[٢]

قدرة بعض مؤشرات الملاءمة الإحصائية في فرز البيانات الخاصة بإعداد المفردات البنكية في ضوء تحليلات نموذج التقدير الجزئي

د. محمد منصور محمد الشافعي

أستاذ باحث مساعد

قسم بحوث الامتحانات– بالمركز القومي للامتحانات والتقويم التربوي

قدرة بعض مؤشرات الملاءمة الإحصائية في فرز البيانات الخاصة بإعداد المفردات البنكية في ضوء تحليلات نموذج التقدير الجزئي

د. محمد منصور محمد الشافعي*

مقدمة:

لقد أسفرت الاتجاهات الحديثة للقياس عن بعض الأساليب المتطورة التي تستخدم في مختلف إجراءات القياس السلوكي، حيث تتمثل تلك الاتجاهات في بنوك الأسئلة التي يمكن استخدامها فى مختلف مجالات وميادين القياس (Anzaldua,2005,36)، وبصفة خاصة بنوك الأسئلة التى يجرى إنشاؤها في ضوء التحليلات المرتبطة بنظرية الاستجابة للمفردة (Item Response Theory (IRT، (Van der Entink, Rinke , Fox, JeanPaul Linden, 2010) ، وتحتوي هذه البنوك أعداداً كبيرة من المفردات التي تقيس متغير أو صفة معينة في مجال ما من مجالات السلوك الانساني، كما يمكن أن تشتمل هذه البنوك على أنواع مختلفة من المفردات (Ingebo,2005,55)، وتوفر هذه البنوك تدريجا لعدداً ضخماً من الأسئلة يمكن الحصول عليها في ضوء التحليلات الإحصائية التي تعتمد على استخدام بعض النماذج الرياضية المختلفة المرتبطة بالنظرية الحديثة للقياس والتي تتلاءم مع نوع وطبيعة الأسئلة البنكية (Thissen, & Orlando. 2006, 102)، ويمكن من خلال هذا

* أستاذ باحث مساعد، قسم بحوث الامتحانات – بالمركز القومي للامتحانات والتقويم التربوي.

التدريج اشتقاق صوراً مختلفة من الاختبارات البنكية التي أثبتت موضوعيتها في مجال القياس (Loken, Kelly,2011) وتتمثل هذه الصور فى اختبارات مسحوبة من هذا التدريج يمكن من خلالها تفادى مشكلات القياس التي واجهت الاختبارات التقليدية والتي تم إعدادها في ضوء النظرية الكلاسيكية للقياس (hambleton and Van Der Linden 1997, 13)، (Andrich, 2004) وتعتمد إجراءات بناء هذه البنوك على تجريب عدد كبير من المفردات البنكية الجديدة على عينة كبيرة من الأفراد من خلال مواقف اختبارية مختلفة تجرى بالواقع التعليمي بهدف الحصول على البيانات اللازمة التي يمكن معالجتها بطرق إحصائية متخصصة تسفر عن مؤشرات إحصائية لكل من المفردات البنكية والأفراد يعتمد عليها في المراحل المختلفة المتعلقة بإنشاء بنك الأسئلة (كاظم، ١٩٨٨أ، ١٠٨) (علام، ٢٠٠٥، ١٥٧)، كما تعتمد إجراءات بناء هذه البنوك أحيانا على البيانات الخاصة بأسئلة الاختبارات التحصيلية التي تستخدم في الاختبارات النهائية بمختلف المراحل التعليمية كأسئلة يمكن إدراجها أيضا بهذه البنوك (David, G, Seung C., 2010)، ونظراً لأن بعض المواقف الاختبارية المختلفة التي تستخدم في جمع هذه البيانات المتعلقة بإعداد هذه البنوك قد لاتمثل أهمية لعينة الأفراد المستخدمين فقد يلجأ بعضهم إلى استخدام أساليب استجابة لا تتصف بالأمانة والدقة تتمثل في مظاهر مختلفة من اللامبالة وعدم الجدية، وقد تمثل بعض المواقف الاختبارية الأخرى أهمية لدى البعض الآخر وبصفة خاصة تلك التي تتعلق بالامتحانات المدرسية فقد تبدو من خلالها استجابات الغش أوالتخمين، وبالتالي فإن حدوث تلك الأنماط غير الملائمة من

الاستجابات يؤدي في النهاية إلى إحباط محاولات الباحثين في الحصول على بيانات دقيقة تمكنهم من مواصلة الإجراءات التي تستهدف إنشاء بنك الأسئلة، (Moore, M. G. 2008) ونظرا لخطورة اختلاط البيانات الخاصة بإعداد بنوك الأسئلة بالبيانات الناجمة عن هذه النوعية غير الملائمة من الاستجابات التي يمكن أن تؤثر بصورة سلبية على استكمال الإجراءات المستخدمة في إعداد هذه البنوك، دعت الضرورة إلى اختبار ودراسة بعض المؤشرات الإحصائية المتضمنة ببعض البرامج الشائع استخدامها في تحليل البيانات الخاصة بتلك البنوك والتي يمكن أن تساعد في فرز تلك البيانات وتمحيصها.

هذا ويستخدم في بناء هذه البنوك بعض البرامج الإحصائية المحوسبة فى تحليل وتمحيص كل من المفردات الاختبارية وكذلك استجابات الأفراد المتلقين لهذه المفردات بهدف استبعاد ما لا يصلح منها وإبقاء ما يصلح لاستكمال الإجراءات الخاصة بإنشاء بنك الأسئلة، ومن بين البرامج الشائعة الاستخدام فى مثل هذه النوعية من الإجراءات برنامج "WINSTEPS" حيث تتضمن التحليلات الإحصائية بهذا البرنامج على مؤشرين لملاءمة كل من المفردات والأفراد لعمليات التحليل وهما مؤشرى الملاءمة التقاربى Infit، والتباعدى Outfit (Liancer, 2005: 127)، وعلى الرغم من أن هذين المؤشرين يشتملان على حدود للملاءمة يمكن من خلالها تحديد البيانات المناسبة لاستكمال إجراءات التحليل، فإن هذين المؤشرين يعوزهما مصداقية الواقع التجريبى والتطبيقى الذي يمكن من خلاله الحكم على صلاحية هذين المؤشرين فى عمليات فرز البيانات المستهدفة من التحليل، وبالتالي كانت هناك حاجة ملحة تستهدف اختبار قدرة هذين المؤشرين

على فرز البيانات الناتجة عن أشكال ملائمة وغير ملائمة من الاستجابات، وهذا مما حدا بالباحث إلى تمحيص هذين المؤشرين من الوجهة الإحصائية من خلال دراسته الحالية وفقا لمواقف اختبارية تطبيقية مختلفة، وبالتالى فإن الدراسة الحالية تحاول استقصاء قدرة هذين المؤشرين في تحديد البيانات المستخرجة من الأفراد غير الملائمين (الذين تم رصدهم من خلال الموقف التطبيقى) وفقا لحالات مختلفة مثل الغش (فردى– جمعى)، وعدم الجدية (الإهمال) بهدف فرز هذه البيانات وتمييزها عن البيانات الأخرى تمهيدا لاستبعادها من عينة البيانات الكلية التي سوف تستخدم في استكمال عمليات التحليل التي تستهدف إنشاء بنك الأسئلة، كما تحاول هذه الدراسة تمحيص تأثير تلك البيانات غير الملائمة (في حال عدم إقصاؤها) على مؤشرات الملاءمة للمفردات وكذلك على تقديرات الصعوبة لها عند استخدام مفردات بنكية يشاع ويكثر استخدامها في الاختبارات التحصيلية بجميع المراحل التعليمية وهي المفردات المقالية (نصار، 2010: 223) وهي النوعية التي تسفر عن استجابات متعددة التدريج.

مشكلة الدراسة:

تنتشر بعض أشكال الاستجابات غير الملائمة بالعديد من المواقف الاختبارية سواء كانت هذه المواقف تتعلق بالاختبارات التحصيلية التي يعتمد عليها في الامتحانات المستخدمة بجميع المراحل التعليمية أو بتلك المواقف الاختبارية الأخرى التي تستخدم بها اختبارات كأدوات لجميع البيانات الخاصة بأغراض البحث العلمي والتي تستهدف إعداد وإنشاء بنوك الأسئلة، فلقد أشارت الكثير من الدراسات، أن سلوك الغش وعدم

الجدية والتخمين من آفات الموقف الاختبارى، ويشكل ظاهرة شائعة ومنتشرة في بلدان كثيرة من العالم، وهي من الظواهر التي يجب التصدى لها ودراستها (Kaufman & Liebling, 2001)، (Dewyer & Michael W., Amanda B and Matthew Jeffrey,2004)، (Nathanson, C, Paulhus, D. L., & B., 2009)، (Williams,K. M.,2010)، (Cizaek,2004).

كما أشارت بعض الدراسات العربية إلى انتشار هذه الظاهرة في الواقع العربي بنسب مختلفة (الخدمة إلإجتماعية ١٩٨٣، ١٩٨٤)، (جاسم الكندرى وآخرون، ١٩٨٩)، (عودة ومقابلة، ١٩٨٩)، مما يشير إلى حقيقة تواجد هذه الظاهرة وخطورتها ووجوب التصدي لها، وعلى جانب آخر فقد أشارت دراسات أخرى مثل دراسات (Bogan,1997) (Linda,1998)، (Ronald,1999)، (Kaufman & Liebling,2001)، (Cizaek,2004) إلى تأثر التحليلات الإحصائية التي تجرى لكثير من المفردات الاختبارية الجديدة بهذه النوعية من الاستجابات مما يعطى صور مزيفة عن حقيقة هذه المفردات، وحيث ان هناك بعض الدراسات مثل دراسات (Handrawan 2004 & Glas, 2006)، (Bay,2006)، (Van der linden & Sotaridona, 2006)، (Liancre, 2006) أشارت إلى فاعلية بعض المؤشرات الإحصائية الجديدة مثل مؤشر (Bm) في الكشف عن حالات الغش بالمقارنة بمؤشرات أخرى تقليدية مثل مؤشر (g2) لفراري frary، ومؤشر ESA لبليزا belleza،وهي مؤشرات ليست متوفرة بالبرامج الإحصائية المتوفرة والمستخدمة لتحليل المفردات الاختبارية من الوجهة الإحصائية.

وبناء على ما نوهت به دراستي (Dorans, 1998)، (Ronald, 1999)، من إمكانية دراسة بعض المؤشرات الإحصائية الجديدة المختلفة باستخدام بعض البرامج الإحصائية لتحليل المفردات، في الكشف عن الاستجابات غير الملائمة، كما اقترحت دراسات أخرى مثل دراسات (Linda, 1998)، (Meijer, 2005)، (Fan & Ping, 2006) إجراء دراسات للبحث عن مؤشرات إحصائية جديدة لهذه الأساليب غير الملائمة من الاستجابات، فإنه يبدو للباحث الحالي أنه قد يمكن اكتشاف حالات الغش الفردى وكذلك حالات الغش الجماعى من خلال المؤشرات الإحصائية المتضمنة ببرنامج Winsteps (وهو برنامج التحليلات الإحصائية للمفردات الاختباريةالأكثر شيوعا واستخدما في العديد من المؤسسات الأكاديمية) مثل مؤشرى الملاءمة (التقاربي والتباعدي)، كما يمكن لنفس المؤشرين اكتشاف حالات عدم الجدية (كأنماط استجابة غير ملائمة) التى تنتشر في العديد من المواقف الاختباراليات التي تستهدف جمع البيانات كما أشارت إليها دراستى,(Bogan 1997)، (1998, Linda)، وهذا مما حدا بالباحث الحالى إلى إجراء دراسته الحالية وذلك اعتماد على مفردات اختبارية متعددة التدريج من نوع المقال باستخدام نموذج تحليل يناسب طبيعة هذه النوعية من المفردات وهو نموذج التقدير الجزئي "Credit Partial Model"، هذا وتلبي هذه الدراسة احتياجات بنوك الأسئلة للمزيد الإجراءات التي تجعل منها أدوات تتمتع بدرجة أعلى من الدقة والموضوعية.

هذا وتتلخص مشكلة الدراسة الحالية فى الإجابة على السؤال العام التالى:

ما قدرة مؤشرات الملاءمة الإحصائية (مؤشرالملاءمة التقاربى، والتباعدى) المتضمنة بتحليلات برنامج Winsteps في الكشف عن الاستجابات غير الملائمة التي تحدث في بعض المواقف الاختبارية (استجابات الغش وعدم الجدية) وهل يؤثر حذف هذه الاستجابات على مؤشرات الملاءمة الإحصائية لمفردات التحليل وتقديرات الصعوبة لها؟

ويتفرع من السؤال العام السابق التساؤلات الآتية:

١- هل يمكن استخدام مؤشرى الملاءمة المتضمنان ببرنامج التحليل في الكشف عن البيانات الناتجة عن استجابات الغش الفردى؟.

٢- هل يمكن استخدام مؤشرى الملاءمة المتضمنان ببرنامج التحليل في الكشف عن البيانات الناتجة عن استجابات الغش الجماعي؟.

٣- هل يمكن استخدام مؤشرى الملاءمة المتضمنان ببرنامج التحليل في الكشف عن البيانات الناتجة عن استجابات عدم الجدية؟.

٤- هل تؤدى استجابات الغش وعدم جدية الاستجابة إلى بيانات يصعب تمييزها من خلال مؤشرى الملاءمة ببرنامج التحليل؟

٥- هل هناك حدود إحصائية لمؤشرات الملاءمة لكل من استجابات الغش وعدم جدية الأداء؟

٦- هل تؤثر البيانات الناتجة عن استجابات الغش الفردى والجماعى وعدم الجدية على مؤشرى الملاءمة الإحصائية للمفردات الاختبارية؟

٧- هل يؤثر حذف البيانات الخاصة باستجابات حالات عدم الملاءمة للأفراد (من واقع حالات الغش وعدم جدية الأداء) على تقديرات الصعوبة للمفردات؟

أهمية الدراسة:

- يعتبر تحليل المفردات (الأسئلة) من الإجراءات الأساسية والهامة فى إعداد بنوك الأسئلة، حيث يجرى تحليل هذه المفردات البنكية باستخدام عينات كبيرة من الأفراد، من خلال مواقف اختبارية متعددة بهدف تدقيقها وفرزها وانتقاء الصالح منها تمهيدًا لاستخدامها فى إعداد الاختبارات التى تستخدم كأدوات فى عمليات التقويم مستقبلا، وبالتالي فإن حدوث حالات من الغش أو عدم جدية الأداء أثناء هذه المواقف قد يؤدى إلى نتائج مضللة، مما يؤدى بدوره إلى عدم صلاحية هذه الاختبارات فى تحقيق الأهداف التى أعدت من أجلها، ومما يلقى بظلاله السيئة وانعكاساته الخطيرة على مجريات العملية التعليمية برمتها، وبالتالى فإن دراسة متغيرات الغش، وعدم الجدية وأثرها على كل من مؤشرات الملاءمة وتقديرات الصعوبة للمفردات قد يؤدى إلى عمل أدق لتحليل هذه المفردات التى سوف تأخذ طريقها إلى الإيداع فى أرصدة هذه البنوك، مما يجعل منها أدوات صادقة وفعالة فى تكوين الاختبارات والأدوات الموضوعية المستخدمة فى عمليات التقويم الخاصة بالمجال التعليمى والتربوى.
- تلقى هذه الدراسة الضوء على أهم المتغيرات التى تؤثر فى سلامة التحليلات الإحصائية والتى تعتمد إلى حد كبير على مؤشرات الملاءمة التقاربى، والتباعدى، فى محاولة التعرف على صدق هذه المؤشرات فى إجراء انتقاء الأفراد الملائمين لاستكمال عمليات التحليل والحدود الإحصائية المسموح بها فى هذا الصدد، وهذا له أهميته

العملية والنظرية فى مجال التقويم بصفة عامة ومجال القياس بصفة خاصة.

- تبدو أهمية هذه الدراسة من الأهمية التى توليها الاتجاهات الحديثة للقياس لبنوك الأسئلة كأدوات ووسائل حديثة انعقدت عليها الآمال فى تحقيق الموضوعية والمرجوة فى مجال قياس السلوك الإنسانى، حيث أن أهمية هذه الدراسة تستهدف التحكم فى المتغيرات التى تعوق هذه البنوك من تحقيق ما تصبو إليه من أهداف.
- قد تفيد نتائج هذه الدراسة العديد من المؤسسات والقطاعات العلمية والأكاديمية (ومن بينها الجهة التي يعمل بها الباحث الحالي بمصر وهي المركز القومي للامتحانات والتقويم التربوي والذي يسعي إلى بناء مثل هذه البنوك وإجراء المزيد من الدراسات العلمية والبحثية المتخصصة التي تتصدى للقضايا المختلفة المتعلقة بها) في تطوير وتحسين الإجراءات التي تستخدمها عند بناء هذه البنوك بهدف التوصل إلى بنوك مختلفة من المفردات يمكن الثقة في دقتها وموضوعيتها.

أهداف الدراسة:

- استقصاء مدى إمكانية استخدام مؤشرات الملاءمة الإحصائية ببرنامج Winsteps في الكشف عن البيانات الناتجة عن الاستجابات غير الملاءمة المحددة بالدراسة.
- التعرف على تأثير استجابات عدم الملاءمة على تقديرات الصعوبة للمفردات الاختبارية.

- الكشف عن تأثير استجابات عدم الملاءمة (غش فردى،غش جماعى، عدم جدية) على المؤشرات الإحصائية لملاءمة كل من الأفراد والمفردات الاختبارية باستخدام برنامج التحليل المستخدم بالدراسة Winsteps.

حدود الدراسة:

- اقتصرت الدراسة على أداة موضوعية تتضمن أسئلة من نوع المقال.
- اقتصرت الدراسة على برنامج التحليل Winsteps.
- اقتصرت الدراسة على استخدام التحليلات الإحصائية الخاصة بنموذج " التقدير الجزئي " المتضمن ببرنامج التحليل.

مصطلحات الدراسة

يقتصر الباحث في الجزء التالي على ذكر المصطلحات التي سوف يلتزم بها في دراسته الحالية دون التعرض للتعريفات الأخرى، حيث يتم مناقشاتها والتعليق عليها فيما بعد.

المتغيرات المستقلة:

نظرا لعدم وجود تعريفات علمية محددة (في حدود علم الباحث) للمتغيرات المستقلة المستخدمة بالدراسة الحالية، لذا قام الباحث بإعطاء تعريفات لكل متغير على حدة وفيما يلي عرض لها:

فرز البيانات Data screening:

هى العملية التي تستخدم في تحديد وتمييز البيانات الناتجة عن استخدام بعض أفراد العينة البحثية لاستجابات غير ملائمة تتمثل في

الغش الفردي والغش الجماعي وعدم الجدية لمفردات الاختبار المستخدم بالدراسة الحالية.

الغش الفردى Single cheating:

هى محاولة فردية من جانب الممتحن تهدف إلى الحصول على الإجابة الصحيحة للأسئلة الذى فشل فى الإجابة عليها، وذلك بأساليب غير مشروعة.

الغش الجماعى Group cheating:

ويبدو هذا النوع من الغش فى محاولة الممتحن أو مجموعة الممتحنين الحصول على بعض الإجابات الصحيحة لبعض الأسئلة الامتحانية بمعاونة فرد أو مجموعة أفراد يقومون بإملائهم الإجابة الصحيحة بصورة جماعية.

عدم الجدية Carelessness:

يبدو في أداء الممتحن الذى لا يلتزم بتعليمات الاختبار، وكذلك عدم الالتزام بالسلوكيات والقواعد المطلوبة داخل الموقف الامتحانى.

المتغيرات التابعة:

وتتلخص المتغيرات التابعة في الدراسة في المصطلحات التالية:

مؤشرات الملاءمة الإحصائية Statistical indexes:

وتتحدد هذه المؤشرات في كل من مؤشري الملاءمة الإحصائية ببرنامج التحليل Winsteps. وهما (laincer,2005):

مؤشرالملاءمة التقاربي Infit (Zstd) index:

وهو يشير إلى مدى اقتراب مستوى قدرات الممتحنين من مستوى صعوبة الأسئلة، كما يشير أيضا إلى مدى ملاءمة المفردات الاختبارية.

مؤشرالملاءمة التباعدى: Outfit (Zstd) index:

ويشير إلى نوعية استجابة الممتحن التي توضح ما إذا كانت إجابته صوابا على أسئلة تزيد فى صعوبتها عن قدرته أو فشله فى الإجابة الصواب على أسئلة تقل فى صعوبتها عن مستوى هذه القدرة، كما يشير هذا المؤشر أيضا إلى احتمال وجود عيب فى صياغة السؤال أو اعتماد مفردة من مفردات الاختبار على بعض المفردات الأخرى بنفس الاختبار.

مؤشر الصعوبة Difficulty index:

هو مؤشر Measure ببرنامج Winsteps لتحديد صعوبة المفردة مقدرة بوحدات اللوجيت (laincer,2005) ويشير إلى موقع المفردة على متصل تدريج الصفة المقاسة من الاختبار ويحدد امتلاكها لمستويات محددة من هذه الصفة (Jason L. Meyers, Edward G. Miller, Walter D. Way, 2009:42).

الخلفية النظرية والأدبيات:

نظراً لأن الدراسة الحالية تهتم بالكشف عن البيانات الاختبارية غير الملائمة بهدف فرزها وإقصائها عن الإجراءات المستخدمة في استكمال عمليات التحليل الخاصة بإعداد مفردات الاختبارات البنكية كان من الضروري التعرف على أنماط الاستجابة غير المنضبطة (غير

الملائمة) التي تحدث من جانب بعض المختبرين بالمواقف الاختبارية المختلفة التي يمكن ان تسفر عن مثل هذه النوعية البيانات لذا كان من الضرورى تناول أنماط الاستجابات غير الملاءمة بالفقرة التالية:

أنماط الاستجابة غير الملائمة:

هناك عدد من انماط الاستجابة غير الملائمة من جانب بعض المختبرين في تعاملهم مع المواقف الاختبارية المختلفة من بينها (Satterlee, A. G., 2011):

١- الغش. ٢- عدم الجدية.

