



دراسة تحليلية لمدى توافق الاختبارات الوطنية مع معايير الاختبار الجيد للمدارس المستقلة في المرحلة الابتدائية بدولة قطر

د. بدرية سعيد المالكي*

أ. د. عائشة أحمد فخرو**

ملخص:

هدفت الدراسة إلى تعرف واقع الاختبارات الوطنية في ضوء معايير جودة التقويم، ومدى تحقيقها للمعايير الخاصة بالمنهاج القطرية في المواد الدراسية الأربع (اللغة العربية، اللغة الإنجليزية، الرياضيات، والعلوم)، بالصفوف من الرابع إلى السادس بالمدارس الابتدائية المستقلة. ولتحقيق أهداف الدراسة تم تصميم سلم مصفوفة الاختبارات الوطنية، ويتضمن أربعة محاور، هي: "تعليمات الاختبار"، "اتفاق مفردات الاختبار مع معايير المنهج"، "اللغة"، و"التنظيم". وتكون مجتمع الدراسة من (١٢) ورقة اختبارية من الاختبارات الوطنية، وفي المواد الدراسية الأربع وللصفوف الثلاثة المتقدمة، التي طبقت خلال السنتين (٢٠١١-٢٠١٢) / (٢٠١٢-٢٠١٣)، وقد أشارت نتائج الدراسة إلى وجود محور في مستوى غير متحقق في مادتي العلوم والرياضيات، وهو محور "تعليمات الاختبار"، حيث وجد القصور في بند "زمن الاختبار الكلي"، و"الدرجة المعطاة لكل سؤال"، و"تناسب عدد الأسئلة مع زمن الاختبار الكلي في المواد الرئيسية الأربعة" ولجميع الصفوف الدراسية، إضافة إلى وجود قصور في معايير خاصة في بعض المواد الدراسية الرئيسية دون الأخرى، وذلك في محوري "اتفاق مفردات الاختبار مع معايير المنهج"، و"التنظيم"، وكان "محور اللغة" في مستوى مرضٍ في المواد الأربع الرئيسية. وفي ضوء هذه النتائج تم تقديم توصيات يمكن أن تساهم في تحسين مستوى نماذج الاختبارات الوطنية بالمدارس الابتدائية.

* أستاذ مشارك، قسم العلوم التربوية، كلية التربية، جامعة قطر، دولة قطر.

** أستاذ بقسم العلوم التربوية، كلية التربية، جامعة قطر، دولة قطر.

المقدمة:

يعد التقويم مكوناً رئيساً من مكونات العملية التدريسية، حيث يقدم مؤشرات عن فاعلية البرامج التربوية، وتوجيه مسارها في أثناء كل مرحلة من مراحل إعدادها وتنفيذها سواء على مستوى الأهداف أو الوظائف أو المحتوى. (علام، ٢٠٠٧: ١٩).

وهناك العديد من الأدوات التي يمكن استخدامها لتقييم مدى تحقق الأهداف التعليمية أو المستويات المتوقعة من الطلاب في جميع مراحل التعليم، وتعتبر الاختبارات التحصيلية التحريرية من أكثر وسائل التقويم شيوعاً وانتشاراً في العملية التعليمية.

وتساعد الاختبارات على فرز المتعلمين وفقاً لمستوياتهم التعليمية (التحصيلية)؛ مما يجعلها مؤشراً للمعلمين والمديرين والآباء عن جوانب الضعف والقصور التي يمكن معالجتها لدى تلاميذهم؛ فهي كالميزان يقاس به مدى الثقة في المخرجات التعليمية، كما أن افتقار الاختبارات التحصيلية إلى المعايير الشكلية والجهرية عند إعدادها وتنفيذها من شأنه الإقلال من مدى موثوقية مخرجات العملية التعليمية. (عبد العزيز والبلوي، ٢٠١١: ٥١).

كما تشكل الاختبارات التحصيلية أهمية لدى واضعي المناهج واختصاصيي إعداد الكتب المدرسية كمنفذين للسياسة التعليمية القومية؛ حيث تمدهم بتغذية راجعة أساسية تساعد في تطوير المناهج، ومن ثم لا بد أن تكون الاختبارات على قدر عال من الدقة في قياس ما وضع لقياسه، وإلا كانت كل الأحكام المبنية على نتائجها أحكاماً غير صادقة ولا يمكن الاعتماد عليها، ولا بد من توافر الشروط الأساسية للقياس الدقيق في اختباراتنا حتى نطمئن إلى نتائجها، ومن ثم إلى التقويم الذي يبنى على تلك النتائج. (سلمان، ٢٠١١: ١٨٣).

(١) This work was made possible by NPRP grant # 4-766-5-117 From the Qatar National Research Fund (A Member of The Qatar Foundation). The statements made herein are solely the responsibility of the authors.

وقد قامت دولة قطر مؤخراً بعمليات إصلاح طموحة، أدت إلى حدوث تغييرات جذرية في المناهج والبرامج التعليمية على مستوى جميع مراحل التعليم العام، بدءاً من التغيرات في مراحل رياض الأطفال إلى الصف الثاني عشر، ويسعى هذا النظام إلى إيجاد نظام تعليمي يتميز بالحدثة والتطور في مضمون التعليم وأسلوبه، لذا تبنت هذه المبادرة نظاماً حديثاً لتطوير المناهج يعرف بمعايير المناهج من أجل رفع مستوى التحصيل العلمي للمتعلمين، وتنوع مخرجات التعليم بما يتناسب مع احتياجات السوق المحلية المستقبلية في دولة قطر، وإتاحة الفرصة للأباء في اختيار البدائل التعليمية لأبنائهم، ومنح الاستقلالية للمدارس، وتطبيق نظام المساءلة لجميع المدارس الحكومية.

وتضمن الإصلاح في بداية مرحلة مبادرة تطوير التعليم وضع معايير للمناهج في المواد الرئيسة الأربع: اللغة العربية، اللغة الإنجليزية، الرياضيات والعلوم، لطلاب المراحل من رياض الأطفال إلى الصف الثاني عشر، وتتوافق المناهج والتقييمات والتطوير المهني مع معايير المناهج، وللتأكد من مدى تحقق المعايير يتطلب الأمر القيام بعمليات تقويم لمعرفة مستوى أداء المتعلمين في تحقيق هذه المعايير. ويهدف نظام التقويم القطري إلى تحقيق الأمور الآتية:

١ - تزويد أصحاب القرارات والسياسات بمعلومات حول مدى التقدم في عملية الإصلاح التعليمي.

٢ - توفير تغذية راجعة للمعلمين، تساعد على تحسين عمليات التدريس لدعم احتياجات المتعلمين.

٣ - توفير معلومات حول أداء المدارس للمجتمع الخارجي.

والتقييم التربوي الشامل لدولة قطر Qatar Comprehensive Educational Assessment يطبق على الطلبة في نهاية كل سنة دراسية، بداية من الصف الرابع إلى الصف الثاني عشر، وهو أول تقييم قومي قائم على المعايير يجري في المنطقة، ويقيس مستوى أداء الطلبة وفقاً لمتطلبات معايير المناهج باستخدام الأسئلة متعددة الخيارات، والأسئلة ذات الإجابات المفتوحة، وطبق التقييم التعليمي الشامل

القطري في عام ٢٠٠٤ على (٨٨,٥٥٥) من الطلبة المسجلين في وزارة التربية والتعليم، والمدارس العربية الخاصة، والمدارس المستقلة، وبدءاً من عام ٢٠٠٧ كان تطبيق التقييم التعليمي الشامل القطري مقصوراً على طلبة المدارس المستقلة. (Gonzalez; Broer; Mariano; Froemel; Goldman. & DaVanzo, 2009).

وتشير نتائج أداء الطلبة في تلك الاختبارات الوطنية في عام ٢٠١٢-٢٠١٣ في المواد الدراسية الرئيسة (اللغة العربية، اللغة الإنجليزية، العلوم والرياضيات) في المرحلة الابتدائية للصفوف من الرابع إلى السادس إلى ما يلي (المجلس الأعلى للتعليم، ٢٠١٣):

- تراوح نسب الذين حققوا المعايير في المواد الدراسية الأربع ما بين ١٣٪ و ٣٣٪، وهي نسبة ضئيلة، ما عدا مادة العلوم للصف الرابع، فقد وصلت إلى ٥١٪.

- أقل نسبة في تحقيق المعايير كانت في مادة الرياضيات للصف الخامس بنسبة ١٣٪، يليها في ذلك الصف السادس ١٨٪، ومن ثم مادة العلوم في الصف السادس بنسبة ١٨٪، يلي ذلك مادة الرياضيات في الصف الرابع بنسبة ٢٠٪.

- تعتبر مادة الرياضيات أقل المواد الدراسية تحقيقاً للمعايير مقارنة مع المواد الدراسية الأخرى، ومن ثم فإن هذه النسب المنخفضة في تحقيق معايير المناهج، تتطلب تعزيز المستوى التحصيلي للطلبة في المرحلة الابتدائية للصفوف الدراسية المتقدمة؛ إذ إنه يلعب دوراً مهماً في تحقيق رسالة المنظومة التعليمية وقيمتها وثوابتها المتعلقة بجودة الأداء، وتحقيق رغبات المستفيدين من العملية التعليمية وتطلعاتهم. (المجتمع، الطلاب، أولياء الأمور).

وتعتبر اختبارات التقييم التربوي الشامل نوعاً من الاختبارات المقننة أو الاختبارات مرجعية المحك، وهي توفر معلومات عن وضع الطالب قياساً على الأهداف التعليمية الخاصة بمجال معين، ولا تركز على ما يركز عليه معلم معين

في تدريسه أو في مدرسة معينة؛ فهي تعطيك نظرة موضوعية عما تعلمه طلابك، وقد تمثل جانباً سلبياً؛ لأن المعارف والمهارات العقلية التي يغطيها الاختبار ربما لم يتعرض المتعلمون لدراستها بدرجة كبيرة، وعليه، من الواجب على الجهة التعليمية مقارنة محتويات الاختبار مع المعايير الوطنية للمناهج. (نبتكو وبروكهارت، ٢٠١٢: ٤٢٧)

وقد أشارت الدراسات إلى أن أداء الطلبة في المناهج القائمة على المعايير أفضل من أداء الطلبة في الاختبارات التقليدية للتحصيل والإنجاز في مادة الرياضيات، ولا سيما في تقييم فهم قدراتهم التصورية على استخدام الرياضيات في حل المسائل (Budak, 2008: 111).

