الاختبارات الموضوعية لبعض مقررات مرحلة البكالوريوس بكلية التربية لإعداد معلمات الابتدائي دراسة تحليلية

سناء أحمد المنصور

قسم التربية وعلم النفس، كلية التربية، جامعة الأميرة نوره بنت عبدالرحمن
الرياض، المملكة العربية السعودية

الملخص:

هدفت هذه الدراسة إلى تحليل أسئلة الاختبارات الموضوعية لمرحلة البكالوريوس بكلية التربية لإعداد معلمات الابتدائي بالرياض واستهدفت ثلاثة محاور رئيسة وهي أولاً: درجة تمثيل الاختبارات لمحتوى المقررات في ضوء الأهداف التدريسية الموثقة في الخطط الدراسية وفق آراء أصحاب الاختصاص، ثانياً: مدى توفر شروط كتابة أسئلة الاختبارات الموضوعية المستخدمة وفق معايير مراجعة الاختبارات الموضوعية في أدبيات التقويم التربوي وبشكل خاص المعايير المعدة من قبل وزارة التربية والتعليم في المملكة العربية السعودية، ثالثاً: التعرف على الفروق ذات الدلالة الإحصائية –إن وجدت – لنتائج التحليل بين الأقسام التعليمية في الكلية. وقد تكونت عينة الدراسة من ٣٢ اختبار موضوعي من المجتمع الأصلي ١٢٧ تمثل أقسام الكلية والمراحل الدراسية الأربعة. استخدمت الباحثة ثلاثة استمارات لرصد عملية التحليل شارك فيها ٢٩ من المحللين من أعضاء هيئة التدريس من مختلف التخصصات، بحيث تم تحليل كل اختبار على حدة من قبل الباحثة وباحثين آخرين في مجال تخصص مادة الاختبار. ومن أبرز نتائج الدراسة تواضع مستوى تمثيل الاختبارات الموضوعية للأهداف السلوكية وافتقاره للشمول حيث بلغ مستوى التمثيل درجة "متوسط" في جميع أقسام الكلية، كما أظهرت النتائج فروق ذات دلالة إحصائية في توفر شروط بناء أسئلة الاختيار من متعدد بين اختبارات قسمي العلوم والرياضيات، ورياض الأطفال لصالح قسم العلوم والرياضيات.

المقدمة:

يحظى التقويم التربوي باهتمام بالغ لدى العاملين في المجالات التربوية ويشكل ركناً أساسياً للتعرف على مدى تحقق الأهداف، كما يلعب دوراً رئيساً في تطوير

المنظومة التربوية. وهو أحد أهم مجالات العلوم التربوية التطبيقية، ظهر في السبعينات كتخصص مستقل وكان له دوراً هاماً في إجراءات إصلاح التعليم وتخطيط المشروعات الهامة، وأصبحت البرامج التعليمية والتدريبية لا تخلو من برنامجاً تقويمياً مصاحباً لها (شعلة، ٢٠٠٠).

وينوه يوسف والرافعي (٢٠٠٥) بأهمية عدم الخلط بين التقويم التربوي والتقويم التعليمي فالأول يشمل كافة الميادين والبرامج والمؤسسات، والثاني يقتصر على المؤسسات التعليمية بهدف تقييم التحصيل الدراسي. ويصنف جوبا ولنكولن Guba & Lincoln (١٩٩٢) وماكملين McMillan (١٩٩٧) تقييم التحصيل الدراسي إلى تقييم ختامي غرضه رصد أداء الطالب في الوحدة الدراسية بعد انتهاء عملية التعلم، وتقييم تكويني غرضه متابعة وتطوير عملية التدريس والتعلُم أثناء الوحدة الدراسية. ويتفق معهما أبولبدة (٢٠٠٥) وتستخدم الاختبارات التحصيلية كأداة هامة للتعرف على مدى تحقق الأهداف التربوية، واستخدامها في قياس تحصيل الطلاب في مؤسسات التعليم العام والتعليم العالي يحكمها شروط وضوابط عند الإعداد تختلف حسب نوعية الاختبار. وتشتمل الاختبارات التحصيلية على الاختبارات الشفوية، والاختبارات المقالية واختبارات الأداء والاختبارات الموضوعية وتتنوع الأخيرة إلى اختبارات الصواب والخطأ، و الاختيار من متعدد والتكميلي و المزاوجة، والترتيب. ويكثر استخدام الاختبارات الموضوعية عندما تكون أعداد الطلاب كبيرة وذلك لسهولة تصحيحها خاصة عندما يكون التصحيح آلياً.