الغش Cheating

من المعروف أن الغش فى الامتحانات من المشكلات التربوية التى ينبغى التخلص منها، حيث أنه قد يسهم فى إعطاء صورة مضللة عن المستوى التحصيلى للطلاب (Nathanson, C, Paulhus, D. L., & Williams,K. M.,2010).

كما أن سلوك الغش يدل على ما يتصف به الطلاب الغشاشون من سوء خلق، ما لم يقتلع من جذوره لديهم سيدوم معهم، ويصبح نمطا سلوكيا سائدا فى شتى مجالات الحياة التى يعملون بها، كما أنه يدعم اللجوء إلى الأساليب الملتوية لتحقيق الغايات المنشودة (عودة ومقابلة، ١٩٨٩)، (Kaufman & Liebling, 2001)، (Camaer, 2003)، (Cizaek, 2004).

عدم الجدية (الإهمال) Carelessness:

تبدو استجابة عدم الجدية أو الإهمال فى المواقف الاختبارية المختلفة والتى يشعر حيالها بعض الممتحنين بعدم الأهمية (Michael W., Amanda B and Matthew B.,2009)، وكثيرا ما تلاحظ هذه الاستجابة أو هذا الأسلوب فى المواقف الاختبارية الخاصة بتطبيق بعض الأدوات البحثية التى تستخدم فى الدراسات العلمية لأغراض جمع بيانات تتعلق بخدمة أهداف علمية بعيدة عن اهتمامات عينة التطبيق من المبحوثين، كما يلاحظ هذا الأسلوب أيضا فى المواقف الاختبارية التى تستهدف تجريب مفردات بنوك الأسئلة فى الواقع الميدانى بهدف الحصول على مؤشرات إحصائية عن هذه المفردات والتى تدون فى الكروت البنكية الخاصة بهذه المفردات (وهذا ما تبين للباحث الحالى من واقع الخبرة التى اكتسبها من خلال العمل كباحث فى مجال بنوك الأسئلة بالمركز القومى للامتحانات والتقويم التربوى)، وقد يؤثر بالطبع هذا الأسلوب فى دقة الإحصاءات الخاصة بهذه المفردات، وبالتالى على موضوعية وسلامة الاختبارات المسحوبة منها (Bay,2006).

هذا وتنتشر أساليب الغش وعدم الجدية في مختلف امتحانات المراحل الدراسية المختلفة ومختلف المواقف الاختبارية الخاصة بتجريب المفردات الاختبارية، وقد أدرك القائمون عن العملية التعليمية وجود هذه الظاهرة، وما يمكن أن تمثله من مشكلات يجب أن تعنى بالبحث والدراسة، وكذلك التأثيرات السلبية البالغة لهذه الظاهرة على التحليلات الإحصائية التي تستهدف الاختبارات المختلفة التي تجرى في ظل هذه المواقف لأغراض الإعدادات الخاصة ببنوك الأسئلة، فقد اهتمت العديد من الدراسات السابقة بالتعرف على مدى انتشار ظاهرة الغش فى

مختلف المراحل التعليمية، فلقد توصلت دراستى الخدمة إلاجتماعية والتى أجريت على طلاب المرحلة الثانوية بدولة الكويت أن حوالى (٦٣%) من المعلمين يرون أن الغش موجود بشكل عام، كما توصلت دراسة فاروق فلية (١٩٨٨)، والتى أجريت فى خمس محافظات مصرية (دمياط،، بور سعيد، الشرقية الإسكندرية، المنيا) أن حوالى نسبة (٦٨.٤%) من المشاركين فى العملية التعليمية، (٣٥.٦%) من أولياء الأمور،، (٢٤%) من الطلاب يرون أن ظاهرة الغش شائعة فى امتحانات المرحلة الثانوية والإعدادية، كما أوضحت دراسة جاسم الكندرى وآخرون (١٩٨٩) التى أجريت على طلاب جامعة الكويت أن الغش منتشر فى المرحلة الثانوية أكثر من المراحل الأخرى، بينما أظهرت دراسة عودة ومقابلة (١٩٨٩)، والتى أجريت على طلاب جامعة اليرموك بالأردن أن نسبة (٨٧%) من الطلبة يرون أن حجم ظاهرة الغش دون المتوسط، كما أوضحت دراسة أندرسون Anderson (1998) والتى أجريت على عينة من طلاب وطالبات جامعة " جورج تاون " أن نسبة (٥٠%) من الطلاب الذكور يلجأون إلى محاولة الغش للحصول على الإجابات الخاصة بالأسئلة الصعبة،كما أن نسبة (٤٣%) من الطالبات ذكرو أنهم يحاولون الغش على نفس الأسئلة إذا ما وردت بالامتحانات. وقد استهدفت دراسة كوفمان وليبلنج Kaufman & Liebling) (2001تقييم الكليات لانتشار وتفشي عدم الأمانة الأكاديمية، وقد اجرت الدراسة مسحا لعدد من الأساتذة الجامعيين في جامعات أمريكية مختلفة بهدف الحصول على تقديرات لانتشار وزيوع عدم الأمانة الأكاديمية وقد استخدمت الدراسة اداة تتكون من عدد (٣٠) فقرة عن ظهور بعض سلوكيات الغش في الامتحانات الجامعية وذلك

على مقياس تساعي لجميع فقرات الأداة (١ سلوك غير واضح– ٩ سلوك واضح ومنتشر) هذا وقد أظهرت نتائج الدراسة أن (٩١%) من عدد الأساتذة الجامعيين واجهوا العديد من حوادث الغش بالجامعة وذلك لعدم وجود إجراءات كافية للحد من هذه الظاهرة كما قرر (٧٢%) منهم ضرورة إتباع سياسات حيال هذه الظاهرة. وقد استهدفت دراسة كيزك (Cizaek, 2004) التعرف على الأسباب التي تؤدي إلى غش الدارسين والاستراتيجيات الخاصة بالحماية من الغش، وقد أظهرت الدراسة من خلال استعراضها لبعض الدراسات العديد التي أجريت لبحث هذه الظاهرة أن نسبة (٥%) من الممتحنين يحاولون الغش في مدارس نيويورك ويؤكدون أن هذا السلوك في تزايد مطرد، كما أشارت داسة كل من مايكل وتشارلز .Michael E., Charles H (2010) إلى أن بعض الطلاب المتدنين من الناحية التحصيلية تتسم استجاباتهم بالإهمال وعدم الجدية وهذا ما أكدته دراسة كل من كيرفي وسامر Kherfi, Samer (2011) والتي توصلت إلى أن الطلاب ضعاف المستوى هم أكثر الطلاب استخدام للأساليب غير المنضبطة في مختلف المواقف الاختبارية والتقويمية.

وحيث أن الدراسة الحالية تهتم باستخدام بعض المؤشرات الإحصائية في الكشف عن البيانات الناجمة عن استجابات الخاصة بسلوك الغش وعدم الجدية بالمواقف الاختبارية المختلفة، كان من الضرورى تناول بعض الأدبيات التى اهتمت بالمظاهر المختلفة والدالة على هذه النوعية من الاستجابات ففي هذا الصدد توصلت دراسة بوجان Bogan (1997) وكذلك دراسة جوردان Jordan A.E (2008) إلى أن أكثر صور الاستجابات الدالة على عدم الجدية تبدو فى الانتهاء من

الإجابة فى وقت مبكر ولا يتناسب مع زمن الإجابة الواقعى للاختبار، والكتابات الساخرة، كما أشارت الدراسة أن هذه الأساليب تكثر فى المواقف الاختبارية التى تستخدم فى تجريب أنواع مختلفة وجديدة من الأسئلة مما يؤثر على التحليلات الإحصائية الخاصة بها، كما توصلت دراستي ليندا ودور أنز Dorans (1998) إلى وجود بعض أنماط الاستجابات الفردية التى تدل على سلوك الغش فى المواقف الاختبارية مثل الإجابة الصحيحة للمفردات الصعبة والإخفاق فى المفردات السهلة أو الأقل صعوبة، أوالتوصل إلى ناتج الحل الصحيح مع اختزال العديد من الخطوات الرئيسة التى تؤدى إلى الحل النهائى، وكذلك شطب الحل الخاطئ ثم إعادة الحل بشكل سريع وردئ، وأشارت هذه الدراسة إلى إمكانية تأثر هذه النوعية من الاستجابات على نتائج التحليلات الإحصائية لهذه المفردات، فى حين توصلت دراسة رونالد Ronald (1999)، وكذلك دراسة إدوارد وجورجينا Edward J., Georgiana S. (2009) إلى بعض الاستجابات التى تدل على الغش الجماعى منها تشابه أسلوب الاستجابات على المفردات الاختبارية خاصة إذا كانت هذه الأخطاء متشابهه، كذلك تشابه ترتيب خطوات الإجابة على المفردة الاختبارية، كما نوهت هذه الدراسة على تأثير هذه الاستجابات على سلامة التحليلات الإحصائية للمفردات التى يتم تجريبها بغرض تمحيصها إحصائياً وهذا ما اكدته أيضا دراسة كل من جينفر ميلاني وسكوت Jennifer Y., Melanie D., and Scott C. (2009). وقد استهدفت دراسة كامير Camaer (2003) إدراك حقيقة سوء استخدام الاختبارات التربوية وأخطاء استخدامها وأخطاء تقدير الدرجات وحوادث الغش المختلفة والتقرير عنها في الاوساط المحلية والدولية، وقد اجرت

الدراسة مسحا لضوابط استخدام الاختبارات والتقنينات الخاصة بها من خلال استعراض جهود الرابطة الأمريكية للأبحاث التربوية والمجلس القومي للقياس في التربية وقد أعطت الدراسة رؤية عامة للقضايا المطروحة لتقنينات الاختبار وضوابط استخدامه ومشكلات التطبيق الخاص به كما استعرضت التطوير الذي طرأ على هذه التقنينات وكيف يمكن ان يستخدم كما استعرضت الدراسة المؤشرات التي يمكن من خلالها كشف حالات الغش والتي تبدو في التسلسل غير المنطقي في خطوات الحل– التعليل الخاطئ للإجابة الصحيحة– الإجابة الصحيحة للسؤال رغم صعوبة السؤال بالنسبة لمستوى قدرة الممتحن وهذا ما أكدته دراسة Satterlee, A. G (2011).

ويتضح من خلال ما سبق عرضه أدبيات مدى انتشار ظاهرة الغش فى مختلف المراحل التعليمية وبصفة خاصة المرحلة الثانوية والإعدادية، وهى ليست ظاهرة محلية فقط بل هى ظاهرة منتشرة فى العديد من البلدان الأخرى أيضا، كما يتضح أيضا من خلال نفس الأدبيات أن أهم المظاهر الدالة على سلوك الغش هي: الحل الصحيح للمفردات الصعبة والإخفاق فى المفردات الأقل صعوبة، التوصل إلى ناتج الحل الصحيح بعد اختزال العديد من الخطوات الرئيسة المؤدية إلى الحل، شطب الحل الخاطئ ثم إعادة الحل بشكل سريع وردئ، تشابه أسلوب الاستجابات على المفردات الاختبارية وبخاصة إذا كانت تتضمن بعض الأخطاء المتشابهه، تشابه شكل وترتيب خطوات الإجابة على المفردات المتضمنة بالاختبار، التسلسل غير المنطقي في الحل.

كما تبين من خلال ما سبق عرضه من دراسات أيضا أن المظاهر الدالة على سلوك عدم الجدية هى: الانتهاء من الإجابة على الاختبار بشكل مبكر وغير مناسب، الكتابات الساخرة، وقد اعتمدت الدراسة الحالية على المظاهر السابقة الدالة على سلوك الغش وعدم الجدية في التعرف حالات الغش المختلفة المحددة بالدراسة الحالية وذلك في إجراءاتها المختلفة المحددة لهذا الأنماط من الاستجابات غير الملائمة.

ونتيجة لانتشار ظاهرة الغش بمختلف المراحل التعليمية وبمختلف المواقف الاختبارية أجريت بعض الدراسات من قبل بعض الباحثين تستهدف الكشف عن استجابات الغش من بينها دراسة Dewyer & Jeffrey (2004) التي استهدفت مراجعة بعض القضايا الخاصة باستخدام البيانات الإحصائية في نتائج الاختبار الذي يحتوي على أسئلة اختيار من متعدد، للبحث عن أدلة تستخدم في الكشف عن حالات الغش للممتحنين بالجامعات، وقد استخدمت الأسلوب المسحي في عرض الدراسات التي تناولت الأساليب الإحصائية التي يمكن استخدامها لتحقيق هذا الغرض وأشارت الدراسة أن نسبة كبيرة من عدم الأمانة الأكاديمية بين طلاب الجامعة الأمريكيين، كما أن هناك بعض الكليات التي كان لها منهجية إحصائية لضبط حالات الغش في الاختبارات التي تحتوي على أسئلة الاختيار من متعدد وقد أثبتت الدراسة انه لا توجد طريقة تقليدية شائعة متاحة يمكن أن تعطي دليل على الغش– كما أشارت الدراسة على أن الدليل الإحصائي بمفرده لا يجب أن يكون دليلا على الغش لاتهام الأفراد المشكوك فيهم كما أن هذه الأساليب الإحصائية يمكن أن تكون حجة في حين أنها لا تستطيع بشكل حاسم

أن تثبت أن الغش قد حدث بالفعل، حيث أن الأساليب الإحصائية يمكن أن تفشل بالفعل في حسم هذه القضايا ولكنها قد تفيد في تصميم الاختبارات والمواقف بشكل أفضل. كما استهدفت دراسة هاندوران وجلاس Handrawan & Glas (2004) التحقق من تأثير عدم ملاءمة الأفراد لنماذج الاستجابة للمفردة على جودة القرار الخاص بتصنيف الأفراد لفئات مختلفة من حيث مستوى القدرة، كما استهدفت الدراسة استقصاء اثر استجابات التخمين كاستجابات غير ملائمة على التقديرات الخاصة بالمفردات الاختبارية، وقد اعتمدت الدراسة على عدة عينات لتحقيق أهداف الدراسة وهي عينات مكونة من أفراد غير ملائمين، (٦٠) مفردة اختبارية، وقد وزعت إجابات التخمين عن طريق محاكاة استجابات تخمين لحالات فعلية، وذلك بنسب مختلفة (٣٠%، ٥٠%)، وتم تحديد الإجابات الصواب عليها بنسبة (٢٠%) داخل العينات المذكورة، وقد تم استخدام إحصاء (t) الملاءمة للأفراد (الملائمين وغير الملائمين) كمعيار لقياس قوة وجودة القرار ونسبة وعدد الأفرد الملائمين وغير الملائمين، ولقد أظهرت نتائج الدراسة أن تأثير الاستجابات غير الملائمة (التخمين) على تصنيف الأفراد كان صغيرا للغاية وانتهت الدراسة أن إحصاء الملاءمة للأفراد المستخدم يمكن استخدامه في تصنيف الأفراد ملائميين وغير ملائمين. وقد استهدفت دراسة ميجر Meijer (2005) استقصاء تأثير الاستجابات غير الملائمة للأفراد على قوة إحصاء الملاءمة للأفراد وقد اعتمدت الدراسة في إجراءاتها على عينة كلية قسمت إلى عينتين فرعيتين إحداها إحتوت على نسبة قدرها ٢٠% من الغشاشين الذين حصلوا على الإجابات الخاصة بالأسئلة الصعبة بالنظر في أوراق إجابة زملائهم مرتفعي المستوى،

والأخرى إحتوت على نسبة ٢٠% مخمنين استخدموا التخمين باستخدام باستخدام احتمال الاستجابة الصواب ٢٥% وذلك لمفردات اختبارية لاختبار يحتوى على (٥٠) مفردة منها (١٧) مفردة صعبة، من نوع الاختيار من متعدد تحتوى على أربعة بدائل للإجابة، وقد أظهرت نتائج الدراسة أن استجابات الغش والتخمين أثرت بصورة سلبية على قوة إحصاء الملاءمة للأفراد واقترحت الدراسة بحث تأثير الاستجابات غير الملاءمة على مؤشرات الملاءمة للمفردات، كما استهدفت دراسة استاريدونا وميجر Sotaridna & Meijer (2005)استقصاء الخصائص الإحصائية لمؤشر K-indx لفحص سلوك الغش في الاختبارات وقد استخدمت الدراسة أسلوب المحاكاة لأسلوب بعض الطلاب الحقيقيين المستخدمين لأسلوب الغش (لتحسين درجاتهم بنقل الإجابات الصحيحة من أقرانهم ذوي القدرات العالية) والآخرين غير المستخدمين لهذا الأسلوب لغرض فحص القدرة التطبيقة للمؤشر وذلك لعدد (صغير – متوسط– كبير) من بيانات الاختبار، وقد قامت الدراسة بتوليد بيانات لدرجات اختبار اختيار من متعدد من خمسة بدائل تتكون من (٨٠) مفردة وعدد من أفراد محاكاة تم محاكاتها من بيانات تجريبية لاختبار فعلي في الرياضيات وقد تم محاكاة استجابات الغش لنسبة (٥%) بشكل عشوائي من حجم كل عينة وقد تم تحديد هذه الأحجام من العينات بهدف مقارنة نتائج الدراسة للحجم الصغير والمتوسط والكبير من بيانات المحاكاة وقد أظهرت نتائج الدراسة أن مؤشر K-indx يكون له فعالية ملحوظة مع العينة كبيرة الحجم وكان له فعالية أقل مع العينتين الصغيرتين. واستهدفت كذلك دراسة باي Bay (2006) الكشف عن فاعلية مؤشر إحصائي جديد (Bm) في الكشف عن حالات الغش

بالمقارنة بمؤشرات أخرى تقليدية مثل مؤشر (g2) لفراري Frary، ومؤشر ESA لبليزا Belleza وذلك على اختبارات الاختيار من متعدد، وقد استخدمت الدراسة مؤشرا حديثا يعتمد على توزيع ذات الحدين ويعتمد في أساسه الرياضي على الأساس الرياضي لافتراض استقلالية، وقد استخدمت الدراسة بيانات تحاكي (١٢) نوع من سلوك الغش الفعلي في الامتحانات تم الحصول عليها من اختبار تقويمي في مادة الرياضيات يتكون من (٥٠) مفردة من نوع الاختيار من متعدد أجل دراسة حساسية المؤشر لفحص حالات الغش والظروف المختلفة التي تؤثر في فاعليته، وقد أظهرت نتائج الدراسة ان المؤشر الحديث كان أكثر فعالية في كشف حالات الغش من خلال المفردات التي حدث فيها وقائع غش كما ان هذا المؤشر كان أكثر فعالية في، كما أن هذا المؤشر كان أكثر فعالية في تمحيص حالات الغش لنسبة (٢٥% إلى ٥٠%) من المفردات التي تم الغش فيها من (٥٠) مفردة اختباريه لكن هذا المؤشر يكون أقل حساسية عندما يكون الاختبار أقصر من ذلك كما هذا المؤشر يكون اقل حساسية أيضا إذا كان عدد حالات الغش قليلة. وأيضا استهدفت دراسة Van der linden & Sotaridona (2006) استخدام اختبار إحصائي لفحص حالات الغش لاختبار الاختيار من متعدد وقد استندت الدراسة إلى أن إجابة الممتحن على المفردة يكون نتيجة لثلاث عمليات ممكنة وهي المعرفة والتخمين والغش، فالممتحنين الذين لا يستطيعون الإجابة على السؤال من خلال العمليتين الأولى والثانية يمكن لهم استخدام العملية الثالثة (الغش) وهذا الافتراض يؤدي على أن التوزيع المقابل للبدائل المختارة بين الممتحنين المستخدمون لأسلوب الغش والممتحنون المتصف سلوكهم بالأمانة ينتمي إلى عائلة

توزيع ذات الحدين، وهذا النموذج الثنائي لتوزيع الإجابة على الخيارات يمكن الحصول عليه بواسطة الانحدار الخطي لنسبة البدائل الخاطئة في كل مجموعة من مجموعات المزاوجة للدرجات لكل من نوعي الممتحنين وقد استخدمت الدراسة مجموعات كبيرة من البيانات الخاصة بإجابات النوعين من الممتحنيين كما استخدمت مؤشر وولاكس Wallack.S, W لكشف حالات الغش وكذلك مؤشر K-indx وكذلك مؤشر (g2) وقد اظهرت نتائج الدراسة وجود نتائج متشابهة بين الاختبارات الإحصائية الثلاثة المذكورة سابقا في الكشف عن نمط استجابات الحالات المختلفة التي تلجأ لأسلوب الغش في التوصل للإجابة الصحيحة لمفردات الاختبار من نوع الاختيار من متعدد.

