الأداء التفاضلي لفقرات ومموهات اختبار محكي المرجع في العلوم للصف السادس الابتدائي مصمم وفق نظرية استجابة الفقرة*

د. فريال محمود محمد الحاج محمود**

*تاريخ التسليم: 2017/2/4م، تاريخ القبول: 2017/5/23م.
**أستاذ مساعد/ جامعة جدة/ المملكة العربية السعودية.

ملخص:

هدفت هذه الدراسة إلى الكشف عن الأداء التفاضلي لفقرات، ومموهات، اختبار محكي المرجع في العلوم، للصف السادس الابتدائي، مصمم وفق نظرية استجابة الفقرة. ولتحقيق هذا الهدف صممت الباحثة اختباراً يتكون من (75) فقرة من نوع الاختيار من متعدد. طبق الاختبار في نهاية العام الدراسي (2015 – 2016م)، على عينة من (520) طالباً وطالبة من طلاب الصف السادس الابتدائي، في مدينة جدة، في المملكة العربية السعودية، منهم (418) من المدارس الحكومية، و (102) من المدارس الاهلية. أظهرت نتائج البرنامج الإحصائي (BILOG-MG3) مطابقة (61) فقرة من فقرات الاختبار للنموذج ثلاثي المعلم، وظهر الأداء التفاضلي في (16) فقرة منها، كما أظهر اختبار مربع كاي ($\chi 2$) للكشف عن الأداء التفاضلي لمموهات الفقرات، وجود أداء تفاضلي لمموهات (6) فقرات منها. وقد وافق (75 %) من المحكمين على ذلك. أوصت الدراسة: بضرورة فحص الأداء التفاضلي لمموهات الفقرة، للتأكد إن كانت هي سبب الأداء التفاضلي للفقرة.

الكلمات المفتاحية: الأداء التفاضلي للفقرة، الأداء التفاضلي للمموهات، نظرية استجابة الفقرة، النموذج ثلاثي المعلم.

Differential Functioning Items and Distractors of a Criterion Referenced Science Test for the Sixth Grade

Abstract:

This study aimed to investigate differential functioning items and distractors of a criterion referenced science test for sixth grade. The test is constructed according to item response theory (IRT). To achieve this goal, the researcher designed a test constitutes of (75) multiple-choice items. The test was applied at the end of the school year (20152016-), on a sample of (520) 6th grade students, in Jeddah, Saudi Arabia, (418) of which are from public schools, and (102) of which are from private schools. The results of the statistical program (BILOG-MG3) verified that 61 test items fit the three parameters model, 16 of which presented item differential functioning (IDF). In addition, chi square test (2) presented distractors differential functioning (DDF) in (6) of these items, and (75%) of the arbitrators agree on this matter. The study recommendation indicates to the importance of checking items distractors differential functioning, to check if it was the reason of item differential functioning.

Keywords : *Item Differential Functioning, Distractors Differential Functioning, Item Response Theory, Three Parameters Mode.*

المقدمة

تتخذ العديد من القرارات – مثل قرارات القبول الجامعي وقرارات التوظيف– في ضوء أداء الأفراد على اختبارات خاصة مصممة لهذه الغاية، ولتكون هذه القرارات صائبة؛ يجب أن تكون بياناتها اصلاً صادقة، أي يجب أن تتمتع اداة القياس –ومنها الاختبارات– التي جمعت بواسطتها البيانات بالصدق (Validity)، والعدالة (Fairness)، (عودة،2010). ويعد فحص الأداء التفاضلي (DIF) (Differential Item Functioning) لفقرات الاختبار من أهم طرق التأكد من كل من الصدق والعدالة، خاصة عندما يطبق هذا الاختبار على مجموعات عرقية، وثقافية مختلفة (Hambleton & Rogers,1995).

ان مصطلح الأداء التفاضلي للفقرة ليس حديثاً؛ فقد استخدم منذ ثمانينيات القرن الماضي كمرادف للتحيز. فالأداء التفاضلي للفقرة يظهر عندما لا يكون احتمال الإجابة بشكل صحيح عن الفقرة، متماثلاً بين مجموعتين، أو أكثر من الأفراد، ممن لهم نفس مستوى القدرة (Herrmann & DeBoe, 2011). ويؤكد كل من هامبلتون وسواميناثان وروجرز (Hambleton, Swaminathan &Rogers, 1991) أن الأداء التفاضلي للفقرة يحدث عندما تكون دوال الاستجابة على الفقرة غير متماثلة في المجموعات الفرعية المختلفة. وفي هذه الحالة: فإن اصحاب نفس القدرة لم يحصلوا على درجات متماثلة على الاختبار، وهذا يتنافى مع الهدف من الاختبار.

هناك العديد من طرق الكشف عن الأداء التفاضلي للفقرة ومنها: الفروق في صعوبة الفقرة في المجموعات الفرعية – وهي الطريقة المستخدمة في هذه الدراسة– وطريقة تحليل التباين، وطريقة مربع كاي (حيث يتم مقارنة عدد الإجابات الصحيحة في المجموعات الفرعية)، وطريقة الانحدار اللوجستي (حيث المتغير التابع هو الاستجابة على الفقرة بينما المتغيرات المستقلة هي: القدرة المشاهدة، والمجموعة الفرعية للمفحوص والتفاعل بينهما)، وطريقة مانتل هانزل (وتستخدم في استقصاء الفروق بين المجموعتين المرجعية والمستهدفة)، وطريقة تحيز الفقرة المتزامن (وهي تشبه طريقة مانتل هنزل ولكنها تناسب المجموعات التي تختلف في مستوى القدرة) (Crocker & Algina,1986).

إن مصمم الاختبار وكما أشار هيرمان وديبو (Herrmann & DeBoe, 2011) لا يريد ظهور الأداء التفاضلي في أي من فقرات اختباره، لذلك يحاول عند كتابته لهذه الفقرات، أن يميز بين المفحوصين من مستويات القدرة المختلفة. أي يريد الوصول إلى الوضع المثالي حيث: احتمال إجابة المفحوص بشكل صحيح عن الفقرة يعتمد على مستواه على السمة التي يتم قياسها، وبذلك يحصل الطلاب ذو نفس القدرة، على نفس الدرجة. وفي محاولة مصمم الاختبار لمنع الطالب الذي لا يعرف الإجابة الصحيحة، من اختيار البديل الصحيح للإجابة، فهو يضع البدائل الخاطئة أو المموهات (Distractors)؛ لتجذب الطلاب الذين لا يملكون المعرفة، أو الذين لديهم معرفة جزئية بالإجابة الصحيحة، وهو يعمل على جعل هذه المموهات جاذبة، بنفس الدرجة للطلاب الذين لا يعرفون الإجابة الصحيحة، ورغم كل هذه المحاولات فإن الطلاب لا

يختارون المموه عشوائياً بل بناءً على المعرفة الجزئية، أو باستخدام الاستدلال الخاطئ. ولكن ماذا لو: جذب أحد مموهات الفقرة الطلاب من ذوي مستوى قدرة معينة، في مجموعة ما، بطريقة لا يجذب بها زملائهم من ذوي نفس القدرة، في مجموعة أخرى؟! إن هذا ما يعرف بالأداء التفاضلي للمموهات (Differential Distractor Function (DDF)).

تستخدم هذه الدراسة اختباراً تم تصميمه وفق نظرية الاستجابة للفقرة (Item Response Theory)، وهي من الاتجاهات الحديثة في القياس، وتتميز بأنه: يجري تقدير قدرة الفرد بشكل مستقل عن عينة الفقرات التي أجاب عنها، كذلك يتم تقدير معالم الفقرات بشكل مستقل عن عينة الأفراد الذين طبقت عليهم هذه الفقرات (Hambleton & Swaminathan,1985,p11). ولهذه النظرية العديد من النماذج ومنها: النموذج اللوجستي ثلاثي المعلم (Three Parameters Logistic Model) (3PL) : الذي يفترض أن الفقرات تختلف في كل من: معلم الصعوبة (b)، ومعلم التمييز (a)، ومعلم التخمين (c) الذي يمثل احتمال توصل الأفراد ذو القدرة المنخفضة إلى الإجابة الصحيحة عن الفقرة عن طريق الصدفة (Hambleton& others, 1991) .

إننا في العادة نهتم بدرجة الطالب الكلية على الاختبار، ولكن وكما أكد كل من ماير وويس (Meyer & Wise,2006) فإن تقدير مستوى قدرة الطالب بناء على عدد اجاباته الصحيحة، غير كاف للحصول على معلومات حول الأداء المعرفي لهذا الطالب، بل يجبّ أيضاً فحص نمط استجاباته غير الصحيحة على فقرات الاختبار.