في عام (٢٠٠٢) قامت كليات البنات (جامعة الأميرة نورة بنت عبد الرحمن حالياً) في المملكة العربية السعودية بفتح باب الانتساب لقبول الطالبات في قسمي الدراسات الاسلامية واللغة العربية، وتم إستخدام الاختبارات الموضوعية لقياس تحصيل (٤٤٢٦٩) طالبة منتسبة في كليات التربية للبنات بمختلف مناطق المملكة (السكران، ٢٠٠٢) وانتشرت كليات البنات في أنحاء المملكة حيث بلغ عددها (١٠٢) كلية يدرس بها (٢٢٣٧٨٨) طالبة (عمادة القبول والتسجيل، ١٤٢٧هـ) وتجدر الإشارة بأن كليات

البنات المنتشرة في مدن المملكة المختلفة قد ضمت إلى الجامعات التي في تلك المدن باستثناء كليات منطقة الرياض التي سميت جامعة الرياض للبنات ثم لاحقاً جامعة الأميرة نورة بنت عبدالرحمن. وتعتبر كلية التربية لإعداد معلمات الابتدائي بالرياض إحدى هذه الكليات، أنشئت عام ١٤١١هـ (١٩٩٠)، و كان عدد طالباتها حينئذ (٤٩٠) طالبة وعدد أعضاء هيئة التدريس (٧٦). وفي العام الجامعي ١٤٢٧- ١٤٢٨هـ بلغ عدد طالباتها (٣٦٦٣) طالبة وارتفع عدد أعضاء هيئة التدريس ليبلغ (١٥٠) عضو (إحصائية الفصل الدراسي الأول، ١٤٢٨هـ).

وتحتوي الكلية على أربعة أقسام تعليمية بعد إلغاء قسم الاقتصاد المنزلي والتربية الفنية وهي: القرآن الكريم والدراسات الإسلامية، واللغة العربية والعلوم الاجتماعية، والعلوم والرياضيات، ورياض الأطفال. ونظراً لأهمية نوعية مخرجات الكلية ودورها المستقبلي في تعليم نشئ المرحلة الابتدائية، فإن أساليب تقويم طالبات الكلية تزداد أهمية وذلك بهدف تخريج كفاءات متميزة إلى الميدان.

مشكلة الدراسة:

أثناء تصحيح أوراق إجابة الاختبارات الفصلية بكلية التربية لإعداد معلمات الابتدائي يطلب من عضو هيئة التدريس تسليم أوراق الإجابة المصححة للجنة الرصد - التي تعني باستخراج نتائج الامتحانات بالكلية - في حدود اثنان وسبعون ساعة. ونظراً للتزايد المطرد لطالبات الكلية أصبحت عملية التصحيح وتسليم أوراق الإجابة خلال الفترة الزمنية المحددة تشكل عبئاً ثقيلاً على كاهل عضو هيئة التدريس إضافة إلى الأعمال المناطة بهم خلال فترة الاختبارات. وفي ضوء الضغوط والإجهاد التي يتعرض لها عضو هيئة التدريس خلال فترة الامتحانات وظهورتقنية المصحح الآلي وتوفرها بالكلية إزاء فتح باب القبول للمنتسبات اشتد إقبال أعضاء هيئة التدريس لاستخدام هذه التقنية الحديثة، وبالتالي الاتجاه لوضع اختبارات موضوعية لقياس تحصيل الطالبات المنتظمات والمنتسبات سوياً. وفي اختبارات الفصل الدراسي الأول للعام الجامعي ١٤٢٧/١٤٢٨هـ بلغت نسبة الاختبارات الموضوعية المصححة آلياً ثمانية

