائية باستخدام برنامج (Muthén & Muthén, 2017 MPLUS) وتم الحكم عليها من خلال عدة مؤشرات.

التحقق من دلالات الصدق والثبات للاختبار (الخصائص السيكومترية)

للإجابة عن سؤال الدراسة الثالث؛ تم ذلك من خلال عدة مؤشرات كما يلي:

أولاً: مؤشرات الصدق: تم التحقق من الصدق الظاهري من خلال عرض التعريف الذي تبناه الباحثان والمجالات التي تتفق مع التعريف حسب الأدب النظري الذي انطلق منه الباحثان، بالإضافة إلى فقرات الاختبار، على ذوي الاختصاص والخبرة. كما تم التحقق من صدق البناء Construct Validity باستخدام برمجية SPSS لمقارنة المتوسطات الحسابية للعلامات لذوي التحصيل الدراسي المتدني والمتفوقين دراسيًا من خلال اختبار t لعينتين مستقلتين، كما تم

استخدام برمجية SPSS لاستخراج دلالة الفروق في الأداء على الاختبار بين الجنسين، وذلك عن طريق اختبار t.

ثانيًا: مؤشرات الثبات: باستخدام برمجيه Bilog-Mg3، تم تقدير ثبات الاختبار النظري (Theoretical Reliability) والثبات التجريبي (الإمبريقي) (Empirical Reliability) للاختبار بصورته النِّهائية وفق النموذج ثلاثي المعْلمات، وباستخدام برمجية SPSS تم تقدير ثبات الاتساق الداخلي (KR-20) للاختبار بصورته النِّهائية.

تقدير معالم الفقرات للنموذج ثلاثي المعْلمات في نظرية الاستجابة للفقرة

للإجابة عن سؤال الدراسة الرابع، تم استخدام برمجية Bilog-Mg3 لتقدير معالم الفقرات باستخدام الأرجحية العظمى الهامشية (Maximum Likelihood Estimaion(ML))، كما تم حساب المؤشرات الإحصائية الوصفية لمعْلمة الفقرة وفقًا للنموذج ثلاثي المعْلمات.

نتائج الدراسة ومناقشتها

النتائج المتعلقة بسؤال الدراسة الأول، والذي يَنص على: *"ما درجة مطابقة الاستجابات عن فقرات الاختبار للنموذج الثلاثي لنظرية الاستجابة للفقرة؟ ومناقشتها"*

أظهرت نتائج برنامج R من خلال برمجية Per-fit وباستخدام مؤشر مطابقة الأفراد L_z لمطابقة الأفراد وجود 64 فردًا غير مطابقين للنموذج ثلاثي المعْلمات عند مستوى الدلالة (α=0.05)، وبالتالي بلغ حجم العينة النِّهائية التي أجريت عليها التحليلات الإحصائية 951 وفقًا للنموذج ثلاثي المعْلمات، كما أظهرت التحليلات عدم مُطابقة 22 فقرة للنموذج ثلاثي المعْلمات كما يظهرها الجدول (3).

الجدول (3)

قيم الاحصائي χ^2 لحسن المطابقة ودرجة الحرية والدلالة الإحصائية لمُطابقة فقرات اختبار التفكير المجرّد وفق النموذج ثلاثي المعلمات

رقم الفقرة	χ^2 لحسن المطابقة	درجة الحرية	الدلالة	رقم الفقرة	χ^2 لحسن المطابقة	درجة الحرية	الدلالة	رقم الفقرة	χ^2 لحسن المطابقة	درجة الحرية	الدلالة
1	0.001	6	39.3	21	0.56	6	4.9	42	0.11	6	10.3
2	0.18	6	8.8	22	0.76	6	3.4	43	0.17	6	9
3	0.001	6	109.6	23	0.38	6	6.4	44	0.98	6	1.2
4	0.001	6	43.7	24	0.08	5	10	45	0.001	6	35
5	0.06	5	10.4	25	0.16	5	8	46	0.96	5	1
6	0.16	6	9.2	26	0.001	6	57.7	47	0.66	6	4.1
7	0.001	6	27.3	27	0.67	6	4.1	48	0.99	6	1
8	0.001	6	34.1	28	0.38	6	6.4	49	0.001	6	22.2
9	0.29	6	7.3	29	0.06	6	12.1	50	0.12	6	10.1
10	0.001	6	39.5	30	0.41	6	6.1	51	0.001	6	23.7
11	0.34	6	6.8	31	0.17	5	7.8	52	0.78	6	3.2
12	0.47	6	5.6	32	0.001	5	27.7	53	0.41	6	6.1
13	0.001	6	49.5	33	0.01	6	17.2	54	0.001	6	25.7
14	0.001	6	35.8	34	0.08	5	10	55	0.18	6	8.8
15	0.001	6	50.2	35	0.24	6	8	56	0.30	5	6.1
16	0.001	6	21	36	0.001	6	20.7	57	0.001	6	21.7
17	0.11	6	10.2	37	0.57	6	4.8	58	0.29	6	7.4
18	0.001	5	46.4	38	0.06	5	10.4	59	0.08	6	11.4
19	0.001	6	24.4	39	0.66	6	4.1	60	0.11	6	10.2
20	0.06	6	12.1	40	0.96	6	1.5	61	0.38	6	6.4
				41	0.51	5	4.3	62	0.001	6	23.8

بناءً على النتائج التي اظهرها الجدول (3)، فقد تم حذف الفقرات غير المطابقة. وعليه، فإن الصورة النِّهائية لاختبار التفكير المجرّد تكونت من 40 فقرة.