في دراسة ابيدي وليون وكاو (Abedi, Leon & Kao,2008) لفحص الأداء التفاضلي لمموهات اختبار تقييم القراءة للطلاب المعاقين، تم فحص المموهات التي اختارها هؤلاء الطلاب في اختبار موحد للقراءة، بحيث إن اختار الطلاب من مجموعة المعاقين والطلاب من مجموعة العاديين ممن هم من نفس القدرة مموهاً مختلفاً، فهذا يدل على أن معنى الفقرة مختلف عند المجموعتين. تكونت عينة الدراسة من مجموعتين من الطلاب العاديين، والطلاب المعاقين من الصف الثالث (6611 طالب منهم 448 معاقين)، ومن الصف التاسع (5287 منهم 522 معاقين)، وذلك من مدارس ولاية لوس انجلوس في الولايات المتحدة الامريكية. صمم الباحثون نموذجان من كل من اختباري الصفين الثالث والتاسع، يتكون كل منهما من (54) فقرة من نوع الاختيار من متعدد. وأظهرت النتائج تفوق الطلاب العاديين، على الطلاب المعاقين في اختبار القراءة. كما ظهر الأداء التفاضلي لمموهات عدة فقرات في الاختبار، وقد تمركز هذه الفقرات في النصف الثاني من الاختبار أكثر من النصف الأول منه، مما يشير – وفق ما ذكره الباحثون–إلى أن المعاقين يحتاجون وقتاً أطول في الإجابة عن الفقرات، مما يحتاجه العاديون. إضافة ذلك كان اختيار الطلاب المعاقين للمموه الشائع (Common Distractor)، اقل من اختيار الطلاب العاديين لهذا المموه؛ مما يشير إلى أنهم كانوا يختارون البدائل عشوائياً. وأكد الباحثون على أن نتائج دراستهم تلقي الضوء على: العوامل المؤثرة في اختبار تقييم القراءة لدى الطلاب المعاقين، وبالتالي يجب تصميم اداة قياس تعمل بشكل صادق، كمقياس موحد لجميع الطلاب.

في دراسة بارناباس (Barnabas، 2012) حاول الباحث كشف الأداء التفاضلي لفقرات اختبار الرياضيات للمرحلة الثانوية، تكون الاختبار من (50) فقرة من نوع الاختيار من متعدد، وطبق على (12) مدرسة، اختيرت عشوائيا، من (20) منطقة تعليمية، وتكونت عينة الدراسة من (825) من الذكور، و(846) من الإناث. واستخدم الباحث طريقة مربع كاي في الكشف عن الأداء التفاضلي، والتي أظهرت وجود أداء تفاضلي في (39) فقرة من فقرات الاختبار. وعلل الأداء التفاضلي لفقرات اختبار الرياضيات، بأنه مقرر اجباري للنجاح في المرحلة الثانوية، فالطلاب ممن لديهم اتجاهات فنية، يدرسون الرياضيات بشكل إجباري، بينما يدرسها الأخرون برغبة ومتعة. وفي نهاية البحث أوصى الباحث بضرورة قيام مصمم أي اختبار بدراسة استطلاعية للاختبار، لاستبعاد الفقرات التي تتضمن أداءً تفاضلياً.

في دراسة كراكايا (Karakaya, 2012) قام الباحث بفحص الأداء التفاضلي، لفقرات الاختبارين الفرعيين الخاصين بتحديد لملمستوى في العلوم والرياضيات، ضمن اختبار (Level Determination Exam (LDE)) وهو اختبار يطبق على طلاب الصفوف: السادس، والسابع، والثامن، في تركيا، لتوزيعهم على فروع التعليم الثانوي، تكونت عينة الدراسة من (6913) طالباً من الصف السادس (3614 من الذكور، 3299 من الإناث)، و (6333) طالباً من الصف السابع (3277 من الذكور، 3066 من الإناث)، و (9374) طالباً من الصف الثامن (4290 من الذكور، 5084 من الإناث)، استخدم الباحث طريقة مانتل هانزل في الكشف عن الأداء التفاضلي لفقرات هذه الاختبارات. وعرض الفقرات ذات الأداء التفاضلي على مجموعة من الخبراء. أظهرت النتائج تحيز (5) فقرات في اختبار العلوم (تحيزت (3) فقرات لصالح الإناث، وفقرتان لصالح الذكور). وتحيز (4) فقرات في اختبار الرياضيات (تحيزت (3) فقرات منها لصالح الذكور، وفقرة واحدة لصالح الإناث). كما لوحظ أن محتوى الفقرات المتحيزة لصالح الإناث، يتعلق بالمعلومات الصحية، بينما كان محتوى الفقرات المتحيزة لصالح الذكور، يتعلق بالطبيعة، والأدوات الصناعية، ودرجات الحرارة. واقترح الباحث فحص تحيز فقرات الاختبارات بالنسبة لمتغيرات أخرى مثل: مكان الإقامة، ونوع المدرسة.

واستخدم اموتش وفان (Amuche & Fan ,2014) طريقة الانحدار اللوجستي في الكشف عن الأداء التفاضلي، تبعاً لكل من موقع المدرسة (قرية، مدينة)، ونوع المدرسة (حكومية، خاصة). ذلك لفقرات الاختبار الوطني في الاحياء لعام 2012 (Nation-al Examinations Council (NECO) Biology questions) تكون الاختبار من (60) فقرة، وطبق على عينة مكونة من (447) طالباً وطالبةً، من مدارس منطقة تارابا في نيجيريا، أظهرت النتائج تحيز (10) فقرات، لنوع المدرسة (حكومية/خاصة). وتحيز (8) فقرات لموقع المدرسة (مدينة/قرية). وأوصت الدراسة بضرورة فحص الأداء التفاضلي لفقرات أي اختبار قبل تطبيقه، وذلك للتأكد من عدالة فقراته، وعدم تحيزها.

أجرى اكسيومي (Xiaomei,2014) دراسة لفحص الأداء التفاضلي لفقرات اختبار بيرسون للغة الانجليزية (Pearson Test of English Academic (PTE) Academic)، وهو اختبار عالمي يستخدم لفحص مستوى الطلاب في اللغة الإنجليزية، قبل التحاقهم

بالجامعة. تكونت عينة الدراسة من: (159) من الإناث و (241) من الذكور. فحص الباحث الأداء التفاضلي للفقرات عبر مرحلتين: المرحلة الاولى (المرحلة الإحصائية)، باستخدام اختبار تحيز الفقرة المتزامن ((Simultaneous Item Bias test (SIBTEST)، وفي المرحلة الثانية (المرحلة التحكيمية) عرض الفقرات ذات الأداء التفاضلي، على مجموعة من المحكمين المختصين باختبار بيرسون للغة الانجليزية، وأظهرت النتائج: تحيز (6) فقرات؛ تحيزت (4) فقرات منها لصالح الإناث، وتحيزت فقرتان لصالح الذكور.

بمتابعة الدراسات السابقة يلاحظ: أن هذه الدراسات تركز على الأداء التفاضلي للفقرات، وقد استخدمت طرقاً مختلفة في الكشف عنه، ولكن أياً منها لم تدرس الأداء التفاضلي للموهات، والتي قد تكون السبب وراء ظهور الأداء التفاضلي للفقرات. وقد يرجع السبب وراء قلة الاهتمام بالأداء التفاضلي للموهات، إلى أن الدرجة الكلية على الاختبار تُحسب بناءً على عدد الاجابات الصحيحة، فحتى وان اختلف الطلاب في اختيار الموه، فهذا لن يؤثر؛ لان جميع الموهات خاطئة وتُعامل بنفس الطريقة. ولكن ما لا يعيره الكثيرون الانتباه هو: أن اختيار الموه قد يكون السبب وراء ظهور الأداء التفاضلي للفقرة. أي هل الموهات هي أحد أسباب ظهور الأداء التفاضلي للفقرة.