وتشير الدراسات أيضاً إلى أن تحصيل الطلبة وإنجازهم يعتمد بشكل كبير على وجود مناهج جيدة التخطيط تتوافق مع كل من معايير الولاية، بالإضافة إلى ذلك لابد من تنفيذ المنهج بصورة متناسقة ومتسلسلة في جميع المراحل الدراسية مع مراعاة الوضوح والتحديد؛ لتجنب التطبيق غير المنسجم مع معايير المناهج. (Metty, 2010: 30)

وعند اتخاذ القرارات المرتبطة بإستراتيجيات التدريس المناسبة، على المعلمين التركيز على معايير المناهج، والعمل على المواءمة بين معايير المنهج والإستراتيجيات التعليمية التي تدعم العملية التعليمية، ومن ثم تتوافر القدرة على تحسين الممارسات التعليمية، بالإضافة إلى ذلك تعد التقييمات المناسبة ضرورية لقياس مدى تقدم الطالب التحصيلي، ويتم إجراء نوعين من التقييم: التقييم البنائي Formative، والتقييم التجميعي Summative. Pemberton; (Rademacher; Tyler & Cereijo. 2006: 284).

وأوصت دراسة (Adomou, 2011) بأهمية إجراء مناقشة حقيقية حول المناهج والتدريس وتعليم الطلبة والتطوير المهني، وتكون البداية بصفة فردية للمدارس، والعناية بمناقشة تعليم الطلبة من أجل إيجاد بدائل أخرى لتنشيطهم، والنظر إلى بنود المساءلة؛ لأنها غير عادلة لكل من المعلمين والطلبة؛ حيث يتم

توجيه اللوم إلى المعلمين من دون حق بشأن الأداء الضعيف للطلبة، ويقوم الطلبة باستخدام اختبارات تفوق استعدادهم الأكاديمي، ومن ثم لابد من استكشاف الدلالات الأخلاقية لمقاييس المساءلة المفروضة لكل من الطلبة والمعلمين.

أما دراسة (الرازحي، ٢٠٠٥) فقد أوصت بإنشاء مركز وطني للاختبارات والتقويم التربوي يتولى أداء الوظائف والمهام الأساسية في مجال الاختبارات والتقويم التربوي، والاهتمام برفع كفاءة الكادر التعليمي المؤهل في مجال التقويم التربوي، وإجراء الدراسات والبحوث اللازمة في مجال الاختبارات بغرض تحسين نوعية الأدوات والأساليب التقويمية المتبعة؛ فضلاً عن ذلك فإن هذا المركز يتولى العديد من المهام المتعلقة بإعداد الاختبارات المقننة، وإنشاء بنوك الأسئلة، لمواجهة احتياجات الجهات المختصة في مراحل التعليم المختلفة.

وتعد عملية الإصلاح القائمة على المعايير وسيلة لتحسين جودة التدريس والتعليم من مرحلة رياض الأطفال حتى الصف الثاني عشر، وهي قضية تعليمية مهمة في التعليم الأمريكي المعاصر، وفي إطار الجهود المبذولة تناولت العديد من الدراسات تقييم الاختبارات التحصيلية على ضوء المعايير العلمية، فقد أجرى (Bishop, 1997) دراسة هدفت إلى وصف سياسات المدرسة ومواردها ومؤثرات معلمي مادتي الرياضيات والعلوم للصف الثامن وعلاقتها بالتقييم الدولي للتقدم التعليمي، ووزعت استبانة على أفراد العينة المكونة من مديري المدارس في كندا، التي تطبق اختبارات التقييم القائمة على معايير المناهج الدراسية. وأظهرت النتائج أن تلك الاختبارات قد أثرت على سلوك الوالدين والمعلمين والإداريين في المدرسة؛ فأصبح الآباء أكثر اهتماماً بتعلم أبنائهم في مادتي الرياضيات والعلوم؛ بالإضافة إلى أن المدارس تقدم المزيد من ساعات التدريس للرياضيات والعلوم، وتكلف المزيد من الواجبات المنزلية ولديها أفضل المعامل، ويطبق المعلمون أفضل الممارسات في الإستراتيجيات التدريسية.

وقام (Campbell & Kyriakides, 2000) بدراسة هدفت إلى تحليل التحصيل التربوي وتقييمه وأثره على السياسات والممارسات التعليمية في المرحلة الابتدائية، وتم استخدام الدراسة التحليلية المقارنة لبعض الدول، من

مثل إنجلترا وقبرص واليابان، وكشفت نتائج الدراسة أنه على الرغم من ارتفاع مستوى أداء طلبة اليابان في التقييم الشامل لمادة الرياضيات، فإن أداءهم في بنود التقييم المرتبطة بالتطبيق والتحليل لم يكن عالياً؛ مما كان له الأثر الإيجابي في تطوير منهج الرياضيات، وتقديم مزيد من الاهتمام على مستوى التطبيق والتحليل، ويمكن لهذه الحقائق أن توفر معلومات قيمة لوضع السياسة في تحديد نقاط القوة والضعف في النظام التعليمي، ولا سيما تطبيق المناهج الدراسية، وذلك بالتركيز المتزايد على نتائج الطلبة.

كما أجرى (Smith, 2001) دراسة تستكشف أنواع الفقرات الاختبارية التي يستخدمها المعلمون والمستويات المعرفية لها، وتكونت عينة الدراسة من (١٧٢) ورقة اختبارية تم تحليلها، وأظهرت النتائج أن فقرات الاختبار من متعدد والمزاوجة والأسئلة ذات الإجابات القصيرة هي الأكثر استخداماً، وأن غالبية المستويات المعرفية التي تقيسها الاختبارات هي في المستوى الإدراكي، كما أظهرت النتائج عدم وجود فروق في نوعية الاختبارات وفقاً لمتغير خبرة المعلمين.

وفي دراسة أجرتها (HaLL, 2009) هدفت إلى الكشف عن أداء الطلبة في منهج الرياضيات القائم على معايير الولاية، وتحديد الطلبة المعرضين لخطر الرسوب، وتخطيط عمليات التدريس، وتم جمع الأدلة من خلال استخدام اختبار الرياضيات والعمل على توافقه مع معايير الولاية، وتطبيقه على (٣٣٧) طالباً، واستطلاع آراء (٢٢) عضواً من هيئة التدريس. أظهرت النتائج أن الاختبار التحصيلي لم يعكس بشكل مناسب أداء الطلبة في نهاية العام الدراسي، ولم يتضمن بنوداً تقيس كلاً من الأهداف الستة المعرفية والمهارات الأساسية وفقاً لمعايير ولاية تكساس، ولا يبدو مناسبة الاختبار كمؤشر لقياس أداء الطلبة في الاختبارات النهائية على مستوى الولايات، وقد كانت وجهة نظر المعلمين في هذا الموضوع أنه لا توجد أداة تشخيصية تقوم بتحديد الطلبة الذين يعانون من صعوبات في مادة الرياضيات، وأنه يتم تخطيط الدروس استجابة لمعايير الولاية أكثر من الاستجابة لاحتياجات الطلبة.

وأجرى (Shapiro. Keller, Lutz, Santoro & Hintz. 2006) دراسة هدفت إلى فحص مدى تقدم الطلبة نحو أهداف المنهج، وإيجاد العلاقة بين معايير الولاية والاختبارات المعيارية في مادة القراءة والرياضيات على مستوى الولاية في ولاية بنسلفانيا. وقد أظهرت النتائج وجود علاقة متوسطة بين الاختبار المنتصف في القراءة والرياضيات وبين الاختبارات المعيارية، وأن الاختبارات المبنية على المعايير تصبح مؤشراً للحكم على مدى نجاح الطلبة على مستوى الولاية.

كما قام مكتب تقييم الأداء التعليمي في نيوزلندا (Education Review office, 2010) بتقييم مدى استعداد قيادات المدارس والمعلمين ومجلس الأمناء للعمل وفقاً لمعايير المناهج الدراسية، وتم إجراء التقييم في (٢٢٨) مدرسة. ووجد أن ٧٨٪ من المدارس حققت تقدماً ملحوظاً، والمدارس التي تم إعدادها بشكل جيد وفق المعايير أحرزت تقدماً في المراجعة الذاتية لمناهجها الدراسية وتصميمها، وكان لمعظم هذه المدارس قيادات مهنية؛ حيث شارك القادة والمعلمون في أنشطة تعليمية مهنية مستمرة ومخططة، ركزت على تصميم ومراجعة المناهج الدراسية، بالإضافة إلى القدرة على جمع بيانات التقييم وتحليلها واستخدامها، أما المدارس التي لم يتم إعدادها جيداً وفقاً لمعايير المناهج، وكانت بنسبة (٢٠٪)؛ فقد وجد مكتب تقييم الأداء التعليمي عدداً من العوامل أسهمت في ذلك، ومنها القيام بقدر كبير في تحسين ممارسات التقييم، ولاسيما المرتبطة بجمع واستخدام المعلومات الخاصة بتحصيل الطلبة كجزء من المراجعة الذاتية للمدرسة، بالإضافة إلى زيادة معدل التغيير والتبديل في فريق العاملين بالمدرسة ووجود معارضة نحو معايير المناهج والقيادة المهنية.