كما استهدفت دراسة Fan & Ping (2006) استقصاء التأثير السلبي لعدم ملاءمة البيانات لنموذج التحليل المستخدم (المرتبط بنظرية الاستجابة للمفردة) على تقديرات نموذج التحليل المستخدم وهي: صعوبة المفردة، تمييز المفردة، تقديرات القدرة للأفراد، وقد استخدمت الدراسة في إجراءاتها مقياس "تكساس للمهارات الأكاديمية" تم تطبيقة عبر عينات مختلفة من طلاب الصف الحادي عشر بولاية تكساس، وقد أظهرت نتائج الدراسة أنه لم يوجد تأثير سلبي لعدم ملاءمة البيانات على تقديرات الصعوبة للمفردات كما أوضحت النتائج أن النموذج الأحادي المعلم أعطى نتائج متقاربة لتقديرات الصعوبة مع البيانات غير الملائمة عبر العينات المختلفة أفضل من النموذج الثلاثي المعلم، كما كان هناك اختلاف بين تقديرات التمييز عبر العينات المختلفة للدراسة باختلاف طبيعة البيانات من حيث الملاءمة،كما بينت الدراسة أن النموذج الثلاثي المعلم أعطى تقديرات قدرة متقاربة مع جميع أنواع البيانات بشكل أفضل

من النموذج الأحادي المعلم. واقترحت دراسة التأثير السلبي لعدم ملاءمة البيانات على مؤشرات الملاءمة الإحصائية لمفردات الاختبار.

كما استهدفت دراسة Liancre (2006) التعرف على تأثير كل من الظروف غير الملائمة (الضوضاء) وكذلك سلوك الغش،وكذلك الصمت التام (بدون مناقشة أو السماح بأي استفسار أو تعليق أو إلقاء تعليمات)على دقة القياس وقد تم تحديد التأثير على دقة القياس باستخدام اختبار مربع المتوسطات للملاءمة ومؤشر(t)، وكذلك مؤشر الثبات في إجراءات الدراسة التي اعتمدت على عينة من تلاميذ المرحلة الابتدائية، طبق عليه اختبار المكعبات وقد استخدمت الدراسة التحليلات الخاصة بالنموذج الأحادي المعلم (نموذج راش) لبيانات الاختبار وقد اظهرت نتائج الدراسة التأثير السلبي للضوضاء والغش والصمت التام على دقة القياس نتيجة التأثير السلبي كذلك على استجابات الاطفال على مفردات الاختبار المستخدم، وفي هذا الصدد قد توصلت دراسة كل من آمي وجريج Amy D., Greg B (2010) على عينة من طلاب كلية التمريض بولاية كاليفورنيا، أن الأساليب الاستجابة غير ملائمة التي تبدو في أساليب الغش التأثير السلبي لاستخدام هذه الأساليب على دقة الأداء الأكاديمي الطلاب، ودقة قياس مهاراتهم باستخدام باستخدام اختبار مربع المتوسطات للملاءمة ومؤشر (t).

كما توصلت دراسة Williams, J. Petridou, A (2011) إلى نفس النتيجة السابقة على باستخدام اختبار في الرياضيات طبيق على عينة كبيرة من طلاب مدارس مانشستر بالتعليم المتوسط، إلا أن هذه الدراسة توصلت أيضا إلى صعوبة التوصل إلى مؤشرات إحصائية تميز استجابات الغش عن الاستجابات المنضبطة.

المفردات البنكية Item Banks:

تهتم الدراسة الحالية بالتعرف على أثر أنماط الاستجابة الخاصة بحالات عدم الملاءمة والتى تحدث فى المواقف الاختبارية، مثل حالات الغش الفردى والجماعى وعدم الجدية، وأثر ذلك على دقة وسلامة التحليلات الإحصائية الخاصة بمفردات بنوك الأسئلة (المفردات البنكية) التى يجرى تطبيقها وتجريبها فى مثل هذه المواقف، حيث تسفر هذه التحليلات عن مؤشرات ملاءمة لكل من الأفراد والمفردات، وكذلك مؤشرات خاصة بتقديرات الصعوبة للمفردات وتقديرات القدرة للأفراد (Glas,2004)، لذا كان من الضرورى التعرف على طبيعة بنك الأسئلة ومحتوياته وكيفية بنائه بالجزء التالي.

بنك الأسئلة:

يتضمن بنك الأسئلة مجموعة من المفردات الاختبارية التى يكون لها خصائص سيكومترية مميزة ومعلومة (علام، ١٩٨٦:١٥٧)، ويضم بنك الأسئلة عدة اختبارات تتدرج بنودها جميعا فى تدريج واحد مشترك بصفر واحد مشترك (Hambleton,1987, 81)،(Richichi, 2003) بحيث تعرف مدى واسعا من مستويات المتغير موضوع القياس (كاظم ب، ١٩٨٨: ١٠٨) حيث يبدأ بنك الأسئلة بتحليل عدد كبير من المفردات على أعداد كبيرة من الأفراد بهدف استبعاد الأفراد غير الملائمين وكذلك المفردات غير الملائمة والإبقاء على ما هو ملائم، ثم دمج هذه المفردات فى تدريج واحد ينتهى بشبكة من عدة اختبارات (تشكل بنكا للأسئلة) يمكن أن تغطى المدى الواسع من المتغير موضوع القياس، (الشافعى،

(Anzaldua, 2005, 55) (١١٢ :٢٠٠٥ ،(علام ،١٣٦ :١٩٩٦)
(Nakamura, 2006, 123)

هذا ويعتمد تكوين البنك على دمج مجموعات مختلفة من الأسئلة فى تدريج واحد مشترك بصفر واحد مشترك، وذلك باستخدام مجموعة من البنود المشتركة بين اختبارين مختلفين فى مستوى صعوبتيهما ويقيسان نفس المتغير، حيث تعمل هذه البنود المشتركة كرابطة تعمل على دمج مفردات هذين الاختبارين فى تدريج واحد مشترك بصفر واحد مشترك(Jahja, 1988, 90) (Anzaldua, 2005, 67) (Nakamura, 2006, 144) وعند تكرار هذا الرباط بين أكثر من اختبار تتكون أعداد كبيرة من المفردات فى تدريج واحد مشترك، تمثل محتويات بنك الأسئلة (Wright & Ston,1979).

خطوات إعداد المفردات البنكية:

يتم إعداد المفردات البنكية وفقاً للخطوات التالية:
(David, G, Seung C., 2010)

١. كتابة أعداد كبيرة من المفردات التي تقيس متغير ما (Item Generation) بالاستعانة بمجموعة من الخبراء والمتخصصين في مجال متغير القياس، وتتنوع أنواع هذه المفردات بين المفردات الموضوعية (المقال) والأخرى المقالية.
٢. إعداد هذه المفردات فى صور اختبارية مختلفة، ثم تطبيقها استطلاعيا (Try Out) على عينة ممثلة للمجتمع الأصلى بالواقع الميدانى بهدف الوقوف على صلاحية هذه المفردات لإدراك ما يعتريها من عيوب وأوجه قصور.

٣. تحليل عينة المفردات على عينة التطبيق الاستطلاعية المشار إليها بالخطوة السابقة، وذلك باستخدام أحد النماذج الرياضية المناسبة المرتبطة بنظرية (IRT)، والتي تتناسب مع نوعية المفردات، وذلك باستخدام برنامج التحليل الإحصائى الذي يتضمن هذه النوعية من التحليلات.

٤. تحديد المفردات غير الملائمة والتى تم رصدها من خلال الملاحظة الدقيقة فى مواقف التطبيق ومن خلال المؤشرات الإحصائية للملاءمة الناتجة من التحليلات الإحصائية الخاصة ببرنامج التحليل (مؤشرى الملاءمة التقاربي والتباعدي المتضمنان ببرنامج Winsteps).

٥. تعديل المفردات فى ضوء نتائج التحليل الموضحة بالخطوة السابقة أو حذفها تمهيدا لعمليات التدريج والإيداع بالبنك.

٦. إعادة تطبيق المفردات التى تم تعديلها على عينة كبيرة بالواقع الميدانى بغرض التعرف على مؤشرات الصعوبة الخاصة بها ثم عمل الإجراءات المتعلقة بتدريج هذه المفردات وفقا لمستوى صعوبتها تمهيدا لدمجها فى تدريج مشترك (بنك الأسئلة).

٧. ويتحدد موقع الدراسة الحالية فى الخطوات أرقام (٤، ٥، ٦)، حيث يحدث غالبا أثناء تطبيق الاختبارات (التى تشتمل على المفردات المطلوب إدخالها بنك الأسئلة) أنماط من الاستجابات غير الجادة فى مثل هذه المواقف الاختبارية المصطنعة، حيث لا يرى المستجيب للاختبار أية أهمية لهذا الموقف، وبالتالي يتعامل مع الاختبار باستهتار وتهاون، كما يحدث فى بعض الأحيان غش فردى وجماعى، والذى يحدث أحياناً نتيجة عدم ضبط الموقف الاختبارى (لعدم اكتراث الممتحنين بالعقوبات المناسبة حيال هذه المواقف، نظرا

لإدراكه بعدم أهمية نتيجة استجاباته على مثل هذه النوعية من الاختبارات على المستوى الشخصى)، وبالتالي فإن هذه النوعية من الاستجابات يمكن أن تؤثر على دقة وسلامة التحليلات الإحصائية للمفردات التى يجرى تجريبها وإعدادها لهذه البنوك، وبالتالي كان من الضرورى معرفة ما إذا كانت هذه البرامج الإحصائية المستخدمة فى إجراءات التحليل لديها القدرة على اكتشاف مثل هذه الحالات ثم استبعادها من استكمال إجراءات التحليل المختلفة، وهذا ما تحاول أن تتناوله الدراسة الحالية بالبحث والتمحيص.

النماذج الرياضية المستخدمة فى تحليل المفردات البنكية:

حيث أن الدراسة الحالية تهتم بفرز البيانات التى يمكن أن تؤثر على سلامة المؤشرات الإحصائية لمفردات بنوك الأسئلة، وحيث أن هذه البنوك لا تقتصر على نوعية محددة من المفردات بل يمكن أن تشتمل على أنواع مختلفة من المفردات التي يمكن أن تكون من النوع الموضوعى أو المقالى (Ingebo, 2005)، (Thissen, & Orlando, 2006) فإن التحليلات الإحصائية الخاصة بهذه البنوك تختلف باختلاف النوع المستهدف من المفردات، مما يتطلب استخدام النماذج الرياضية اللوغاريتمية المناسبة المرتبطة بنظرية الاستجابة للمفردة (IRT) والتي تتناسب مع نوعية المفردات المكونة للبنك (Glas,2004)، فهناك النماذج الرياضية التي تتناسب مع المفردات الموضوعية مثل نموذج "راش Rasch"، وهناك النماذج الرياضية والتي تتناسب مع المفردات المقالية مثل نموذج التقدير الجزئي "Partial Credit" وحيث ان هذا النموذج سوف تعتمد عليه الدراسة في تحليل البيانات الخاصة بالمفردات

البنكية المستهدفة فسوف يتم تناول هذا النموذج بصورة مختصرة فيما يلي:

نموذج التقدير الجزئي Partial Credit Model:

لقد وضع "ماسترز" Masters (1982) نموذج التقدير الجزئي كتعميم لنموذج "راش" أحادي البارامتر؛ حيث تتحول مفرداته من التصحيح الثنائي إلى مفردات متعددة التصحيح. وتصنف الاستجابة على المفردات في فئات مرتبة بدرجات متتابعة من صفر، ١، ٢، ٣ ...الخ (De Ayala, 2009, 163)، وقد افترض Masters درجة لكل مهمة أو خطوة ينجزها الفرد بنجاح.

ويمكن أن يعبر عن احتمالية أن يجيب المفحوص ذو مستوى قدرة معين على فئة معينة من الدرجات بالنموذج التالي (Yan,1997)، Tie Liang; Wells, Craig (Roberts et.al., 2003) .(S..,2009)

$$P_{xi}(\theta) = \frac{\exp\left[\sum_{j=0}^{x_i}(\theta - b_{xi})\right]}{\sum_{K=0}^{mi}\exp\left[\sum_{j=0}^{K}(\theta - b_{xi})\right]}$$

وهذا النموذج هي الشكل العام للحصول على المنحنى المميز لدرجة أي فئة في نموذج التقدير الجزئي. والرمز b_{xi} هو بارامتر صعوبة الخطوة step المتزامل مع درجة الفئة xi و θ هي مستوى السمة أوالسمة الكامنة.

ففي سياق حل المشكلات–على سبيل المثال– فإنθ هنا مرتبطة بمستوى قدرة المفحوص. وعلى هذا فهناك b واحدة متزاملة مع كل

خطوة أنجزت للاستجابة على المفردة. ولذلك أطلق Masters (1982) على البارامتر b صعوبة الخطوة، وليس صعوبة المفردة؛ ل أنها تشير إلى مدى إتمام المفحوص لكل خطوة بنجاح في سبيل حل المشكلة. فالخطوة هي مرحلة مطلوبة لإتمام المفردة. فعلى سبيل المثال ({٣/٦} + ٢)٢ تحتوي على ثلاث خطوات، لأن هناك ثلاث مراحل مختلفة يجب أن تنجز بترتيب معين للإجابة على المسألة بصورة صحيحة (De Ayala, R., & Koch, W.,1992, 20).

ويتطلب نموذج التقدير الجزئى أنجاز الخطوات داخل كل مفردة بالتتابع (بنظام ثابت أو بترتيب معين). بيد أن هذا لا يعني أن الخطوات متساوية الصعوبة. كما أنه ليس بالضرورة أن تكون آخر مفردة هي الأصعب في حلها؛ أو أن الخطوة الأولى هي الأسهل. فمدى صعوبة المفردات يتفاوت وليس له ترتيب محدد كما هو الحال مع نموذج الاستجابة المتدرجة. وإذا احتوت المفردة على فئتين فقط تحول النموذج إلى نموذج "راش", (Koch & DeYala 1992:338)، (Wang& Wang,2002).

والنموذج السابق يستخدم فى تحليلات برنامج Winsteps لأسئلة المقال، وسوف يقتصر الباحث الحالى على استخدام نموذج " Partial Credi" لماسترز، نظراً لأن هذا النموذج يناسب تحليل مفردات الاختبار المستخدم بالدراسة الحالية، الذى يحتوى على مفردات مقالية تتسم الاستجابات عليها ب أنها متعددة التقسيم، يحصل الطالب بمقتضاها على درجة كلية أودرجة جزئية منها، كما أن هذا النموذج هو نموذج الاستجابات متعددة التقسيم الوحيد المتضمن بتحليات البرنامج المذكور المتوافر فى زمن إجراء هذه الدراسة، كما أن للباحث خبرة ومهارة فى

استخدام هذا البرنامج من خلال ممارسته الوظيفية فى أنشاء بنوك الأسئلة (من النوع المقالى) وإجراء التحليلات الخاصة بها مستخدما هذا البرنامج.

برنامج تحليل مفردات بنوك الأسئلة "Winsteps":

يشتمل هذا البرنامج على عدة إجراءات من بينها تحليل وإعداد مفردات بنوك الأسئلة، ويتضمن هذا البرنامج كل من نموذجى "راش"، و"بارشال" فى التحليلات الخاصة بأنواع مختلفة من المفردات، ويعطى هذا البرنامج تقديرات لصعوبة المفردات بوحدة اللوجيت، وكذلك تقديرات لقدرة الأفراد بنفس الوحدة.

ويشتمل هذا البرنامج على مؤشرات إحصائية لملاءمة كل من الأفراد والمفردات بهدف الإبقاء على ما يصلح منها واستبعاد مالا يصلح لاستكمال إجراءات التحليل، وسوف يتم تناولها تفصيلاً كما يلى (Liancer, 2005, 127):

أولاً: إحصاءات الملاءمة للأفراد:

يستخدم البرنامج في تحليلاته للأفراد مؤشرى الملاءمة التقاربى، والتباعدى لغرض فرز الأفراد الصالحين لاستكمال عمليات وإجراءات التحليل وفيما يلي توضيح مفصل لكل مؤشر على حدة:

مؤشرالملاءمة التقاربى Infit (zstd):

يستخدم هذا المؤشر للتعرف على الأنماط غير الملاءمة لاستجابات الأفراد على مغردات اختبار معين وفقا للنحو التالى (Liancer, 2005, 197):

إذا كانت إحصاءات الملاءمة المتقاربة لها قيم موجبة كبيرة تجاوزت الحد (+٢) فإن هذا يشير إلى أن الطلاب يفشلون فى الإجابة الصواب على مفردات تقترب فى مستوى صعوبتها من مستوى قدراتهم (حيث يعتبر الباحث الحالى أن هذه الأنماط من الاستجابة تميز الأفراد الذى يتسم أدائهم على الاختبار بعدم الجدية أو الغش الفردى أوالتخمين).

أما إذا قلت إحصاءات الملاءمة عن الحد (−٢) فقد يشير ذلك إلى تشابه الاستجابات على الاختبار (ويعتبر الباحث الحالى أن هذه الأنماط من الاستجابة قد تميز الطلاب الذين يتسم أدائهم بالغش الجماعى أوالتخمين).

مؤشر الملاءمة التباعدى (zstd) Oufit:

كما يستخدم هذا المؤشر أيضا فى التعرف على الأنماط غير الملاءمة لاستجابات الأفراد على المفردات الاختبارية كالأتى (Liancer, 2005, 197):

إذا كانت إحصاءات الملاءمة المتباعدة لها قيمة كبيرة موجبة (تجاوزت الحد +٢) فإن هذا يدل على أن الطلاب يستطيعون الإجابة صوابا على مفردات تزيد فى مستوى صعوبتها عن مستوى قدراتهم أويفشلون فى الإجابة الصواب على مفردات تقل فى مستوى صعوبتها عن مستوى قدراتهم (ويمكن اعتبار ذلك النمط من الأنماط التى قد تميز استجابات الغش الفردى والتخمين) أما إذا قلت إحصاءات الملاءمة عن الحد (−٢) فقد يشير ذلك إلى تشابه إجابات الأفراد (وهذا قد يميز استجابات حالات الغش الجماعى).

ثانياً: إحصاءات الملاءمة لمفردات الاختبار:

إذ تجاوزت إحصاءات الملاءمة للمفردات الاختبارية لكل من المؤشرين السابقين عن القيمة (+٢) لقيمة z-standardized (zstd) الناتجة من التحليلات الخاصة بهذين المؤشرين، فإن هذه المفردات تعتبر ضعيفة من حيث مستوى الملاءمة،وهذا يشير على وجود عيبا فى صياغتها، أوعدم صدقها فى قياس ما تقيسه باقى المفردات، أما إذا كانت إحصاءات الملاءمة لنفس المؤشرين تقل عن القيمة (−٢) فيمكن اعتبارها متجاوزة لحدود الملاءمة، مما يدل على أن هذه المفردات شديدة التشابه فيما بينها، أو تعتمد على مفردات الاختبار الأخرى فهى غير مستقلة عنها وهذا يتعارض مع الفروض الخاصة بالنموذج، ولذلك يجب حذفها (Liancer, 2005, 197).

وقد اعتبر "لينكر" (المرجع السابق) أنه ليس هناك تقدير مطلق لحدود الملاءمة لكل من المؤشرين السابقين بل يجب أن يكون الحكم على الملاءمة متوافقا مع أهداف الموقف وقد اعتبر الباحث الحالى أن حدود الملاءمة المناسبة (±٢) تقريبا لكل من المؤشرين حتى يمكن الوصول إلى أنماط الاستجابات غير الملاءمة بشكل واضح ودقيق.

يتبين من خلال ما عرض سابقا من ادبيات وأطر نظرية وجود بعض الدراسات السابقة التي اهتمت ببعض جوانب المتغيرات المستهدفة من الدراسة الحالية ولكنها لم تتعلق بها تعلقا مباشرا، نظراً لعدم تمكن الباحث الحالي من الحصول على أية دراسات تتناول المؤشرات الإحصائية لبعض برامج التحليلات الإحصائية يمكن استخدامها للكشف عن البيانات المتعلقة بأساليب الغش والتخمين وعدم الجدية في المواقف

الاختبارية التي تستخدم في إعداد المفردات البنكية، تلك البيانات التي تساهم في إعطاء مؤشرات وصورة مزيفة عن هذه المفردات.

وعلى الرغم من عدم وجود دراسات تتعلق بمتغيرات الدراسة الحالية واهدافها بشكل صريح، فقد تبين من خلال هذه الدراسات أن سلوك الغش وعدم الجدية من آفات الموقف الاختبارى والتي تشكل ظاهرة شائعة ومنتشرة في بلدان كثيرة من العالم وهي من الظواهر التي يجب التصدى لها ودراستها (2001) Kaufman & Liebling، Dewyer & Jeffrey (2004)، Cizaek (2004)، وقد أشارت دراسات Bogan (1997)، Linda (1998)، Ronald (1999)، Kaufman & Liebling (2001)، Cizaek (2004)، إلى تأثر التحليلات الإحصائية لكثير من المفردات بهذه النوعية من الاستجابات مما يعطى صور مزيفة عن العديد منها، ويبدو للباحث الحالى من خلال دراستى Dorans (1998)، Ronald (1999) والدراسات التي أثبتت فعالية بعض المموشرات في الكشف عن الاستجابات غير الملائمة مثل دراسات Handrawan & Glas (2004)، Van ,Bay (2006) der lindend & Sotaridona (2006)، Liancre (2006)، Amy D., Greg B. (2010)، فقد اتضح من خلال النتائج التي توصلت إليها تلك الدراسات أن هناك عدد من المؤشرات الإحصائية التي تستخدم في كشف استجابات الغش مثل مؤشرات (t) (K-indx)، (Bm)، (Wallack.S,W)، (g2)، (ESA)، في حين أشارت دراسة Petridou, A(., Williams, J., 2011) صعوبة وجود مؤشرات إحصائية يمكن من خلالها تحديد الاستجابات غير الملائمة، وعلى الرغم من قدرة بعض المؤشرات التي اشارت إليها بعض الدراسات السابقة في

الكشف عن هذه النوعية من الاستجابات غير الملائمة التي أثرت سلبا في مؤشرات الملاءمة الإحصائية ودقة القياس بشكل عام إلا أنها ليست دارجة في البرامج الإحصائية التي تستخدم في تحليل بيانات الاختبارات في ضوء النماذج الرياضية المرتبطة بنظرية الاستجابة للمفردة، والمتداولة والمتوفرة بمصر والعالم العربي، مثل برامج Winsteps، بايلوج Bilog، راسكال Rascal، رم Rumm، مايكروسكيل Microscale والتي تستخدم في العديد من المؤسسات العلمية المعنية بتحليل المفردات وإعداد بنوك الأسئلة، هذا ويتضح أيضا من خلال تفقد هذه الدراسات أن جميعها لم تستخدم مفردات اختبارية تحتلف عن مفردات الاختيار من متعدد مثل مفردات من نوع المقال، كما أنها لم تستخدم أيضا نماذج تعتمد على الاستجابات متعددة التقسيم مثل نموذج "التقدير الجزئي" وهذا تعني به الدراسة الحالية.