وثمانون في المائة ٨٨٪ من مجمل اختبارات الكلية. ويوضح جدول (١) الأقسام التعليمية بالكلية وعدد مواد الفرق الدراسية لكل قسم، وعدد المواد التي استخدمت فيها الاختبارات الموضوعية والاختبارات المقالية ونسبهم المئوية. وتصدّر قسم اللغة العربية والعلوم الاجتماعية جميع الأقسام في استخدامه للاختبارات الموضوعية من نوع الصواب والخطأ، والاختيار من متعدد حيث بلغت نسبة استخدامها ٩٧٪ يليه قسم القرآن الكريم والدراسات الإسلامية بنسبة ٩٥٪ ثم قسم رياض الأطفال بنسبة ٩٤٪ بينما حصل قسم العلوم والرياضيات على النسبة الأدنى ٦٦٪.

يوضح جدول (١) عدد ونسبة الاختبارات الموضوعية والمقالية بجميع المقررات الدراسية في الأقسام التعليمية بالكلية.

جدول (١)

العدد الكلي والنسب المئوية للاختبارات الموضوعية والمقالية بجميع المواد الدراسية في الأقسام التعليمية بالكلية

نوع الاختبار / القسم	المجموع		الاختبارات الموضوعية		الاختبارات المقالية	
	عدد المواد	النسبة المئوية	عدد المواد	النسبة المئوية	عدد المواد	النسبة المئوية
القرآن الكريم والدراسات الإسلامية	٣٧	١٠٠	٣٥	٩٥	٢	٥
اللغة العربية والعلوم الاجتماعية	٣٨	١٠٠	٣٧	٩٧	١	٣
العلوم والرياضيات	٣٢	١٠٠	٢١	٦٦	١١	٣٤
رياض الأطفال	٣٦	١٠٠	٣٤	٩٤	٢	٦
المجموع	١٤٣	١٠٠	١٢٧	٨٨	١٦	١٢

وتختصر مشكلة الدراسة في الأسئلة التالية:

١. ما مستوى تمثيل أسئلة الاختبارات الموضوعية لمحتوى المقررات في ضوء الأهداف التدريسية الموثقة في الخطط الدراسية وفق آراء أصحاب الاختصاص ؟
٢. ما مدى الالتزام بشروط بناء فقرات الصواب والخطأ عند إعداد الاختبارات وفق معايير تقييم أسئلة الصواب والخطأ في أدبيات التقويم التربوي وبشكل خاص المعايير المعدة من قبل وزارة التربية والتعليم؟
٣. ما مدى الالتزام بشروط بناء فقرات الاختيار من متعدد عند إعداد الاختبارات وفق معايير تقييم أسئلة الاختيار من متعدد في أدبيات التقويم التربوي وبشكل خاص المعايير المعدة من قبل وزارة التربية والتعليم ؟
٤. هل توجد فروق ذات دلالة إحصائية بين الأقسام التعليمية عند تحليل أسئلة الاختبارات الموضوعية؟

ومن هذا المنطلق ارتأت الباحثة أهمية إجراء دراسة تحليلية للاختبارات الموضوعية لمواد الفصل الدراسي الأول للعام الجامعي ١٤٢٧هـ/١٤٢٨هـ لمرحلة البكالوريوس بكلية التربية لإعداد معلمات الابتدائي بالرياض وتسليط الضوء على مدى جودة إعدادها والتعرف على مواطن الخلل – إن وجدت - في محاولة لإصلاح طرق قياس تحصيل الطالبات وإيجاد نوعية ذات كفاءة عالية من الخريجات.

أهداف الدراسة:

تهدف الدراسة للتعرف على:.

١. تمثيل الاختبارات الموضوعية لمحتوى المقررات في ضوء الأهداف التدريسية الموثقة في الخطط الدراسية.
٢. الالتزام بشروط بناء فقرات الصواب والخطأ عند بناء الاختبارات الفصلية.
٣. الالتزام بشروط بناء فقرات الاختيار من متعدد عند بناء الاختبارات الفصلية.
٤. التعرف على الفروق ذات الدلالة الإحصائية - إن وجدت – بين تحليل أسئلة الاختبارات الموضوعية في مواد البكالوريوس تبعاً لمتغيرات الأقسام التعليمية.