وأجرى (Adamou, 2011) دراسة حول إدراك معلمي اللغة الإنجليزية والرياضيات بالمدارس المتوسطة والعالية لقيمة وآثار عملية الإصلاح القائمة على المعايير، فيما يخص المناهج والممارسات التدريسية وجودة تعليم الطلبة. وأشارت نتائج الدراسة إلى أن الإصلاح القائم على المعايير لا يعطي قيمة واهتماماً بالتعليم، وإنما ينصب اهتمامه على المعايير ولا يتواءم مع حاجات

المتعلمين وقدراتهم؛ مما سيفقد التعليم أهميته وقيمته. ووجد أن سيادة المعايير تتجاوز قيمة المهارات الفكرية لدى الطلبة وعدم ابتكار الأنشطة التي تشرك الطلبة وتثير إبداعاتهم، وتجاهل واضعي السياسات الاحتياجات الحقيقية لكل من الطلبة والمعلمين. وأضافت الدراسة أن تشديد واضعي السياسات على نتائج الاختبارات سيفقد التعليم قيمته، ولن يكون التعليم بذلك عملية للنمو وللاكتشاف الذاتي.

وقام (عبد العزيز والبلوي، ٢٠١١) بدراسة تقويم الاختبارات التحصيلية لمنهج الرياضيات للصف الأول بالمرحلة المتوسطة وفقاً لمعايير التقويم في المملكة العربية السعودية. واقتصرت عينة البحث على بعض المدارس المتوسطة بمدينة تبوك وعددها (٩) مدارس، واعتمد البحث على إعداد جدول مواصفات الاختبار التحصيلي النموذجي، ومقياس تقدير المعايير الشكلية للاختبارات التحصيلية وأظهرت نتائج الدراسة التنوع بين الأسئلة المقالية والموضوعية، وزيادة عدد الأسئلة في مستوى التطبيق، وتلاشي الأخطاء العلمية، أما نقاط الضعف فقد اتضح عدم توافر معيار الصدق والشمولية والتميز في جميع الاختبارات لمقرر الرياضيات، وإغفالها لتنوع الأهداف.

وهدف دراسة أجراها (الطراونة، ٢٠١٢) إلى تعرف مدى توافق الاختبارات التحصيلية بمديرية التربية والتعليم في محافظة الكرك مع معايير وتصميم الاختبار التحصيلي الجيد، وقام الباحث بتطبيق الدراسة على عينة مؤلفة من (١٤١) ورقة اختبارية توزعت على ثلاثة مباحث دراسية، وقد اعتمدت الدراسة في جمع بياناتها على بطاقة تتضمن معايير الاختبار. وتوصلت إلى توافق الاختبارات مع معايير بناء وتصميم الاختبار الجيد بنسبة (٧٠٪)، واحتلت معايير كتابة تعليمات الاختبار بنسبة (٧٣٪)، وجاءت المعايير المطلوبة في كتابة فقرات الاختبار في المرتبة الأخيرة بنسبة (٥٢٪). واتضح عدم وجود فروق دالة إحصائية تبعاً لمتغير جنس معد الاختبار أو تخصصه الأكاديمي.

وهدف دراسة (Abu-Hamour, 2013) إلى تقييم مدى كفاءة اختبار التهجئة الذي صمم وفق معايير المناهج في قياس تحصيل الطلبة في مستوى النطق.

تكونت العينة من (٩٥) تلميذاً من تلاميذ الصف الثالث، واستخدم مع المجموعة التجريبية طريقة المنهج التجريبي والتوجيهات المباشرة للتهجئة، أما المجموعة الضابطة فقد استخدم معها الطريقة التقليدية في التدريس، وكانت المدة الزمنية نحو (١٨) أسبوعاً، وأظهرت النتائج تفوق أداء المجموعة التجريبية على المجموعة الضابطة، وجودة الاختبار في تحقيق الهدف منه.

من خلال الاطلاع على الدراسات السابقة تبين: حرص التربويين على أن تكون الاختبارات ذات كفاءة عالية، وتتصف بالموضوعية والصدق والثبات، وحسن الإعداد والإخراج؛ بحيث تكون عوناً للمعلم والطالب على حد سواء، وبما أن هدف الاختبارات معرفة مستوى أداء المتعلم، وما يترتب على نتائجها من قرارات مهمة تؤثر في مستقبل المتعلم؛ فلا بد أن تكون على درجة عالية من الدقة في القياس؛ فهي المعنية أكثر من غيرها بدقة التخطيط، وتحتاج إلى مجهود كبير في انتقاء وحداتها.

واستناداً إلى ما سبق، ونظراً لأهمية الاختبارات التحصيلية المقننة كمحك للحكم على جودة مخرجات التعلم، وتعلم المواد الدراسية، فإن الدراسة الحالية تتطلع للوقوف بشكل علمي دقيق يستند إلى معايير علمية في جوانب القوة والضعف في الاختبارات الوطنية بدولة قطر، في ضوء معايير التقويم الجيد.

مشكلة الدراسة وتساولاتها:

انطلاقاً من أن الاختبارات التحصيلية المقننة تعتبر من أفضل وسائل القياس والتقويم المعرفي إذا ما أحسن إعدادها، وأكثرها شيوعاً واستخداماً في المجالات الأكاديمية، وإعداد الاختبارات بشكل منظم وفعال يعكس الصورة الحقيقية للتعليم المدرسي، ويعتمد عليها في قياس قدرات المتعلمين على اختلاف مستوياتهم، ويمكن استخدامها في البرامج الاختبارية التي تصمم لتحديد مدى تحقيق المدارس لمستويات تربوية متوقعة.

ومن ثم تقوم هذه الدراسة على محاولة تعرف واقع الاختبارات الوطنية بدولة قطر، وفي ضوء معايير جودة التقويم، ومدى تحقيقها للمعايير الخاصة بالمنهج القطرية في المواد الدراسية الأربع (اللغة العربية، اللغة الإنجليزية،

الرياضيات، والعلوم)، في الصفوف من الرابع إلى السادس بالمدارس الابتدائية المستقلة، وتحديد مواطن القوة وتدعيمها، وتحديد مواطن الضعف وعلاجها بالتغذية الراجعة الفعالة، وذلك من خلال الإجابة عن السؤال الرئيس الآتي:

– ما مدى تحقيق الاختبارات الوطنية في المرحلة الابتدائية بدولة قطر لمعايير بناء وتصميم الاختبار التحصيلي الجيد ؟
ويتفرع عن هذا السؤال التساؤلات الآتية:

س ١: ما مدى توافق أسئلة الاختبارات الوطنية في مادة اللغة العربية بالصفوف الثلاثة المتقدمة (الرابع، الخامس، السادس) مع معايير وتصميم الاختبار الجيد ؟

س ٢: ما مدى توافق أسئلة الاختبارات الوطنية في مادة اللغة الإنجليزية بالصفوف الثلاثة المتقدمة (الرابع، الخامس، السادس) مع معايير تصميم الاختبار الجيد ؟

س ٣: ما مدى توافق أسئلة الاختبارات الوطنية في مادة العلوم بالصفوف الثلاثة المتقدمة (الرابع، الخامس، السادس) مع معايير تصميم الاختبار الجيد ؟
س ٤: ما مدى توافق أسئلة الاختبارات الوطنية في مادة الرياضيات بالصفوف الثلاثة المتقدمة (الرابع، الخامس، السادس) مع معايير تصميم الاختبار الجيد ؟

س ٥: هل توجد فروق ذات دلالة إحصائية عند مستوى (٠,٠٥) في مدى توافق الاختبارات الوطنية مع معايير تصميم الاختبار تعزى لمتغيرات التخصص، الصف الدراسي، سنة الاختبار المطبق؟

أهداف الدراسة:

تهدف الدراسة إلى:

– تعرف مدى توافق الاختبارات الوطنية للصفوف الثلاثة المتقدمة (الرابع، الخامس، السادس) بالمرحلة الابتدائية مع معايير تصميم الاختبار التحصيلي الجيد.

- تعرف إذا ما كانت هناك فروق في مدى توافق الاختبارات الوطنية مع معايير تصميم الاختبار التحصيلي الجيد تعزى لمتغيرات التخصص، الصف الدراسي، وسنة تطبيق الاختبار.

أهمية الدراسة:

تحدد أهمية الدراسة في النقاط الآتية:

- ١ - تعطي دليلاً واضحاً على واقع الاختبارات الوطنية المطبقة في مدارس المرحلة الابتدائية المستقلة بدولة قطر، وتكشف عن مدى اشتغالها على معايير بناء وتصميم الاختبار الجيد، وجوانب القوة والضعف فيها، بهدف تحسينها وتطويرها؛ لتصبح على درجة عالية من الكفاءة والجودة في تحقيق الأهداف المنشودة.
- ٢ - نتائج هذه الدراسة تعد ذات أهمية كبيرة للمعلمين؛ لأنها تساعدهم في تحديد المعايير التي نجح طلابهم في إتقانها، والمعايير التي يجدون صعوبة في إتقانها؛ مما يساعد في تحسين برامجهم التعليمية وتطويرها، وإستراتيجيات التعليم لتصبح أكثر ملائمة لتلبية احتياجات طلابهم.
- ٣ - تقدم للمسؤولين في المجلس الأعلى للتعليم معلومات قيمة عن مدى كفاءة الاختبارات الوطنية المطبقة في المدارس القطرية بالمرحلة الابتدائية، ومدى الحاجة إلى تطويرها.
- ٤ - نتائج هذه الدراسة قد تشجع باحثين آخرين على القيام بدراسات أخرى في مجال تقييم وتطوير الاختبارات الوطنية في مواد دراسية ومراحل أخرى، وبشكل أوسع وأعمق.
- ٥ - تعتبر هذه الدراسة أول دراسة تجرى في المنطقة - على حد علم الباحثين - في هذا المجال، مما قد يساهم في تحسين وتطوير واقع الاختبارات الوطنية في منطقة الخليج لتحقيق الأهداف المرجوة بمستوى أكثر كفاءة ودقة، ومن ثم إصلاح التعليم وتطويره في المنطقة.

حدود الدراسة:

اقتصرت الدراسة على:

- الاختبارات الوطنية القطرية الخاصة بالمرحلة الابتدائية من الصف الرابع إلى السادس الابتدائي؛ حيث إن الاختبارات الوطنية القطرية تطبق على المتعلمين بداية من الصف الرابع الابتدائي دون الصفوف الدنيا.
- الاختبارات الوطنية الخاصة بالمواد الدراسية الأربع (اللغة العربية، اللغة الانجليزية، العلوم، الرياضيات)؛ حيث إن الاختبارات الوطنية في بداية ظهور نظام المدارس المستقلة كانت خاصة بهذه المواد الدراسية فقط.
- الاختبارات الوطنية المطبقة في السنوات من (٢٠١١ - ٢٠١٢) و(٢٠١٢ - ٢٠١٣).
- معايير الاختبار الجيد المشار إليها في أداة الدراسة.