مشكلة الدراسة وأهدافها وأسئلتها

أولا: مشكلة الدراسة

إن تصنيف الطلاب إلى مجموعات تبعا لقدرتهم، هو من أهم أهداف بناء الاختبارات، فعندما يكتب المعلم فقرات الاختبار، فهو يتوقع أن الطلاب من نفس مستوى القدرة سيحققون نفس الأداء على الاختبار. ولكن هذا الوضع المثالي لا يتحقق دوماً، فعلى الرغم من خضوع الطلاب من نفس القدرة لنفس الفقرة، وفي ظل نفس الظروف، فقد يختلف مستوى أدائهم عليها، وتصبح المشكلة أكثر تعقيداً، ان كانت مجموعة الأداء الأعلى ممن ينتمون لمجموعة: ثقافية، أو عرقية، أو ديموغرافية معينة. إن هذا ما يعرف بالأداء التفاضلي للفقرة. والذي قد يشير لتحيز الفقرة، فبعد ثبوت الأداء التفاضلي لفقرة اختبارية ما، يعرض المعلم هذه الفقرة على مجموعة من المحكمين، والذين يقررون ان كانت الفقرة متحيزة – يجري استبعادها– أما أنها غير متحيزة، وفي هذه الحالة لا تحذف الفقرة رغم ثبوت الأداء التفاضلي. والسؤال المطروح هل يمكن أن يقرر المحكمون أن فقرة ما ليست متحيزة مع انها ليست كذلك؟! ان ملاحظة عدد من المحكمين، وهم يقومون بفحص الفقرة ذات الأداء التفاضليً، توضح انهم يفحصون بالدرجة الأولى: نص الفقرة، ونص البديل الصحيح، ويعطون القليل من الاهتمام للبدائل غير الصحيحة (المموهات) والتي قد تكون السبب في ظهور الأداء التفاضلي. فقد يجذب أحد المموهات أفراد إحدى المجموعات من ذوي قدرة معينة –لاختياره كإجابة صحيحة –أكثر مما يجذب أفراد مجموعة أخرى من ذوي نفس القدرة، وهذا يؤدي بدوره لظهور الأداء التفاضلي للفقرة، وبمعنى اخر احتمال أن تكون الفقرة متحيزة، بسبب الأداء التفاضلي للموهات، وهو ما لا يتم عادةً فحصه احصائياً، كما لا يُعطي المحكمون الاهتمام الكافي للموهات عند فحص الفقرة، للحكم على التحيز. ومن هنا جاءت هذه الدراسة لتؤكد على أهمية فحص الموهات إحصائيا، عند ثبوت الأداء التفاضلي للفقرة، وضرورة فحصها من قبل المحكمين عند الحكم على تحيز الفقرة وعدم الاكتفاء – كما يحدث عادة– بفحص نص الفقرة والبديل الصحيح فقط.

ثانيا: أهداف الدراسة

تهدف هذه الدراسة للكشف عن الأداء التفاضلي لفقرات اختبار محكي المرجع في العلوم للصف السادس الابتدائي مصمم وفق نظرية استجابة الفقرة، وبالتالي فهي تحاول الكشف عن:

- مدى تحقيق بيانات اختبار العلوم للصف السادس الابتدائي لافتراضات نظرية استجابة الفقرة.
- وجود فقرات ذات أداء تفاضلي في فقرات اختبار العلوم للصف السادس الابتدائي.
- دور الأداء التفاضلي للموهات في جعل الفقرة ذات أداء تفاضلي.

ثالثا: أسئلة الدراسة

تحاول هذه الدراسة الإجابة عن الأسئلة الآتية:

◄ السؤال الأول: ما مدى تحقيق بيانات اختبار العلوم للصف السادس لافتراضات نظرية استجابة الفقرة ؟

◄ السؤال الثاني: هل يتضمن اختبار العلوم للصف السادس فقرات ذات أداء تفاضلي؟

◄ السؤال الثالث: هل كان للموهات دور في اظهار الأداء التفاضلي لبعض فقرات الاختبار؟

أهمية الدراسة ومحدداتها:

أولا: أهمية الدراسة

تتمثل أهمية هذه الدراسة في الجانب النظري: بكونها نموذجا يحتذى به، عند الكشف عن اسباب ظهور الأداء التفاضلي للفقرة. أما في الجانب التطبيقي: فإن هذه الدراسة توفر مؤشرات نوعية، وإرشادات تساعد المهتمين في بناء الاختبارات في الوقوف على الأسباب التي تجعل بعض الفقرات تظهر أداءً تفاضلياً، مما يشير بعد موافقة المحكمين، على أنها فقرات متحيزة يجب التخلص منها.

ثانيا: حدود الدراسة

اقتصرت هذه الدراسة على:

- طلبة الصف السادس الابتدائي في مدارس منطقة جدة في المملكة العربية السعودية، ممن هم على مقاعد الدراسة في الفصل الدراسي الثاني من العام الدراسي (1436 – 1437 هـ) الموافق (2015 – 2016 م).
- استخدام النموذج ثلاثي المعلم، واستخدام برمجية (Bilog-MG3) .
- مقرر (العلوم)، وهو أحد المقررات التي يدرسها طلاب الصف السادس الابتدائي.

الطريقة وإجراءات الدراسة:

أولا: مجتمع وعينة الدراسة

ـ مجتمع الدراسة

تكون مجتمع الدراسة من جميع طلبة الصف السادس الابتدائي في مدارس منطقة جدة في المملكة العربية السعودية، والبالغ عددهم (15929) طالباً وطالبةً، وفق إحصائيات وزارة التربية والتعليم السعودية للعام الدراسي (1436 – 1437 هـ) الموافق (2015 – 2016 م)، موزعين على متغيرات الدراسة كما يلي: (11821) طالباً وطالبة في المدارس الحكومية، و(4108) طالباً وطالبة في المدارس الاهلية.

♦ **عينة الدراسة**

طبق الاختبار بصورته النهائية على عينة الدراسة، والمكونة من (600) طالباً وطالبةً، تم اختيارهم بطريقة العينة العشوائية الطبقية: (445) من المدارس الحكومية، و(155) من المدارس الحكومية، وذلك من (8) مدارس حكومية، و(5) مدارس أهلية، تم اختيارها عشوائيا من قائمة المدارس في مدينة جدة. وبعد جمع الأوراق تم حذف الاوراق ذات الإجابات النمطية، والإجابات غير المنطقية، وهي الاوراق التي لوحظ فيها أن الطالب يقدم اجابات عشوائية – كأن يختار دوما نفس البديل، أو يجيب على شكل رسم معين – وقد بلغ عددها (80) ورقة اختبارية، وهكذا تكونت العينة الرئيسية النهائية من (520) طالباً وطالبة: (418) من المدارس الحكومية، و(102) من المدارس الاهلية.

ثانياً: مصطلحات الدراسة

– **اختبار محكي المرجع**: اختبار يستخدم في تقييم أداء الفرد بالنسبة لمحك، أو مستوى أداء، دون مقارنته بأداء الاخرين.

ويعرف اجرائياً: مجموعة من الفقرات من نوع الاختيار من متعدد (أربعة بدائل أحدها صحيح)، أعدت وفق نظرية استجابة الفقرة، بما يحقق الدقة، والموضوعية، في قياس تحصيل طلاب الصف السادس الابتدائي في مقرر العلوم.

– **نظرية استجابة الفقرة**: نظرية حديثة في القياس النفسي والتربوي، ويتم فيها تحديد العلاقة بين أداء المفحوص والسمة الكامنة موضوع القياس وفق دالة رياضية محددة.

ثالثاً: أداة الدراسة

الأداة المستخدمة في هذه الدراسة هي اختبار تحصيلي محكي المرجع، في العلوم للصف السادس الابتدائي، مصمم وفق نظرية استجابة الفقرة. ويتكون من (75) فقرة من نوع الاختيار من متعدد. وسيتم توضيح إجراءات بناء هذا الاختبار عند عرض إجراءات الدراسة.

رابعاً: صدق الأداة وثباتها

تم التحقق من صدق الاختبار التحصيلي محكي المرجع من خلال:

1. **صدق المحتوى**: حيث تم عرض الفقرات بصورتها الأولية، على مجموعة من المحكمين، وهم من المدرسين الذين يدرسون مقرر العلوم للصف السادس، وعددهم (10) مدرسين. ومن المختصين بأساليب وطرق تدريس العلوم، وعددهم (8) محكمين. وبناء على اقتراحاتهم أجريت تعديلات على بعض فقرات الاختبار، وحذفت (18) فقرة، اجمع على الأقل (75 %) من المحكمين على أنها مكررة، أو غير مناسبة.

2. الصدق العاملي: استخدم التحليل العاملي الاستكشافي (Exploratory Factor Analyses) باستخدام تحليل المكونات الأساسية (Principal Component Analyses)، وحساب نسبة التباين المفسر (Explained Variance) لكل عامل من العوامل، وهذا ما سيتم عرضه بالتفصيل خلال عرض النتائج.