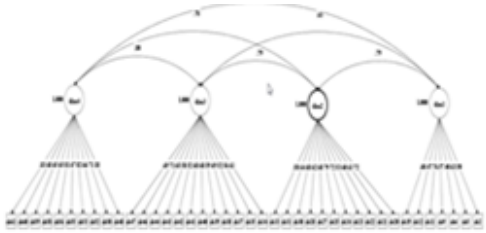
النتائج المتعلقة بسؤال الدراسة الثاني، والذي يَنص على: **"***هل يختلف البناء العاملي المقترح لاختبار قياس قدرة التفكير المجرّد بصورته النِّهائية باختلاف جنس الطالب؟* "*ومناقشتها*"

أظهرت نتائج سؤال الدراسة الثاني تحقق تكافؤ القياس بين الذكور والإناث للصورة النهائية لاختبار التفكير المجرّد المكّون من 40 فقرة، وذلك من خلال استخدام التحليل العاملي التوكيدي متعدد المجموعات (MGCFA). وقد توصلت الدراسة الحالية إلى هذه النتيجة من خلال مدخل مكون من ثلاث خطوات. تضمنت الخطوة الأولى إيجاد أفضل نموذج قياس يتطابق مع بيانات

الدراسة. ولتحقيق ذلك، تمت مقارنة أربعة نماذج قياس: النموذج أُحادي العامل (Uni-factor Model)، الشكل رقم (1)، حيث تتشبع جميع الفقرات على عامل واحد، ونموذج من الدرجة الأولى مُكون من أربعة عوامل مرتبطة (First-order Factor Model). الشكل رقم (2)، ونموذج من الدرجة الثانية (Second-order Factor Model)، الشكل (3)، حيث تتشبع العوامل الأربعة من الدرجة الأولى على عامل من الدرجة الثانية، ونموذج ثنائي العوامل (Bi-factor Model)، الشكل (4) حيث تتشبع جميع الفقرات على عامل عام وفي الوقت نفسه تتشبع كل مجموعة من الفقرات على عامل من العوامل الأربعة. ويبين الجدول (4) نتائج مؤشرات المطابقة بين النماذج الأربعة.

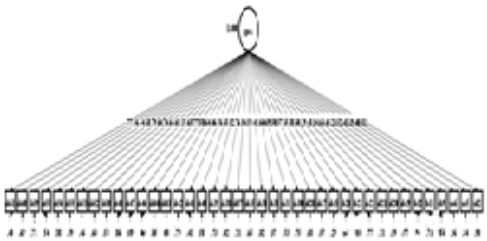
الشكل (1)

نموذج أحادي العامل (Uni-Factor Model)



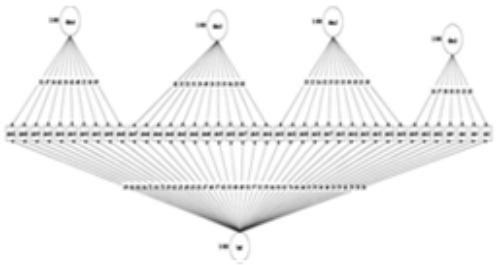
الشكل (2)

نموذج من الدرجة الأولى (First-order Factor Model)



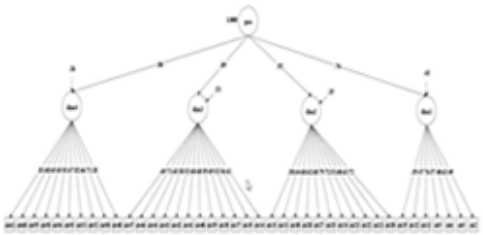
الشكل (3)

نموذج من الدرجة الثانية (Second-order Factor Model)



الشكل (4)

نموذج ثنائي العوامل (Bi-factor Model)



الجدول (4)

مؤشرات جودة مطابقة نماذج قياس الصورة النهائية (40 فقرة) لاختبار التفكير المجرد .