مصطلحات الدراسة:

١ - الاختبارات الوطنية National Test: هي اختبارات مقننة وأداة علمية موضوعية، طُورت بغرض تطوير التعليم، وتحسين مخرجاته؛ فهي تبنى على المعايير والمعارف ومهارات التفكير التي ينبغي للمتعلمين امتلاكها، للتعامل بكفاءة عالية في حل المشكلات والمستجدات التي يواجهونها داخل المدرسة وخارجها، وتطبق بشكل دوري في نهاية الصفوف الدراسية، ويمكن من خلالها الحصول على مؤشرات ذات درجة عالية من الصدق والثبات عما يحدث في الميدان التعليمي؛ وذلك من خلال ما يمتلكه الطلاب والطالبات من معارف ومهارات، وما يستطيعون إنجازه في المواد الدراسية المحددة في التعليم العام، التي يمكن تقديمها إلى صانعي القرار ومخططي المناهج. (وزارة التربية والتعليم، ٢٠١١: ١).

٢ - معايير تصميم الاختبار الجيد Good test standards: هي خصائص وصفات مطلوب توافرها في الاختبار التحصيلي الجيد؛ بحيث يصلح لأداء الغرض الذي صمم من أجله على الوجه الأكمل؛ أي من خلالها يمكن

الحكم على مدى جودة الاختبار في قياس المطلوب بدرجة عالية من الدقة والكفاءة.

٣ - المدارس المستقلة بدولة قطر Independent Schools: هي مدارس ممولة حكومياً، تشمل جميع المراحل التعليمية، ولها الحرية في تحقيق رؤيتها ورسالتها وأهدافها التربوية الخاصة بها، مع الالتزام بالبنود المنصوص عليها في العقد المبرم بينها وبين المجلس الأعلى للتعليم، الذي يتطلب مشاركة جميع طلاب المدارس في الاختبارات الوطنية التي تعتمد في محتواها على معايير المناهج القطرية، وهي تعتبر تقييمات رسمية توفر للمسؤولين في الدولة معلومات تفصيلية عن مستوى الجودة في تحقيق معايير المناهج في هذه المدارس، وقد تم تحويل جميع المدارس الحكومية إلى مدارس مستقلة تدريجياً بدولة قطر، وذلك بداية من عام ٢٠٠٢/٢٠٠٣. (المجلس الأعلى للتعليم، ٢٠٠٥).

منهج الدراسة:

اعتمدت الدراسة على المنهج الوصفي التحليلي للوقوف على مدى تحقيق الاختبارات الوطنية القطرية بالمرحلة الابتدائية لمعايير بناء وتصميم الاختبار التحصيلي الجيد، وتعرف جوانب القوة والضعف فيها من أجل تحسينها وتطويرها، حتى تحقق الغرض المطلوب منها على الوجه الأكمل.

إجراءات الدراسة:

أولاً - مجتمع الدراسة وعينتها:

تكون مجتمع الدراسة من مجموعة من الاختبارات الوطنية الخاصة بالصفوف من الرابع إلى السادس بالمدارس القطرية الابتدائية المستقلة، وفي المواد الدراسية الأربع (اللغة العربية، اللغة الإنجليزية، العلوم، الرياضيات)، التي طبقت خلال السنتين (٢٠١١-٢٠١٢ / ٢٠١٢-٢٠١٣)، بمعدل ١٢ ورقة اختبارية تمثل عينة الدراسة. والجدول (١) يوضح العينة الخاصة بالاختبارات الوطنية.

الجدول (١)

عينة الاختبارات الوطنية بحسب نوع المادة الدراسية، نوع المدرسة،
سنة التطبيق، الصف الدراسي

		الصف							
		رابع		خامس		سادس		المجموع	
		العدد	%	العدد	%	العدد	%	العدد	%
	المجموع	١٢	٪١٠٠,٠	١٢	٪١٠٠,٠	١٢	٪١٠٠,٠	٣٦	٪١٠٠,٠
نوع	لغة عربية*	٤	٪٣٣,٣	٤	٪٣٣,٣	٤	٪٣٣,٣	١٢	٪٣٣,٣
المادة	لغة إنجليزية*	٤	٪٣٣,٣	٤	٪٣٣,٣	٤	٪٣٣,٣	١٢	٪٣٣,٣
	علوم	٢	٪١٦,٧	٢	٪١٦,٧	٢	٪١٦,٧	٦	٪١٦,٧
	رياضيات	٢	٪١٦,٧	٢	٪١٦,٧	٢	٪١٦,٧	٦	٪١٦,٧
سنة تطبيق	٢٠١٢-٢٠١١	٦	٪٥٠,٠	٦	٪٥٠,٠	٦	٪٥٠,٠	١٨	٪٥٠,٠
الاختبار	٢٠١٣-٢٠١٢	٦	٪٥٠,٠	٦	٪٥٠,٠	٦	٪٥٠,٠	١٨	٪٥٠,٠

* تحتوي الاختبارات الوطنية لمادتي اللغة العربية واللغة الإنجليزية نمونتين:
(استماع، ومهارات الكتابة) خلال الأعوام التي تم تطبيق سلم التقييم عليها، ولذلك
نجد أن لكل من اللغة العربية واللغة الإنجليزية أربعة نماذج بدلاً من اثنين كما في
مادتي العلوم والرياضيات.

ثانياً - أداة الدراسة :

تتمثل أداة الدراسة في إعداد سلم مصفوفة الاختبارات الوطنية، واشتملت
المصفوفة على (٣٣) معياراً، موزعة على أربعة محاور، هي:
المحور الأول: تعليمات الاختبار، ويتضمن (٥) معايير.
المحور الثاني: اتفاق مفردات الاختبار مع معايير المنهج، ويتضمن (١٥) معياراً.
المحور الثالث: اللغة، ويتضمن (٣) معايير.
المحور الرابع: التنظيم، ويتضمن (١٠) معايير.
وقد تم اختيار مقياس رباعي متدرج من أربع درجات للإجابة عن عبارات

المصفوفة، على النحو الآتي: (٣) درجات لعبارة (بدرجة مرضية)؛ (درجتان) لعبارة (بدرجة متوسطة)؛ (١) درجة لعبارة (بدرجة ضعيفة)؛ (صفر) لعبارة (لا يتحقق).
صدق الأداة: للتأكد من صدق الأداة عرضت على مجموعة من المحكمين، من أعضاء هيئة التدريس بكلية التربية / جامعة قطر، ومجموعة من اختصاصيي مركز تطوير التربويين بكلية التربية في المواد الدراسية الأربع، وقد تم الأخذ بأرائهم فيما يتعلق بوضوح العبارات، ومدى مناسبتها وتمثيلها للمحور الذي تنتمي إليه، وفي ضوء ملاحظاتهم تم تعديل الأداة.

الصدق الداخلي لسلم تقييم الاختبارات الوطنية:

نتيجة لقلّة عدد نماذج الاختبارات الوطنية المستخدمة في الدراسة؛ فقد تم حساب الصدق الداخلي لهذه الاختبارات بدلاً من الصدق الإحصائي، ويقاس الصدق الداخلي بحساب الجذر التربيعي لمعامل ثبات الاختبار، وباستخدام هذه الطريقة فإن أكبر قيمة وصل إليها معامل صدق الاختبارات الوطنية هي (٠,٩٥) وهو معامل يوحي بالثقة في صدق المقياس.

ثبات سلم مصفوفة الاختبارات الوطنية: لإيجاد ثبات سلم مصفوفة الاختبارات الوطنية تم تقييم عينة عشوائية من الاختبارات الوطنية في المواد الدراسية الأربع " اللغة العربية، اللغة الانجليزية، العلوم، والرياضيات، للصفوف الثلاثة (الرابع، الخامس، السادس) " باستخدام مصفوفة الاختبار، ومن ثم أعيد التصحيح بعد شهر، وحسبت معاملات الارتباط بين درجات التصحيح الأول والتصحيح الثاني لكل المواد مجتمعة، ولكل مادة على حدة، ووجد أن جميع قيم معاملات الارتباط مرتفعة وذات دلالة إحصائية عند مستوى دلالة (٠,٠١)، وقد راوحت قيم معاملات الارتباط بين (٠,٨٢ - ٠,٩٧)؛ مما يدل على أن سلم مصفوفة الاختبارات الوطنية يتمتع بثبات عال.

المعالجة الإحصائية:

كون البيانات المستخدمة في هذا البحث من النوع الرتبي؛ فقد استخدمت أساليب الإحصاء الوصفي (التكرارات، الوسيط)، وحدد معيار لقبول العبارات في مصفوفة تقييم الاختبارات الوطنية، وهو: مستوى مرضٍ (وسيط = ٣)،

ومستوى متوسط (وسيط=٢)، أما العبارات التي كانت مستوياتها أقل من ذلك (وسيط = ١، ٠)؛ ففيها أوجه قصور. ولكون البيانات لا تتبع التوزيع الاعتيادي كما هو مبين أدناه فقد استخدمت أساليب الإحصاء الاستدلالي اللابارامتري (اختبار مان وتني واختبار كروسكال- واليز)؛ لمعرفة دلالة الفروق بين المتغيرات (Kinnear & Gray, 2008)، على النحو الآتي:

- اختبار كروسكال- واليز Kruskal- Wallis Test (لكون البيانات لا تتبع التوزيع الاعتيادي)؛ لمعرفة دلالة الفروق بين المتغيرات التي تحتوي أكثر من مستويين:

- الصف الدراسي (الرابع، الخامس، السادس).
- المادة الدراسية (اللغة العربية، اللغة الإنجليزية، العلوم، الرياضيات).