ونظرا لأن برنامج Winsteps هو أحد البرامج الإحصائية الأكثر استخداما في تحليل المفردات الاختبارية وكذلك إنشاء بنوك الأسئلة التي تعتمد على النماذج الرياضية اللوغاريتمية المرتبطة بالنظرية الحديثة للقياس (IRT) في مصر والعالم العربي، فيرى الباحث أنه قد يمكن اكتشاف وفرز البيانات غير الملائمة والناجمة عن استخدام أساليب الغش الفردى وكذلك حالات الغش الجماعى، من خلال المؤشرات الإحصائية المتوافرة ببرنامج Winsteps مثل مؤشرى الملاءمة (التقاربي والتباعدي) والذي يعتمد عليهما البرنامج في اختبار ملاءمة البيانات لنموذج التحليل، كما يمكن لنفس المؤشرات اكتشاف وفرز البيانات الاختبارية الناتجة عن استجابات الاهمال وعدم الجدية (كأنماط استجابة غير ملائمة) وهي تلك الاستجابات التى أشارت إليها دراستى Bogan

(1997)، Linda (1998)، وبناء على المقترحات التي وردت ببعض الدراسات مثل دراسة Linda (1998)، Meijer (2005)، Fan & Ping (2006) (بشأن إجراء دراسات للبحث عن مؤشرات إحصائية لهذه الأساليب غير الملائمة من الاستجابات، وهذا مما حدا بالباحث إلى محاولة التعرف على هذه النوعية من الاستجابات الدالة على استجابات الغش من خلال بعض المؤشرات الإحصائية المتضمنة ببعض البرامج الإحصائية مثل برنامج Winsteps، وهو البرنامج الأكثر استخداما فى تحليل مفردات الاختبارات المختلفة وفقا لنظرية الاستجابة للمفردة (IRT) وذلك فى وقت وزمن إجراء هذه الدراسة.

فروض الدراسة:

من خلال ما سبق عرضه من دراسات سابقة يمكن للباحث صياغة الفروض التالية:

الفرض الأول:

تختلف مؤشرات الملاءمة الإحصائية (الناتجة من التحليل باستخدام برنامج Winsteps) لاستجابات الممتحنين الذين يتسم أدائهم بالغش الفردي والغش الجماعي وعدم جدية في الموقف الاختباري عن الممتحنين المتميزين بانضباط الأداء.

الفرض الثاني:

تختلف مؤشرات الملاءمة الإحصائية لمفردات اختبار المقال (الناتجة من التحليل باستخدام برنامج WINSTEPS) قبل وبعد حذف استجابات الممتحنين الذين يتسم أدائهم بالغش الفردى،الجماعى، عدم الجدية فى الموقف الاختبارى.

الفرض الثالث:

لا تختلف تقديرات الصعوبة لمفردات اختبار المقال الناتجة من تحليلها وفقا لاستجابات العينة المستخدمة بالدراسة الحالية قبل وبعد حذف جميع حالات عدم الملاءمة (غش فردى، غش جماعى، عدم جدية)، باستخدام برنامج Winsteps.

إجراءات الدراسة:

يتضمن هذا الجزء العينة المستخدمة فى إجراءات الدراسة الحالية، وكذلك الأداة التى تم بإعدادها بهدف جمع البيانات المتعلقة بالدراسة الحالية، وفى النهاية الإجراءات الخاصة باختبار الفرضيات البحثية.

أولاً: عينة الدراسة:

تكونت عينة الدراسة الحالية من (٧٤٦) طالب من الصف الثانى الإعدادى من التعليم العام دون النوعى، بهدف اختبار فروض الدراسة المختلفة، والجدول التالى يوضح توزيع هذه العينة وفقا لإجراءات اختبار هذه الفروض.

جدول رقم (١)

توزيع أفراد العينة الكلية كما استخدمت فى اختبار الفروض المختلفة للدراسة الحالية

العينة	عدد الأفراد
منضبطين	٥١٤
غش فردى	١٤٤
غش جماعى	٦٠
عدم جدية	٢٢
المجموع	٧٤٠

مبررات اختيار العينة:

لقد وقع اختيار الباحث الحالى لعينة الدراسة وفقا للمبررات الآتية:

- طلاب هذه المرحلة على درجة من النضج من استيعاب التعليمات الخاصة بالموقف الاختبارى بسهولة.
- الصف الثانى الإعدادى يمثل منتصف الحلقة الثانية من التعليم الأساسى مما يمثل استقرارا للطلاب.
- تكثر حالات الغش الفردى والجماعى فى امتحانات هذه المرحلة، وقد توصل الباحث إلى هذه المعلومة من خلال خبرته أثناء عمله مدرسا للرياضيات بالتعليم العام.

ثانيا: الأداة المستخدمة بالدراسة:

استخدمت بالدراسة الحالية أداة واحدة فقط وهى اختبار يتضمن أسئلة من نوع المقال حيث تضمن هذا الاختبار فى صورته النهائية (١٣) مفردة فى مادة جبر الصف الثانى الإعدادي (أنظرملاحق الدراسة– ملحق رقم ١)، وقد تم إعداد الاختبار وفقا لجدول مواصفات مادة جبر الصف الثانى الإعدادى والذى تم أعداده بالمركز القومى للامتحانات والتقويم التربوى، وبعد إعداد الاختبار فى صورته المبدئية (مشتملا على ١٥ مفردة) تم عرضه على مجموعة من المحكمين (أنظر قائمة بأسماء المحكمين بملاحقة الدراسة– ملحق رقم ٢)، وقد استقر الاختبار فى صورته النهائية (أنظر ملاحق الدراسة– ملحق رقم ١) بعد إجراء التعديلات المقترحة من قبل المحكمين ووفقا للملاحظات التى تم جمعها من الموقف الاختبارى الاستطلاعى الذى تم من خلاله تجريب الاختبار على عينة مكونة من (١٥١) طالب من طلاب الصف الثانى الإعدادى

من المدارس الحكومية بالمقطم، ومن خلال تحليل الاستجابات الخاصة بهؤلاء الطلاب باستخدام تحليلات برنامج Winsteps (نموذج بارشال) تم استبعاد بعض المفردات غير الملائمة وفقا للمؤشرات الإحصائية للملاءمة المتضمنة بتحليلات هذا البرنامج (وهى المفردات التى تجاوزت الحدود المقبولة لمؤشرى الملاءمة التقاربى، والتباعدى، وهى المفردات التى تقل مؤشرات الملاءمة لها عن (−٢) والأخرى التى تزيد مؤشرات الملاءمة لها عن (+٢) لكل من مؤشرى الملاءمة التقاربى والتباعدى، وقد بلغ عدد المفردات التى تم استبعادها وفقا لهذين المؤشرين عدد (٢) من المفردات.

التحقق من الخصائص السيكومترية للاختبار:

قام الباحث بحساب معامل الثبات للاختبار فى صورته النهائية باستخدام معادلة "ألفا كرونباخ"، وقد بلغ معامل ثبات الاختبار (٠.٨٧٥) وهو معامل ثبات مرتفع يمكن من خلاله الوثوق بما يسفر عنه الاختبار من نتائج، وحيث أن الاختبار هو اختبار تحصيلى فى مادة الرياضيات فقد اكتفى الباحث بصدق المحكمين للتحقق من صدق الاختبار، وقد تم تحديد نسبة (٨٥%) على الأقل من تكرارات موافقة المحكمين لكل مفردة اختبارية لوضعها بالصورة النهائية للاختبار كحد أدني لقبول المفردة ولضمان صدق ما أعدت لقياسه

ثالثاً: إجراءات اختبار الفرضيات البحثية:

نظرا لأن فروض الدراسة الحالية تختبر العلاقة بين حدوث حالات الغش وعدم الجدية ودقة اكتشافها باستخدام مؤشرى الملاءمة الإحصائية (التقاربي والتباعدي)، المتضمنان بالتحليلات الخاصة ببرنامج Winsteps

ونظراً للاستخدام الواسع الانتشار لهذا البرنامج وبصورة خاصة فى بعض المراكز العلمية المتخصصة والمعنية بهذا الأمر، ونظرا لأهمية هذين المؤشرين فى تمحيص المفردات الاختبارية الصالحة وغير الصالحة وكذلك الأفراد، فإن الباحث الحالى يرى أنه لمزيد من الدقة فى اختبار حساسية هذين المؤشرين لل أنواع المختلفة من الاستجابات الخاصة بتلك الحالات غير الملائمة، يجب دراسة جميع الاحتمالات التى يمكن تصورها لعينة الدراسة الحالية (كعينة تحليل) والتى يمكن أن تحدث فى العديد من المواقف الاختبارية، وبالتالى يمكن أن تتشكل عينة التحليل بالدراسة الحالية على النحو التالى:

- عينة المنضبطين– عينة المنضبطين + حالات الغش الفردى– عينة المنضبطين + حالات الغش الجماعى– عينة المنضبطين + حالات عدم الجدية– عينة المنضبطين + حالات الغش الفردى + حالات الغش الجماعى– عينة المنضبطين + حالات الغش الفردى + حالات عدم الجدية– عينة المنضبطين + حالات الغش الجماعى + حالات عدم الجدية– عينة المنضبطين + حالات الغش الفردى + حالات الغش الجماعى + حالات عدم الجدية، وفيما يرى عرض للإجراءات الخاصة بتحديد حالات عدم الملاءمة المطروحة بالدراسة الحالية:

اولاً: الإجراءات الخاصة بتحديد حالات عدم الملاءمة (غش– عدم جدية):

١– الإجراءات الخاصة بتحديد حالات الغش الفردى:

- تطبيق اختبار الدراسة على العينة الكلية المحددة بالدراسة الحالية مع ضبط اللجنة بشدة وعدم السماح بحدوث أى حالة غش نهائيا.

- يقوم الباحث بتسجيل حالات الغش الفردى (السماح لبعض الممتحنين بالحصول على الإجابات الصحيحة لبعض الأسئلة التي أجابو عنها إجابة خاطئة) تبعا للخطوات التالية:

- إعداد مفتاح إجابة لجميع أسئلة الاختبار يتضمن جميع رموز البدائل التي تمثل الإجابات الصحيحة لجميع أسئلة اختبار المقال المستخدم بالدراسة (والذي يتألف من عدد ١٣ سؤال) وطبع أكثر من نسخة منه وتوزيعها على بعض مدرسى الرياضيات بالمدرسة الجارى تتطبيق الاختبار بها، وذلك لقيامهم بأدوار محددة ومرسومة لتحقيق أهداف الدراسة.
- الاستعانة بالمدرسين المشار إليهم بالخطوة السابقة للمرور على الممتحنين بالموقف الاختبارى التطبيقى للاختبار بهدف التعرف على الحالات التي لم تتمكن من الإجابة بشكل صحيح عن نصف العدد الكلي لأسئلة الاختبار أو الحالات التي لم تتمكن من مجرد الإجابة على نفس العدد أى عن عدد من الأسئلة يقارب (٦) أسئلة، وذلك قبل نهاية الوقت المخصص للاختبار بنصف ساعة تقريبا لتنفيذ ما هو مطلوب لتكوين حالات الغش الفردى بالأسلوب التالى:

أ. إعطاء الإجابة الصحيحة للطلاب الذين لم يتمكنون من الإجابة عن أكثر من نصف عدد أسئلة الاختبار كما سبقت الإشارة وذلك لعدد لا يزيد عن سؤالين للممتحنين الذين لم تتجاوز إجاباتهم عن عشرة مفردات.

ب. إعطاء الإجابة الصحيحة للممتحنين الذين لم يتمكنون من الإجابة إلا عن عدد (٨) من أسئلة الاختبار، وذلك لعدد لا يزيد عن سؤال واحد.

ج. وضع علامات محددة (أكواد) متفق عليها على أوراق إجابة الممتحنين المتلقين للإجابات الصحيحة كحالات للغش الفردى وفقا لعدد الأسئلة الذين تلقوا الإجابة الصحيحة عنها.

د. مراعاة إعطاء وتوزيع الإجابة عن الأسئلة كما تم توضيحه بالخطوات السابقة بشكل عشوائي غير مرتب وليست كما ورد ترتيبها بالاختبار.

والجدول التالي رقم (٢) يعرض عدد حالات الغش الفردي على جميع أسئلة الاختبار.

جدول رقم (٢)

توزيع حالات الغش الفردى على جميع أسئلة اختبار المقال

رقم السؤال	عدد حالات الغش
١	٧
٢	٩
٣	٨
٤	١٥
٥	١٠
٦	٩
٧	١٠
٨	١٣
٩	١٤
١٠	١١
١١	١٢
١٢	١٤
١٣	١٢
المجموع	١٤٤

٢ – الإجراءات الخاصة بالغش الجماعي:

يقوم الباحث بتسجيل حالات الغش الجماعي (السماح لمجموعات من الممتحنين بالحصول على الإجابات الصحيحة لبعض الأسئلة التي أجابوا عنها إجابة غير صحيحة) تبعا للخطوات التالية:

- تحديد الأسئلة الأكثر صعوبة عن غيرها في اختبار الدراسة (اختبار المقال) لاختيار عدد قدرة ستة أسئلة من بينها وهى المفردات التي نتجت تقديرات الصعوبة الخاصة بها من تحليل عينة المنضبطين فقط بعد استبعاد جميع حالات عدم الملاءمة.
- إملاء الإجابات الصواب لكل مفردتان مختلفتان منها على ثلاثة لجان امتحانية مختلفة بشكل جماعي في الربع ساعة الأخيرة من زمن الإجابة على الاختبار وبلجان لم يتم فيها حدوث حالات الغش الفردى، ويتم هذا الإجراء من قبل مدرسي الرياضيات المتفق معهم لتحقيق أهداف الدراسة كما سبقت الإشارة من قبل، وقد تم توزيع حالات الغش الجماعى على الأسئلة الكثر صعوبة في الاختبار وفقا للجدول التالى:

جدول رقم (٣)

حالات الغش الجماعى موزعة على الأسئلة الأكثر صعوبة باختبار المقال

أرقام الأسئلة الأكثر صعوبة	عدد حالات الغش الجماعي	أرقام الأسئلة الأكثر صعوبة	عدد حالات الغش الجماعي	أرقام الأسئلة الأكثر صعوبة	عدد حالات الغش الجماعي
٢	٢٤	٦	٢٢	٨	١٤
٥		٧		٩	

٣- الإجراءات الخاصة بتحديد حالات عدم الجدية:

يقوم الباحث بتطبيق نفس الاختبار فى لجنة أخرى تختلف عن اللجان المستخدمة مع كل من الإجراء الأول والثانى، مع ضبط اللجنة وتدوين الحالات غير الجادة ويمكن التعرف عليها من خلال الاستجابات الدالة عليها وهى محددة كالآتى:

أ- تسليم ورقة الإجابة فى زمن لا يتجاوز ربع الوقت المخصص للإجابة.

ب- وجود بعض الكتابات الساخرة.

ج- ظهور أى سلوك آخر يدل على عدم الجدية.

د- عدم الالتزام بتعليمات الموقف الاختبارى وفقا لرؤية الملاحظ أو مقدر الدرجات.

والجدول التالي يوضح عدد حالات عدم الجدية وفقا للمحددات السابقة

جدول رقم (٤)

حالات عدم الجدية موزعة وفقا لمحدداتها

المحددات	عدد الحالات
تسليم ورقة الإجابة فى زمن لا يتجاوز ربع الوقت المخصص للإجابة.	٨
وجود بعض الكتابات الساخرة	٣
ظهور أى سلوك آخر يدل على عدم الجدية	٥
عدم الالتزام بتعليمات الموقف الاختبارى وفقا لرؤية الملاحظ أو مقدر الدرجات	٦
المجموع	٢٢

٤- تحديد وحصر عدد استجابات حالات عدم الملاءمة لكل سؤال من أسئلة الاختبار:

تم حصر حالات عدم الملاءمة السابقة (غش فردي- غش جماعي- عدم الجدية)، وفقا لكل سؤال من أسئلة الاختبار لتحديد النسب المئوية لها، وقد اتضح من خلال الجداول السابقة أن أعلى نسبة مئوية للاستجابات الخاصة بحالات عدم الملاءمة المختلفة لأسئلة الاختبار، كانت للسؤال رقم (٥) حيث بلغت حالات استجابات عدم الملاءمة له (غش الفردى (١٠) حالة، غش جماعى (٢٤) حالة)، وبالتالي بلغ المجموع الكلي لاستجابات الحالات غير الملائمة على هذا السؤال عدد (٣٤) حالة عدم ملاءمة من جملة (٧٤٠) ممتحن أجابوا على هذه المفردة بنسبة (٤.٦%) تقريباً من إجمالى عدد استجابات العينة الكلية، هذا بالإضافة على نسبة (٣ %) تقريباً من استجابات حالات عدم الجدية البالغ عددها (٢٢) حالة من جملة استجابات العينة الكلية (٧٤٠) والتي لم يظهر تأثيرها على أسئلة محددة ولكن كاستجابة غير ملائمة للموقف الاختباري ككل، وبالتالي بلغ عدد نسبة الاستجابات غير الملائمة المفترضة على نفس السؤال (٧.٦ %) ثم جاءت عدد استجابات غير الملائمة لجميع الأسئلة المتبقية بالاختبار بنسب مئوية أقل مما يدل على قبول ما تسفر عنه التحليلات الإحصائية من تقديرات ومؤشرات لأسئلة الاختبار المستخدم بالدراسة بدرجة كبيرة من الثقة، نظراً لأن عدد استجابات الحالات غير الملاءمة في المواقف الاختبارية الخاصة بتجريب الأسئلة تزيد بكثير عن نظائرها في الاختبارات الخاصة بالامتحانات الرسمية وهذا ما اتضح للباحث من خلال خبرته للعمل بهذا

المجال سواء من خلال عمله مدرسا للرياضيات بالتعليم العام أو عمله باحثاً بالمركز القومي للامتحانات والتقويم التربوى.

ثانياً: الإجراءات الخاصة باختبار الفرضيات البحثية:

١- تحليل الاستجابات الخاصة بجميع الممتحنين على اختبار المقال المستخدم بالدراسة الحالية بعد دمجها فى ملف واحد باستخدام برنامج Winsteps.

٢- حصر الأفراد الذين تجاوزوا حدود الملاءمة الإحصائية وفقا لمؤشرات الملاءمة ببرنامج التحليل.

٣- مقارنة الحالات المرصودة من اللجنة (حالات الغش الفردى، والجماعى، عدم الجدية) مع الحالات التى أسفرت عنها التحليلات الإحصائية لبرنامج التحليل من خلال مؤشرات الملاءمة الخاصة به.

٤- إيجاد النسب المئوية لحالات عدم الملاءمة التى تم التعرف عليها من خلال المؤشرات المذكورة سابقا منسوبة إلى عدد الحالات الفعلية التى تم رصدها من خلال الموقف الاختبارى.

٥- اختبار دلالة فروق النسب المئوية (باستخدام اختبار Z) المعدل باستخدام بنفروني Binferroni للمقارنات المتعددة (لتجنب تضخم الخطأ من النمط اللأول) (User's Guide. SPSS Inc., 2010) لعدد الحالات التى تجاوزت حدود الملاءمة الإحصائية لبرنامج التحليل، وذلك بين الحالات المختلفة لعدم الملاءمة المحددة بالدراسة الحالية (غش فردى- غش جماعى- عدم الجدية) وعينة المنضبطين.

٦- حذف الاستجابات الخاصة بحالات الغش الفردى من عينة التحليل الكلية المستخدمة فى الدراسة الحالية ثم تحليل مفردات المقال وفقا لاستجابات العينة المتبقية (منضبطين + غش جماعى + عدم جدية) باستخدام برنامج التحليل.

٧- تكرار نفس الخطوة السابقة ولكن مع حذف الحالات الأخرى على التوالى (غش جماعى، عدم جدية).

٨- تحديد عدد مفردات الاختبار التى تجاوزت الحدود المقبولة لمؤشرات الملاءمة الإحصائية وحساب النسبة المئوية لها منسوبة إلى العدد الكلى لمفردات الاختبار نتيجة تحليل استجابات العينة المحددة لكل خطوة من الخطوات السابقة.

٩- اختبار دلالة فروق النسب المئوية (باستخدام اختبار Z) لعدد المفردات التى تجاوزت الحدود المقبولة لمؤشرات الملاءمة الإحصائية للعينات التى تم تحديدها فى الخطوات (١١ إلى ١٢) لبرنامج التحليل.

١٠- تحليل اختبار الدراسة وفقا لاستجابات عينة المنضبطين بالإضافة إلى استجابات عينة الغش الفردى باستخدام برنامج التحليل.

١١- تكرار الخطوة رقم (١٠) ولكن مع استخدام الاستجابات الأخرى لحالات عدم الملاءمة على التوالى (غش جماعى، عدم جدية).

١٢- تحديد عدد مفردات الاختبار التى تجاوزت الحدود المقبولة لمؤشرات الملاءمة الإحصائية وحساب النسبة المئوية لها منسوبة إلى العدد الكلى لمفردات الاختبار.