مؤشرات جودة المطابقة							
النموذج	مربع كآي	درجة الحرية df	نسبة χ^2 /df لحسن المطابقة	مؤشر المطابقة المقارن CFI	مؤشر توكر-لويس TLI	الجذر التربيعي لمربعات أخطاء التقريب RMSEA	الجذر المعياري لمتوسطات البواقي SRMR
أحادي العامل	2398.85	740	3.24	0.91	0.91	0.049	0.045
الدرجة الأولى	1910.43	734	2.60	0.92	0.92	0.41	0.044
الدرجة الثانية	1953.62	736	2.65	0.92	0.92	0.042	0.044
ثنائي العوامل	1874.90	700	2.68	0.92	0.92	0.042	0.043

يتضح من الجدول (4) أن أكثر النماذج مطابقة للبيانات هو النموذج من الدرجة الأولى (First-order Factor Model). ولتحديد ما إذا كانت مجالات المقياس تضيف إلى نموذج القياس فوق ما يقدمه العامل العام، تمت مقارنة النموذج ثنائي العوامل (Bi-factor Model) مع النموذج أحادي العامل (Uni-factor Model)، وذلك بحساب التباين المشترك المُفسر (Explained Common Variance) بواسطة العامل العام، وأوميجا الهرمية (Hierarchical Omega) للعامل العام والعوامل الفرعية. وقد أوضحت النتائج أن العامل العام يفسر منفردًا 76% من التباين بين فقرات اختبار التفكير المجرد، في حين أن المجالات الفرعية تفسر

6%، و6% و7% و5% على التوالي من التباين بين فقرات الاختبار. ووفقًا لما قدمه رايس ومور وهافيلاند (Reise, Moore & Haviland, 2010)، فإن التباين المشترك المُفسر يُعد مؤشرًا على أحادية البُعد، وكلما كانت قيمة التباين المشترك المُفسر بواسطة العامل العام كبيرة، دل ذلك على سيطرة العامل العام. وبناءً على ذلك، فإنَّ أفضل نموذج يُطابق البيانات هو نموذج العامل العام (Uni-factor Model)، وقد استُخدم هذا النموذج لاختبار تكافؤ القياس بين الذكور والإناث. ويوضح الجدول (5) جودة مطابقة نموذج العامل العام لكل من الذكور والإناث على حدة.

الجدول (5)

مؤشرات جودة مطابقة النموذج أحادي العامل (نموذج العامل العام) لقياس النسخة النهائية (40 فقرة) لاختبار التفكير المجرّد للذكور والإناث.

مؤشرات جودة المطابقة							
النموذج	مربع كآى	درجة الحرية df	النسبة	مؤشر المطابقة المقارن CFI	مؤشر توكر-لويس TLI	الجذر التربيعي لمربعات أخطاء التقريب RMSEA	الجذر المعياري لمتوسطات البواقي SRMR
الذكور	1459.50	740	1.97	0.91	0.91	0.045	0.05
الإناث	1875.14	740	2.54	0.91	0.90	0.045	0.05

يتضح من الجدول (5) تطابق نموذج قياس العامل العام الوحيد لعينة الذكور والإناث. وبالتالي فقد تحققت الخطوة الأولى من اختبار تكافؤ القياس بين الذكور والاناث، وهي اللاتغير الشكلي (التكويني) (Configural Invariance)، مما يدل على عدم اختلاف البناء العاملي المقترح لاختبار قياس قدرة التفكير المجرد. ويمكن عزو هذه النتيجة إلى أن العوامل الكامنة تحمل النسق نفسه من التشبعات الحرة عبر المجموعات؛ أي أن العوامل الكامنة في المجموعات المختلفة يرتبط بها عدد الفقرات نفسه في المجموعات كافة، وهذا يؤكد تحقق صدق الاختبار، وهذا يتفق مع دراسة ديمتروف ودراسة إيونيس (Dimitrov, 2013; Ioannis, 2015) التيْن دلّتا على مناسبة التحليل العاملي التوكيدي للكشف عن ملاءمة بيانات اختبار القدرات العامة للنموذج المقترح، كما دلّتا على صحة البناء العاملي للاختبار. وتتمثل الخطوة الثانية في اختبار تكافؤ القياس بين الذكور والإناث باختبار اللاتغير المتري (Metric Invariance). وقد تمَّ تثبيت تشبعات فقرات الاختبار بين العينتين (الذكور والإناث)، وجاءت النتائج كما يوضحها الجدول (6).

الجدول (6)

المقارنة بين النموذج التكويني (الشكلي) والنموذج المتري لعينتي الذكور والإناث

مؤشرات جودة المطابقة							
النموذج	مربع كآي	درجة الحرية df	النسبة	مؤشر المطابقة المقارن CFI	مؤشر توكر-لويس TLI	الجذر التربيعي لمربعات أخطاء التقريب RMSEA	الجذر المعياري لمتوسطات البواقي SRMR
الشكلي	3334.64	1480	2.25	0.91	0.91	0.045	0.05
المتري	3044.94	1520	2.00	0.92	0.91	0.044	0.05