أهمية الدراسة:

تكمن أهمية الدراسة فيما يلي:

١. تشخيص نقاط القوة والضعف في إعداد الاختبارات الفصلية بكلية التربية لإعداد معلمات الابتدائي.
٢. إحاطة أعضاء هيئة التدريس عن مدى جودة الاختبارات الموضوعية التي يقومون بإعدادها وإتاحة الفرصة لهم بالتعرف على أهم مواطن الضعف و كيفية علاجها.
٣. تقديم معلومات وافية عن واقع إعداد الاختبارات الموضوعية النهائية لمتخذي القرار في الكلية لاتخاذ الإجراءات اللازمة بهدف تطويرها.

حدود الدراسة:

تقتصر حدود الدراسة الحالية على:

١. الاختبارات الموضوعية لمرحلة البكالوريوس بكلية التربية لإعداد معلمات الابتدائي بالرياض.
٢. الفصل الدراسي الأول من العام الجامعي ١٤٢٧-١٤٢٨هـ.

مصطلحات الدراسة:

تتعرض الدراسة للمصطلحات التالية:

الاختبارات الموضوعية:

أسئلة لقياس التحصيل الدراسي تكون الإجابة عليها إما بالتعرف والاختيار أو بكتابة كلمة/كلمات لا مرادف لها أو عدد أو رمز معين. اكتسبت اسمها من طريقة التصحيح لخلوها من ذاتية وحكم المصحح (أبولبدة، ٢٠٠٥).

الاختبار التحصيلي:

"أداة من أدوات القياس التي يستخدمها المعلم بطريقة منظمة لتحديد تحصيل الطلاب في مادة دراسية معينة" (القرني، ١٤٢٠هـ، ص٥).

جذر أو مقدمة السؤال:

الجزء الأول من سؤال الاختيار من متعدد ويأتي على شكل عبارة ناقصة أو سؤال كامل.

البدائل:

الجزء الثاني من سؤال الاختيار من متعدد، ويتكون من ٢- ٥ عبارات أو كلمات أو أرقام أو رموز إحداها تمثل الإجابة الصحيحة.

المموهات:

هي بدائل قابلة للتصديق في سؤال الاختيار من متعدد لا تمثل الإجابة الصحيحة.

الإطار النظري:

يحتوي الإطار النظري على تعريف مفهوم الاختبار وخطوات إعداده وعلاقته بصدق المحتوى وصدق الاختبار ومعايير تقييمه، والاختبارات الموضوعية مع التركيز على اختباري الصواب والخطأ والاختيار من متعدد لارتباطهما الوثيق بموضوع الدراسة وما يتعلق بهما من تعريف ومزايا وعيوب ومعايير تقييم.

مفهوم الاختبار:

يعرف هوجان Hogan الاختبار بأنه إجراء مقنن يقدم معلومات عن عينة من السلوك أو المعرفة بشكل كمي (٢٠٠٧) كما يعرفه نانالي Nunnaly بأنه حالة مقننة يخرج منها الفرد بدرجة (١٩٧٢) ويعرفه قاموس المصطلحات لتكنولوجيا التربية "أنه أداة لتقدير مستوى أداء التلاميذ في أحد الموضوعات" (فتح اللّه، ٢٠٠٠، ص٢٣٨) ويعرفه منسي (٢٠٠٢) بأنه "أداة قياس يتم إعدادها بخطوات منظمة للخروج بخصائص مرغوبة في هذا الاختبار بحيث يتم الحصول بواسطتها على بيانات كمية تساعد في إنجاح الهدف الذي استخدم من أجله" (ص٨٢) ويرى الكبيسي (٢٠٠٧) أن الاختبار هو "أداة قياس يتم إعدادها وفق طريقة منظمة يتكون من عدة خطوات تتضمن مجموعة من الإجراءات التي تخضع لشروط وقواعد محددة بغرض تحديد درجة امتلاك الفرد لسمة