3. **الصدق** بدلالة المحك (الصدق التلازمي): وذلك بتطبيق الاختبار على عينة استطلاعية، مكونة من (70) طالباً وطالبةً من خارج عينة الدراسة. وحساب معامل ارتباط بيرسون، بين درجاتهم على الاختبار، ودرجاتهم على المحك (درجاتهم على مقرر العلوم في كشف درجات المدرسة للفصل الدراسي الأول). وقد بلغت قيمة معامل الارتباط (0.85)، وهي قيمة دالة احصائياً مما يشير لتوفر الصدق بدلالة محك.

تم حساب ثبات الاتساق الداخلي لفقرات الاختبار باستخدام معامل ثبات كرونباخ الفا، بعد التطبيق على العينة الاستطلاعية السابقة، وبلغت قيمته (0.89). أما قيمة معامل ثبات الاختبار محكي المرجع ووفقاً لمعادلة ليفنجستون، وكما أشار عودة (2010)، فقد بلغت (0.88) وهي قيمة مرتفعة. كذلك تم استخدام برنامج البيولوج (BILOG-MG3) في حساب معامل ثبات الاختبار، وفق نظرية استجابة الفقرة، بعد تطبيقه على عينة الدراسة الكلية (491) طالباً وطالبة، وكانت قيمته –بناء على نتائج البرنامج– (0.95)، مما يؤكد ان الاختبار يتمتع بدرجة عالية من الثبات.

خامساً: اجراءات الدراسة

لبناء اداة الدراسة اتبعت خطوات بناء الاختبارات التحصيلية محكية المرجع، إذ جرى:

1. تحديد الغرض من الاختبار: وهو قياس تحصيل الطلاب في مقرر العلوم للصف السادس.

2. تحديد المحتوى المراد قياسه: فبعد حصر المفردات التي يتم تدريسها في مقرر العلوم للصف السادس الابتدائي، وعرض هذه المفردات على: مجموعة من مدرسي مقرر العلوم ، والمشرفين التربويين للأخذ بآرائهم. تم الاتفاق على أن مفردات المقرر هي (تنوع الحياة، عمليات الحياة، الانظمة البيئية، الفضاء، المادة، القوى والطاقة).

3. صياغة الأهداف السلوكية المتعلقة بمفردات المقرر، وذلك بالاستعانة بمجموعة من المختصين في القياس و التقويم.

4. إعداد جدول المواصفات: والذي يربط الوزن النسبي لكل موضوع من موضوعات المقرر، بالوزن النسبي لكل مستوى من مستويات الأهداف.

5. صياغة فقرات الاختبار: بالاستعانة بجدول المواصفات، تم وضع (100) فقرة اختباريه من نوع الاختيار من متعدد، يتكون كل منها من أربعة بدائل، حيث تم صياغة هذه الفقرات وفقاً

للأهداف السلوكية، بحيث لا تعتمد إجابة أي فقرة منها، على إجابة فقرة أخرى في الاختبار، وفق ما تتطلب نظرية استجابة الفقرة (Hattie,1985).

6. التحقق من صلاحية الفقرات (صدق المحتوى): كما عرض سابقاً عند الحديث عن صدق الأداة.

7. إخراج كراسة الاختبار: تم إخراج كراسة الاختبار، والتي تكونت من (82) فقرة من نوع الاختيار من متعدد، لكل منها (4) بدائل تمثل إحداها إجابةً صحيحةً، كما تم تجهيز نموذج الإجابة، لنقل الإجابة إليه.

8. التجريب الأولي للاختبار: طبق الاختبار بصورته الأولية، والتي تكونت من (82) فقرة، على عينة تجريبية – وهي عينة جديدة، وتختلف عن العينة الاستطلاعية المذكورة سابقا– مكونة من (80) طالباً وطالبةً، من خارج عينة الدراسة، ومن طلبة الصف السادس الابتدائي، وذلك في الفصل الدراسي الثاني، من العام الدراسي (2015 – 2016م) الموافق (1436 – 1437 هـ). وذلك بهدف التعرف على صعوبة، وتمييز الفقرات، ووضوح الصياغة اللغوية لها .

بعد فحص أوراق الإجابات استبعدت أوراق إجابة (7) طلاب، لوحظ فيها تكرار اختيار الطالب لأكثر من بديلين، عند الإجابة عن بعض الأسئلة. كما لوحظ في ورقة أحد الطلاب اتباع نمط معين في الإجابة، مما يدل على عدم الجدية في الإجابة عن الأسئلة. وبالتالي تكونت العينة التجريبية النهائية من أوراق اختبار (72) طالب وطالبة.

استخدمت مبادئ النظرية الكلاسيكية في تحليل بيانات التجريب، باستخدام برنامج التحليل الإحصائي (SPSS). حيث تم حساب كلٍ من: معامل الصعوبة (نسبة الطلاب الذين أجابوا إجابة صحيحة عن الفقرة)، ومعامل التمييز (معامل الارتباط بين الدرجة على الفقرة والدرجة الكلية على الاختبار). وذلك لكل فقرة من فقرات الاختبار بصورته المطبقة على عينة التجريب. وتراوحت قيم معاملات الصعوبة بين (0.24 – 0.94)، بمتوسط حسابي (0.53)، كما تراوحت قيم معاملات التمييز بين (0.13 – 0.94)، بمتوسط حسابي (0.60)، وتبعا لما أورده عودة (2010) يجري حذف الفقرات الصعبة جدا، لذلك حذفت الفقرة (9)، ومعامل صعوبتها (0.24)، كما تم الاحتفاظ بالفقرات التي تزيد قيمة معامل تمييزها عن (0.39)، وبالتالي حذفت الفقرات (14، 22، 30، 41، 62، 73)، والتي قيمة معامل تمييزها على الترتيب (0.25، 0.32، 0.13، 0.34، 0.29، 0.35)، أي تم حذف (7) فقرات والاحتفاظ بـ (75) فقرة. إضافة لذلك لم يتذمر أي طالبٍ، خلال تطبيق الاختبار من عدم وضوح الصياغة اللغوية، لأي فقرة من فقرات الاختبار.

9. التطبيق النهائي للاختبار: تم تطبيق الاختبار بصورته المكونة من (75) فقرة، على عينة الدراسة النهائية، وذلك في الفصل الدراسي الثاني، من العام الدراسي (2015 – 2016) الموافق (1436 – 1437 هـ).

10. التحليل الإحصائي، باستخدام كل من برنامج (Bilog-MG3) وبرنامج (SPSS).

سادساً: طريقة تحليل البيانات

بعد تطبيق الاختبار على عينة الدراسة، صححت جميع الاوراق وادخلت لبرنامج الحزمة الإحصائية للعلوم الاجتماعية (SPSS)، مرتان ففي المرة الاولى تم إعطاء الدرجة (1) للإجابة الصحيحة، والدرجة (صفر) للإجابة غير الصحيحة عن الفقرة. وبذلك كانت الدرجة الكلية على الاختبار مساوية لمجموع عدد الفقرات التي اجاب عنها الطالب بصورة صحيحة. وفي المرة الثانية ادخلت البدائل بحيث تم اعطاء الدرجة (1) عند اختيار البديل (أ)، والدرجة (2) عند اختيار البديل (ب)، والدرجة (3) عند اختيار البديل (ج)، والدرجة (4) عند اختيار البديل (د).

استخدم برنامج (SPSS) في حساب كل من التحليل العاملي، والاوساط الحسابية، ومعامل الارتباط، ومعامل ثبات كرونباخ الفا، مربع كاي. كما استخدم برنامج (BILOG-MG3) في حساب كل من: احصائي مربع كاي للمطابقة، ومعالم الصعوبة، والتمييز، والتخمين للفقرات.

نتائج ومناقشة نتائج الدراسة:

أولا- النتائج والمناقشة المتعلقة بالسؤال الأول:

للإجابة عن السؤال (ما مدى تحقيق بيانات اختبار العلوم للصف السادس لافتراضات نظرية استجابة الفقرة ؟) تم التحقق من افتراضات نظرية استجابة الفقرة وهي:

1. **أحادية البعد**: وذلك باستخدام التحليل العاملي (Hambleton & Swaminathan,1985)، بعد التأكد من تحقق شروط التحليل العاملي وهي:

– أن تختلف قيمة محدد مصفوفة معاملات الارتباط بين فقرات الاختبار (Correlation Coefficient Sample Determinant) عن الصفر، وقد أظهرت نتائج برنامج (SPSS) أن قيمة محدد المصفوفة= 0.049، وهي قيمة لا تساوي الصفر، مما يؤكد تحقق هذا الشرط.