يتضح من الجدول (6) أن وضع قيد على معْلمات النموذج بحيث نجعل القيم المعيارية لتشبعات فقرات الاختبار متساوية بين الذكور والإناث أدى إلى تحسن في قيم مربع كآي، إلا أن قيمة الفرق في درجة الحرية لكاي المعدلة = 269.67 دالة عند مستوى الدلالة 0.01. وقد أظهرت مؤشرات التعديل وجود 13 فقرة تشبعاتها غير متكافئة عبر عينتي الذكور والإناث، وقد تم ترك تشبعات تلك الفقرات حرة بين الذكور والإناث، وبناءً على ذلك انخفضت قيمة $X^2_{dif} = 35.02$ ودرجة الحرية 27 بمستوى دلالة 0.14، وهذه القيمة أكبر من 0.05، مما يدل على وجود 27 فقرة تطابقت تشبعاتها بين الذكور والإناث و13 فقرة اختلفت تشبعاتها بين العينتين. وبهذا يتحقق اللاتغير المتري. أما الخطوة الثالثة من خطوات اللاتغير في القياس فهي اختبار اللاتغير العددي (Scalar Invariance)، حيث يُختبر في هذه الخطوة تساوي العتبات الفارقة (Thresholds) لكل فقرة من فقرات اختبار التفكير المجرّد بين عينتي الذكور والإناث. وقد تم اختبار الفرق بين جعل العتبات الفارقة لعينة الذكور تختلف عن العتبات الفارقة لعينة الإناث مقابل افتراض تساوي تلك العتبات بين العينتين. وقد جاءت النتائج كما يوضحها الجدول (7).

الجدول (7)

المقارنة بين النموذج التكويني (الشكلي) والنموذج المتري والنموذج العددي لعينتي الذكور والإناث.

مؤشرات جودة المطابقة							
النموذج	مربع كاي	درجة الحرية df	النسبة	مؤشر المطابقة المقارن CFI	مؤشر توكر-لويس TLI	الجذر التربيعي لمربعات أخطاء التقريب RMSEA	الجذر المعياري لمتوسطات البواقي SRMR
الشكلي	3334.64	1480	2.25	0.91	0.91	0.045	0.05
المتري	3044.94	1520	2.00	0.92	0.91	0.044	0.05
المتري المعدل	3018.25	1507	2.00	0.92	0.91	0.044	0.05
العددي	3173.61	1547	2.05	0.91	0.90	0.046	0.06

يتضح من الجدول (7) تأثر النموذج العددي بصورة سلبية بإضافة افتراض تطابق قيم العتبات الفارقة لفقرات اختبار التفكير المجرّد بين الذكور والإناث؛ فقد زادت قيمة مربع كآي وانخفضت قيم مؤشر المطابقة المقارن ومؤشر توكر-لويس، وزادت قيم جذر متوسطات مربعات الأخطاء التقريبية، وبالمثل زادت قيم الجذر المعياري لمتوسطات البواقي. وقد قام الباحثان بإعادة الخطوة السابقة بعد السماح لتلك الفقرات أن تكون حرة بين المجموعتين؛ بمعنى أن يقوم البرنامج بتقدير قيم مختلفة للعتبات الفارقة لتلك الفقرات بين عينتي الذكور والإناث، كما يوضحها الجدول (8).

الجدول (8)

المقارنة بين النموذج المتري والنموذج العددي الجزئي لعينتي الذكور والإناث.

مؤشرات جودة المطابقة							
النموذج	مربع كاي	درجة الحرية df	النسبة	مؤشر المطابقة المقارن CFI	مؤشر توكر-لويس TLI	الجذر التربيعي لمربعات أخطاء التقريب RMSEA	الجذر المعياري لمتوسطات البواقي SRMR
الشكلي	3334.64	1480	2.25	0.91	0.91	0.045	0.05
المتري	3044.94	1520	2.00	0.92	0.91	0.044	0.05
المتري المعدل	3018.25	1507	2.00	0.92	0.91	0.044	0.05
العددي	3173.61	1547	2.05	0.91	0.90	0.046	0.06
العددي الجزئي	3027.17	1529	1.98	0.92	0.92	0.044	0.05

يتضح من الجدول (8) تطابق النموذج العددي الجزئي؛ أي أن هناك تكافؤًا جزئيًا بين الذكور والإناث في قيم العتبات الفارقة لفقرات الاختبار. فمن بين 40 فقرة، تطابقت العتبات الفارقة لـ 22 فقرة واختلفت العتبات الفارقة لـ 18 فقرة، وهذا يؤكد أن للاختبار التركيب العاملي نفسه، والتشبعات نفسها لدى فئات الطلبة المختلفة من ذكور وإناث. ويمكن عزو نتيجة الفروق بين الذكور والإناث لاختلاف خصائص تلك الفئات وقدراتها؛ إذْ يلاحظ في هذه الفئة العمرية شيء من الاستهتار وعدم التعامل مع المواقف بجدية، وهي مرحلة حساسة من عمر الفرد (عمر المراهقة) تحصل فيها للفرد تغيرات نوعية، وتحدث فيها صراعات داخل الفرد لا تعود لاختلاف خصائص الاختبار عبر المجموعات المختلفة.