أو قدرة معينة من خلال إجاباته على عينة من الأسئلة التي تمثل السمة المرغوب قياسه مثل تحصيل المتعلم لمعلومات ومهارات في مادة دراسية تعلمها مسبقاً" (ص١٠٧) وتعرفه إحسان شعراوي بأنه "مجموعة من المثيرات تقدم للمفحوص لاستخراج إجابات يعطى بناءً عليها تقديراً عددياً" (خضر، ١٩٨٧، ص ١٧) ويحلل ملحم (٢٠٠٥) تعريف براون Brown للاختبار الذي ينص بأنه "إجراء منظم لقياس سمة ما من خلال عينة من السلوك" (٤٧ص) ويفيد أن هذا التعريف يشير إلى عدد من الأمور وهي الإجراء المنظم للاختبار والسلوك والعينة والتقدير والسمة الخاصة بالاختبار وما يقيسه وهل إجراءاته تقيس القدرة المراد قياسها.

خطوات بناء الاختبار التحصيلي وعلاقته بصدق المحتوى:

يتفق أخصائيي التقويم التربوي على أهمية التخطيط للاختبارات التحصيلية جميعها بما فيها الاختبارات النهائية أو الفصلية التي يعدها المعلمون لغرض إصدار القرارات الإدارية كالنجاح والرسوب ووضع تقديرات الطلاب. ويؤكد الدوسري (٢٠٠١) على أهمية إتباع أساليب بناء الاختبارات لضمان صدق المحتوى. ويوضح الزيود وعليان (٢٠٠٥) خطوات بناء الاختبار إلى ثمان خطوات وهي "تحديد الأهداف، وتحليل المحتوى، وإعداد جدول المواصفات، وتنظيم المخطط التحليلي، وتحديد شكل الاختبار وموضوعيته، وكتابة فقرات الاختبار، وكتابة تعليمات الاختبار، وبيان طريقة التصحيح وتسجيل النتائج" (ص٩٠) ويتفق معهما هاشم (٢٠٠٦) و تنوه الباحثة إلى استخدام عناوين خطوات إعداد الاختبارات التحصيلية للزيود وعليان كإطار لتنظيم هذه الخطوات لوضوحها وسهولة تسلسلها. يتصدر قائمة تخطيط الاختبار أولاً: تحديد أهداف الاختبار وكتابتها بصورة سلوكية. ويعرف المركز العربي للبحوث التربوية لدول الخليج العربي الهدف السلوكي بأنه "التغير المرغوب المتوقع حدوثه في سلوك المتعلم والذي يمكن تقويمه بعد مرور المتعلم بخبرة تعليمية معينة" (الجلبي، ٢٠٠٥، ص٣٩) ويعلق العبيدي والجبوري (١٩٨١) على أهمية صياغة الهدف السلوكي بحيث يكون له بعدان: بعد سلوكي وبعد محتوى. وهناك شروط ومواصفات يجب أن

تؤخذ بالاعتبار عند صياغة الهدف السلوكي. ويرى عريفج و مصلح (١٩٩٩) أن الهدف السلوكي يتسم بدلالته على نتاج التعلم والقابلية للقياس والتحقيق في فترة زمنية محددة وصياغته بصورة مباشرة غير مركبة وشموله لمجالات الأهداف العقلية والوجدانية والحركية و توافقه مع الفلسفة التربوية واشتماله على سلوك ومحتوى ومناسبته لقدرات الطلاب وحاجاتهم. ثانياً: تحديد محتوى المادة العلمية، ويوضح الدوسري (٢٠٠١) "أهمية ربط المجال أو المحتوى الذي يغطيه من المعارف أو الموضوعات التحصيلية بالقدرات الذهنية اللازمة للتعامل مع تلك المعارف والموضوعات" (ص ١٦٤) ثالثاً: إعداد جدول المواصفات، حيث يهدف إلى تحديد مجالات الاختبار وربط أهدافه بالمحتوى، ويقترح الصراف (٢٠٠٢) أن يكون إعداد الجدول قبل بدء العملية التعليمية لما في ذلك من فائدة تنعكس على أداء المعلم في إعداد الخطة الدراسية وتوفير المناخ الملائم للتعلم. ويتسم جدول المواصفات ببعدين أفقي ورأسي، يوضح الأول المواضيع التي يغطيها الاختبار، بينما يوضح الثاني العمليات الذهنية المرتبطة بالموضوعات، ويتكون من تقاطع البعدين الرأسي والأفقي خلايا تستخدم لوضع اوزان اعتبارية أو نسب مئوية تخص كل موضوع وعملية ذهنية (الدوسري، ٢٠٠١) وبمعادلة رياضية يتم تحويل النسب المئوية إلى فقرات تعرض في جدول آخر مماثل. ويعلق على أهميته أبو لبدة (٢٠٠٥) بأنه يؤمن صدق الاختبار من خلال تمثيل أجزاء المادة و توزيع الأسئلة عليها، وتوظيف مستويات العمليات الذهنية وإعطاء جزئيات المادة وزنها الذي تستحقه في ضوء أهميته والفترة الزمنية المتاحة للإجابة عليه. ويتفق معه أبو زينه (١٩٩٢) ويتضح أن هذه الخطوة تتصل اتصالاً مباشراً بدرجة تمثيل الاختبار لأهداف المقرر الدراسي. رابعاً: تنظيم المخطط التحليلي، ويصنفه البعض كجزء من الخطوة الثالثة وهو جدول ذو بعدين: بعد الأهداف السلوكية وبعد المحتوى، يهدف إلى ربط أجزاء المحتوى بالصيغ السلوكية التي تناسبها ثم عند كتابة الاختبار يتم تحويل الأهداف السلوكية إلى صيغ أسئلة (الزيود وعليان، ٢٠٠٥) خامساً: تحديد شكل الاختبار وموضوعيته، يرى عودة (٢٠٠٢) أهمية اختيار الأسلوب الملائم لقياس الأهداف السلوكية. و يشير