– فحص ملاءمة المعاينة (Sampling Adequacy): باستخدم معامل كايزر– مايز–اولكن (Kaiser-Meyer-Olkin (KMO)) في فحص ملاءمة المعاينة، ووفق ما ذكره كايزر (Kaiser,1974) فان قيمة هذا المعامل يجب أن تكون (0.5) على الأقل، ويحكم على ملاءمة المعاينة وفق الدلالات التالية، فالملاءمة مقبولة إن وقعت قيمة (KMO) بين (0.5– 0.7)، وجيدة إن وقعت في المدى (0.7 – 0.8)، أما القيمة (0.8 – 0.9) فهي قيمة كبيرة، والقيمة فوق (0.9) كبيرة جداً. وأظهر برنامج (SPSS) أن قيمة معامل (KMO) في هذه الدراسة (0.95) وهي قيمة كبيرة جداً، وفق ما ذكره كايزر، مما يؤكد تحقق شرط ملاءمة المعاينة.

– فحص تجانس العينة بالنسبة لحجم العينة: وهو ما يوضحه مستوى دلالة مربع كاي ($\chi 2$) لاختبار بارتليت (Bartlett›s Test of Sphericity)، وأظهرت نتائج برنامج (SPSS) أن قيمة مربع كاي لاختبار بارتليت (15263)، وذلك عند درجة حرية (2775)، ومستوى دلالة ($\alpha = 0.000$) مما يؤكد تحقق هذا الشرط.

بعد التأكد من تحقق شروط التحليل العاملي، استخدم

للتحليل العاملي الاستكشافي بطريقة المكونات الأساسية (Principal Component Method)، وبتدوير العوامل على محاور متعامدة بطريقة فاريماكس (Varimax)، للتأكد من أن الاختبار يقيس سمة واحدة. كما تم حساب قيمة الجذر الكامن (EignValue)، ونسبة التباين المفسر (Explained Variance)، والتباين المفسر التراكمي (Cumulative Explained Variance) لكل عامل من العوامل، ويظهر جدول (1) قيم الجذر الكامن، والتباين المفسر للعوامل المستخلصة من التحليل العاملي والتي قيمة جذرها الكامن واحد فأكثر.

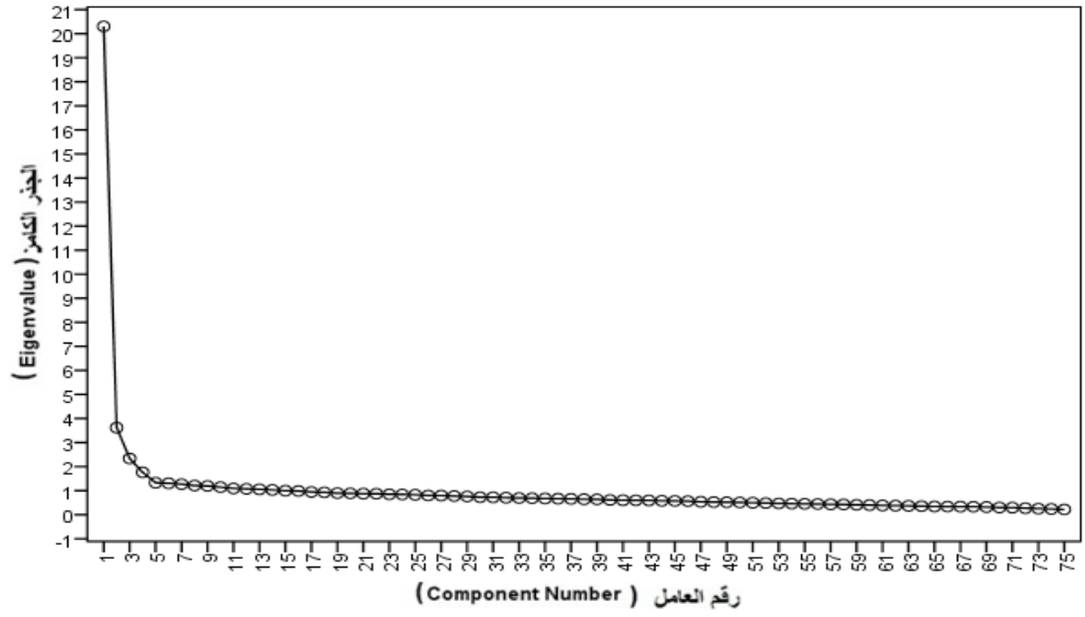
جدول (1)

قيم الجذر الكامن والتباين المفسر للعوامل المستخلصة من التحليل العاملي والتي قيمة جذرها الكامن واحد فأكثر

رقم العامل	الجذر الكامن	نسبة التباين المفسر	نسبة التباين التراكمي المفسر	رقم العامل	الجذر الكامن	نسبة التباين المفسر	نسبة التباين التراكمي المفسر
1	20.305	27.073	27.073	8	1.214	1.619	44.186
2	3.617	4.823	31.896	9	1.195	1.594	45.780
3	2.337	3.116	35.012	10	1.145	1.527	47.307
4	1.763	2.351	37.363	11	1.094	1.458	48.765
5	1.331	1.774	39.138	12	1.073	1.431	50.196
6	1.305	1.740	40.878	13	1.055	1.406	51.603
7	1.267	1.689	42.567	14	1.030	1.373	52.975

على الرغم من أن جدول (1) يظهر أن هناك (14) عاملا كانت قيمة الجذر الكامن لكل منها أكبر من (1) – وهذه العوامل هي العوامل التي يتم الاحتفاظ بها (Crocker & Algina,1986) – وهي تفسر معا ما نسبته (52.98 %) من تباين الدرجات على الاختبار؛ فان ذلك لا ينفي تحقق شرط أحادية البعد، فوفق ما اشار إليه ريكاس(Reckase,1979) فان شرط أحادية البعد يتحقق إن لم تقل نسبة التباين المفسر للعامل الأول عن (20 %)، وهذا الشرط متحقق، كما يُظهر جدول (1) أن نسبة التباين المفسر للعامل الاول (27 %)، مما يؤكد أن العامل الأول هو العامل المسيطر. أي أن هناك سمة كامنة واحدة هي المسؤولة عن تفسير التباين، في درجات الاختبار (Hambelton & Swaminathan,1985;Hattie,1985).

كذلك وتبعا لما أورده لورد (Lord,1980) إن كانت نسبة قيمة الجذر الكامن للعامل الأول، إلى قيمة الجذر الكامن للعامل الثاني تزيد عن (2) فهذا مؤشر لأحادية البعد. وهذا الشرط متحقق أيضاً في هذه الدراسة، فنسبة الجذر الكامن للعامل الأول إلى الجذر الكامن للعامل الثاني (5.6) وهي أكبر من (2). ويظهر شكل (1) رسماً بيانياً لقيم الجذور الكامنة للعوامل المرشحة للاستخلاص، من تحليل المكونات الأساسية. حيث يظهر ارتفاع قيمة الجذر الكامن للعامل الأول، مقارنة بباقي العوامل التي يمكن استخلاصها وهذا مؤشر أيضا على تحقق أحادية البعد. بناء على ما سبق يمكن قبول الافتراض بان: الاختبار يقيس سمة أحادية البعد، وهو أول افتراضات النموذج ثلاثي المعلم.



شكل (1)

الرسم البياني (scree plot) للعوامل المستخلصة من تحليل المكونات الأساسية

2. الاستقلال الموضعي :إن تحقق افتراض أحادية البعد، يتضمن تحقق افتراض الاستقلال الموضعي (Hambleton & Swaminathan,1985).

3. الاختبار ليس اختبار سرعة: يمكن التأكد من أن الاختبار ليس اختبار سرعة، من خلال نسبة الطلبة الذين أكملوا الاختبار. فوفقا لما ذكره هامبلتون وآخرون (Hambleton et al,1991)، إن أكمل (75 %) من الطلبة الإجابة عن فقرات الاختبار، أو إن تم الإجابة عن (80 %) من الفقرات، فإن عامل السرعة عندئذ لن يكون عاملاً مهماً في الأداء على الاختبار. وبفحص أوراق إجابة الطلاب لاحظت الباحثة أن جميع الطلاب اجابوا عن جميع الأسئلة، ولم يتركوا أي سؤال دون إجابة، مما يؤكد أن عامل السرعة، لم يكن عاملاً مهماً في الأداء على الاختبار. كذلك أكمل جميع الطلاب الإجابة عن الاختبار قبل انتهاء الزمن المحدد للإجابة، ولم يتذمر أيٌ منهم من ضيق الوقت.