١٣- اختبار دلالة فروق النسب المئوية لعدد المفردات التى تجاوزت الحدود المقبولة لمؤشرات الملاءمة الإحصائية للعينات التى تم تحديدها فى الخطوتين (١٠- ١١) لبرنامج التحليل.

١٤- الحصول على تقديرات الصعوبة لاختبار الدراسة نتيجة تحليل استجابات العينة الكلية قبل وبعد حذف جميع الأفراد غير الملائمين.

١٥- إيجاد الفروق فى تقديرات الصعوبة للمفردات المتناظرة الناتجة من التحليل قبل وبعد إجراء الحذف المشار إليه سابقا.

١٦- تحديد الفروق الجوهرية بين تقديرات الصعوبة المتناظرة لمفردات الاختبار على النحو التالى:

- إذا كان الفرق بين تقديرى الصعوبة المتناظرين لكل مفردة أكبر من مجموع الخطأين المعياريين لهما كان هذا الفرق جوهريا.
- إذا كان الفرق فى تقديرى الصعوبة المتناظرين لكل مفردة أقل من أو يساوى مجموع الخطأين المعياريين لهما كان فرقا غير جوهريا. (Lidwine B.,2006)، (Aimee,M, 2008).

١٧- حساب النسبة المئوية للفروق الجوهرية لتقديرات الصعوبة المتناظرة لكل من التحليلين (قبل وبعد إجراء الحذف) فإذا كانت هذه النسبة تقل عن أو تساوى (٥%) دل ذلك على عدم وجود اختلاف جوهرى بين تقديرات الصعوبة المتناظرة أما إذا ازدادت عن هذه النسبة دل ذلك على وجود اختلافات جوهرية لهذه التقديرات.

عرض ومناقشة نتائج الدراسة:

أولاً: عرض ومناقشة نتائج اختبار الفرض الأول:

أظهرت نتائج اختبار الفرض الأول والذى ينص على: تختلف مؤشرات الملاءمة الإحصائية (الناتجة من التحليل باستخدام برنامج Winsteps) لاستجابات الممتحنين الذين يتسم أدائهم بالغش الفردي والغش الجماعي وعدم جدية في الموقف الاختباري عن الممتحنين المتميزين بانضباط الأداء.

وقد تم اختبار الفرض السابق وذلك وفقا للإجراءات الثلاث التالية:

١- تحليل استجابات العينة الكلية والتى تشتمل على حالات المنضبطين وحالات الغش الفردى والجماعى وحالات عدم الجدية، على اختبار المقال باستخدام برنامج التحليل Winsteps.

٢- حصر العدد والنسب المئوية للحالات التى تجاوزت حدود الملاءمة الإحصائية لكل من مؤشرى الملاءمة الإحصائية التقاربى والتباعدى المتضمنان ببرنامج Winsteps.

٣- اختبار دلالة فروق النسب المئوية للحالات التى تجاوزت حدود الملاءمة وفقا للفئات المحددة للدراسة الحالية (منضبطين غش فردى، غش جماعى، عدم جدية) وفقا لبرنامج التحليل السابق.

وسوف يتم فيما يلى عرض النتائج التى تم التوصل إليها وفقا للإجراءات الثلاث السابقة:

١ – النتائج التى أسفر عنها الإجراء الأول وهى نتائج تحليل استجابات العينة الكلية على اختبار المقال باستخدام برنامج Winsteps:

يوضح الجدول رقم (٦) مؤشرات الملاءمة الإحصائية التقاربي والتباعدي لحالات غير الملائمة التى أسفرت عنها التحليلات الخاصة باستجابات عينة التحليل الكلية باستخدام برنامج Winsteps على اختبار الدراسة.

جدول رقم (٦)

الحالات التى تجاوزت حدى الملائمة الإحصائية لمؤشر الملاءمة التقاربى والتباعدى وفقا لاستجابتهم على اختبار الدراسة (باستخدام برنامج Winsteps)

أرقام حالات عدم الملائمة بالعينة الكلية	تقدير القدرة باللوجيت	مؤشر الملائمة التقاربى infit	مؤشر الملائمة التباعدى outfit	الخطا المعيارى للقياس	النوع
١٨٦	– ٠.٨١	٢.٢٠	٢.٦٠	٠.٢٦	غش فردى
١٨٨	– ٠.٤٢	٢.٦٠	٢.٦١	٠.٢٣	غش فردى
١٨٥	– ٠.١٩	٣.٥٤	٣.٢٧	٠.٢٣	غش فردى

ويتضح من الجدول رقم (٦)، أن التحليلات التى أسفرت عن استخدام برنامج التحليل أدت إلى اكتشاف حالات محدودة لعدم الملاءمة (تجاوزت الحدود المقبولة للملاءمة) من عينة الغش الفردى فقط حيث استطاع البرنامج إكتشاف حالت أن للغش الفردى رقمى (١٨٥، ١٨٨) تلقتا إجابتان لسؤالان من أسئلة الاختبار، وحالة غش فردى واحدة تلقت إجابة لسؤال واحد وهى رقم (١٨٦)، بينما لم تُكتشف أى حالة من الحالات الأخرى (غش جماعى، عدم الجدية) وجميع حالات عدم الجدية.

٢- النتائج التى أسفر عنها الإجراء الثانى وهى نتائج حصر العدد والنسب المئوية للحالات التى تجاوزت حدود الملاءمة الإحصائية المتضمنة ببرنامج Winsteps:

يوضح الجدول التالى رقم (٧) عدد ونسبة الممتحنين المتجاوزين لحدود الملاءمة الإحصائية (لتقاربي، والتباعدي) فى كل من الحالات الأربع (انضباط- غش فردى- غش جماعى- عدم جدية) وفقا لاستجاباتهم على اختبار المقال.

جدول رقم (٧)

الحالات التى تجاوزت مؤشرى الملاءمة الإحصائية (التقاربي، والتباعدي) والنسب المئوية لها الناتجة من تحليل أداء العينة الكلية على اختبار المقال

الحالة	العدد الكلى	عدد الممتحنين المتجاوزين لحدى مؤشر Infit (± ٢)				عدد الممتحنين المتجاوزين لحدى مؤشر Out (± ٢)				جملة حالات غير الملائمين	
		٢	- ٢	المجموع	النسبة المئوية	٢	- ٢	المجموع	النسبة المئوية	العدد	النسبة المئوية
إنضباط	٥١٤	–	–	–	–	–	–	–	–	–	–
غش الفردى	١٤٤	٢	–	٢	١.٤	٣	–	٣	٢.٠٨	٣	٢.٠٨
غش الجماعى	٦٠	–	–	–	–	–	–	–	–	–	–
عدم الجدية	٢٢	–	–	–	–	–	–	–	–	–	–

يبدو من النتئج الموضحة بالجدول السابق رقم (٧) عدم تجاوز أية حالة من حالات عدم الجدية أوالغش الجماعى وأغلب حالات الغش الفردى حدود الملاءمة الإحصائية، بينما تجاوزت ثلاث حالات للغش

الفردى بنسبة (٢.٠٨%)، وهى أرقام (١٨٥، ١٨٦، ١٨٨) الحد (+٢) لكل من لمؤشر الملاءمة التقاربى، وكذلك مؤشر الملاءمة التباعدى، حيث تجاوزت الحالتان (١٨٥، ١٨٨) نفس الحد لمؤشر الملاءمة التقاربى، وتجاوزت الثلاث حالات نفس الحد مع المؤشر التباعدى.

٣- النتائج التى أسفر عنها الإجراء الثالث والخاص باختبار دلالة فروق النسب المئوية للحالات التى تجاوزت حدود الملاءمة وفقا للفئات المحددة للدراسة الحالية (منضبطين غش فردى، غش جماعى، عدم جدية) وفقا لبرنامج التحليل.

يوضح الجدول رقم (٨) قيمة (Z) ودلالتها لاختبار فروق النسب المئوية بين حالات الغش وعدم الجدية وحالات المنضبطين التى تجاوزت حد الملاءمة التقاربى الناتجة من تحليل الاستجابات الخاصة بها على اختبار المقال باستخدام برنامج.

جدول رقم (٨)

قيمة (Z) ودلالتها لاختبار فروق النسب المئوية بين حالات الغش وعدم الجدية وحالات المنضبطين التى تجاوزت حد الملائمة التقاربى (اختبار المقال)

الحالات	منضبطة	غش فردى	غش جماعى	عدم جدية
منضبطة	-			
غش فردى	٢.٧*	-		
غش جماعى	-	٠.٩٢	-	
عدم جدية	-	٠.٥٦	-	-

* دال عند مستوى (٠.٠٥).

يتضح من الجدول رقم (٨) عدم وجود فروق دالة إحصائيا بين النسب المئوية الخاصة بحالات المنضبطين وحالات الغش الجماعي وعدم الجدية نظرا لعدم تجاوز أى حالة من هذه الحالات حدود الملاءمة الإحصائية لمؤشر الملاءمة التقاربى، فى حين أظهرت النتائج وجود فروق دالة إحصائية بين النسب المئوية لعدد حالات المنضبطين وعدد حالات الغش الفردى الذين تتجاوزت استجاباتهم حد الملاءمة الإحصائية للمؤشر التقاربى وفقا لاستجابتهم على اختبار الدراسة، وهذه النتيجة تثير الدهشة حيث أنه من المعروف وفقا لتعليمات برنامج Winsteps (Liancer, 2005, 127)، أن نسبة تجاوز حدود الملاءمة الإحصائية لحالات عدم الملاءمة (غش– عدم جدية) وفقا لهذا المؤشر الإحصائى سوف تزيد عن نسب حالات المنضبطين، وهذا ما لم يتحقق إجرائيا، إلا مع حالات قليلة جداً من حالات الغش الفردى كما سبقت الإشارة من قبل، وبرغم عدم وجود هذا الفرق الدال إلا أنه جاءت النتائج عكس المتوقع، حيث يتبين أن البرنامج اكتشف بعض الحالات القلية للغاية من حالات عدم الملاءمة (غش فردى) إلا أنه لم يفلح فى اكتشاف الغالبية العظمى منها وبصفة خاصة حالات الغش الجماعى وعدم الجدية.

وهذه النتيجة إن دلت على شئ فإنما تدل على أن هذه المؤشرات الإحصائية لا تستطيع الكشف والتعرف بشكل حاسم ومؤكد على حالات الغش التي يمكن ان تحدث بالفعل في المواقف الاختبارية وهذا ماأوضحته وأشارته إليه دراسة دوير وجيفري (Dewyer & Jeffrey, 2004) التي نوهت عن صعوبة الكشف عن مثل هذه الاستجابات من خلال بعض المؤشرات الإحصائية.

جدول رقم (٩)

قيمة (Z) ودلالتها لاختبار فروق النسب المئوية بين حالات الغش وعدم الجدية وحالات المنضبطين التى تجاوزت حد الملاءمة التباعدى (اختبار المقال)

الحالات	منضبطة	غش فردى	غش جماعى	عدم جديه
منضبطين	–			
غش فردى	٣.٣*	–		
غش جماعى	–	١.١	–	
عدم جديه	–	٠.٢٥		–

* دال عند مستوى ٠.٠٥

كما يتضح أيضا من الجدول رقم (٩) وجود فروق داله إحصائياً بين بعض النسب المئوية لعدد الحالات التى تجاوزت حدى الملاءمة الإحصائية التباعدى وفقاً لاستجابتهم على نفس الاختبار، فعند مقارنة نسب تجاوز عدد حالات الغش الفردى بنسب الحالات الأخرى التى تجاوزت هذين الحدين، يتبين وجود فروق دالة فى جانب حالات الغش الفردى فقط إذا ما قورنت بعينة المنضبطين، وهذه النتيجة تتشابه مع النتائج التي توصلت إليها دراسات Handrawan & Glas (2004)، Meijer (2005)، Bay (2003) والتي أشارت إلى إمكانية التمييز بين الأفراد الملائمين وغير الملائمين من خلال بعض مؤشرات الملاءمة الإحصائية، وبرغم هذه النتيجة إلا أنه يتضح عدم قدرة البرنامج أيضا فى كشف الغالبية العظمى من هذه الحالات، فبرغم اكتشافه لعدد ضئيل من حالات الغش الفردى إلا أنه لم يتمكن من التعرف على بقية الحالات الأخرى من عدم الملاءمة (غش جماعى، عدم جديه) مما يثير العديد

من الشكوك حول دقة هذه البرنامج، والذى يستخدم حالياً فى العديد من المؤسسات العلمية،، وهذا ما توصلت إليه أيضا دراسة ديوير وجيفري Dewyer & Jeffrey, (2004) وأيضا دراسة Williams, J., Petridou, A. (2011) التي اشارت بصعوبة العثور على ادلة إحصائية يمكن من خلالها تمييز الاستجابات غير الملائمة.

ومن خلال نتائج اختبار الفرض الأول السابق يتبين أن الفرض لم يتحقق إجرائيا فيما يتعلق باختلاف بعض مؤشرات الملاءمة الإحصائية بين جميع العينات المستهدفة من الدراسة باستخدام برنامج التحليل WINSTEPS، حيث أظهرت النتائج عدم وجود دلالة فى فروق النسب المئوية لعدد الحالات التى تجاوزت حدود مؤشر الملاءمة التقاربى، إلا مع حالات قليلة جدا للغش الفردى عند مقارنتها بحالات المنضبطين فقط، كما ظهرت فروق دالة إحصائياً بين نسب أعداد تجاوز هذه الحالات لحدود مؤشر الملاءمة التباعدى وذلك فى جانب عينة الغش الفردى عند مقارنتها بعينة المنضبطين أيضا، وهذه النتائج غير المتوقعة هى نفس النتائج تقريبا التى تم التوصل إليها مع المؤشر السابق، مما يدل على هذين المؤشرين فشلا فى التمييز بين استجابات حالات الغش الجماعى وعدم الجدية وأغلب حالات الغش الفردى واستجابات المنضبطين كما هو مفترض ومتوقع وهذا ما يتناقض مع ما ورد عن هذا البرنامج فى اكتشاف هذه الحالات من عدم الملاءمة (Liancer, 2005: 197)

من خلال مناقشة نتائج الفرض السابق تبين وجود نتائج غير متوقعة تتمثل فى عدم وجود اختلافات جوهرية بين أغلب عينات الدراسة الخالية والتى تشتمل على حالات عدم ملاءمة، وذلك فيما يتعلق بتجاوز

هذه الحالات لحدود الملاءمة الإحصائية المتضمنان ببرنامج Winsteps، وهذا مما حدا بالباحث لتحليل أنماط الاستجابة الخاصة بحالات الغش الفردى التى أسفرت عنها تحليلات هذا البرنامج وفقا لاستجاباتهم لاختبار المقال المكون من ثلاث عشر مفردة، والجدول التالى رقم (١٠) يوضح الأنماط المختلفة من الحالات التى تم اكتشافها من قبل البرنامج (حالات الغش الفردى وهى أرقام ١٨٨،١٨٦،١٨٥) وثلاث حالات مثالية هى أرقام (١١١،٤٩،٤٢) تم اختيارها من عينة المنضبطين (تقترب قيم مؤشرات الملاءمة لها من الصفر)، وثلاث حالات أخرى من عينات عدم الملاءمة والتى لم يتعرف عليها البرنامج وهى أرقام (١٥٣، ١٨٣، ٢١٢) وهى حالات للغش الجماعى والغش الفردى وعدم الجدية على الترتيب، تم اختيارها من الواقع التطبيقى للاختبار، وتبين من هذا الجدول أن حالات الغش الفردى (التى أكتشفت) كانت أغلب استجاباتها الصواب والتى حصلت فيها على درجات مرتفعة كانت على أسئلة الاختبار مرتفعة الصعوبة (بالخلايا المظللة) بينما قلت الاستجابة الصواب وقلت الدرجات المرتفعة على الأسئلة السهلة، وبالتالى كانت أنماط استجابة غير متوقعة وفقا لفلسفة البرنامج (Liancer, 2005: 197)، فتعرف عليها البرنامج جميعا، أما بالنسبة للحالات المتبقية (المثالية، وحالات غير الملائمة التى لم يتعرف عليها البرنامج كما هى موجودة بالجدول السابق)، كانت اغلب استجابات الصواب والدرجات المرتفعة لها على الأسئلة السهلة بينما قلت هذه النوعية من الاستجابات على المفردات الصعبة (للحالات المثالية أرقام ٤٢، ٤٩، ١١١)، أما حالات غير الملائمة والتى لم يتعرف عليها البرنامج (أرقام ١٥٣، ١٨٣، ٢١٢) تبين أن أغلب استجابات الصواب الخاصة بها ودرجاتها

المرتفعة كانت موزعة على جميع الأسئلة السهلة والصعبة والمتوسطة المستوى مما أدى إلى صعوبة التعرف على هذه الحالات من قبل البرنامج.

جدول رقم (١٠)

أنماط الاستجابة لعدد من حالات الملاءمة والحالات غير الملاءمة وفقا لتحليل استجاباتهم على اختبار المقال

(باستخدام برنامج Winsteps) مرتبة وفقا لصعوبة المفردات

رقم المفرد مرتبة وفقا لمستوى الصعوبة	تقديرات الصعوبة باللوجيت	أرقام الحالات التى نجاوزت حدود الملاءمة			أرقام الحالات التى لم تتجاوز حدود الملاءمة					
		١٨٥	١٨٦	١٨٨	١١١	٤٩	٤٢	١٥٣	٢١٢	١٨٣
		قدرات القدرة بوحدة اللوجيت								
		-٠.١٩	-٠.٨	-٠.٤٢	-١.٨	-٠.٦	-٠.٥	٤.٤	٠.٠٣٩	-٠.٢
		غش فردى	غش فردى	غش فردى	منضبط	منضبط	منضبط	غش جماعى	غش فردى	عدم جدية
١٠	-٠.٦٩	٣	٠	٠	١	٠	٠	٣	١	٣
١٣	-٠.٥٩	٠	٠	١	٠	٢	١	٢.٥	٣	٣
١٢	-٠.٥٧	٠	٠	٣	٠	٣	٠	٣	٣	٣
٣	-٠.٥٦	٠	٠	١	٠	٣	٢	٣	١	٢
١	-٠.٢٤	٣	٠	٠	٠	١	٣	٣	٢	٠
١١	-٠.٨	٠	٠	٠	١	٠	١	٣	٢	٣
٤	-٠.٠٦	٣	٠	٠	٠	٠	٠	٣	٣	٠
٢	٠.٠٣	٣	٣	٣	٠	١	٢	٣	١	٠
٩	٠.٢٨	٣	٢	٣	٠	٠	٠	٣	٠	٣
٧	٠.٣٧	٣	٣	٣	٠	٢	١	٣	٠	٠
٦	٠.٣٨	١	١	١	٠	٠	٠	٣	١.٥	١
٨	٠.٥٢	٠	٠	٠	٠	٠	٠	٣	١.٥	١
٥	١.٢١	٠	٠	٠	٠	٠	٠	٣	٢.٥	٠

ثانياً: عرض ومناقشة نتائج الفرض الثانى:

أظهرت نتائج اختبار الفرض الثاني والذى ينص على: تختلف مؤشرات الملاءمة الإحصائية لمفردات اختبار المقال (الناتجة من التحليل باستخدام برنامج Winsteps) قبل وبعد حذف استجابات الممتحنين الذين يتسم أدائهم بالغش الفردى،الجماعى، عدم الجدية فى الموقف الاختبارى.

وقد تم اختبار الفرض السابق وفقاً للإجراءات الثماني التالية:

١- تحليل مفردات اختبار الدراسة وفقاً لاستجابات العينة الكلية والتى تشتمل على حالات المنضبطين وحالات الغش الفردى والجماعى وحالات عدم الجدية باستخدام برنامج التحليل Winsteps.

٢- حصر العدد والنسب المئوية للمفردات التى تجاوزت حدود الملاءمة الإحصائية لكل من مؤشرى الملاءمة الإحصائي التقاربي، والتباعدي المتضمنان ببرنامج Winsteps.

٣- حذف حالات (الغش الفردى– الغش الجماعى– عدم الجدية) كل على حده من العينة الكلية للممتحنين، ثم تحليل مفردات الاختبار على استجابات العينة المتبقية بعد إجراء الحذف فى كل مره باستخدام برنامج Winsteps ثم تحليل العينة الكلية مرة أخرى أيضا بعد حذف جميع حالات عدم الملاءمة منها للحصول على تحليلات خاصة بعينة المنضبطين.

٤- حصر العدد والنسب المئوية للمفردات التى تجاوزت حدود الملاءمة الإحصائية لكل من مؤشرى الملاءمة الإحصائية التقاربي والتباعدي لكل خطوه موضحه بالإجراء السابق.

٥- اختبار دلالة فروق النسب المئوية للمفردات التى تجاوزت حدود الملاءمة وذلك للعينة الكلية بدون حذف أى حالة من ناحية، والعينة الكلية بعد حذف حالات الغش الفردى والغش الجماعى وعدم الجدية كل على حده مره، ثم العينة الكلية بعد حذف جميع هذه الحالات مرة أخرى وفقاً لبرنامج التحليل السابق.

٦- حذف كل حالات عدم الملاءمة والإبقاء على نوعية واحدة فقط من هذه الحالات.

(على سبيل المثال حذف حالات الغش الفردى والغش الجماعى وعدم الجدية والإبقاء على حالات التخمين فقط إلى جانب حالات المنضبطين الذين تم تثبيتهم مع الإجراءات المختلفة المتعلقة بالحذف) ثم تحليل مفردات اختبار الدراسة وفقاً لاستجابات العينة بعد إجراء الحذف كل مره باستخدام برنامج Winsteps.