الصراف (٢٠٠٢) أن هنالك عوامل تؤثر في شكل الاختبار وهي: الغرض منه، والوقت المخصص لإعداده وتصحيحه، وعدد الطلبة المفحوصين، وتوفر الموارد المالية لطباعته وتصويره، ومهارة المعلم في إعداد فقراته المختلفة. ويشير الزيود وعليان (٢٠٠٥) ويتفق معهما الظاهر وتمرجيان و عبد الهادي (١٩٩١) أن طبيعة المحتوى المراد قياسه تؤثر في شكل الاختبار، وعلى الرغم من إمكانية استخدام أي نوع من الاختبارات لأية مادة إلا أن هناك اختبارات تكون أكثر ملائمة من غيرها لقياس التحصيل في بعض المواد الدراسية كاستخدام الاختبار التكميلي في مادة التاريخ، واختبار المزاوجة في مادة الجغرافية، و أن مستوى الدقة في التعبير عن واقع التحصيل يؤثر في شكل الاختبار ومدى موضوعيته، فإذا استخدم اختبار الصواب والخطأ يكون مستوى الدقة أقل من اختبار الاختيار من متعدد حيث تكون نسبة التخمين في الاختبار الأول ٥٠٪ بينما تكون ٢٥٪ عند وجود أربعة بدائل. ويضيف الزيود وعليان (٢٠٠٥) أن عدد فقرات الاختبار تتأثر بنوعية الاختبارات المستخدمة والفترة الزمنية المتاحة للاجابة وعمر المفحوص ومستوى القدرة لديه ومستوى العملية الذهنية المطلوبة والمعالجات الحسابية التي تتطلبها إن وجدت. سادساً: كتابة فقرات الاختبار، في ضوء الخطوات السابقة يتم تحويل الأهداف السلوكية إلى فقرات أو أسئلة اختبار ثم يقوم المعلم بتجميعها وترتيبها. ويصنف باشموس و خيري و مهني (١٩٨٥) يؤيدهم عبد السلام وطاهر ومهني (١٩٩٢) ويتفق معهم هاشم (٢٠٠٦) ترتيب فقرات الاختبار تبعاً للصعوبة، حيث تتدرج من السهل للصعب ويؤخذ على هذه الطريقة بأنها قد تشعر المفحوص بالإحباط والتوقف عن الإجابة عند تعثره في الإجابة على فقرة ظناً بأنه لا يستطيع الإجابة على الفقرات التالية، كما يؤخذ عليها عدم انتظام فقراتها حسب تنظيم الموضوعات الدراسية، أو الترتيب الدوري بالنسبة للصعوبة حيث يبدأ بفقرات تتدرج من السهولة للصعوبة ويعود مرة أخرى بنفس الطريقة والتالي تتجمع الفقرات الصعبة في وسط الاختبار، ويؤخذ على هذه الطريقة الانتقال المفاجئ من موضوع لآخر مما قد يشتت انتباه المفحوص والتأثير على صحة إجابته. ويشير ملحم (٢٠٠٥) أن تجميع فقرات الاختبار يشتمل على