لفحص مؤشرات المطابقة الخاصة بالأفراد، تم تقدير: القدرة، والخطأ المعياري، وإحصائي المطابقة لكل فرد من أفراد الدراسة. وقد أظهرت النتائج عدم مطابقة اجابات (29) طالباً للنموذج ثلاثي المعلم. فعند فحص مطابقة أي نموذج من نماذج استجابة الفقرة، نجد أن نمط استجابة بعض الأفراد لا يطابق افتراضات النموذج، وهؤلاء الافراد يجب حذفهم قبل حساب معالم الفقرات، لأنها ستؤثر على دقة تقديرات معالم الفقرات، وتعلل الباحثة عدم مطابقة هؤلاء الافراد بعدم مصداقيتهم خلال الإجابة عن الفقرات، وبعد حذف الأفراد غير المطابقين، تم فحص مطابقة الفقرات باستخدام إحصائي مربع كاي، وذلك بتجميع المفحوصين في (9) فئات، بناء على قدراتهم (Zimowski, Muraki, Mislevy & Bock, 2003). حيث يفحص هذا الإحصائي الفرضية الصفرية القائلة باتساق الدرجات على الفقرات مع النموذج، وبالتالي تتحقق مطابقة الفقرة للنموذج، إن كانت قيمة مربع كاي غير دالة إحصائياً (Traub & Lam, 1985).

كما تم تقدير قيم كل من: معلم الصعوبة (b)، ومعلم التمييز (a)، ومعلم التخمين (c)، والخطأ المعياري لكل من هذه القيم، ومستوى الدلالة لاختبار مربع كاي لفحص مطابقة فقرات الاختبار. وأظهرت النتائج عدم مطابقة (14) فقرة من فقرات الاختبار للنموذج ثلاثي المعلم – الذي يأخذ بعين الاعتبار: اختلاف قيم معالم الصعوبة، والتمييز، وتخمين الفقرات، مما يجعله نموذجا ناجحا من الناحية العملية – وهي الفقرات: (2، 6، 7، 18، 20، 22، 24، 29، 30، 31، 47، 49، 51، 65)، وبالتالي طابقت (61) فقرة النموذج. وبعد حذف الفقرات غير المطابقة تم حساب القدرة والخطأ المعياري في تقدير قدرة الطلاب الذين تطابقوا مع النموذج ثلاثي المعلم وعددهم (491). وقد كانت قيمة المتوسط الحسابي لتقدير قدرة الطلاب (0.066–)، بانحراف معياري (0.84).

ثانيا: النتائج والمناقشة المتعلقة بالسؤال الثاني

للإجابة عن السؤال (هل يتضمن اختبار العلوم للصف السادس فقرات ذات أداء تفاضلي؟) تم تحليل فقرات كل مجموعة من مجموعتي (المدارس الاهلية والمدارس الحكومية)، باستخدام برنامج البيولوج (Bilog-MG3)، وذلك للكشف عن الأداء التفاضلي لفقرات الاختبار، بحساب الفرق بين معاملي الصعوبة لمجموعتي: المدارس الاهلية والمدارس الحكومية، على كل فقرة من الفقرات. ووفقا لما ذكره الاجومالي وكيفيس (Alagumalai & Keeves,1999) فان الفقرة تعتبر ذات أداء تفاضلي إن كان الفرق المطلق بين قيمتي معلم الصعوبة للمجموعتين، أكبر من أو يساوي (0.5). ويظهر جدول (2) الفرق بين معلمي الصعوبة لمجموعتي المدارس الاهلية والحكومية.

جدول (2)

الفرق بين معلمي الصعوبة لمجموعتي المدارس الاهلية والحكومية

رقم الفقرة	معلم الصعوبة للمجموعة الأولى (الحكومية)	معلم الصعوبة للمجموعة الثانية (الاهلية)	الفرق المطلق بين معلمي الصعوبة	قرار فحص الأداء التفاضلي	رقم الفقرة	معلم الصعوبة للمجموعة الأولى (الحكومية)	معلم الصعوبة للمجموعة الثانية (الاهلية)	الفرق المطلق بين معلمي الصعوبة	قرار فحص الأداء التفاضلي
1	-0.634	-0.832	0.198	لا يوجد	42	-1.650	-1.697	0.047	لا يوجد
3	-1.765	-2.346	0.581	يوجد	43	-6.704	-4.889	1.815	يوجد
4	-1.127	-1.634	0.506	يوجد	44	-1.111	-0.956	0.155	لا يوجد
5	-0.675	-1.527	0.852	يوجد	45	-0.987	-1.302	0.316	لا يوجد
8	-1.135	-1.606	0.471	لا يوجد	46	-1.478	-1.204	0.274	لا يوجد
9	-0.634	-1.335	0.702	يوجد	48	-2.645	-3.255	0.610	يوجد
10	-0.904	-1.191	0.288	لا يوجد	50	-1.836	-1.928	0.092	لا يوجد
11	-4.656	-4.471	0.185	لا يوجد	52	-1.906	-1.451	0.456	لا يوجد
12	-0.641	-0.923	0.282	لا يوجد	53	-0.564	-0.865	0.301	لا يوجد
13	-0.965	-1.591	0.626	يوجد	54	-0.832	-0.606	0.227	لا يوجد
14	-0.989	-1.418	0.430	لا يوجد	55	-1.030	-0.962	0.068	لا يوجد
15	-1.559	-1.981	0.422	لا يوجد	56	-0.937	-1.507	0.569	يوجد

رقم الفقرة	معلم الصعوبة للمجموعة الأولى (الحكومية)	معلم الصعوبة للمجموعة الثانية (الاهلية)	الفرق المطلق بين معلمي الصعوبة	قرار فحص الأداء التفاضلي	رقم الفقرة	معلم الصعوبة للمجموعة الأولى (الحكومية)	معلم الصعوبة للمجموعة الثانية (الاهلية)	الفرق المطلق بين معلمي الصعوبة	قرار فحص الأداء التفاضلي
16	-3.368	-3.382	0.013	لا يوجد	57	-0.514	-0.918	0.403	لا يوجد
17	-2.477	-1.982	0.494	لا يوجد	58	-0.364	-1.255	0.891	يوجد
19	-0.524	-0.871	0.347	لا يوجد	59	-2.643	-2.625	0.018	لا يوجد
21	-4.658	-5.986	1.328	يوجد	60	-0.662	-0.644	0.018	لا يوجد
23	-4.125	-5.099	0.974	يوجد	61	-2.356	-2.033	0.323	لا يوجد
25	-0.387	-0.857	0.470	لا يوجد	62	0.088	-0.078	0.166	لا يوجد
26	-1.243	-0.878	0.366	لا يوجد	63	-1.274	-2.926	1.652	يوجد
27	-0.669	-0.971	0.303	لا يوجد	64	-0.662	-1.008	0.346	لا يوجد
28	-3.052	-3.272	0.219	لا يوجد	66	-0.836	-0.804	0.032	لا يوجد
32	-0.524	-0.817	0.293	لا يوجد	67	-1.081	-1.339	0.259	لا يوجد
33	-0.917	-1.181	0.264	لا يوجد	68	-0.695	-0.615	0.081	لا يوجد
34	-1.641	-2.762	1.121	يوجد	69	-0.856	-0.731	0.125	لا يوجد
35	-3.406	-3.573	0.167	لا يوجد	70	-0.947	-1.003	0.057	لا يوجد
36	-2.747	-3.217	0.470	لا يوجد	71	0.437	0.494	0.057	لا يوجد
37	-1.126	-0.882	0.244	لا يوجد	72	-1.257	-1.168	0.089	لا يوجد
38	-3.401	-4.510	1.109	يوجد	73	-0.899	-1.351	0.452	لا يوجد
39	-1.328	-2.395	1.067	يوجد	74	-2.838	-3.928	1.090	يوجد
40	-3.334	-3.812	0.477	لا يوجد	75	-1.196	-1.550	0.354	لا يوجد
41	-1.713	-1.642	0.071	لا يوجد					

يظهر جدول (2) تحقق شرط الأداء التفاضلي في (16) فقرة من فقرات الاختبار، وهي الفقرات (3، 4، 5، 9، 13، 21، 23، 34، 38، 39، 43، 48، 56، 58، 63، 74)، ويمكن القول مبدئياً أن هذه الفقرات، هي فقرات متحيزة، ولكن وبعد عرض هذه الفقرات على مجموعة من المحكمين المختصين بمناهج العلوم والقياس والتقويم – دون لفت انتباههم لأهمية فحص المموهات–أكد هؤلاء المحكمون أن هذه الفقرات ليست فقرات متحيزة. ومن ثمّ، فإن ظهور بعض الفقرات المتحيزة في الاختبار، يؤكد أن الاكتفاء بحساب معاملات الصعوبة والتمييز للحكم على فقرات الاختبار غير كافٍ، وهذا يتفق مع ما أوره كل من كراكايا (Karakaya, 2012)، وبارناباس (Barnabas, 2012)، واكسيومي (Xiaomei,2014) حول ضرورة قيام مصمم أي اختبار بدراسة استطلاعية للاختبار؛ لاستبعاد الفقرات التي تتضمن أداءً تفاضلياً، كما تتفق أيضا مع دراسة اموتش وفان (Amuche & Fan ,2014) التي أظهرت تحيز عدد من فقرات اختباره في الرياضيات لنوع المدرسة (حكومية/ خاصة)، مما يؤكد مرة أخرى ضرورة فحص الأداء التفاضلي لفقرات أي اختبار، قبل تطبيقه بصورة نهائية.