٧- حصر العدد والنسب المئوية للمفردات التى تجاوزت حدود الملاءمة الإحصائية لكل من مؤشرى الملاءمة الإحصائية التقاربي والحصول المتضمن ببرنامج Winsteps وذلك نتيجة تحليل عينة المنضبطين ثم تم تكرار تحليل هذه العينة بالإضافة إلى نوعية مختلفة من حالات عدم الملاءمة (غش فردى- غش جماعى- عدم جديه).

٨- اختبار دلالة فروق النسب المئوية للحالات التى تجاوزت حدود الملاءمة وفقاً للحالات المحددة للدراسة الحالية (منضبطين- عينة كلية بدون حالات الغش فردى- عينة كلية بدون حالات الغش جماعى- عينة كلية بدون حالات عدم جديه).

النتائج التى أسفرت عنها إجراءات اختبار الفرض السابق وهى مرتبة على النحو التالى:

١- النتائج التى أسفر عنها الإجراء الأول الذى يتعلق بتحليل مفردات اختبار الدراسة (المقال) وفقا استجابات العينة الكلية:

يتبين من الجدول رقم (١١) مفردات اختبار الدراسة (المقال) غير الملائمة والتى نتجت من تحليلات برنامج Winsteps لاستجابات العينة الكلية ومؤشرات الملاءمة الإحصائية لها (التقاربي، والتباعدي) وفقاً لترتيب المؤشر التقاربي.

جدول رقم (١١)

مؤشرى الملاءمة الإحصائية التقاربي والتباعدي لمفردات اختبار الدراسة التى تجاوزت حدود الملاءمة والناتجة من تحليل استجابات العينة الكلية التى تشمل على حالات المنضبطين بالإضافة إلى الغش الفردى والجماعى وعدم الجدية

رقم المفردة	تقديرات الصعوبة بوحدة اللوجيت (measure)	مؤشر الملاءمة التقاربى infit	مؤشر الملاءمة التباعدى outfit	الخطأ المعيارى للقياس
١٠	٠.٧٣	٤.٧٦	٢.٢٧	٠.٠٥
١٣	٠.٦٠	٢.١٥	٠.٧٢	٠.٠٥
٣	٠.٥٩	٣.٢٢	١.٢٠	٠.٠٥
١٢	٠.٥٤	١.١٧	-٠.٠١	٠.٠٥
١	٠.٢٢	٠.٠٩	٠.٢٥	٠.٠٥
٤	٠.٠٨	٠.٥٩	٠.٢٥	٠.٠٥
١١	٠.٠٥	٠.٦٧	-٠.١٢	٠.٠٥
٢	٠.٠٥	١.١٦	١.٧٢	٠.٠٥
٩	٠.٣٠	١.٩٩	٠.٣٨	٠.٠٥
٧	٠.٣٧	-٠.٤١	-٠.٢٠	٠.٠٥
٦	٠.٣٨	-١.٢٥	-٠.٧٢	٠.٠٦
٨	٠.٥٤	-٤.١٢	-٢.٩٧	٠.٠٦
٥	١.١٥	-٣.٣٤	٠.٨٥	٠.٠٤

ويتضح من الجدول السابق أن هناك أربعة مفردات تجاوزت حدى الملاءمة التقاربي وهى المفردات أرقام (٣، ٥، ٨، ١٠) كما أن هناك مفردة واحدة تجاوزت الحد السالب لمؤشر الملاءمة التباعدى وهي المفردة رقم ٨.

٢- النتائج التي أسفر عنها الإجراء الثانى الذي يتعلق بحصر العدد والنسب المئوية للمفردات التى تجاوزت حدود الملاءمة الإحصائية لكل من مؤشرى الملاءمة الإحصائية التقاربي والتباعدي والتى أسفرعنها تحليل استجابات العينة الكلية لاختبار الدراسة:

يوضح الجدول رقم (١٢) العدد والنسب المئوية للمفردات التى تجاوزت حدود الملاءمة الإحصائية لكل من مؤشرى الملاءمة التقاربي والتباعدي

جدول رقم (١٢)

عدد المفردات التى تجاوزت حدود الملاءمة الإحصائية والنسب المئوية لها نتيجة تحليل استجابات العينة الكلية التى تشمل على حالات المنضبطين بالإضافة إلى الغش الفردى والجماعى وعدم الجدية

عينة التحليل	عدد مفردات الاختبار	عدد المفردات التي تجاوزت حدود الملاءمة التقاربية infit	النسبة المئوية	عدد المفردات التي تجاوزت حد الملاءمة التباعدي outfit	النسبة المئوية	جملة المفردات غير الملائمة	النسبة المئوية
العينة الكلية	١٣	٤	٣٠.٨	١	٧.٦٩	٤	٣٠.٨

ويلاحظ من الجدول السابق أن العدد الكلي للمفردات غير الملائمة التى تجاوزت حدى مؤشر الملاءمة التقاربى، والتباعدى (٤) مفردات من جملة مفردات الاختبار والبالغ عددها (١٣) مفردة.

٣- النتائج التى أسفر عنها الإجراء الثالث والذى يتعلق بحذف حالات الغش الفردى والجماعى وعدم الجدية كل على حدة من العينة الكلية للدراسة (على سبيل المثال حذف حالات الغش الفردى والإبقاء على عينة تشمل حالات منضبطين بالإضافة إلى حالات الغش الجماعى وحالات عدم الجدية....وهكذا)، ثم تحليل مفردات الاختبار على استجابات العينة المتبقية بعد إجراء الحذف فى كل مرة باستخدام برنامج Winsteps.

الجداول التالية أرقام (١٣)، (١٤)، (١٥)، (١٦) توضح نتائج تحليل مفردات اختبار الدراسة المتمثلة بتقديرات الصعوبة ومؤشرات الملاءمة الإحصائية لكل مفردة، وذلك بعد حذف حالات الغش الفردى والغش الجماعى وعدم الجدية على الترتيب.

جدول رقم (١٣)

تقديرات الصعوبة ومؤشرات الملاءمة الإحصائية التقاربي والتباعدي لمفردات اختبار الدراسة والناتجة من تحليل العينة الكلية بعد استبعاد حالات الغش الفردى والإبقاء على الغش الجماعى وعدم الجدية (وفقا لترتيب مؤشر الملاءمة التقاربي)

رقم المفردة	تقديرات الصعوبة بوحدة اللوجيت (measure)	مؤشر الملاءمة التقاربى infit	مؤشر الملاءمة التباعدى outfit	الخطأ المعيارى للقياس
٨	٠.٥٤	٤.١٢	٢.٩٧	٠.٠٥
٣	٠.٥٩	٣.٢٢	١.٢٠	٠.٠٥
١٣	٠.٦٩	٢.١٥	٠.٧٢	٠.٠٥
٩	٠.٣٠	١.٩٩	-٠.٣٨	٠.٠٥
٢	٠.٠٥	١.١٦	١.٧٢	٠.٠٥
١١	٠.٠٥	٠.٦٧	-٠.١٢	٠.٠٥
٤	٠.٠٨	٠.٥٠	٠.٢٥	٠.٠٥
١	٠.٢٢	-٠.٠٩	٠.٢٥	٠.٠٥
٧	٠.٣٧	-٠.٤١	-٢٠.	٠.٠٥
١٢	٠.٥٤	-١.١٧	-٠.٠١	٠.٠٥
٦	٠.٣٨	-١.٢٥	-٠.٧٢	٠.٠٦
٥	١.١٥	-٣.٣٤	-٠.٨٥	٠.٠٤
١٠	٠.٧٣	-٤.٧٦	-٢.٧٢	٠.٠٥

جدول رقم (١٤)

تقديرات الصعوبة ومؤشرات الملاءمة الإحصائية التقاربي والتباعدي لمفردات اختبار الدراسة والناتجة من تحليل العينة الكلية بعد استبعاد حالات عدم الجدية والإبقاء على الغش الفردى والجماعى (وفقا لترتيب مؤشر الملاءمة التقاربي)

رقم المفردة	تقديرات الصعوبة بوحدة اللوجيت (measure)	مؤشر الملاءمة التقاربى infit	مؤشر الملاءمة التباعدى outfit	الخطأ المعيارى للقياس
٨	٠.٥١	٣.٥٥	٢.٧٧	٠.٠٦
٣	٠.٥٧	٣.٤٠	١.٢٥	٠.٠٥
١٣	٠.٦٦	٢.٠١	٠.٨٢	٠.٠٥
٩	٠.٣٣	١.٨٩	-٠.٤٤	٠.٠٥
٢	٠.٠٤	١.٢٢	١.٥٥	٠.٠٥
١١	٠.٠٧	٠.٦٩	-٠.٣٣	٠.٠٥
٤	٠.٠٨	٠.٥٢	٠.٥٠	٠.٠٥
١	٠.٢٥	-٠.١١	٠.٤٧	٠.٠٥
٧	٠.٣٦	-٠.٣٨	-٢.٠٠	٠.٠٥
١٢	٠.٥٦	-١.١٩	-٠.٢٥	٠.٠٤
٦	٠.٤١	-١.٢٨	-٠.٥٢	٠.٠٥
٥	١.٠٨	-٣.١١	-٠.٩١	٠.٠٤
١٠	٠.٨٨	-٤.٦٥	-٢.٨٧	٠.٠٥

جدول رقم (١٥)

تقديرات الصعوبة ومؤشرات الملاءمة الإحصائية التقاربي والتباعدي لمفردات اختبار المقال الناتجة من تحليل العينة الكلية بعد استبعاد حالات الغش الجماعى والإبقاء على الغش الفردى وعدم الجدية

(وفقا لترتيب مؤشر الملاءمة التقاربي)

رقم المفردة	تقديرات الصعوبة بوحدة اللوجيت (measure)	مؤشر الملاءمة التقاربى infit	مؤشر الملاءمة التباعدى outfit	الخطأ المعيارى للقياس
٨	٠.٥٧	٤.٢٣	٢.٩٤	٠.٠٥
٣	٠.٥٦	٣.٣٤	١.٢٠	٠.٠٥
١٣	٠.٦٠	٢.٠٣	٠.٥٦	٠.٠٥
٩	٠.٣٤	١.٥٨	-٠.٤٤	٠.٠٥
٢	٠.١٠	١.٢٤	١.٥٠	٠.٠٥
١١	٠.٠٩	٠.٨٥	-٠.١٢	٠.٠٥
٤	٠.٠٨	٠.٥٥	٠.٢٦	٠.٠٥
١	٠.١٧	-٠.١٨	٠.٢٥	٠.٠٥
٧	٠.٣٧	-٠.٣٦	-١.٩	٠.٠٥
١٢	٠.٥٥	-١.٢١	-٠.٠٥	٠.٠٥
٦	٠.٤٨	-١.٢٦	-٠.٢٣	٠.٠٥
٥	١.٢٠	-٣.٣٩	-٠.١٨	٠.٠٤
١٠	٠.٦٩	-٤.٨٧	-٢.٧٧	٠.٠٦

جدول رقم (١٦)

تقديرات الصعوبة ومؤشرات الملاءمة الإحصائية التقاربي والتباعدي لمفردات اختبار الدراسة والناتجة من تحليل عينة المنضبطين بعد استبعاد جميع حالات الغش الفردى والغش الجماعى وعدم الجدية من العينة الكلية (وفقا لترتيب مؤشر الملاءمة التقاربي)

رقم المفردة	تقديرات الصعوبة بوحدة اللوجيت (measure)	مؤشر الملاءمة التقاربى infit	مؤشر الملاءمة التباعدى outfit	الخطأ المعيارى للقياس
٨	٠.٥٥	٤.٥٦	٢.٧٤	٠.٠٥
٣	٠.٥٧	٣.٦٦	١.١٤	٠.٠٥
١٣	٠.٦١	٢.٠٠	٠.٦٢	٠.٠٥
٩	٠.٢٨	١.٢٧	-٠.٤٨	٠.٠٥
٢	٠.٠٥	١.١٢	١.٧٥	٠.٠٥
١١	٠.١٥	٠.٦١	-٠.١٧	٠.٠٤
٤	٠.٠٩	٠.٥٦	٠.٢٦	٠.٠٥
١	٠.٢٢	-٠.٠٩	٠.٢٥	٠.٠٤
٧	٠.٣٧	-٠.٣١	-١.٤٠	٠.٠٥
١٢	٠.٥٢	-١.٢٥	-٠.٢١	٠.٠٥
٦	٠.٣٠	-١.٦٥	-٠.٥٢	٠.٠٥
٥	١.١٩	-٣.١٤	-٠.٨٥	٠.٠٤
١٠	٠.٧٦	-٤.٢٦	-٣.٤٢	٠.٠٦

يتبين من البيانات الموضحة بالجداول السابقة أن عدد مفردات اختبار المقال المستخدم بالدراسة الحالية التى تجاوزت الحدود المقبولة (± ٢) لمؤشرى الملاءمة الإحصائية التقاربي والتباعدي، والناتجة من التحليلات الخاصة باستجابات عينات التحليل السابقة هو عدد قليل لم

يتجاوز (٤) مفردات وهى المفردات أرقام (٣، ٥، ٨، ١٠) من جملة مفردات اختبار المقال والبالغ عددها (١٣) مفردة.

٤-النتائج التي أسفر عنها الإجراء الرابع والذي يتعلق بحصر العدد والنسب المئوية للمفردات التى تجاوزت حدود الملاءمة الإحصائية لكل من مؤشرى الملاءمة الإحصائية التقاربي والتباعدي لكل خطوة متضمنة بهذا الإجراء

الجدول رقم (١٧) يوضح عدد المفردات التى تجاوزت حدود الملاءمة الإحصائية والنسب المئوية لها نتيجة تحليل استجابات العينة الكلية لاختبار المقال قبل وبعد إجراء الحذف لحالات الغش وعدم الجدية.

جدول رقم (١٧)

عدد المفردات التي تجاوزت حدود الملاءمة الإحصائية والنسب المئوية لها نتيجة تحليل استجابات العينة الكلية لاختبار المقال قبل وبعد إجراء الحذف لحالات الغش وعدم الجدية.

الحالات	عدد المفردات التي تجاوزت حدود الملاءمة التقاربيةinfit				عدد المفردات التي تجاوزت حدود الملاءمة التباعدية outfit				المفردات غير الملائمة	
	٢	٢-	المجموع	النسبة المئوية	٢	٢-	المجموع	النسبة المئوية	العدد	النسبة المئوية
العينة الكلية	٢	٢	٤	٣٠.٨	-	١	١	٧.٦٩	٤	٣٠.٨
كلية- غش فردى	٢	٢	٤	٣٠.٨	-	١	١	٧.٦٩	٤	٣٠.٨
كلية- غش جماعى	٢	٢	٤	٣٠.٨	-	١	١	٧.٦٩	٤	٣٠.٨
كلية- عدم جدية	٢	٢	٤	٣٠.٨	-	١	١	٧.٦٩	٤	٣٠.٨
منضبطين	١	١	١	٧.٧	١	١	٢	١٥.٤	٢	١٥.٤

يبدو من الجدول السابق بشكل عام أن جملة عدد المفردات التى تجاوزت كل من حدى الملاءمة (التقاربى والتباعدى) بلغ أربعة مفردات بنسبة (٣٠.٨%) فى حالة استخدام العينة الكلية، والعينة الكلية بعد استبعاد كل من حالات الغش الفردى والجماعى وعدم الجدية على التوالى كعينات تحليل للاستجابات على اختبار المقال، بينما اتضح من خلال نفس الجدول أنه بعد استبعاد جميع حالات عدم الملاءمة المحددة للدراسة الحالية، وتحليل الاستجابات على عينة المنضبطين فقط تقلص عدد المفردات التى تجاوزت مؤشرى الملاءمة الإحصائية إلى النصف حيث بلغ عدد هذه المفردات (٢) مفردة فقط بنسبة (١٥.٤%) من جملة مفردات الاختبار.

كما يبدو أن الأفراد غير الملائمين (غش فردى، غش جماعى، عدم جدية) بعد إجراء الحذف لكل منهم على حدة والاستبعاد من العينة الكلية كان له تأثيراً متساوياً على مؤشرات الملاءمة الإحصائية لمفردات الاختبار (على الرغم من عدم قدرة هذه المؤشرات فى التعرف على هؤلاء الأفراد) حيث بلغ عدد المفردات التى تجاوزت حدى مؤشر الملاءمة التقاربى (٤) مفردات بنسبة (٣٠.٨%)، وهى المفردات أرقام (٣، ٥، ٨، ١٠)، وقد تجاوزتا منهما المفردتان (٣، ٨) الحد (−٢)، والمفردتان رقما (٥، ١٠) الحد (+ ٢)، أما المفردات التى تجاوزت حدى مؤشرى الملاءمة التباعدى بلغ عددها مفردة واحدة بنسبة (٧.٦٩%) وهى المفردة رقم (٨)، وهى نفس النتائج ونفس المفردات التى تم التوصل إليها بعد إجراء الحذف لحالات الغش الفردى وكذلك إجراء حذف حالات الغش الجماعى وأيضاً حالات عدم الجدية، هذا وقد انخفضت أعداد المفردات التى تجاوزت حدود الملاءمة الإحصائية لعينة المنضبطين وهى العينة

الناتجة من إجراء حذف جميع حالات عدم الملاءمة من العينة الكلية، حيث بلغ عدد المفردات التى تجاوزت حدى مؤشر الملاءمة التقاربى (٢) مفردة بنسبة (١٤.٨%) وهما رقما (٣، ١٠)، حيث تجاوزت المفردة رقم (٣) الحد (−٢) لكل من مؤشرى الملاءمة التقاربى والتباعدى، بينما تجاوزت المفردة رقم (١٠) حد الملاءمة التباعدى (+٢)، وتشير هذه النتيجة إلى التأثير الشديد لاستجابات حالات عدم الملاءمة المحددة بالدراسة الحالية على مؤشرات الملاءمة لمفردات اختبار المقال المستخدم حيث انخفض عدد المفردات التى تجاوزت كل من حدي مؤشرى الملاءمة الإحصائية التقاربى والتباعدى إلى النصف بعد حذف جميع استجابات حالات عدم الملاءمة (التى لم يتمكن البرنامج من التعرف عليها) من استجابات العينة الكلية للممتحنين على مفردات هذا الاختبار، وهذا ما نوهت إليه بعض الدراسات السابقة من احتمال تأثير استجابات هذه النوعية من الحالات على المؤشرات الإحصائية للمفردات الاختبارية (Bogan,1997)، (Dorans,1998)، (Ronald, 1999)، (Handrawan & Glas,2003)، (Van der linden & Sotaridona, 2006) مما يشير إلى خطورة عدم اكتشاف هذه الحالات من قبل هذا البرنامج مما يؤدى إلى إعطاء صورة غير حقيقية للعديد من المفردات، من خلال مؤشرات الملاءمة لها والتى يعتمد عليها بشكل رئيس فى حذف العديد منها، مما يلقى بظلال من الشكوك أيضاً حول الرؤية الصادقة لهذه المفردات من خلال تلك المؤشرات.

٥- النتائج التى أسفر عنها الإجراء الخامس والذي يتعلق باختبار دلالة فروق النسب المئوية للمفردات التي تجاوزت حدود الملاءمة، نتيجة تحليل العينة الكلية بدون حذف أى حالة من ناحية، والعينة

الكلية بعد الحذف المتتالى لحالات الغش الفردي والجماعة وعدم الجدية من ناحية أخرى، ثم إعادة التحليل بعد حذف جميع حالات عدم الملاءمة (عينة المنضبطين) باستخدام برنامج Winsteps.

الجداول أرقام (١٩)، (٢٠) توضح قيمة (z) ودلالتها لاختبار فروق النسب المئوية للمفردات التى تجاوزت حدود الملاءمة الإحصائية لكل من المؤشرين التقاربى والتباعدى على الترتيب والناتجة من تحليل استجابات العينة الكلية لاختبار المقال والتحليل بدون حالات الغش وعدم الجدية على التتابع.

جدول رقم (١٩)

قيمة (z) ودلالتها لاختبار فروق النسب المئوية لعدد مفردات اختبار المقال التى تجاوزت حد الملاءمة التقاربى

والناتجة من تحليل العينة الكلية والتحليل بدون حالات الغش وعدم الجدية على التتابع

عينة التحليل	العينة الكلية	العينة الكلية بدون (غ.ف)	العينة الكلية بدون (غ.ج)	العينة الكلية بدون (ع.ج)	عينة المنضبطين
العينة الكلية	–				
العينة الكلية بدون (غ.ف)	٠.٠٠٠	–			
العينة الكلية بدون (غ.ج)	٠.٠٠٠	٠.٠٠٠	–		
العينة الكلية بدون (ع.ج)	٠.٠٠٠	٠.٠٠٠	٠.٠٠٠	–	
عينة المنضبطين	١.٤٩٤	١.٤٩٤	١.٤٩٤	١.٤٩٤	–

* دال عند مستوى (٠.٠٥).

(غ.ف) حالات الغش الفردي (غ.ج) حالات الغش الجماعى (ع.ج) حالات عدم الجدية.

يتضح من الجدول السابق عدم وجود فروق دالة إحصائياً بين النسب المئوية لعدد المفردات التى تجاوزت حدى الملاءمة الإحصائية لمؤشر الملاءمة التقاربى، لجميع عينات التحليل المبينة بالجدول السابق، مما يدل على أن حذف استجابات حالات عدم الملاءمة (غش فردى، غش جماعى، عدم جدية) من العينة الكلية بالترتيب ليس لها أى تأثير على مؤشر الملاءمة التقاربى لمفردات الاختبار. وهذا ما يدل إلى عدم حساسية هذا المؤشر لهذه النوعية من الاستجابات إذا ما استخدم برنامج Winsteps فى مثل هذه النوعية من التحليلات، وهذا بخلاف ما يدعيه هذا البرنامج من قدرته على كشف هذه النوعية غير الملاءمة من الاستجابات (Liancer, 2005, 127) وتتسق هذه النتيجة مع ماتوصلت إليه دراسة Williams, J., Petridou, A (2011)، والتي أفادت بصعوبة وجود مؤشرات إحصائية يمكن من خلالها الكشف عن أنماط الاستجابات غير الملائمة في الموقف الاختباري، وهذا لا يتفق كذلك مع ما نوهت إليه وبعض الدراسات السابقة من احتمال تأثير استجابات هذه النوعية من الحالات على المؤشرات الإحصائية للمفردات الاختبارية (Bogan,1997)، (Linda & Dorans, 1998)، (Ronald, 1999).