النتائج المتعلقة بالسؤال الثالث، والذي ينص على: **"***ما الخصائص السيكومترية لفقرات اختبار قياس القدرة على التفكير المجرّد (الصدق والثبات)* ؟**"** *ومناقشتها*

أولاً: دلالات الصدق

تم التحقق من الصدق المنطقي (Logical Validity) من خلال التحليل النظري، وذلك عن طريق تحليل مفهوم التفكير المجرد، وكتابة فقرات تمثل أبعاد التفكير المجرد، وتحكيمها من مجموعة من المحكمين (صدق المحتوى). وبناءً على آرائهم، تم تعديل فقرات المقياس. وتُعد طريقة التحليل المنطقي (Logical Analysis) إحدى الطرق والإجراءات التي تستخدم للتحقق من صدق بناء أدوات القياس (Construct Validity) (Cronbach, 1971). واستنادًا إلى الأدب السابق، وللتحقق من صدق التكوين الفرضي، فقد بُنيت فرضيتان: تنص الأولى على عدم وجود فروق بين الذكور والإناث في قدرة التفكير المجرد، وتنص الثانية على وجود فروق في قدرة التفكير المجرّد بين ذوي التحصيل المتدني والمتفوقين. وللتحقق من الفرضية الأولى، تم اشتقاق عينة عشوائية مكونة من 24 طالبًا و24 طالبة من العينة الكلية، واستخدم اختبار t لعينتين مستقلتين لمقارنة الدرجة الكلية على مقياس التفكير المجرّد بين الجنسين، وجاءت النتائج كما يوضحها الجدول (9).

الجدول (9)

نتائج اختبار t للفروق بين الذكور والإناث على اختبار التفكير المجرد .

الجنس	العدد	المتوسط	الانحراف المعياري	قيمة *t*	درجة الحرية	الدلالة الإحصائية
الذكور	24	28.00	13.61	0.683	46	0.498
الاناث	24	30.88	15.47			

يتضح من الجدول (9) عدم وجود فروق ذات دلالة إحصائية بين متوسط درجات الذكور ومتوسط درجات الإناث على اختبار التفكير المجرد، مما يدل على تحقق الفرضية الأولى. وهذا يتفق مع العديد من الدراسات مثل دراستي (Burney, 1974; Bakir & Bicer, 2015)، بينما يخالف دراسة (Sarsani, 2008) التي أشارت إلى وجود فروق ذات دلالة إحصائية بين الذكور والإناث لصالح الذكور.

وللتحقق من الفرضية الثانية، تم اشتقاق عينة مكونة من 30 طالبًا وطالبة من مدرسة الملك عبد الله الثاني للتميز (من المتفوقين)، وعينة عشوائية مكونة من 30 طالبًا وطالبة من المدارس الأخرى (من ذوي التحصيل المتدني). وقد تم استخدام اختبار*t* لعينتين مستقلتين لاختبار صحة هذه الفرضية، وقد جاءت النتائج كما يوضحها الجدول (10).

الجدول (10)

نتائج اختبار t للفروق بين ذوي التحصيل المتدني والمتفوقين على اختبار التفكير المجرد .

الطلبة	العدد	المتوسط	الانحراف المعياري	قيمة *t*	درجة الحرية	الدلالة الإحصائية
ذوو التحصيل المتدني	30	30.37	16.62	4.88	58	0.01
المتفوقون	30	45.87	5.15			

يتضح من الجدول (10) وجود فروق ذات دلالة إحصائية بين متوسط درجات الطلاب ذوي التحصيل المتدني والمتفوقين لصالح الطلبة المتفوقين، وهذا يؤكد توافر صدق التكوين الفرضي في اختبار التفكير المجرد (62 فقرة).

ثانيًا: دلالات الثبات

قُدرت معاملات الثبات للاختبار بطريقتين: أولاهما باستخدام الطرق التقليدية في القياس، والثانية باستخدام نظرية الاستجابة للفقرة. أما بالنسبة للطرق التقليدية، فقد تم استخراج معامل ثبات الاتساق الداخلي باستخدام معادلة كرونباخ ألفا. وقد بلغت قيمة ألفا باستخدام هذه الطريقة (0.91)، وهي قيمة عالية تشير إلى أن المقياس يتمتع بثبات عال يتيح استخدامه في مواقف ذات علاقة بمتغير التفكير المجرد، إلى جانب تمتع الاختبار بدلالات اتساق داخلي (كمؤشر على الثبات) عالية. أما باستخدام نظرية الاستجابة للفقرة، فقد استُخْدمت برمجية Bilog-Mg3 لاستخراج الثبات النظري والثبات التجريبي للفقرات الـ 40 وفق النموذج ثلاثي المعْلمات، حيث بلغ معامل الثبات النظري والتجريبي (0.91 و 0.89) على التوالي. ويُلاحظ تقارب قيم معاملات الثبات النظري التي تعتمد على استخدام طريقة الأرجحية العظمى في تقدير قدرات الأفراد مع قيم معاملات الثبات التقليدي، الأمر الذي يؤكد الثبات المرتفع للاختبار. وبهذا تم التوصل إلى تمتع اختبار التفكير المجرّد بدلالات صدق وثبات مرتفعة.

النتائج المتعلقة بالسؤال الرابع، والذي ينص على: *"ما قيم تقديرات معالم الفقرات على الاختبار وفقاً للنموذج ثلاثي المعلمات في نظرية الاستجابة للفقرة؟ ومناقشتها"*

للإجابة عن سؤال الدراسة الرابع، قام الباحثان بحذف الفقرات التي لم تطابق النموذج، وتكون الاختبار بصورتة النهائية من 40 فقرة، وتم تقدير معالم الفقرات وفقًا للنموذج ثلاثي المعْلمات باستخدام طريقة الأرجحية العظمى الهامشية Marginal Maximum Likelihood (MML) التي تستخدم في برنامج Bilog-Mg3. ويبين الجدول (11) قيم معْلمات الصعوبة والتمييز والتخمين والخطأ المعياري لتقدير كل منها وفقًا للنموذج ثلاثي المعْلمات وذلك للصورة النهائية للاختبار المكونة من 40 فقرة.