مناقشة نتائج الدراسة وتحليلها:

ستعرض النتائج في ضوء أسئلة الدراسة المطروحة:

وللإجابة عن هذه الأسئلة حساب الوسيط لكل بند ولكل محور على حدة، وللمحاور مجتمعة للبيانات المجمعة من الأداة (مصفوفة تقييم الاختبارات الوطنية)، وتم تحديد مستوى كفاية الاختبارات الوطنية بناء على السلم الآتي:

قيمة الوسيط	٣	٢	١	٠
المعيار	مستوى مرضٍ	مستوى متوسط	مستوى ضعيف	لا يتحقق

علماً بأنه عند تفسير النتائج تم احتساب قيمة الوسيط (من ٢-٣) على أن المعيار متحقق، وأقل من ذلك يعتبر فيه قصور، ويحتاج إلى مزيد من الاهتمام. والجدول (٢) يوضح النتائج المتعلقة بالأسئلة الثلاثة الأولى للدراسة.

النتائج المتعلقة بالسؤال الأول، ونصه "ما مدى توافق أسئلة الاختبارات الوطنية في مادة اللغة العربية بالصفوف الثلاثة المتقدمة مع معايير تصميم الاختبار الجيد"؟ الجدول (٢) يوضح هذه النتائج.

تابع / الجدول (٣)
الوسيط لكل معيار على حدة وللمعايير كل لأداة الاختبارات الوطنية للمواد الدراسية (اللغة العربية، اللغة الإنجليزية، العلوم، الرياضيات) في المرحلة الابتدائية للمصنف الثلاثة (الرابع، الخامس، السادس)

الرياضيات		العلوم		اللغة الإنجليزية		اللغة العربية	
الصف		الصف		الصف		الصف	
الرابع	الخامس	الرابع	الخامس	الرابع	الخامس	الرابع	الخامس
الوسيط	الوسيط	الوسيط	الوسيط	الوسيط	الوسيط	الوسيط	الوسيط
الكل	الكل	الكل	الكل	الكل	الكل	الكل	الكل
للمعيار	للمعيار	للمعيار	للمعيار	للمعيار	للمعيار	للمعيار	للمعيار

المحور الثاني – اتفاقها مع معايير المنهج: تراعي أسئلة بنود الاختبار الجوانب الآتية:

الارتباط بمعايير المنهج.

تغطية أوزان مجالات المادة

كما تنص عليها المعايير.

الملاءمة مع المنهج الدراسي للمادة.

تغطية عناصر المحتوى (المعرفي، المهاري

الحركي) بشكل متوازن.

إقبال من الأسئلة يقيس المستويات الدنيا من

التفكير (مستوى التفكير).

الكثير من الأسئلة يركز على المستويات

العقلية العليا (كأنهم والقراءة على

تطبيق المعلومات وحل المشكلات).

تابع / الجدول (٣)
الوسيط لكل معيار على حدة وللمعايير كل أداة الاختبارات الوطنية للمواد الدراسية (اللغة العربية، اللغة الإنجليزية، الرياضيات، العلوم، الرياضيات) في المرحلة الابتدائية للمصنف الثلاثة (الرابع، الخامس، السادس)

اللغة العربية			اللغة الإنجليزية			العلوم			الرياضيات		
الصف			الصف			الصف			الصف		
الوسيط الخامس السادس			الوسيط الرابع الخامس			الوسيط الرابع الخامس			الوسيط الرابع الخامس		
الكل			الكل			الكل			الكل		
الوسيط			الوسيط			الوسيط			الوسيط		
المعيار			المعيار			المعيار			المعيار		
٢	١	٢	٢	٣	٣	٢	٣	٣	٢	٢	٢
النوع، منها: أ- المقالة ذات (الوجزة)											
الاستجابة القصيرة (الوجزة)											
ب- الموضوعية بأواعها.											
٢	٢	٢	٢	٢	٢	٢	٢	٢	٢	٢	٢
أ- الالتزام ببند كتابة وصناعة											
الأسئلة المثالية وفق الشروط العلمية.											
٢	٢	٢	٢	٢	٢	٢	٢	٢	٢	٢	٢
الالتزام ببند كتابة وصناعة الأسئلة											
الموضوعية وفق الشروط العلمية لكل منها.											
٢	٢	٢	٢	٢	٢	٢	٢	٢	٢	٢	٢
اشتغال بعض الأسئلة على مستوى											
مناسب من الصعوبة.											
٢	٢	٢	٢	٢	٢	٢	٢	٢	٢	٢	٢
تركيز الأسئلة على المصطلحات											
والمفاهيم الرئيسة في الموضوع القيس.											
٢	٢	٢	٢	٢	٢	٢	٢	٢	٢	٢	٢
١٢- خلو الأسئلة من المفاهيم الغامضة											
والألفاظ غير الواضحة.											

الوسيط لكل معيار على حدة وللمعايير ككل لإداة الاختبارات الوطنية للمواد الدراسية (اللغة العربية، اللغة الإنجليزية، العلوم، الرياضيات) في المرحلة الابتدائية للصفوف الثلاثة (الرابع، الخامس، السادس) تابع / الجدول (٢)

يتضح من الجدول (٢) النتائج الآتية:

- المستوى العام لكفاية إعداد الاختبارات الوطنية ولكل المحاور في مادة اللغة العربية للصفوف الثلاثة المتقدمة (الرابع، الخامس، السادس) كانت مرضية (وسيط=٣).
 - مستوى المحور الأول "تعليمات الاختبار" كانت محققة في الصنفين الرابع والخامس (وسيط=٣)، وفي الصف السادس (وسيط=٢).
 - البندان "زمن الاختبار الكلي" و"الدرجة المعطاة لكل سؤال"، في المحور الأول (تعليمات الاختبار) في مستوى غير متحقق (وسيط = ٠) في الصفوف الثلاثة (من الرابع إلى السادس). وحصل البند "نموذج يبين كيفية الإجابة عن أسئلة الاختبار" على درجة متوسطة (وسيط=٢) في الصف السادس.
 - البندان التاليان في المحور الثاني "اتفاقها مع معايير المنهج" في مستوى غير متحقق بشكل مرضٍ؛ فقد حصل البند "يتوافر عدد قليل من الأسئلة يقيس المستويات الدنيا من التفكير" على مستوى ضعيف (وسيط=١) للصفوف الثلاثة، كما حصل البند "التنوع في الأسئلة المقالية والموضوعية بأنواعهما" على مستوى ضعيف (وسيط=١) للصنفين الرابع والسادس، وعلى درجة متوسطة (وسيط = ٢) بالنسبة للصنف الخامس.
 - البند: "تناسب عدد الأسئلة مع زمن الاختبار الكلي" في المحور الرابع (التنظيم) في مستوى غير متحقق (وسيط = ٠) للصفوف الثلاثة، كما حصل البند "وضوح الصور والأشكال المتضمنة في الاختبار" على مستوى غير متحقق (وسيط = ٠) في الصف السادس فقط.
- في ضوء النتائج السابقة يتضح أن هناك القليل من البنود التي لم تتحقق في الاختبارات الوطنية في مادة اللغة العربية للصفوف الثلاثة المتقدمة (الرابع، الخامس، السادس)، بالمرحلة الابتدائية بدولة قطر، في كتابة تعليمات الاختبار،

وفي تغطيته لمعايير مناهج اللغة العربية للصفوف الثلاثة، وفي تنظيم محتوى الاختبار؛ حيث بلغ إجمالي البنود التي لم تتحقق (٦) بنود من المجموع الكلي للبنود وعددها (٣٣) بنوداً، ومع ذلك فإن هذا القصور يشير إلى أن هناك بعض المهارات يفتقدها واضعو الاختبارات الوطنية في إعدادهم لهذه الاختبارات، وأنهم لم يتقيدوا في إعدادهم لها بمواصفات الاختبار التحصيلي الجيد ومعاييرهم، ويمكن إرجاع هذا القصور إلى نقص خبراتهم في بعض فقرات إعداد الاختبارات التحصيلية المقننة، أو قصور في البرامج التدريبية التي تقدم لهم في هذا المجال، أو قصور في معرفتهم بالمعايير الوطنية القطرية لمنهج اللغة العربية.

وأياً كان مستوى هذا القصور فإنه سوف يؤثر سلباً على تقييم مستوى أداء المتعلمين في هذه الاختبارات؛ من حيث إنه لن يساعد في إعطاء صورة صادقة عن مستوى أدائهم فيها، ومن ثم صعوبة تحديد معايير المنهج التي أتقنها المتعلمون، والمعايير التي يجدون صعوبة فيها، بما لا يساعد في تزويد المسؤولين واختصاصيي المنهج بمعلومات مفيدة تساهم في صناعة قرارات تعليمية صائبة تساعد في تطوير التعليم وتحسينه في مجال تعليم اللغة العربية بمرحلة التعليم الابتدائي. وهذا يتطلب من المسؤولين حسن اختيار من يتم ترشيحهم لإعداد هذه النوعية من الاختبارات، والعمل على تطوير مهاراتهم فيما يتعلق بمواصفات الاختبار التحصيلي الجيد ومعاييرهم. وقد أوصت دراسة الرازحي (٢٠٠٥) بإنشاء مركز للاختبارات والتقييم، يتولى رفع كفاءة الكوادر التعليمية في مجال التقييم التربوي وإعداد الاختبارات المقننة.

النتائج المتعلقة بالسؤال الثاني، ونصه "ما مدى توافق أسئلة الاختبارات الوطنية في مادة اللغة الإنجليزية بالصفوف الثلاثة المتقدمة مع معايير تصميم الاختبار الجيد"؟ بالرجوع إلى جدول (٢) يتبين أنه يوضح هذه النتائج.

– المستوى العام لكفاية إعداد الاختبارات الوطنية العامة، وللمحاور الثلاثة (تعليمات الاختبار، اللغة، والتنظيم) لمادة اللغة الإنجليزية للصفوف الثلاثة المتقدمة (الرابع، الخامس، السادس) كانت مرضية (وسيط = ٣)،

وحصل محور "اتفاقها مع معايير المنهج" على مستوى متوسط (وسيط = ٢) للصفين الرابع والسادس، أما في الصف الخامس فقد حصل على مستوى مرضٍ (وسيط = ٣).