كما أن هذه النتيجه لا تتشابه مع النتائج التي توصلت إليها بعض الدراسات الأخرى والتي أفادت بالتأثير السلبي للاستجابات غير الملامة على المؤشرات الإحصائية للمفردات (Meijer,2005)، (Sotaridna & Meijer, 2005)، (Bay, 2006)، (Van der Linden & Sotaridna, 2006)، (Liancer, 2006).

جدول رقم (١٨)

قيمة (Z) ودلالتها لاختبار فروق النسب المئوية لعدد مفردات اختبار المقال التى تجاوزت حدى الملاءمة التباعدى والناتجة من تحليل العينة الكلية والتحليل بدون حالات الغش وعدم الجدية على التتابع

عينة التحليل	العينة الكلية	العينة الكلية بدون (غ.ف)	العينة الكلية بدون (غ.ج)	العينة الكلية بدون (ع.ج)	عينة المنضبطين
العينة الكلية	–				
العينة الكلية بدون (غ.ف)	٠.٠٠٠	–			
العينة الكلية بدون (غ.ج)	٠.٠٠٠	٠.٠٠٠	–		
العينة الكلية بدون (ع.ج)	٠.٠٠٠	٠.٠٠٠	٠.٠٠٠	–	
عينة المنضبطين	٠.٦١٥	٠.٦١٥	٠.٦١٥	٠.٦١٥	–

* دال عند مستوى (٠.٠٥)

يتضح من النتائج الموضحة بالجدول السابق عدم وجود فروق ذات دلالة إحصائية بين النسب المئوية لعدد المفردات التى تجاوزت حدى الملاءمة التباعدى، الناتجة من تحليل العينة الكلية والتحليل بدون حالات الغش (فردى، جماعى) وعدم الجدية على مفردات اختبار المقال، وذلك لجميع العينات المشار إليها بالجدول السابق، مما يشير إلى عدم حساسية المؤشر أيضاً لاستجابات هذه الحالات، وذلك فيما يتعلق بعدد المفردات غير الملائمة التى أسفرت عنها تحليلات هذا البرنامج والتى تتجاوز حدى الملاءمة المشار إليه وتتسق هذه النتيجة مع

ماتوصلت إليه دراسة Williams, J., Petridou, A (2011)، والتي أفادت بصعوبة وجود مؤشرات إحصائية يمكن من خلالها الكشف عن أنماط الاستجابات غير الملائمة في الموقف الاختباري.

وبرغم النتائج السابقة الخاصة بعدم وجود فروق ذات دلالة إحصائياً بين نسب تجاوز المفردات غير الملائمة لكل من مؤشرى الملاءمة الإحصائية المشار إليهما فيما سبق بين جميع عينات التحليل كما هو موضح بالجدولين رقمى (١٧، رقم ١٨)، إلا أن النتائج أوضحت أنه فى حالة حذف جميع الاستجابات الخاصة بحالات عدم الملاءمة من استجابات العينة الكلية أنخفض عدد المفردات غير الملائمة إلى النصف مما يشير إلى خطورة عدم التعرف على هذه الحالات من قبل هذا البرنامج، وبالتالى خطورة الاستناد إلى هذه المؤشرات وقبولها بدرجة عالية من الثقة فيما يتعلق بهذا الجانب. وهذا يتفق مع ما نوهت إليه بعض الدراسات السابقة من احتمال تأثير استجابات هذه النوعية من الحالات على المؤشرات الإحصائية للمفردات الاختبارية (Bogan, 1997)، (Dorans, 1998)، (Ronald, 1999)، (Meijer, 2005)، (Sotaridna & Meijer, 2005)، (Bay, 2006)، (Van der linden & Sotaridona, 2006)، (Liancer, 2006).

٦- النتائج التى أسفر عنها الإجراء السادس والخاص بحذف كل حالات عدم الملاءمة، والإبقاء على نوعية واحدة فقط من هذه الحالات (على سبيل المثال حذف حالات الغش الفردى والغش الجماعى والإبقاء على حالات عدم الجدية بالإضافة إلى حالات المنضبطين) ثم تحليل مفردات اختبار المقال لاستجابات العينة بعد إجراء الحذف كل مرة باستخدام برنامج التحليل Winsteps.

الجداول أرقام (٢٠)، (٢١)، (٢٢) توضح نتائج تحليل استجابات العينة الكلية بعد حذف حالات عدم الملاءمة كما ورد بالخطوات الموضحة فى هذا الإجراء.

جدول رقم (٢٠)

تقديرات الصعوبة ومؤشرات الملاءمة الإحصائية التقاربي والتباعدي لمفردات اختبار المقال والناتجة من تحليل استجابات عينة المنضبطين بالإضافة إلى استجابات عينة الغش الفردى

(وفقا لترتيب مؤشر الملاءمة التقاربي)

رقم المفردة	تقديرات الصعوبة بوحدة اللوجيت (measure)	مؤشر الملاءمة التقاربى infit	مؤشر الملاءمة التباعدى outfit	الخطأ المعيارى للقياس
٨	٠.٥٣	٤.٠٥	٢.٧٧	٠.٠٥
٣	٠.٥٨	٣.١١	١.٥٥	٠.٠٥
١٣	٠.٦٥	٢.٠١	٠.٥٦	٠.٠٥
٩	٠.٢٥	١.٨٠	-٠.٤٤	٠.٠٥
٢	٠.٠٩	١.٠٦	١.٨٨	٠.٠٥
١١	٠.٠٥	٠.٦٩	-٠.٢٥	٠.٠٥
٤	٠.٠٨	٠.٨٩	٠.٦٢	٠.٠٥
١	٠.٢٠	-٠.١٢	٠.٣٠	٠.٠٥
٧	٠.٣٩	-٠.٤٤	-٠.٢٦	٠.٠٥
١٢	٠.٥٥	-١.٠٧	-٠.٦١	٠.٠٦
٦	٠.٤٠	-١.٠٥	-٠.٩٢	٠.٠٦
٥	١.٣٠	-٣.٦٤	-٠.٨٥	٠.٠٥
١٠	٠.٨٩	-٤.٢٢	-٢.٨٨	٠.٠٦

جدول رقم (٢١)

تقديرات الصعوبة ومؤشرات الملاءمة الإحصائية التقاربي والتباعدي لمفردات اختبار المقال والناتجة من تحليل استجابات عينة المنضبطين بالإضافة إلى استجابات عينة الغش الجماعى

(وفقا لترتيب مؤشر الملاءمة التقاربي)

رقم المفردة	تقديرات الصعوبة بوحدة اللوجيت (measure)	مؤشر الملاءمة التقاربى infit	مؤشر الملاءمة التباعدى outfit	الخطأ المعيارى للقياس
٨	٠.٥٥	٣.٩٩	٢.٩٧	٠.٠٥
٣	٠.٥٩	٢.٩٦	١.٢٠	٠.٠٥
١٣	٠.٦٨	٢.٢٠	٠.٧٢	٠.٠٥
٩	٠.٣١	١.٨٠	-٠.٣٨	٠.٠٥
٢	٠.٠٨	١.٠٦	١.٧٢	٠.٠٥
١١	٠.٠٥	٠.٣٣	-٠.٥٥	٠.٠٥
٤	٠.٠٨	٠.٨٥	٠.٣٥	٠.٠٥
١	٠.٢٦	-٠.٠٥	٠.١٩	٠.٠٥
٧	٠.٣٧	-٠.٣٣	-٠.٢٧	٠.٠٥
١٢	٠.٦٦	-١.٣٧	-٠.١١	٠.٠٥
٦	٠.٣٣	-١.٢٢	-٠.٦٥	٠.٠٦
٥	١.١٧	-٣.٠٤	-٠.٥٥	٠.٠٤
١٠	٠.٧٤	-٣.٨٦	-٢.٨٨	٠.٠٥

جدول رقم (٢٢)

تقديرات الصعوبة ومؤشرات الملاءمة الإحصائية التقاربي والتباعدي لمفردات اختبار المقال والناتجة من تحليل استجابات عينة المنضبطين بالإضافة إلى استجابات عينة عدم الجدية

(وفقا لترتيب مؤشر الملاءمة التقاربي)

رقم المفردة	تقديرات الصعوبة بوحدة اللوجيت (measure)	مؤشر الملاءمة التقاربى infit	مؤشر الملاءمة التباعدى outfit	الخطأ المعيارى للقياس
٨	٠.٥٠	٤.٢٥	٢.٩٠	٠.٠٥
٣	٠.٦٣	٤.٠٢	١.٢١	٠.٠٥
١٣	٠.٧١	٢.٠١	٠.٨٩	٠.٠٥
٩	٠.٢٧	١.٥٠	٠.٤٧-	٠.٠٥
٢	٠.٠٤	١.١٥	١.٠٥	٠.٠٥
١١	٠.٠٦	٠.٩٦	٠.٥٥-	٠.٠٥
٤	٠.٠٨	٠.٣٣	٠.٢٨	٠.٠٥
١	٠.٣٤	٠.١٩-	٠.٢٢	٠.٠٥
٧	٠.٣١	٠.٣٦-	١.٣٢-	٠.٠٥
١٢	٠.٥٤	١.٥٨-	٠.٥١-	٠.٠٥
٦	٠.٤١	١.١٤-	٠.٢٤-	٠.٠٥
٥	١.٨٠	٢.٩٢-	٠.٨٧-	٠.٠٤
١٠	٠.٦٩	٣.٨٨-	٣.١٢-	٠.٠٦

٧- النتائج المتعلقة بالإجراء السابع والذي يتعلق بحصر العدد والنسب المئوية للمفردات التي تجاوزت حدود الملاءمة الإحصائية لكل من مؤشرى الملاءمة الإحصائية التقاربي والتباعدي، وذلك نتيجة تحليل عينة المنضبطين ثم تكرار تحليل استجابات هذه العينة بالإضافة

إلى نوعية مختلفة من حالات عدم الملاءمة (غش فردية، غش جماعي، عدم جدية)

الجدول رقم (٢٢) يوضح العدد والنسب المئوية للمفردات التي تجاوزت حدود الملاءمة الإحصائية لكل من مؤشري الملاءمة الإحصائية التقاربي والتباعدي، نتيجة تحليل عينة المنضبطين ثم تكرار تحليل استجابات هذه العينة بالإضافة إلى نوعية مختلفة من حالات عدم الملاءمة (غش فردى، غش فردى، عدم جدية)

جدول رقم (٢٢)

العدد والنسب المئوية للمفردات التي تجاوزت حدود الملاءمة الإحصائية لكل من مؤشري الملاءمة الإحصائية التقاربي والتباعدي نتيجة تحليل عينة المنضبطين ثم تكرار تحليل استجابات هذه العينة بالإضافة إلى نوعية مختلفة من حالات عدم الملاءمة (غش فردى، غش فردى، عدم جدية)

الحالات	عدد المفردات التي تجاوزت حدود الملاءمة التقاربيةinfit				عدد المفردات التي تجاوزت حدود الملاءمة التباعدية outfit				المفردات غير الملائمة	
	٢	-٢	المجموع	النسبة المئوية	٢	-٢	المجموع	النسبة المئوية	العدد	النسبة المئوية
العينة الكلية	-	١	١	٧.٧	١	١	٢	١٥.٤	٢	١٥.٤
كلية- غش فردى	٢	٢	٤	٣٠.٨	-	١	١	٧.٦٩	٤	٣٠.٨
كلية- غش جماعى	٢	٢	٤	٣٠.٨	-	١	١	٧.٦٩	٤	٣٠.٨
كلية- عدم جدية	٢	٢	٤	٣٠.٨	-	١	١	٧.٦٩	٤	٣٠.٨
منضبطين	٢	٢	٤	٣٠.٨	-	١	١	٧.٦٩	٤	٣٠.٨

يبدو من الجدول السابق أيضاً أن اختلاف الأفراد غير الملائمين (غش فردى، غش فردى، عدم جدية) قبل وبعد إجراء الإضافة لعينة المنضبطين على التوالى لم يؤثر تقريباً على مؤشرات الملاءمة لمفردات الاختبار، حيث بلغ قصى عدد من المفردات أقصى

أقصى عدد من المفردات التى تجاوزت حدى الملاءمة التقاربى (٤) مفردات بنسبة (٣٠.٨%) وهى المفردات أرقام (٣، ٥، ٨، ١٠) وقد تجاوزتا منهما المفردتان رقما (٣، ٨) الحد (– ٢) والمفردتان رقما (٥، ١٠) الحد (+ ٢).

أما المفردات التى تجاوزت حد الملاءمة التباعدى بلغ عددها مفردة واحدة بنسبة (٧.٦٩%) وهى المفردة رقم (٨)، وهى نفس النتائج ونفس المفردات التى تجاوزت حد الملاءمة التقاربى، وكذلك التباعدى، لجميع العينات الأخرى.

هذا وقد انخفضت أعداد المفردات التى تجاوزت حدود الملاءمة الإحصائية لعينة المنضبطين، حيث بلغ عدد المفردات التى تجاوزت حد الملاءمة التقاربى مفردة واحدة بنسبة (٧.٦٩%) وهى المفردة رقم (٣)، بينما بلغ عدد المفردات التى تجاوزت حد الملاءمة التباعدى مفردتان وهما رقما (٣، ١٠)، وهى نفس النتيجة المحيرة التى تم الإشارة إليها من قبل مع عرض ومناقشة الإجراء الرابع، مما يؤكد خطورة عدم اكتشاف هذه الحالات من قبل هذا البرنامج والذي قد يترتب عليه إعطاء صورة غير حقيقية للعديد من هذه المفردات، من خلال هذه المؤشرات من الملاءمة والتى يعتمد عليها بشكل رئيس فى حذف العديد منها فى الإجراءات الخاصة بمثل هذه التحليلات، مما يلقى بظلال من الشكوك

أيضاً حول الرؤية الصادقة لهذه المفردات من خلال تلك المؤشرات وهى نفس الشكوك التى أثارتها دراسة Petridou, A. (2001) ., Williams, J، والتي أشارت بصعوبة وجود مؤشرات إحصائية يمكن من خلالها الكشف عن أنماط الاستجابات غير الملائمة في الموقف الاختباري.

كما تتشابه هذه النتيجة مع النتائج التي توصلت إليها دراسة دوير وجيفري Dewyer & Jeffrey (2004) والتي تفيد بعدم قدرة المؤشرات الإحصائية في كشف وقائع الغش في المواقف الاختبارية بشكل مؤكد.

٨- **النتائج التى أسفر عنها الإجراء الثامن والذى يتعلق باختبار دلالة فروق النسب المئوية للحالات التى تجاوزت حدود الملاءمة الإحصائية وفقاً للحالات المحددة للدراسة الحالية (منضبطين، منضبطين بالإضافة إلى حالات الغش الفردى، منضبطين بالإضافة إلى حالات الغش الجماعى، منضبطين بالإضافة إلى حالات عدم الجدية).**

الجداول أرقام (٢٤)، (٢٥) توضح قيمة (z) ودلالتها لاختبار فروق النسب المئوية لعدد المفردات التى تجاوزت حدى الملاءمة التقاربى، وحدى الملاءمة التباعدى، على الترتيب، وفقاً للحالات المحددة للدراسة:

- (منضبطين، منضبطين بالإضافة إلى حالات الغش الفردى، منضبطين بالإضافة إلى حالات الغش الجماعى، منضبطين بالإضافة إلى حالات عدم الجدية).

جدول رقم (٢٤)

قيمة (Z) ودلالتها لاختبار فروق النسب المئوية لعدد مفردات اختبار المقال التى تجاوزت حد الملاءمة التقاربى والناتجة من تحليل عينة المنضبطين، وعينة المنضبطين بالإضافة إلى حالات الغش وعدم الجدية على التتابع

عينة التحليل	منضبطين	منضبطين+ غش فردى	منضبطين+ غش جماعى	منضبطين+ عدم جدية
منضبطين	–			
منضبطين+ غش فردى	١.٤٩٤	–		
منضبطين+ غش جماعى	١.٤٩٤	٠.٠٠٠	–	
منضبطين+ عدم جدية	١.٤٩٤	٠.٠٠٠	٠.٠٠٠	–

* دال عند مستوى (٠.٠٥).

يتضح من الجدول السابق عدم وجود فروق دالة إحصائياً بين النسب المئوية لعدد مفردات اختبار المقال التى تجاوزت حدى الملاءمة لمؤشر الملاءمة التقاربى لجميع عينات التحليل المحددة بالجدول السابق، مما يدل على استجابات جميع حالات عدم الملاءمة مضافاً إليها استجابات المنضبطين لمفردات اختبار المقال ليس لها أية تأثير على مؤشر الملاءمة التقاربى مما يدل على عدم حساسية هذا المؤشر لهذه النوعية من الاستجابات إذا ما استخدم برنامج فى مثل هذه النوعية من التحليلات، وهذه النتائج لا تتفق مع نتائج دراسة وتتسق هذه النتيجة مع ما توصلت إليه دراسة (Williams, J. Petridou, A., 2011) والتي أفادت بصعوبة وجود مؤشرات إحصائية يمكن من خلالها الكشف عن أنماط الاستجابات غير الملائمة في الموقف الاختباري، وتتفق هذه النتيجة مع

ما نوهت إليه بعض الدراسات السابقة من احتمال تأثير استجابات هذه النوعية من الحالات على المؤشرات الإحصائية للمفردات الاختبارية (Bogan,1997)، (Dorans,1998)،(Ronald, 1999) كما أن هذه النتيجه تتشابه مع النتائج التي توصلت إليها بعض الدراسات الأخرى والتي أفادت بالتأثير السلبي للاستجابات غير الملامة على المؤشرات الإحصائية للمفردات (Meijer,2005)، (Sotaridna & Meijer, 2005)، (Bay, 2006)، (Van der Linden & Sotardina, 2006)، (Lincer, 2006).

جدول رقم (٢٥)

قيمة (Z) ودلالتها لاختبار فروق النسب المئوية لعدد مفردات اختبار المقال التى تجاوزت حد الملاءمة التباعدى والناتجة من تحليل عينة المنضبطين، وعينة المنضبطين بالإضافة إلى حالات الغش وعدم الجدية على التتابع

عينة التحليل	منضبطين	منضبطين+ غش فردى	منضبطين+ غش جماعى	منضبطين+ عدم جدية
منضبطين	–			
منضبطين+ غش فردى	٠.٦١٥	–		
منضبطين+ غش جماعى	٠.٦١٥	٠.٠٠٠	–	
منضبطين+ عدم جدية	٠.٦١٥	٠.٠٠٠	٠.٠٠٠	–

* دال عند مستوى (٠.٠٥)

يتضح من الجدول السابق عدم وجود فروق دالة إحصائياً بين النسب المئوية لعدد مفردات اختبار المقال التى تجاوزت حدى الملاءمة

لمؤشر الملاءمة التباعدى لجميع عينات التحليل المحددة بالجدول السابق، مما يدل على استجابات جميع حالات عدم الملاءمة مضافاً إليها استجابات المنضبطين لمفردات اختبار المقال ليس لها أية تأثير على مؤشر الملاءمة التباعدى، مما يدل على عدم حساسية هذا المؤشر أيضاً (كما هو الحال مع مؤشر الملاءمة التقاربى) لهذه النوعية من الاستجابات إذا ما استخدم البرنامج فى مثل هذه النوعية من التحليلات، وهذه النتائج تتفق مع نتائج دراسة وتتسق هذه النتيجة مع ماتوصلت إليه دراسة (Williams, J. Petridou, A., 2011) والتي أفادت بصعوبة وجود مؤشرات إحصائية يمكن من خلالها الكشف عن أنماط الاستجابات غير الملائمة في الموقف الاختباري.

كما أن هذه النتيجة لا تتفق مع ما نوهت إليه بعض الدراسات السابقة من تأثير استجابات هذه النوعية من الحالات على المؤشرات الإحصائية للمفردات الاختبارية (Bogan,1997)،(Dorans,1998)، (Ronald, 1999).

كما أن هذه النتيجه لا تتشابه مع النتائج التي توصلت إليها بعض الدراسات الأخرى والتي أفادت بالتأثير السلبي للاستجابات غير الملامة على المؤشرات الإحصائية للمفردات (Meijer,2005)، (Sotaridna & Meijer, 2005)، (Bay,2006)، (Van der Linden & Sotardina, 2006)، (Lincer, 2006).

وبرغم النتائج الخاصة بعدم وجود فروق دلالة إحصائية بين نسب تجاوز المفردات غير الملاءمة لكل من مؤشرى الملاءمة الإحصائية المشار إليهما سابقاً بين جميع عينات التحليل المبينة بالجداول السابقة،

إلا أن النتائج أوضحت أنه فى حالة تحليل استجابات عينة المنضبطين فقط بدون إضافة من استجابات حالات عدم الملاءمة أنخفض عدد المفردات غير الملائمة التى تجاوزت مؤشرى الملاءمة الإحصائية إلى النصف مما يشير إلى خطورة عدم تعرف البرنامج على هذه الحالات، مما يوحى بضرورة عدم الاستناد على نتائج تحليلات هذا البرنامج فى هذا الشأن، وهذا هو نفس الاستنتاج الذى تم التوصل إليه من خلال الإجراء السابق.