الجدول (11)

قيم معلمات صعوبة الفقرات ومعلمات تمييزها ومعلمات تخمينها والخطأ المعياري في تقديرها للفقرات الـ 40 المكونة للاختبار النهائي وفق النموذج اللوجستي ثلاثي المعلمة

رقم الفقرة	التمييز	الخطأ المعياري للتمييز	الصعوبة	الخطأ المعياري للصعوبة	التخمين	الخطأ المعياري للتخمين	رقم الفقرة	التمييز	الخطأ المعياري للتمييز	الصعوبة	الخطأ المعياري للصعوبة	التخمين	الخطأ المعياري للتخمين
2	2.00	0.16	0.69	0.08	0.35	0.03	37	1.61	0.16	0.74	0.05	0.17	0.02
5	1.76	0.16	0.2	0.23	0.36	0.07	38	2.26	0.28	0.39	0.04	0.23	0.03
6	1.71	0.16	1.01	0.12	0.26	0.03	39	2.6	0.16	0.31	0.07	0.2	0.03
9	2.49	0.16	0.97	0.08	0.19	0.02	40	2.39	0.16	0.53	0.07	0.2	0.03
11	2.22	0.16	0.27	0.09	0.17	0.04	41	2.42	0.16	0.49	0.05	0.12	0.02
12	1.61	0.16	0.79	0.12	0.22	0.04	42	1.82	0.16	0.33	0.07	0.16	0.03
20	2.16	0.29	0.51	0.07	0.06	0.03	43	1.86	0.16	0.9	0.06	0.15	0.02
21	2.26	0.16	0.55	0.06	0.19	0.03	44	2.55	0.16	0.54	0.09	0.16	0.04
22	2.31	0.16	0.46	0.09	0.21	0.03	46	2.22	0.16	0.29	0.05	0.18	0.03
23	2.15	0.16	0.69	0.07	0.25	0.03	47	2.68	0.16	0.26	0.06	0.2	0.02
24	2.73	0.16	0.18	0.06	0.13	0.03	48	0.48	0.16	0.42	0.07	0.17	0.03
25	2.3	0.16	0.14	0.06	0.11	0.03	50	2.76	0.18	2.09	0.66	0.22	0.04
27	2.17	0.16	0.28	0.07	0.13	0.03	52	2.49	0.16	0.56	0.06	0.1	0.02
28	2.43	0.16	0.37	0.09	0.18	0.04	53	2.67	0.16	0.5	0.06	0.13	0.03
29	2.06	0.27	0.08	0.11	0.10	0.05	55	2.28	0.16	0.05-	0.08	0.13	0.04
30	2.65	0.16	0.5	0.06	0.14	0.03	56	2.64	0.16	0.35	0.08	0.16	0.03
31	2.69	0.16	0.48	0.07	0.19	0.03	58	2.6	0.16	0.2	0.07	0.12	0.02
33	2.66	0.16	0.6	0.06	0.16	0.03	59	1.77	0.16	0.29	0.08	0.22	0.03
34	1.68	0.28	0.02-	0.15	0.22	0.05	60	2.54	0.27	0.33	0.1	0.1	0.01
35	2.4	0.16	0.75	0.06	0.15	0.02	61	2.65	0.16	0.55	0.06	0.12	0.02

يتضح من الجدول (11) أن قيم معلمة التمييز تراوحت بين 0.48 و2.76، وقد كانت الفقرة 50 الأكثر تمييزًا، والفقرة 48 الأقل تمييزًا، وأن قيم معلمة الصعوبة تراوحت بين -0.05 و2.09، حيث كانت الفقرة 50 أصعب فقرة في الاختبار، والفقرة 55 أسهل فقرة، كما يلاحظ أن قيم معلمة التخمين تراوحت بين 0.06 و0.36 وكانت الفقرة 5 الأكثر تخمينًا، والفقرة 20 الأقل تخمينًا. ويلاحظ من القيم السابقة أن فقرات الاختبار تتمتع بمعلمات جيدة وضمن المدى العملي المقبول في القياس التربوي والنفسي. وهذا يتفق مع النتيجة التي توصل اليها لسون وفلتشر (Lesson & Fletcher, 2003) من أنّ استخدام النموذج الذي يتضمن معلمة التمييز يُحسن من المطابقة. وقد أكد كروكر وألجينا وهاتي (Crocker & Algina, 1986; Hattie, 1985) أن استخدام عينات كبيرة من المفحوصين يضمن لنا الحصول على دقة عالية في تقديرات معلمة التمييز.