- حصل البندان "زمن الاختبار الكلي" و "الدرجة المعطاة لكل سؤال" في محور "تعليمات الاختبار" على مستوى غير متحقق (وسيط = ٠).
- حصل بند "الملاءمة مع المنهج المدرسي للمادة" في المحور الثاني على مستوى ضعيف (وسيط = ١) للصفين الرابع والخامس، وعلى مستوى متوسط (وسيط = ٢) للصف السادس، وحصل البند "التنوع في الأسئلة بين المقالية والموضوعية بأنواعهما" على مستوى غير متحقق (وسيط = ٠) للصفين الرابع والخامس، ومستوى متوسط (وسيط = ٢) للصف السادس، وحصل البند "الالتزام ببنود كتابة وصياغة الأسئلة الموضوعية وفق الشروط العلمية لكل منها" على مستوى ضعيف (وسيط = ١) للصف الرابع، ومستوى متوسط (وسيط = ٢) للصفين الرابع والسادس.

- حصل البندان "تناسب عدد الأسئلة مع زمن الاختبار الكلي" و "تحديد مستوى معيار الأداء المطلوب في السؤال العام للفقرات" في محور "التنظيم" على مستوى غير متحقق (وسيط = ٠) في الاختبارات الوطنية لمادة اللغة الإنجليزية لجميع الصفوف من الرابع حتى السادس.

في ضوء النتائج السابقة يتضح أن هناك بعض جوانب القصور في بعض بنود بناء الاختبارات الوطنية لمادة اللغة الإنجليزية للصفوف الثلاثة من الرابع إلى السادس الابتدائي، في كتابة تعليمات الاختبار، وفي تغطية فقرات الاختبار لمعايير منهج اللغة الإنجليزية، وفي تنظيم محتوى الاختبار؛ أي أن إجمالي البنود التي لم تتحقق (٧) بنود من المجموع الكلي للبنود وعددها (٣٣) بنوداً، وهذا يعني أن واضعي هذه الاختبارات في مجال اللغة الإنجليزية يفتقدون لبعض مهارات إعداد الاختبارات الجيدة، وفق مواصفات ومعايير محددة؛ الأمر

الذي يتطلب منهم بذل المزيد من الدقة والجهد في إعداد فقرات الاختبارات الوطنية، المتعلقة بتحديد الزمن الكلي للاختبار، وتحديد درجة كل سؤال؛ حتى يتمكن المتعلم من توزيع الوقت المحدد له للإجابة عن فقرات الاختبار بشكل متوازن ومتناسق، دون ضياع لجهد، كما يجب عليهم الحرص على الملاءمة بين فقرات الاختبار ومعايير منهج اللغة الإنجليزية؛ بحيث يغطي جزءاً كبيراً منها، والاهتمام بمراعاة معيار التنوع في الأسئلة المقالية والموضوعية، وصياغتها وفق الشروط العلمية لكل نوع فيها، كذلك التنوع فيها من حيث السهولة والصعوبة، حتى تناسب قدرات المتعلمين، وتراعي الفروق الفردية بينهم، وهذا يتفق مع نتائج دراسة كل من (Hall, 2009)، ودراسة (Smith, 2001)؛ حيث أظهرت نتائجهما أن الاختبار التحصيلي لم يعكس بشكل مناسب أداء الطلبة في نهاية العام الدراسي، ولم يتضمن بنوداً تقيس كلاً من الأهداف الستة المعرفية والمهارات الأساسية وفقاً لمعايير ولاية تكساس، ولا يبدو مناسبة الاختبار كمؤشر لقياس أداء الطلبة في الاختبارات النهائية على مستوى الولايات.

النتائج المتعلقة بالسؤال الثالث، ونصه: ما مدى توافق أسئلة الاختبارات الوطنية في مادة العلوم بالصفوف الثلاثة المتقدمة (الرابع، الخامس، السادس) مع معايير تصميم الاختبار الجيد؟

- تم التعامل مع تحليل الاختبارات الوطنية بالطريقة نفسها التي تمت في الأسئلة السابقة، وجدول (٢) يوضح النتائج الخاصة بهذا الصدد.
- المستوى العام لكفاية إعداد الاختبارات الوطنية لمادة العلوم، وللمحاور الثلاثة (اتفاقها مع معايير المنهج، اللغة، والتنظيم)، وللصفوف الثلاثة المتقدمة (الرابع، الخامس، السادس) كانت مرضية (وسيط = ٣)، وفي محور (تعليمات الاختبار) في مستوى غير متحقق (وسيط = ٠).
- البنود التالية في المحور الأول "تعليمات الاختبار" في مستوى غير متحقق في الاختبارات الوطنية لمادة العلوم للصفوف الثلاثة من الرابع

حتى السادس: "نوع الأسئلة المتضمنة في الاختبار" زمن الاختبار الكلي " الدرجة المعطاة لكل سؤال " (الوسيط = ٠)

البند "الملاءمة مع المنهج المدرسي للمادة" في المحور الثاني "اتفاقها مع معايير المنهج" كان في مستوى غير متحقق (وسيط = ٠)، بينما حصل البند "تغطية عناصر المحتوى (المعرفي، المهارى الحركي) بشكل متوازن" على مستوى ضعيف (وسيط = ١). كذلك البند "الكثير من الأسئلة تركز على المستويات العقلية العليا، كالفهم والقدرة على تطبيق المعلومات وحل المشكلات" حصل على مستوى ضعيف (وسيط = ١) للصف الرابع، وعلى مستوى متوسط (وسيط = ٢) للصفين الخامس والسادس.

البنود التالية في مستوى غير متحقق (وسيط = ٠) في المحور الرابع (التنظيم) من الاختبارات الوطنية لمادة العلوم للصفوف الثلاثة من الرابع حتى السادس، هي: "تناسب عدد الأسئلة مع زمن الاختبار الكلي"، "تجميع الفقرات التي تقيس نفس المحتوى أو المهارة وعرضها معاً"، "تحديد مستوى معيار الأداء المطلوب في السؤال العام للفقرات".

وفي ضوء ما سبق يتضح أن هناك بعض القصور في بنود الاختبارات الوطنية لمادة العلوم في كتابة "تعليمات الاختبار"، وفي "التغطية لمعايير المنهج"، و"تنظيم الاختبار"، وأن إجمالي البنود التي لم تتحقق (٨) بنود من المجموع الكلي للبنود وعددها (٣٣) بنوداً، ومن ثم لا بد من العناية بمواصفات ومبادئ بناء الاختبار، ومدى أهميته في تحسين التعلم والتدريس. ومن المفيد مقارنة مفردات الاختبار ومراجعتها بمعايير المنهج المقصود قياسه، وبناء أنواع الأسئلة المختلفة من أسئلة الاستدعاء والتفكير. ولقد أكدت دراسة (Adomou, 2001) أن على الإصلاح القائم على المعايير مراعاة توافق إستراتيجيات التدريس مع

بنود الاختبار؛ مما يكون له الأثر الكبير على تحسين أداء المتعلمين في الاختبارات المقننة.

النتائج المتعلقة بالسؤال الرابع، ونصه: ما مدى توافق أسئلة الاختبارات الوطنية في مادة الرياضيات بالصفوف الثلاثة المتقدمة (الرابع، الخامس، السادس) مع معايير تصميم الاختبار الجيد؟

والجدول (٢) يوضح النتائج المتعلقة بهذا السؤال على النحو التالي:

- كفاية إعداد الاختبارات الوطنية العامة وللمحاور الثلاثة (اتفاقها مع معايير المنهج، اللغة، والتنظيم)، لمادة الرياضيات للصفوف الثلاثة المتقدمة (الرابع، الخامس، السادس) كانت في المستوى العام مرضية (وسيط = ٣)، وفي محور (تعليمات الاختبار) في مستوى غير متحقق (وسيط = ٠).

- البنود التالية في المحور الأول "تعليمات الاختبار" في مستوى غير متحقق (وسيط = ٠) في الاختبارات الوطنية لمادة الرياضيات للصفوف الثلاثة من الرابع حتى السادس وهي: "عدد الأسئلة الكلية في الاختبار"، "زمن الاختبار الكلي"، "الدرجة المعطاة لكل سؤال".

- البند التالي "تغطية أوزان مجالات المادة كما تنص عليها المعايير..". في المحور الثاني: (اتفاقها مع معايير المنهج) في مستوى غير متحقق (وسيط = ٠)، وحصل البند "يتوافر عدد قليل من الأسئلة تقيس المستويات الدنيا من التفكير (مستوى التذكر)" على مستوى ضعيف (وسيط = ١) للصفوف الثلاثة. كما حصل بند "التنوع، منها: أ- المقالية ذات الاستجابة القصيرة (الموجزة) ب- الموضوعية بأنواعها" على مستوى ضعيف (وسيط = ١) للصف السادس، وفي مستوى متوسط (وسيط = ٢) للصفين الرابع والخامس.

- البنود التالية في مستوى غير متحقق (وسيط = ٠)، في المحور الرابع (التنظيم) من الاختبارات الوطنية لمادة الرياضيات للصفوف الثلاثة،

هي: "تناسب عدد الأسئلة مع زمن الاختبار الكلي"، "تجميع الفقرات التي تقيس نفس المحتوى أو المهارة وعرضها معاً"، "تحديد مستوى معيار الأداء المطلوب في السؤال العام للفقرات".