ثالثا: نتائج الفرض الثالث:

أوضحت نتائج اختبار الفرض الثالث والذي ينص على لا تختلف تقديرات الصعوبة لمفردات اختبار المقال الناتجة من تحليلها وفقا لاستجابات العينة المستخدمة بالدراسة الحالية قبل وبعد حذف جميع حالات عدم الملاءمة (غش فردى، غش جماعى، عدم جدية)، باستخدام برنامج Winsteps.

أسفرت إجراءات اختبار هذا الفرض عن النتائج الموضحة بالجدول التالى رقم (٢٦) الذى يوضح دلالة الفروق فى تقديرات الصعوبة المتناظرة لمفردات اختبار المقال قبل وبعد حذف الأفراد الملائمين.

جدول رقم (٢٦)

دلالة الفروق فى تقديرات الصعوبة المتناظرة لمفردات اختبار المقال قبل وبعد حذف الأفراد غير الملائمين باستخدام برنامج Winsteps

رقم المفردة	تقديرات الصعوبة قبل الحذف	الخطأ المعياري قبل الحذف	تقديرات الصعوبة قبل الحذف	الخطأ المعياري بعد الحذف	الفروق المطلقة فى تقديرات الصعوبة قبل وبعد الحذف	مجموع الخطأين المعياريين لتقديرات الصعوبة قبل وبعد الحذف	دلالة الفروق فى تقديرات الصعوبة
١	-٠.٢٢	٠.٠٥	-٠.٣٧	٠.٠٧	٠.١٥	٠.١٢	*
٢	٠.٠٥	٠.٠٥	-٠.١٤	٠.١٧	٠.١٩	٠.١٢	*
٣	-٠.٥٩	٠.٠٥	-٠.٤١	٠.٠٧	٠.١٨	٠.١٢	*
٤	-٠.٠٨	٠.٠٥	٠.١٠	٠.٠٧	٠.١٨	٠.١٢	*
٥	١.١٥	٠.٠٤	-٠.٣١	٠.٠٧	١.٤٦	٠.١١	*
٦	٠.٣٨	٠.٠٦	٠.٤٤	٠.٠٨	٠.٠٦	٠.١٤	
٧	٠.٣٧	٠.٠٥	٠.٥٤	٠.٠٩	٠.١٧	٠.١٤	*
٨	٠.٥٤	٠.٠٦	١.٥١	٠.١٤	٠.٩٧	٠.١٩	*
٩	٠.٣٠	٠.٠٥	٠.٤٢	٠.٠٨	٠.١٢	٠.١٣	
١٠	-٠.٧٣	٠.٠٥	-٠.٥٩	٠.٠٧	٠.١٤	٠.١٢	*
١١	-٠.٠٥	٠.٠٥	-٠.٠٥	٠.٠٨	٠.٠٠	٠.١٣	
١٢	-٠.٥٤	٠.٠٥	-٠.٦١	٠.٠٧	٠.٠٧	٠.١١	
١٣	-٠.٦٠	٠.٠٥	-٠.٥٣	٠.٠٧	٠.٠٧	٠.١٢	

يلاحظ من الجدول السابق أن عدد المفردات التى لها فروق جوهرية فى تقديرات الصعوبة المتناظرة (٨) مفردات بنسبة (٦١.٥%) من إجمالى عدد مفردات الاختبار البالغ عددها (١٣) مفردة حيث تجاوزت هذه النسبة الحد المقبول وهو (٥%)، مما يشير أن حذف الأفراد غير الملائمين الذين تم رصدهم من خلال الواقع التطبيقى للاختبار أدى إلى تغيرات جوهرية فى تقديرات الصعوبة لمفردات اختبار

المقال، مما أدى إلى رفض الفرض السابق والذى ينص على: لا تختلف تقديرات الصعوبة لمفردات اختبار المقال التاتجة من تحليلها وفقا لاستجابات العينة المستخدمة بالدراسة الحالية قبل وبعد حذف جميع حالات عدم الملاءمة (غش فردى، غش جماعى، عدم جدية)، باستخدام برنامج Winsteps.

وتتسق هذه النتيجة مع ما نوهت إليه بعض الدراسات السابقة من تأثير استجابات هذه النوعية من الحالات على المؤشرات الإحصائية للمفردات الاختبارية (Bogan,1997)، (Dorans,1998)، (Ronald, 1999)، وكذلك دراسة (Van der Linden & Sotardina, 2006) التي أشارت إلى تأثر تقديرات صعوبة المفردات بالاستجابات غير الملائمة، مما يدل على أن تقديرات صعوبة المفردات تتأثر بشكل جوهرى بمثل هذه الاستجابات غير الملاءمة بصرف النظر عن نوع الاختبار سواء كان اختبار يحتوى على أسئلة موضوعية أو أسئلة مقالية، على الرغم من عدم قدرة هذا الاختبار على كشف وتحديد الاستجابات غير الملاءمة من خلال مؤشرات الملاءمة التى يفترض من تعليمات البرنامج أن لها القدرة على ذلك (Wright & Liancer, 1987, 4-20).

خاتمة:

من خلال ما أسفرت عنه الدراسة الحالية من نتائج يمكن التوصل إلى الاستنتاجات التالية:

١- عدم قدرة برنامج التحليل المستخدم بالدراسة الحالية "Winsteps" فى اكتشاف الغالبية العظمى من حالات عدم الملاءمة المحددة للدراسة الحالية (غش فردى، غش جماعى، عدم جدية) من خلال

مؤشرى الملاءمة المتضمنان بالبرنامج (التقاربي والتباعدي) والذي افترض فعاليتهما فى كشف مثل هذه الحالات.

٢- التأثير السلبى العكسى لاستجابات حالات عدم الملاءمة على مؤشرات الملاءمة الإحصائية لمفردات اختبار التحليل المستخدم (اختبار المقال) مما يعطى صورة غير حقيقية عن هذه المفردات.

٣- إجراءات الحذف الخاصة باستجابات حالات عدم الملاءمة والتى حُددت بالدراسة الحالية (والتى لم يتعرف عليها البرنامج) أثرت بشدة على تقديرات الصعوبة الخاصة بالاختبار المستخدم.

توصيات الدراسة:

من خلال النتائج التى تم التوصل إليها يمكن للباحث الحالى أن يوصى بما يلى:

- يجب عدم الاستناد إلى تحليلات هذا البرنامج بشكل أساسى فى التعرف على حالات الملاءمة بالمواقف الاختبارية المختلفة واللجوء إلى وسائل أخرى بديلة تعتمد على دقة الملاحظة لهذه المواقف الاختبارية أو أى إجراءات أخرى بديله.
- يجب عدم الاستناد بشكل مطلق إلى مؤشرات الملاءمة الإحصائية للمفردات الناتجة من تحليلات هذا البرنامج بشكل أساسى فى حذف أو تعديل بعض المفردات.

دراسات مقترحة:

١- دراسة المؤشرات الملاءمة الإحصائية التى يمكن من خلالها التعرف على حالات عدم الملاءمة (غش فردى- غش جماعى- عدم جدية-

تخمين) عند استخدام اختبار مقال باستخدام برامج أخرى تختلف عن برنامج Winsteps.

٢- دراسة المؤشرات الإحصائية التى يمكن من خلالها التعرف على حالات عدم الملاءمة (غش فردى- غش جماعى- عدم جدية- تخمين) باستخدام نماذج تحليل مختلفة عن نموذج "بارشال كريدت".

المراجع:

- أحمد عودة، ونصر مقابلة (١٩٨٩). ظاهرة الغش فى الامتحانات حجمها وأسبابها كما يدركها طلبة جامعة اليرموك بالأردن. المجلة التربوية. جامعة الكويت. المجلد السادس. العدد الحادى والعشرون. ص ص ٩٣-١٠٧.
- إدارة الخدمة النفسية (١٩٨٤). دراسة بعض أنماط السلوك الطلابى المرتبطة بالامتحانات. الكويت: وزارة التربية.
- أمينة محمد كاظم (١٩٨٨ أ). استخدام نموذج راش فى بناء نموذج تحصيلى فى علم النفس وتحقيق التفسير الموضوعى للنتائج. الكويت. جامعة الكويت. مطبوعات جامعة الكويت.
- أمينة محمد كاظم (١٩٨٨ب). دراسة نظرية نقدية حول التفسير الموضوعى للسلوك (نموذج راش). ط٢. الكويت. مؤسسة الكويت للتقدم العلمى.
- جاسم الكندرى، غانم النجار، بدر العيسى (١٩٨٩). قياس اتجاهات أعضاء هيئة التدريس والطلبة حول ظاهرة الغش بجامعة الكويت. المجلة التربوية. جامعة الكويت. المجلد السادس. عدد (٢٢). ص ص ١٧٣-٢٢٠.
- حنان حسين عبد الرسول (١٩٩٧). دراسة لبعض العوامل المعرفية وغير المعرفية المرتبطة بسلوك الغش فى الامتح أنات لدى تلاميذ المرحلة الإعدادية. جامعة عين شمس. كلية البنات. رسالة ماجستير غير منشورة.
- زكريا توفيق أحمد (١٩٨٦). قلق إلامتحان وعلاقته بالمهارات الدراسية لدى طلاب المرحلة الثانوية. الجمعية المصرية للدراسات النفسية. الكتاب السنوى فى علم النفس. المجلد الخامس. ص ص ١٤٥-١٦٤.

- زكريا توفيق أحمد (١٩٩١). الغش الدراسى وعلاقته ببعض أساليب المعاملة الوالدية- والدافع للإنجاز- والتخصص الدراسى لدى طلبة كليات إعداد المعلمين. بحوث المؤتمر السنوى الرابع للطفل المصرى. مركز دراسات الطفولة. جامعة عين شمس. ص ص ١٩٥٥-١٩٨٣.

- صلاح أحمد مراد، ومحمد منصور الشافعى (١٩٩٨). أثر حجم العينة فى دقة وكفاءة ضم اختبارين فى تدريج مشترك. مجلة البحوث النفسية والتربية. جامعة المنصورة. كلية التربية. العدد الثانى. السنة الثالثة عشر. ص ص ٩٦-١٤٤.

- صلاح الدين محمود علام (١٩٨٥). دراسة موزنة ناقدة لنماذج السمات الكامنة والنماذج الكلاسكية فى القياس النفسى والتربوى. المجلة العربية للعلوم الإنسانية. جامعة الكويت. المجلد الخامس. ص ص ٩-٤٠.

- صلاح الدين محمود علام (١٩٨٦). تطورات معاصرة فى القياس النفسى والتربوى. الكويت: مطابع القبس التجارية.

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

- عبد الله سليمان (١٩٩٤). الغش فى الامتحانات وعلاقته ببعض المتغيرات النفسية وأساليب التعلم. مجلة دراسات تربوية. رابطة التربية الحديثة. المجلد التاسع (ع) ٢٦٤. ص ص١٥٥-٢١٠.

- عبد المجيد نشواتى (١٩٩٨). علم النفس التربوى. ط٩. بيروت: مؤسسة الرسالة.

- فاروق عبد الفتاح موسى (١٩٩٠). القياس النفسى والتربوى للأسوياء والمعوقين. ط١. القاهرة: مكتبة النهضة المصرية.

- فاروق عبدة فلية (١٩٨٨). ظاهرة الغش فى الامتحانات التشخيص والعلاج. القاهرة: مكتبة النهضة المصرية.

- ليونا أ تايلر (١٩٨٣). الاختبارات والمقاييس. ط١. ترجمة: سعد عبد الرحمن. القاهرة: دار الشروق.

- محمد زياد حمدان (١٩٨٦). الغش فى الاختبارات أداء الواجبات المدرسية ماهيته وأصوله وتشخيصه وعلاجه. سلسلة المكتبة التربوية. الاردن – عمان: دار التربية الحديثة.

- محمد منصور الشافعى (١٩٩٦). أثر طرق معادلة درجات الاختبار وضوابط اختيار العينة على تدريج بنوك بنك الأسئلة باستخدام نموذج راش. رسالة دكتوراه غير منشورة. جامعة المنصورة. كلية التربية.

- وفاء محمود نصار (٢٠١٠). القياس والتقويم في العملية التعليمية. ط١. الرياض: درا الزهراء.

- **Aimee, M, (2008). A Comparison of (CAT) Scores Based on Based on Different Measurement Models for Testlets. Educational and Psychological Measurement. 31(3). PP. 371-398.**

- **Amy D., Greg B. (2010). Physician Assistant Students-Attitudes and Behaviors Toward Cheating and Academic Integrity. Journal of Physician Assistant Education. 21(1). PP.214-140.**

- **Anderson, J. (1998). Male And Female Attiudes Towards Cheating In Examinations. Paper Presented at the Annual Meeting of the American Educational Research Association. 75th. Los Angeles. Ca. April 13-17.**

- **Andrich, D. (2004). Controversy and the Rasch Model: A Characteristic of Incompatible Paradigms? Medical Care. 42. PP. 1-16.**
- **Anzaldua, R.M. (2005). Item Banks: What Were Why And How. Paper Persented at The Annual Meeting Southwest Educational Research Association. (25th). Febuarary 18-20.**
- **Bay, L. (2006). Detection of Cheating on Multiple-Choice Examination, Paper Presented at the Annual Meeting of the American Educational Research Association, San Francisco. Ca. April. 20-24. Eric Accession Number Ed 384685.**
- **Bogan, Evelyn D. (1997). Careless Response Indications in Testing Setuation. Paper Presented at the Annual Meeting of the American Educational Research Association. 77th. Montered. Quebec. April 11-15.**
- **Camara, W.J. (2003). Professional Testing Students: What Educators Need To Do. Journal Of Educational Measurenent. 20(3). PP. 320-44**
- **Cizaek, G.j. (2004). Educational Testing, Integrity: Why Educators and Students Cheat and How to Prevent it. Dissertation Abstracts International. P.645,. ed 825423.**
- **David, G. Seung C. (2010). The Future of Outcomes Measurement: Item Banking. Tailored Short-Forms. Quality of Life Research, Vol.(1). N(6). PP.133-141.**
- **De Ayala, R., & Koch, W. (1992). A Comparison of Partial Credit and Graded Response Model in Computerized Adaptive Testing. Applied Psychological Measurement. 5. 17-34.**

- **De Ayala, R.J. (2009). The Theory and Practice of Item Response Theory. NY.: The Guilford Press.**
- **Dorans, Neil J. (1998). Response Pattern for Individual Cheating. Journal of Educational Measurenent. 28 (5). PP. 245-67.**
- **Dwyer, D.J.; Jeffrey, B.K. (2004). Cheating Detection Statistical Legally and Policy Implications. Illinois Stat Univ. Eric Accession Number Ed 384685.**
- **Edward J., Georgiana S. (2009). Study of a Cognitive Dissonance Intervention to Address High School Students- Cheating Attitudes and Behaviors. Ethics & Behavior. 19(3). PP. 218–226.**
- **Fan,X. G.; Ping, Y.B. (2006). Assessing the Effect of Model Data Misfit on Invariance Property of IRT Parameter Estimates. Paper Presented at the Annual Meeting of the American Educational Research Association. April 22-26. Eric Accession Number Ed 586247.**
- **Glas, C.A. (2004). CUSUM, Statistics for Larg Item Banks. ERIC. ED478820.**
- **Hambleton, Ronald K. (1987). Item Response Theory Principles And Application: NY. Nighoff Publishing.**
- **Hambleton, Ronald K. and Van Der Linden Wim J. (1997). Handbook of Modern Item Response Theory: NY Springer Veriag Inc.**
- **Hendrawen, I,H.; Glas, A.W. (2004). The Effect of Person Misfit on Classification Decisions. Research- Report, Tewent Univ., Enscheade, Netherlond, Faculty of Educational Science and Technology. Eric Accession Number Ed 523641.**

- **Ingebo, G. (2005). Item Response Theory: Cosistency of Rasch Calibration in Basik Skills Item Banks. Chicago. Mesa Press.**
- **Jahja, V. (1988). Robustness of the Simple Linking Procedure in Item Banking Using The Rasch Model. Dissertation Abstracts International. P.174. ed395687.**
- **Jason L. Meyers, G. Edward Miller, Walter D. Way (2009). Item Position and Item Difficulty Change in an IRT-Based Common Item Equating Design. Applied Measurement In Education. 22(3). PP. 38–60.**
- **Jennifer Y., Melanie D., and Scott C. (2009). True Confessions?: Alumni's Retrospective Reports on Undergraduate Cheating Behaviors. Ethics & Behavior. 19(1) PP. 1–14.**
- **Jordan, A. E. (2008). College Student Cheating: The Role of Motivation. Perceived Norms. Attitudes. and Knowledge of Institutional. Policy. Ethics & Behavior. 14. PP. 342–359.**
- **Kaufman, A.M.; Libling, D.E. (2001). Assessment of Prevalence of Academic Dishonesty, Paper Presented at the Annual Meeting of the American Psychological Association. San Francisco. Ca. August 2-6.**
- **Kherfi, Samer (2011). Top of Form Whose Opinion is it Anyway? Determinants of Participation in Student Evaluation of Teaching, Journal of Economic Education. 42 (1). PP. 19-30.**
- **Liancre, J.M. (2006). The Effect of Misfit on Measurement, Paper Presented at he Annual Meeting of the International Objective Measurement Workshop. 8th. Barkeley. Ca. April.**

- **Lidwine B.(2006).Improving the Quality and Applicability of the Dutch Scales of the Communication Profile for the Hearing Impaired Using Item Response Theory. Applied Psychological Measurement. 3(2). PP.98-122.**
- **Linda l. (1998). A Flixible Alternative To Conventional Methods for Solving Practical Testing Problems. Paper Presented at the Annual Meeting of the American Educational Research Association 75th. Los Angeles. Ca. April 13-17.**
- **Loken, E., Kelly, L. (2010). Estimation of a Four-Parameter Item Response Theory Model, British Journal of Mathematical and Statistical Psychology. 63(2). PP. 509–525.**
- **Masters, G.B. (1982). Rating Scale Analysis. Chicago, Mesa Press,.**
- **Meijer, R.R. (2005). The Influence of the Presence Item Score Patterns, University of Tewente. Research- Report. Enscheade. Netherlond. Faculty of Educational Science and Technology. November. Eric Accession Number Ed 6235478.**
- **Michael E., Charles H. (2010). The Effects of Grading and Teaching Practices on Students' Perceptions of Grading Fairness. College Teaching. Vol.58. PP. 93–98.**
- **Michael W., Amanda B and Matthew B. (2009). Cognitive Responses of Students Who Witness Test Cheating. Journal of Instructional Psychology. 42 (3). PP.255-271.**
- **Moore, M. G. (2008). Cheating. American Journal of Distance Education. 20. PP. 1-3.**
- **Nakamura,Y. (2006).Rasch Measurent And Item Banking: Theory And Practics. Mahwah. NJ: Erlbaum**

- **Nathanson, C, Paulhus, D. L., & Williams,K. M. (2010). Predictors of a Behavioral Measure of Scholastic Cheating: Personality and Competence But Not Demographics. Contemporary Educational Psychology. 31. PP.91-122.**

- **Petridou, A., Williams, J. (2011). Accounting for Unexpected Test Responses Through Examinees'. Assessment in Education: Principles. Policy & Practice. 17 (4). PP.357-382.**

- **Richichi,R.V. (2003). Analysis of Test Bank Multiple-Choice Items Using Item Response Theory. Report- Research. Eric. ED404248.**

- **Roberts, J.S.; Bao, H.C.; Gagne,C.P. (2003). Exploring Altrnative Prameter Estates From the Generalized Partial Credit Model, Paper Presented at the Annual Meeting of National Council On Measurment In Educational. April 22-24.**

- **Ronald T. (1999). Cheating in School Examinations How to Avoid it. Paper Presented at the Annual Meeting of the American Educational Research Association 34th. Los Angeles. Ca. April 25-37. Eric Accession Number Ed 395687.**

- **Satterlee, A. G. (2011). Academic Dishonesty Among Students: Consequences and interventions. (Eric Document Reproduction Service No. ED668477.**

- **Thissen, D. & Orlando, M. (2006). Item Response Theory for Items Scored in Two Categories. in D. Thissen & Wainer. H. (Eds.). Test Scoring. PP. 73-140. Mahwah. NJ.**

- **Tie Liang; Wells, Craig S..(2009). A Model Fit Statistic for Generalized Partial Credit Model. Educational & Psychological Measurement. 69 (6). PP.913-928**
- **User's Guide. SPSS Inc., (2010). IBM SPSS Statistics 19 Core System. IBM Company.**
- **Van der linden, W.J. (2006). Sotaridona, L.S.:A Statistical Index For Detecting Answer Copying On Multiple-Choice Tests. Research Report. Twent Univ. Enscheade. Netherlond. Faculty of Educational Science And Technology. Eric Accession Number Ed 562489.**
- **Van Der Linden, Wim J. Entink, Rinke H, Klein, Fox, Jean-Paul (2010). IRT Parameter Estimation With Response Times As Collateral Information. Applied Psychological Measurement. Jul. 34 (5). PP.327-347.**
- **Wang, S.; Wang, T. (2002). Relative Precision of Ability Estemats In Polytomous CAT. A Comparison Under The Generalized Partial Credit Model And Gradad Response Model. Research-Report. (ERIC). ED 487926.**
- **Wright B.D. and Stone, M.H. (1979). Best Test Design. Chicago. Mesa Press.**
- **Yan,J.W. (1997). Examining Local Item Dependence Effects in a Larg- Scale Science Assessment By the Rasch Credit Model. Paper Presented at the Annual Meeting of American Educational Research Association (Chicago). IL. March 24-28.**