الصورة النهائية للاختبار

تكوّن الاختبار بصورته النهائية من (40) فقرة، تقيس القدرة على التفكير المجرد. وإذا تم تطبيق المقياس، فإن التعليمات تشير إلى أن الزمن المخصص لذلك لا يزيد على 65 دقيقة. والمطلوب من المستجيب اختيار بديل واحد من بين أربعة بدائل، وبذلك تتراوح الدرجة الكلية (Raw Score) على الاختبار بين (0-40). ولتفسير الدرجات على المقياس، يقترح الباحثان أن يتم تصنيف المستجيبين وفق الدرجات الكلية على المقياس في ثلاث فئات على النحو الآتي: المستجيبون الذين تقل درجاتهم الخام عن (14) ويقابلها وفق نظرية الاستجابة للفقرة على متصل القدرة (0.99-) وحدة لوجيت يصنفون بأنهم ما زالوا في مرحلة التفكير المحسوس (المادي)، والمستجيبون الذين تقع درجاتهم الخام بين (15-24) ووفق متصل القدرة بين (0.99-1.02) وحدة لوجيت يصنفون بأنهم في المرحلة الانتقالية بين التفكير المحسوس والتفكير المجرد. أمّا المستجيبون الذين تزيد درجاتهم الخام على (24) ووفق متصل القدرة تزيد على (1.02) وحدة لوجيت؛ فيصنفون بأنهم وصلوا مرحلة التفكير المجرد.

الاستنتاجات والتوصيات

هدفت هذه الدراسة إلى بناء اختبار لقياس القدرة على التفكير المجرّد وفق نظرية الاستجابة للفقرة. ووفقًا للنتائج التي تم عرضها، والتحقق من مدى مطابقة هذه الاستجابات لافتراضات النموذج ثلاثي المعلمات، تمتع الاختبار بدرجة مقبولة من الموضوعية، وكانت فقرات المقياس ذات خصائص سيكومترية مقبولة تبرر استخدامه بصورته النهائية المكونة من (40) فقرة لقياس القدرة على التفكير المجرّد التي يمتلكها طلبة الصف العاشر الأساسي الممثلون لعينة الصدق والثبات المستخدمة في تطوير المقياس، حيث يمكن الاستفادة من هذه الأداة في التخطيط للمناهج والاختيار للقبول في التخصصات والفروع الأكاديمية المختلفة، بحيث تضيف معلومات مهمة عن الطالب إلى جانب ما تقدمه أدوات القياس المعرفية. وبذلك يمكن معرفة مستوى قدرة التفكير المجرّد التي يمتلكها الطلبة. ونظرًا لاقتصار عينة الدراسة على طلبة الصف العاشر الأساسي، فإن الباحثين يوصيان بإجراء المزيد من الدراسات على الصورة الأولية المكونة من (62) فقرة، والصورة النهائية المكونة من (40) فقرة لاختبار قدرة التفكير المجرّد لتشمل طلبة من مستويات أخرى وعينات أوسع، وفئات مختلفة من المجتمعات كون الاختبار متحررًا من أثر الخلفية التعليمية والثقافية، وذلك من أجل تأكيد الثقة بالخصائص السيكومترية لفقرات المقياس لاستخدامه بدرجة عالية من الثقة في الكشف عن مستوى قدرة التفكير المجرّد للطلبة للقيام بالإجراءات المناسبة عندئذ.

References

Abdel-Wahab, M. (2010). *The use of response models for the test paragraph in grading the vocabulary of some cognitive tests.* Unpublished Doctoral Dissertation, Minia University, Egypt.

Abdullah, A. & Azzawi, A. (2019). The effect of Ady and Shire model on fourth-grade students' achievement in physics. *Journal of Tikrit University for Humanities*, 26(10), 260-286.

Abu Jadu, S. & Nofal, M. (2010). *Teaching theoretical and practical thinking.* Third Edition, Amman: Dar Al-Masirah for Publishing and Distribution.

Abu Roman, K. (1991). *The relationship between formal thinking and scientific skills and achievement of first-grade scientific secondary students*. Unpublished Master Thesis, Yarmouk University.

Al-Atoum, A. (2004). *Cognitive psychology: Theory and practice.* Dar Al-Masirah, Amman, Jordan.

Al-Azawi, W. (2016). Constructing formal thinking test in physics for the students of fifth scientific year. *Basic Education College Magazine for Educational and Humanities Sciences*, 25, 661-681.

Al-Otaibi, K. (2001). *The effectiveness of a proposed program to develop inferential-thinking skills for a sample of secondary students in Riyadh*. Unpublished Master Thesis, King Saud University, Saudi Arabia.

Al-Tammi, A. (2015). *The development of the abstract reasoning by Paul Newton on the students of preparatory stage by the use of one-parameter model according to the item theory*. Unpublished Master Thesis, University of Baghdad.

Alter, C. & Egan, M. (1997). Logic modeling: A tool for teaching critical thinking in social-work practice. *Journal of Social Work Education*, 33(1), 85-102.

Baker, F. B. (2001). *The basics of item response theory*. Retrieved from: http://www.ericae. net/irt/baker.

Bakir, S. & Bicer, O. (2015). Logical thinking and cognitive development levels of pre-service science teachers. *J. Educ. Res.*, 5(1), 149-165.