وفي ضوء ما سبق يتضح أن هناك بعض القصور في بنود الاختبارات الوطنية لمادة الرياضيات في كتابة "تعليمات الاختبار"، وفي "التغطية لمعايير المنهج"، "وتنظيم الاختبار"، وأن إجمالي البنود التي لم تتحقق (٩) بنود من المجموع الكلي للبنود وعددها (٣٣) بنداً، وهذا يعني أن واضعي الاختبارات الوطنية في مادة الرياضيات في حاجة للعناية والاهتمام بتعليمات الاختبار، ومنها: تحديد عدد الأسئلة في الاختبار، زمن الاختبار، ودرجته، وهذه البنود تساعد المتعلم في توزيع الوقت المسموح للإجابة عن كل سؤال؛ مما يساعده على التوافق بين عدد الأسئلة والزمن الكلي للاختبار، وهناك بنود مهمة لابد من أن يوجه الاهتمام بها من قبل واضعي الأسئلة، منها: التنوع في صياغة الأسئلة المقالية والموضوعية وفق الشروط العلمية، وأن يكون هناك توازن في تركيز الأسئلة على المستويات العقلية المتنوعة والعمل على مطابقة مستوى صعوبة الأسئلة مع نواتج التعلم وقدرات الطلبة، والتوازن في تغطيتها لمجالات المادة، وتحديد معيار الأداء لكل سؤال. وأكدت كل من دراسة (Smith, 2001) ودراسة (Hall, 2009) أن الاختبار التحصيلي لم يتضمن بنوداً تقيس المستويات المعرفية المتنوعة، ومن ثم لابد من المراجعة العامة للاختبارات للحكم على جودتها ومدى ملاءمتها، والعمل على التحليل الدقيق للسلوك والأداء المحدد في معايير المناهج للاسترشاد به في بناء المفردات أو الأسئلة الاختبارية؛ مما يساعد على جعل محتوى الاختبار صادقاً، بالإضافة إلى ذلك العمل على تحليل إجابات الطلبة في الاختبار لكل مفردة من مفردات الاختبار ودرجاتهم الكلية ومن ثم تحسينها باستمرار.

النتائج المتعلقة بالسؤال الخامس، ونصه: "هل توجد فروق ذات دلالة إحصائية عند مستوى (٠,٠٥) في مدى توافق الاختبارات الوطنية مع معايير تصميم الاختبار باختلاف المادة الدراسية (العربي، الإنجليزي، العلوم، الرياضيات)، الصف الدراسي (الرابع، الخامس، السادس) والسنة الدراسية التي طبق فيها؟" للإجابة عن هذا السؤال استخدمت الاختبارات اللابارامترية كاختبار كروسكال - واليز واختبار (مان وتيني) والجدولان التاليان (٣)، (٤)، يبينان ذلك.

الجدول (٣)

نتائج اختبار كروسكال - واليز Kruskal-Wallis Test ومستوى الدلالة باختلاف المادة الدراسية والصف الدراسي

المتغير	النوع	العدد (ن)	متوسط الرتب	مربع كاي	درجات الحرية	مستوى الدلالة
المادة الدراسية	لغة عربية	١٢	١٨,٥	٠,٠٠	٣,٠٠	١,٠٠
	لغة إنجليزية	١٢	١٨,٥			
	علوم	٦	١٨,٥			
	رياضيات	٦	١٨,٥			
الصف الدراسي	رابع	١٢	١٨,٥	٠,٠٠	٢,٠٠	١,٠٠
	خامس	١٢	١٨,٥			
	سادس	١٢	١٨,٥			

من جدول (٣) يتضح أنه:

- لا توجد فروق ذات دلالة إحصائية في مدى توافق الاختبارات الوطنية مع معايير تصميم الاختبار باختلاف المادة الدراسية (العربي، الإنجليزي، العلوم، الرياضيات)؛ حيث كان مربع كاي صغيراً جداً (٠,٠٠) وبمستوى دلالة = ١ (أكبر من ٠,٠٥).
- لا توجد فروق ذات دلالة إحصائية في مدى توافق الاختبارات الوطنية مع معايير تصميم الاختبار باختلاف الصف الدراسي (الرابع، الخامس،

السادس)؛ حيث كان مربع كاي صغيراً جداً (٠,٠٠) وبمستوى دلالة = ١ (أكبر من ٠,٠٥).

الجدول (٤)

نتائج اختبار مان ويتني Mann-Whitney Test ومستوى دلالة الفروق لكل مادة دراسية ولكل صف دراسي بحسب السنة الدراسية

السنة	العدد (ن)	متوسط الرتب	مجموع الرتب	مان ويتني ويلكيسون (U)	Z	مستوى الدلالة
٢٠١٢-٢٠١١	١٨	١٨,٥٠	٣٣٣	١٦٢,٠٠	٣٣٣,٠٠	١,٠٠٠

من جدول (٤) نجد أنه لا توجد فروق ذات دلالة إحصائية في مدى توافق الاختبارات الوطنية مع معايير تصميم الاختبار باختلاف السنة (٢٠١٢-٢٠١١ و ٢٠١٣-٢٠١٤)؛ حيث بلغت قيمة اختبار مان ويتني ($U = 212.50$)، وبمستوى دلالة = ١,٠٠ أكبر من ٠,٠٥.

ويدل عدم وجود فروق ذات دلالة إحصائية بين متغيرات الدراسة على تماثل خصائص واضعي الاختبارات الوطنية في اكتساب المعارف والخبرات في بناء أسئلة الاختبار ومفرداته، ومن ثم لا تختلف خبراتهم ومؤهلاتهم ونموهم المهني باختلاف المادة الدراسية والصفوف الدراسية وسنة تطبيق الاختبارات الوطنية؛ حيث يصبح الاختبار الوسيلة الرئيسة لتحديد ما اكتسبه المتعلم، وقد أكدت دراسة (Campbell & Kyriakides, 2000) أنه على الرغم من أداء طلبة اليابان في التقييم الشامل لمادة الرياضيات كان عالياً، فإن أداءهم في بنود التقييم المرتبطة بالتطبيق والتحليل لم يكن عالياً؛ مما كان له الأثر في تطوير منهج الرياضيات وتقديم مزيد من الاهتمام على مستوى التطبيق والتحليل؛ حيث إن التركيز على نتائج الطلبة يساعد واضعي السياسة على تحديد نقاط القوة والضعف في النظام التعليمي، والعمل على تطويرها.

ملخص النتائج:

كشفت نتائج الدراسة عن الآتي:

– المستوى العام لكفاية إعداد الاختبارات الوطنية العامة في مادة اللغة العربية للصفوف الثلاثة المتقدمة في جميع المحاور كان في مستوى مرضٍ، ولكن توجد بنود فيها قصور، وحصلت على مستوى غير متحقق في الصفوف الثلاثة، وهي: في محور تعليمات الاختبار في البندين: " زمن الاختبار الكلي " و " الدرجة المعطاة لكل سؤال " ، وفي محور التنظيم البند "تناسب عدد الأسئلة مع زمن الاختبار الكلي". وفي محور "اتفاقها مع معايير المنهج" ، البند "يتوافر عدد قليل من الأسئلة يقيس المستويات الدنيا من التفكير"؛ فقد حصلت هذه البنود على مستوى ضعيف للصفوف الثلاثة، كما حصل البند "التنوع في الأسئلة المقالية والموضوعية بأنواعهما" على مستوى ضعيف للصفين الرابع والسادس.

– المستوى العام لكفاية إعداد الاختبارات الوطنية العامة، وللمحاور الثلاثة (تعليمات الاختبار، اللغة، والتنظيم) في مادة اللغة الإنجليزية للصفوف الثلاثة المتقدمة كان في مستوى مرضٍ، وفي مستوى متوسط في محور "اتفاقها مع معايير المنهج" ، ولكن توجد بنود فيها قصور؛ ففي محور "اتفاقها مع معايير المنهج" حصل بند "الملاءمة مع المنهج المدرسي للمادة" على مستوى ضعيف للصفين الرابع والخامس، وحصل البند "التنوع في الأسئلة بين المقالية والموضوعية بأنواعهما" على مستوى غير متحقق، وضعيف للصفين الرابع والخامس، وحصل البند "الالتزام بنود كتابة وصياغة الأسئلة الموضوعية وفق الشروط العلمية لكل منها" على مستوى ضعيف للصف الخامس. أما في محور "التنظيم" فحصل البنودان "تناسب عدد الأسئلة مع زمن الاختبار الكلي" و "تحديد مستوى معيار الأداء المطلوب في السؤال العام للفقرات" على مستوى غير متحقق لجميع الصفوف الثلاثة.

المستوى العام لكفاية إعداد الاختبارات الوطنية في مادة العلوم، وللمحاور الثلاثة (اتفاقها مع معايير المنهج، اللغة، والتنظيم)، للصفوف الثلاثة المتقدمة كان في مستوى مرضٍ وفي مستوى غير متحقق في محور تعليمات الاختبار. ولكن توجد بنود فيها قصور، ففي محور تعليمات الاختبار حصلت البنود "نوع الأسئلة المتضمنة في الاختبار"، "زمن الاختبار الكلي" الدرجة المعطاة لكل سؤال "على مستوى غير متحقق، وفي محور اتفاقها مع معايير المنهج حصل البنودان "الملاءمة مع المنهج المدرسي للمادة". و"تغطية عناصر المحتوى المعرفي، المهاري والحركي بشكل متوازن"، في الصفوف الثلاثة على مستوى غير مرضٍ، كذلك البند "الكثير من الأسئلة تركز على المستويات العقلية العليا، كالفهم والقدرة على تطبيق المعلومات وحل المشكلات" حصل على مستوى ضعيف في مادة العلوم للصف الرابع فقط، أما في محور التنظيم فالبنود التالية في مستوى غير متحقق للصفوف الثلاثة، وهي "تناسب عدد الأسئلة مع زمن الاختبار الكلي"، "تجميع الفقرات التي تقيس نفس المحتوى أو المهارة وعرضها معاً"، "تحديد مستوى معيار الأداء المطلوب في السؤال العام للفقرات".

المستوى العام لكفاية إعداد الاختبارات الوطنية العامة وللمحاور الثلاثة (اتفاقها مع معايير المنهج، اللغة، والتنظيم)، في مادة الرياضيات للصفوف الثلاثة المتقدمة كان في مستوى مرضٍ، وفي مستوى غير متحقق في محور (تعليمات الاختبار). ولكن توجد بنود فيها قصور وفي مستوى غير متحقق للصفوف الثلاثة ومنها على سبيل المثال: في محور تعليمات الاختبار، وهي: "عدد الأسئلة الكلية في الاختبار"، "زمن الاختبار الكلي" الدرجة المعطاة لكل سؤال. وفي محور "اتفاقها مع معايير المنهج" حصل البند "تغطية أوزان مجالات المادة كما تنص عليها المعايير" على مستوى غير متحقق، بينما حصل البند "يتوافق

عدد قليل من الأسئلة تقيس المستويات الدنيا من التفكير (مستوى التذكر) " على مستوى ضعيف. كما حصل البند " التنوع، منها: أ- المقالة ذات الاستجابة القصيرة (الموجزة) ب- الموضوعية بأنواعها " على مستوى ضعيف في مادة الرياضيات للصف السادس. أما محور " التنظيم " فقد كانت هناك بنود في مستوى غير متحقق للصفوف الثلاثة، وهي: "تناسب عدد الأسئلة مع زمن الاختبار الكلي"، "تجميع الفقرات التي تقيس نفس المحتوى أو المهارة وعرضها معاً"، "تحديد مستوى معيار الأداء المطلوب في السؤال العام للفقرات".