ثالثا: النتائج والمناقشة المتعلقة بالسؤال الثالث

للإجابة عن السؤال الثالث: (هل كان للمموهات دور في اظهار الأداء التفاضلي لبعض فقرات الاختبار؟). وللكشف عما إن كان الأداء التفاضلي للفقرات – التي أظهرت الأداء التفاضلي عند الإجابة عن السؤال السابق – ناتجا عن نص الفقرة، أم عن البدائل (المموهات) تم تقسيم أفراد العينة – في كل من مجموعتي المدارس الحكومية، والمدارس الاهلية – تبعا لمتوسط الدرجات على الاختبار في كل مجموعة، إلى مجموعتين: مجموعة الأداء الدنيا وعددهم (192، 33) طالباً من المدارس الحكومية والمدارس الاهلية على الترتيب، ومجموعة الأداء العليا وعددهم (205، 61) طالباً من المدارس الحكومية والمدارس الاهلية على الترتيب. استخدم اختبار مربع كاي ($\chi 2$) للكشف عن أي فروق دالة احصائيا، بين اعداد الطلاب الذين اختاروا كل بديل من البدائل، في كل من مجموعتي: الأداء الدنيا، والأداء العليا في كل من مجموعتي: المدارس الحكومية، والمدارس الاهلية. ويظهر جدول (3) نتائج اختبار مربع كاي للكشف عن الأداء التفاضلي لمموهات الفقرات، التي أظهرت النتائج تحيزها.

يظهر جدول (3) وجود أداء تفاضلي للفقرات (4، 23، 39،

43)، وذلك بين افراد فئة الأداء العليا من طلاب المدارس الحكومية وطلاب المدارس الاهلية، ووجود أداء تفاضلي للفقرات (13، 39، 43، 58) بين أفراد فئة الأداء الدنيا من طلاب المدارس الحكومية وطلاب المدارس الاهلية. ويظهر جدول (4) تحليل إجابات أفراد لكل من فئة الأداء العليا وفئة الأداء الدنيا، على بدائل الفقرات ذات الأداء التفاضلي للمموهات.

جدول (3)

اختبار مربع كاي للكشف عن الأداء التفاضلي لمموهات الفقرات

م	مجموعة الأداء	المجموعة المرجعية (الحكومية)				المجموعة المستهدفة (الاهلية)				قيمة مربع كاي	درجة الحرية	مستوى الدلالة	قرار الأداء التفاضلي للمموهات
		البديل أ	البديل ب	البديل ج	البديل د	البديل أ	البديل ب	البديل ج	البديل د				
3	الدنيا	24	38	27	103*	2	9	6	16*	1.80	2	0.41	لا يوجد
	العليا	18	13	11	163*	6	1	3	51*	1.88	2	0.39	لا يوجد
4	الدنيا	119*	25	27	21	23*	3	5	2	0.68	2	0.71	لا يوجد
	العليا	190*	7	1	7	58*	0	3	0	12.60	2	0.00	يوجد
5	الدنيا	31	94*	43	24	2	24*	4	3	0.49	2	0.78	لا يوجد
	العليا	6	185*	8	6	0	60*	0	1	2.10	2	0.35	لا يوجد
9	الدنيا	32	95*	41	24	1	25*	5	2	1.68	2	0.43	لا يوجد
	العليا	5	188*	7	5	0	59*	1	1	0.86	2	0.65	لا يوجد
13	الدنيا	28	29	121*	14	0	3	25*	5	8.65	2	0.01	يوجد
	العليا	3	4	196*	2	0	1	60*	0	1.11	2	0.57	لا يوجد
21	الدنيا	138*	17	19	18	25*	5	2	1	3.09	2	0.21	لا يوجد
	العليا	183*	8	9	5	56*	1	2	2	0.80	2	0.67	لا يوجد
23	الدنيا	22	124*	23	23	6	21*	4	2	1.87	2	0.39	لا يوجد
	العليا	10	176*	11	8	0	54*	0	7	12.17	2	0.00	يوجد
34	الدنيا	104*	29	29	30	25*	3	3	2	0.27	2	0.87	لا يوجد
	العليا	176*	8	7	14	51*	2	2	6	0.42	2	0.81	لا يوجد
38	الدنيا	14	16	20	142*	1	5	2	25*	2.82	2	0.24	لا يوجد
	العليا	7	7	6	185*	2	0	1	58*	1.73	2	0.42	لا يوجد
39	الدنيا	33	31	33	95*	1	1	10	21*	10.87	2	0.00	يوجد
	العليا	7	14	13	171*	7	1	1	52*	10.60	2	0.01	يوجد
43	الدنيا	9	17	143*	23	6	3	22*	2	6.63	2	0.04	يوجد
	العليا	12	6	181*	6	2	4	48*	7	4.77	2	0.09	يوجد
48	الدنيا	13	159*	13	7	1	29*	2	1	0.32	2	0.85	لا يوجد
	العليا	4	193*	3	5	1	58*	1	1	0.10	2	0.95	لا يوجد
56	الدنيا	22	31	32	107*	2	4	3	24*	0.22	2	0.90	لا يوجد
	العليا	7	10	6	182*	4	1	2	54*	2.30	2	0.32	لا يوجد
58	الدنيا	43	76*	40	33	7	15*	10	1	5.09	2	0.08	يوجد
	العليا	21	130*	22	32	11	37*	8	5	4.20	2	0.12	لا يوجد
63	الدنيا	99*	35	31	27	23*	5	3	2	0.65	2	0.72	لا يوجد

م	مجموعة الأداء	المجموعة المرجعية (الحكومية)				المجموعة المستهدفة (الاهلية)				قيمة مربع كاي	درجة الحرية	مستوى الدلالة	قرار الأداء التفاضلي للمموهات
		البديل أ	البديل ب	البديل ج	البديل د	البديل أ	البديل ب	البديل ج	البديل د				
	العليا	178*	11	7	9	59*	1	0	1	0.71	2	0.70	لا يوجد
74	الدنيا	154×	12	12	14	29×	2	1	1	0.56	2	0.76	لا يوجد
	العليا	193×	5	5	2	59×	1	0	1	1.75	2	0.42	لا يوجد

* الإجابة الصحيحة

فيما يتعلق بطلاب فئة الأداء العليا، والذين كانت اجابتهم غير صحيحة على الفقرات، التي أظهرت مموهاتها أداء تفاضلياً، فكما يتضح من جدول (4) توجد فروق في اختيار المموه بين طلاب المدارس: الحكومية والأهلية، ففي الفقرة (4) اختار طلاب المدارس الحكومية المموهين (ب) و (د)، بينما لم يختر أي من طلاب المدارس الاهلية هذين المموهين فجميعهم اختاروا المموه (ج)، على انه الإجابة الصحيحة. وفيما يتعلق بالفقرة (23)، فلم تجذب المموهات (أ) و (ج) أياً من طلاب المدارس الاهلية، الذين اختاروا جميعاً المموه (د) كإجابة صحيحة للفقرة. وفي الفقرة (39) كان المموه (أ) هو الأكثر جذباً لطلاب المدارس الاهلية، بينما توزعت إجابات طلاب المدارس الحكومية، على المموهات الثلاث (أ) و (ب) و (ج). وأخيرا بالنسبة للفقرة (43)، كان المموه (د)، هو الأكثر جذباً لطلاب المدارس الاهلية، بينما كان المموه (أ) هو الأكثر جذباً لطلاب المدارس الحكومية.