Bautista, R. P. M. (2013). *Rasch modeling of abstract reasoning in project TALENT.* Unpublished Doctoral Dissertation, University of Southern California.

Bryon, M. (2008). *How to pass diagrammatic-reasoning tests: Essential practice for abstract input type and spatial-reasoning tests.* Kogan Page Publishers.

Burney, G. M. (1974). *The construction and validation of an objective formal-reasoning instrument.*

Cai, L., Du Toit, S. H. C. & Thissen, D. (2011). *IRTPRO: Flexible, multi-dimensional, categorical IRT modeling* [computer software]. Chicago, IL, Scientific Software International.

Chan, F. (2018). *Development of matrices of abstract-reasoning items to assess fluid intelligence.* Doctoral Dissertation, University of Cambridge.

Crocker, L. & Algina, J. (1986). *Introduction to classical and modern test theory*. Holt, Rinehart and Winston, 6277 Sea Harbor Drive, Orlando, FL, 32887.

Cronbach, L. J. (1971). Test validation. In: R. L. Thorndike (Ed.), *Educational measurement* (2nd Edn., pp. 443-507). Washington, DC, Arrerican Council on Education.

Dimitrov, D. M. (2013). *GAT-verbal: Testing for dimensionality and validation of factorial structure*. (Research Report No. RR-13-1). Riyadh, SA: National Center for Assessment in Higher Education.

Drager, K. W. (2014). *The relationship between abstract reasoning and performance in high-school algebra*. Doctoral Dissertation, University of Kansas.

Gilhooly, K. J. (2004). Working memory and reasoning. In: J. P. Leighton & R. J. Sternberg (Eds.), *The nature of reasoning* (pp. 49–77). Cambridge: Cambridge University Press.

Groves, J. (2015). *The predictive validity of abstract-reasoning test and the Raven's advanced progressive matrices test for the academic results of first-year engineering students.* Doctoral Dissertation, University of Witwatersrand, Johannesburg.

Hambleton, R. K. (1989). *Principles and selected applications of item response theory.* New York: Macmillan Publishing Company.

Hambleton, R. K., Swaminathan, H. & Rogers, H. J. (1991). *Fundamentals of item response theory*. Newbury Park: Sage.

Hamel, R. & Schmittmann, V.D. (2006*). The 20-minute version as a predictor of the Raven advanced progressive matrices test. Educational and Psychological Measurement,* 66, 1039-1046.

Hattie, J. (1985). Methodology review: Assessing uni-dimensionality of tests and items. *Applied Psychological Measurement,* 9(2), 139–164.

Hazzan, O. & Zazkis, R. (2005). Reducing abstraction: The case of school mathematics. *Educational Studies in Mathematics*, 58(1), 101-119.

Helland, F. (2016). Explanatory item response modelling of an abstract-reasoning assessment: A case for modern test design. Master's Thesis, University of Oslo. Retrieved from: https://www.duo.uio.no/handle/10852/52720.

Hunt, E. (2011). *Human intelligence.* Cambridge: Cambridge University Press.

Ioannis, T. (2015). Factorial invariance and latent mean differences of scores on SAAT across gender. Technical Report TR092-2015. National Center for Assessment in Higher Education.

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

Muthén, L. K. & Muthen, B. (2017). *Mplus user's guide: Statistical analysis with latent variables*. Muthén & Muthén.

Newton, P. & Bristoll, H. (2009). *Psychometric success: Abstract reasoning.*

Nordstrom, C., Rendel, G. & Tavares, R. (2015). *Mastering the UKCAT*. CRC Press.

Odeh, A. (2015). *Measurement and evaluation in the teaching process*. Irbid: Dar Al-Amal for Publishing and Distribution, Jordan.

Piaget, J. (1970). *Science of education and the psychology of the child*. Trans. D. Coltman.

Psytech SA. (2019). *Manual with questions for the abstract-reasoning test.* Retrieved on 15/08/2019 from: www.psytech.co.za.

Reckase, M. D. (1979). Uni-factor latent trait models applied to multi-factor tests: Results and implications. *Journal of Educational Statistics*, 4(3), 207-230.

Reise, S. P., Bonifay, W. E. & Haviland, M. G. (2013). Scoring and modeling psychological measures in the presence of multi-dimensionality. *Journal of Personality Assessment*, 95(2), 129-140.

Reise, S. P., Moore, T. N. & Haviland, M. G. (2010). Bifactor models and rotations: Exploring the extent to which multi-dimensional data yields univocal scale scores. *Journal of Personality Assessment*, 92, 544–559.

Sarsani, M. R. (2008). A study of the reasoning abilities of ninth grade standard students with respect to their gender and type of the institution. *Journal of Educational Psychology*, 1(3), 44-54.

Slavin, R. E. (1996). Research on cooperative learning and achievement: What we know, what we need to know. *Contemporary Educational Psychology*, 21(1), 43-69.

Solso, R. L., MacLin, M. K. & MacLin, O. H. (2005). *Cognitive psychology*. Pearson Education, New Zealand.