- الاختبارات الوطنية في مادة اللغة العربية كانت أقلها قصوراً في بنود الاختبار؛ حيث بلغ إجمالي البنود التي فيها قصور (٦) بنود من المجموع الكلي لعدد بنود الاختبار وهو (٣٣) بنداً، يليها مادة اللغة الإنجليزية؛ حيث بلغ إجمالي البنود التي فيها قصور (٧) بنود، يليها مادة العلوم حيث بلغ عدد البنود التي فيها قصور (٨) بنود، وآخرها مادة الرياضيات؛ حيث بلغ عدد البنود الضعيفة (٩) بنود.

- عدم وجود فروق ذات دلالة إحصائية عند مستوى (٠,٠٥) في الاختبارات الوطنية باختلاف المادة الدراسية (العربي، الإنجليزي، العلوم، الرياضيات)، والصف الدراسي (الرابع والخامس والسادس)، واختلاف السنة الدراسية المطبق فيها الاختبار.

- يلاحظ تشابه الاختبارات الوطنية في المواد الدراسية الأربع في معظم البنود التي فيها قصور في جميع المحاور، كما يلاحظ اتفاقها في محور " اللغة " من حيث سلامة بنوده من جوانب القصور في جميع المواد الدراسية، وهذا التشابه في جوانب القصور قد يشير إلى القصور في البرامج التدريبية التي تقدم لواجعي هذه الاختبارات في المواد الدراسية المختلفة، وإلى عدم معرفتهم بمعايير إعداد الاختبار الجيد والالتزام بها عند إعداد هذه الاختبارات.

التوصيات والمقترحات:

- من خلال عرض ومناقشة النتائج واستنتاجاتها توصي الدراسة بما يأتي:
 - الاهتمام بالأسئلة التي تقيس المستويات العقلية العليا من فهم وتطبيق وتحليل وتركيب وتقويم، وعدم التركيز على تقويم الحفظ والاستظهار، والتنوع في الأسئلة المقالية، والموضوعية، والعمل على تغطية مجالات المحتوى وأوزانها بشكل مناسب.
 - العمل على تناول تعليمات الاختبار بجميع الجوانب المغطاة فيه، ومنها: نوع الأسئلة، وعددها، وزمن الاختبار، ودرجة كل سؤال، ونموذج الإجابة، ووضوح التعليمات، حتى يمكن تقديم رؤية واضحة للطلبة عما يتضمنه الاختبار ويهدف إلى قياسه.
 - الاهتمام بالتنظيم في ورقة الاختبارات الوطنية، من حيث تناول تحديد مستوى معيار الأداء المطلوب وتناسب الأسئلة مع زمن الاختبار، وتجميع الفقرات التي تقيس المستوى نفسه.
 - اختيار العاملين في مجال إعداد الاختبارات الوطنية وفق معايير دقيقة، تبين مدى قدراتهم وخبراتهم في مجال بناء الاختبارات التحصيلية المقننة على المستوى الوطني.
 - تدريب العاملين في مجال التقويم التربوي على أسس بناء الاختبارات الوطنية المرتبطة بمخرجات التعلم وفق مواصفات الاختبار الجيد ومعاييره، والاستمرار في هذه الدورات وفقاً للمستجدات في مجال الاختبارات الوطنية.
 - مراعاة تحكيم الاختبارات الوطنية بعد الانتهاء منها من قبل متخصصين في مجال إعداد الاختبارات الوطنية، والمناهج الدراسية لضمان سلامتها ومناسبتها لتحقيق الهدف منها.
 - وضع مخطط لجدول مواصفات الاختبار، يربط الأهداف التعليمية للموضوعات الدراسية بمحتوى هذه الموضوعات، وارتباط مفردات

- أسئلة الاختبار بخلايا الجدول، وعلى ذلك يمكن التأكد من التوازن في تغطية أسئلة الاختبار، والتوازن في تغطية مستويات الأسئلة المختلفة.
- مراجعة الاختبارات المقننة ذات المرجعية المعيارية بصورة منتظمة؛ للحفاظ على تحديثها من حيث أولويات المحتوى، والعمل على المطابقة بين المناهج الدراسية والبنود التي يغطيها الاختبار، وتحليل إجابات الطلبة على كل مفردة من مفردات الاختبار ودرجاتهم الكلية؛ مما يمكن واضعي الاختبارات من التحقق من فاعلية هذه المفردات وتحسينها بصورة مستمرة.
- إجراء مزيد من الدراسات حول تحليل الاختبارات الوطنية للمراحل والمواد الدراسية الأخرى التي لم تشملها هذه الدراسة.

المراجع

أولاً - المراجع العربية:

- الرازحي، عبدالوارث. (٢٠٠٥). معايير تقويم كفاءة نظام إعداد الاختبارات التحصيلية بجامعة صنعاء. المؤتمر العلمي السابع عشر. مناهج التعليم والمستويات المعيارية. مصر. (١). ص ص ٤٢٢-٤٣٤.
- الطراونة، عيسى. (٢٠١٢). تقييم الاختبارات التحصيلية بمدارس محافظة الكرك وفق معايير الاختبار الجيد. مجلة رابطة التربية الحديثة. مصر. ٥ (١٥). ص ص ١٨٣-٢٢٨.
- سلمان، سلمان. (٢٠١١). الاختبار التحصيلي وكيف نعهده؟ وكيف نقدمه؟ دراسات تربوية. السودان. ١٢ (٢٤). ص ص ١٨٢-١٩١.
- المجلس الأعلى للتعليم. (٢٠٠٥). معايير المناهج الوطنية لدولة قطر، نشرة جديدة. الدوحة. قطر: هيئة التعليم.
- المجلس الأعلى للتعليم. (٢٠١٣). التعليم في مدارس دولة قطر. التقرير السنوي للعام الدراسي ٢٠١٢-٢٠١٣. الدوحة. قطر: هيئة التعليم.
- عبد العزيز، حمدي؛ البلوي، عبد الله. (٢٠١١). دراسة تحليلية للاختبارات التحصيلية لمنهج الرياضيات المطور بالصف الأول المتوسط بمدينة تبوك في ضوء أهداف منهج الرياضيات ومعايير التقويم بالملكة العربية السعودية. مجلة كلية التربية. جامعة طنطا. مصر. ٤٣. ص ص ٤٩-١٠٨.
- علام، صلاح الدين. (٢٠٠٧). القياس والتقويم التربوي في العملية التدريسية. عمان: دار المسيرة للنشر والتوزيع.
- نيتكو، أنتوني؛ بروكهارت، سوزان. (٢٠١٢). التقييم التربوي للطلبة. الطبعة الخامسة. ترجمة مكتب التربية العربي لدول الخليج. نيوجرسي: بيرسون للتعليم.
- وزارة التربية والتعليم. (٢٠١١). مساهمة فعالة في تطوير التعليم.

قسم التقويم والجودة التربوية. صحيفة صوت الأخدود. المملكة العربية السعودية. المنطقة الشرقية

ثانياً – المراجع الأجنبية:

- Abu-Hamour, B. (2013). Arabic spelling and curriculum based measurement. *The Australian Educational and Developmental Psychologist*, 30(02), 140-156.
- Adomou, D. S. (2011). *Teachers' perceptions of the effects of standards-based reform on curriculum instructional practices, and the quality of student teaching: Keeping the debate alive*. (Unpublished doctoral dissertation). California State University, Long Beach, CA.
- Bishop, John H. (1997). The effect of national standards and curriculum-based exams on achievement. *American Economic Review*, 87 (2), 260-264.
- Budak, A. (2008). *The impact of a standards-based curriculum and teaching practices on students' mathematic achievement* (Unpublished doctoral dissertation). Indiana University, Bloomington, Indiana.
- Campbell, R. J. & Kyriakides, L. (2000). The national curriculum and standards in primary schools: A comparative perspective. *Comparative Education*, 36(4), 383-395.
- Education Review Office. (2010). *Working with the national standards within the New Zealand curriculum*. Retrieved from
- Gonzalez, G. C., Le, V., Broer, M., Mariano, L. T., Froemel, E., Goldman, C. A. & DaVanzo, J. (2009). *Lessons learned from developing and implementing the Qatar student assessment system*. RAND-Qatar Policy Institute, Doha, Qatar. Retrieved from < http://www.rand.org/pubs/research_briefs/RB9421-1.html > .
- Hall, L. D. (2009). *Using mathematics curriculum based measurement as an indicator of student performance on state standards* (Unpublished doctoral dissertation). Texas A & M University, College Station, Texas.

- Kinnear, P. R., & Gray C. D. (2008). *SPSS 16 made simple*. New York: Psychology Press.
- Metty, J. M. (2010). *The effects of a standards-based curriculum on science teachers' instructional decisions*. (Unpublished doctoral dissertation). Texas A & M University, College Station, Texas.
- Pemberton, J. B., Rademacher, J. A., Tyler-Wood, T., & Cereijo, M. V. P. (2006). Aligning assessments with state curriculum standards and teaching strategies. *Intervention in School & Clinic*, 41, 283-289.
- Shapiro, E. S., Keller, M.A., Lutz, J.G., Santoro, L. E., & Hintz, J. M. (2006). Curriculum based measures and performance on state assessment and standardized tests reading and math performance in Pennsylvania. *Journal of psychology educational Assessment*, 24(1), 19-35.
- Smith. E. L. (2001). An analysis of teacher made tests. *Contemporary Educational Psychology*, 16(3), 279-286.