فيما يتعلق بطلاب فئة الأداء الدنيا، والذين كانت اجاباتهم غير صحيحة على الفقرات، التي أظهرت مموهاتها أداءً تفاضلياً. فكما يتضح من جدول (4): بالنسبة للفقرة (13) فقد جذب المموهان (أ) و (ب) غالبية طلاب المدارس الحكومية، بينما اختار غالبية طلاب المدارس الاهلية المموه (د). وبالنسبة للفقرة (39): فقد جذب المموه (ج) طلاب المدارس الاهلية بشكل ملفت للنظر، بينما توزعت خيارات طلاب المدارس الحكومية على المموهات الثلاث (أ) و (ب) و (ج). وفيما يتعلق بالفقرة (43): كان المموه (أ) أكثر جذباً لطلاب المدارس الاهلية، بينما كان المموه (د) هو الأكثر جذباً لطلاب المدارس الحكومية. وأخيرا وبتدقيق اختيار طلاب فئة الأداء الدنيا، لمموهات الفقرة (58) فيلاحظ عدم جذب المموه (د) لطلاب المدارس الاهلية بنفس درجة جذبه لطلاب المدارس الحكومية.

جدول (4)

تحليل إجابات أفراد فئتي الأداء العليا والدنيا على بدائل الفقرات ذات الأداء التفاضلي للمموهات

رقم الفقرة	البديل	فئة الأداء العليا				رقم الفقرة	البديل	فئة الأداء الدنيا			
		المدارس الحكومية		المدارس الاهلية				المدارس الحكومية		المدارس الاهلية	
		العدد	النسبة المئوية	العدد	النسبة المئوية			العدد	النسبة المئوية	العدد	النسبة المئوية
4	أ*	190	93.5	58	95	13	أ	28	15	0	0
	ب	7	3	0	0		ب	29	15	3	9
	ج	1	0.5	3	5		ج*	121	63	25	76
	د	7	3	0	0		د	14	7	5	15
23	أ	10	5	0	0	39	أ	33	17	1	3
	ب*	176	86	54	89		ب	31	16	1	3
	ج	11	5	0	0		ج	33	17	10	30
	د	8	4	7	11		د*	95	50	21	64
39	أ	7	3	7	12	43	أ	9	5	6	18
	ب	14	7	1	2		ب	17	9	3	9
	ج	13	6	1	2		ج*	143	74	22	67
	د*	171	84	52	84		د	23	12	2	6

رقم الفقرة	البديل	فئة الأداء العليا				رقم الفقرة	البديل	فئة الأداء الدنيا			
		المدارس الحكومية		المدارس الاهلية				المدارس الحكومية		المدارس الاهلية	
		العدد	النسبة المئوية	العدد	النسبة المئوية			العدد	النسبة المئوية	العدد	النسبة المئوية
43	أ	12	6	2	3	58	أ	43	22	7	21
	ب	6	3	4	6		ب*	76	40	15	46
	ج*	181	88	48	79		ج	40	21	10	30
	د	6	3	7	12		د	33	17	1	3

* الإجابة الصحيحة

بعد عرض جدول (4) على مجموعة من المحكمين: من المختصين بمناهج تدريس العلوم ومن المشرفين التربويين لمقرر العلوم، والذين يتابعون كلاً من: المدارس الحكومية والأهلية في مدينة جدة، أكد (75 %) منهم صحة ما يتعلق بالأداء التفاضلي لمموهات هذه الفقرات. وعزوا اختيار طلاب إحدى مجموعتي المدارس الحكومية، أو الاهلية لمموه معين إلى: نوعية الوسائل التعليمية التي تستخدمها تلك المدارس، فبعضها يعتمد فقط على الوسائل التعليمية السمعية، أو البصرية، بينما يعتمد بعضها على توظيف الانترنت. وعزا البعض سبب اختلاف اختيار المموه بين أصحاب نفس القدرة إلى اشتراط بعض المدارس، أن يقدم المعلمون أسئلة مقترحة للطلاب. إن هذه النتيجة تتفق مع دراسة ابيدي وزملاؤه (Abedi et al , 2008) التي أكدت على تأثر الأداء التفاضلي للفقرات بالأداء التفاضلي للمموهات، مما يؤكد على أهمية فحص الأداء التفاضلي للمموهات عند الحكم على الأداء التفاضلي للفقرة.

التوصيات والمقترحات:

في ضوء النتائج السابقة فان الدراسة توصي بما يلي:

- تنظيم برنامج لتبادل الخبرات حول طريقة التدريس بين معلمي المدارس الحكومية والمدارس الاهلية.
- ضرورة فحص الأداء التفاضلي لمموهات الفقرات ذات الأداء التفاضلي، قبل عرضها على المحكمين، لتقديم قرارهم بشأن تحيز الفقرة.
- اجراء مزيد من الدراسات على الطرق المختلفة للكشف عن الأداء التفاضلي للمموهات، والمقارنة بين هذه الطرق.
- اجراء دراسة للمقارنة بين طرق الكشف عن دالة الفقرة التفاضلية في حالة وجود أداء تفاضلي للمموهات.

المصادر والمراجع:

أولاً- المراجع العربية:

1. عودة، أحمد (2010). القياس والتقويم في العملية التدريسية، الطبعة الرابعة، دار الأمل، الأردن.

ثانياً- المراجع الأجنبية:

1. *Abedi, J., Leon, S., & Kao, J. (2008). Examining Differential Distractor Functioning in Reading Assessments for Students with Disabilities. (CRESST Rep. No. 743). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).*
2. *Alagumalai S, Keeves J. (1999). Distractors – Can they be biased too? Journal of Outcome Measurement, 3 (1), 89-102.*
3. *Amuche C. & Fan A. (2014). an Assessment of Item Bias Using Differential Item Functioning Technique in Neco Biology Conducted Examinations in Taraba State Nigeria, American International Journal of Research in Humanities, Arts and Social Sciences, 6(1), March -May, 2014. 95-100.*
4. *Barnabas C. (2012). Analysis of Gender-Related Differential Item Functioning In Mathematics Multiple Choice Items Administered By West African Examination Council (WAEC), Journal of Education and Practice, 3(8),71-78.*
5. *Crocker,L. & Algina, J.(1986). Introduction to classical and modern test theory. New York: CBS College Publishing.*
6. *Hambleton, R. & Swaminathan, H. (1985). Item response theory principles and applications. Boston: Kluwer Nijhoff Publishing.*
7. *Hambleton , R and Swaminithan , H. and Rogers,J.(1991). Fundamentals of item response theory. Sage publications , Newbury perk.*
8. *Hambleton, R. & Rogers, J. (1995). Item bias review. Practical Assessment, Research & Evaluation, 4(6). Retrieved August 1. 2016 from http://PAREonline.net/getvn.asp?v=4&n=6.*
9. *Hattie, J. (1985).Methodology review : Assessing unidimensionality of test and items. Applied Psychological Measurement, 9 , 139-164.*
10. *Herrmann-Abell, C. & DeBoer, G. (2011). Using distractor-driven standards-based multiple-choice items and Rasch modeling to investigate hierarchies of chemistry misconceptions and detect structural problems with individual items. Chemistry Education Research and Practice, 12. 184–192.*
11. *Kaiser, H.F. (1974). An index of factorial simplicity. Psychometrika, 39, 31-36.*
12. *Karakaya, I. (2012). An Investigation of Item Bias in Science & Technology Subtests and Mathematic Subtests in Level Determination Exam (LDE), Educational Sciences: Theory & Practice, 12(1), 222-229.*

13. *Lord, F. (1980). Application of item response theory to practical testing problems. NJ: Lawrence Erlbaum Associates Inc.*
14. *Meyer, P. & Wise, S. (2006). Including item response time in distractor analysis via multivariate kernel smoothing. Paper presented at the 2006 meeting of the National Council on Measurement in Education, San Francisco, CA.*
15. *Reckase, M. (1979). Unifactor trait models applied to multifactor tests: Results and application. Journal of Educational Statistics, 41. 207-230.*
16. *Traub,R. & Lam, R.(1985). Latent Structure and item sampling models for testing. Annual Review of Psychology, 36, 19-48.*
17. *Xiaomei, S. (2014). Research Note: DIF investigations with Pearson Test of English Academic ,unpublished dissertation, Queens' University, Canada,*
18. *Zimowski,M., Muraki, E., Mislevy, R. & Bock, R.(2003). BILOG-MG (Version 3.0). IN: Scientific Software Inc.*