اختيار ها أولاً ثم ترتيبها، واختيار الفقرات يجب أن يأخذ بالاعتبار وحدة الموضوع ووحدة الهدف ونوع السؤال ومستوى الصعوبة، كمايشير ملحم ويتفق معه عودة (٢٠٠٢) أن ترتيب فقرات الاختبار يتم حسب النوع أو المحتوى أو مستوى الصعوبة أو حسب العملية الذهنية التي تقيسها الفقرة. سابعاً: كتابة تعليمات الاختبار، وترى الجلبي (٢٠٠٥) أهمية شرح فكرة الاختبار للمفحوصين شرحاً جيداً وإبراز التعليمات العامة بخط أكبر واضح ووضع أمثلة قبل كل اختبار وصياغة تعليمات متسقة لأجزاء الاختبار المختلفة إن أمكن، وتحديد طريقة الإجابة ومكانها. أما عريفج ومصلح (١٩٩٩) فهم يشيران إلى أهمية اهتمام المعلم بطباعة الاختبار بحيث يخلو من الأخطاء الإملائية والمطبعية. ثامناً: بيان طريقة تصحيح الاختبار وتسجيل النتائج. ويقترح الدوسري (٢٠٠١) قراءة عينة من الأوراق للتعرف على مدى جودة وكفاية الإجابة وأيضاً لتكوين فكرة عامة عن مستوياتها، كما يعتبر مقارنة نموذج الإجابة النموذجي بعيّنة من أوراق الإجابة ضرورياً للتأكد من تطابقهما والتعرف على درجة التفاوت بين الإجابة النموذجية وإجابة المفحوصين ويتعلق بهذا الجانب توزيع درجات الاختبار فربما يتعين عمل تعديلات على توزيعها وأوزانها إن تطلب الأمر ذلك.

وفي هذا الشأن يصدر مكتب دعم التدريس بجامعة كورنيل (Cornell University) وثيقة بناء الاختبارات الموضوعية، ويصنف خطوات إعداد الاختبارات التحصيلية بثلاث مراحل وهي التخطيط للاختبار وتشتمل على اختيار الاختبار المناسب لقياس مخرجات التعلم وفقاً لما تم وضعه في جدول المواصفات، ثم إعداده، ثم تحليله وتعديله (Devine & Yaghlian، n.d.) .

وتجدر الإشارة أن الأدبيات التي تعالج مواضيع الاختبارات تؤكد أن اتباع خطوات إعداد الاختبار لها تأثيراً هاماً على جودته وما يفترض أن يتسم به من صدق وثبات وموضوعية.

صدق الاختبار:

الصدق: ويعرفه منسي (٢٠٠٢) أنه "الدرجة التي يقيس فيها ما أعد لقياسه أي يحقق الأهداف التي من أجلها وضع الاختبار" (ص٩٢) ويذكر الدوسري (٢٠٠١) أن مفهوم الصدق أصبح موحداً في الأدبيات الحديثة ولكنه يصنف لتبسيطه وهو بينة المحتوى أو صدق المحتوى وبينة المحك وبينة السمة، ويعرف صدق المحتوى بأنه "مدى تمثيل عناصر الاختبار وملائمتها للمحتوى الذي يقيسه" ويضيف "يتم توفير البينة التي تدل على صدق المحتوى عن طريق تقييم المختصين لعناصر الاختبار وعلاقاتها بمكونات المحتوى وأهم سماته ومدى تمثيل العناصر بشكل شامل لها" ص٤٩. ويفسر الصراف (٢٠٠٢) بأن الحكم على صدق المحتوى ليس تعبير رقمي وبأنه يتم عبر تحديد الجوانب المعرفية التي سيمثلها الاختبار ثم فحص كل سؤال من أسئلة الاختبار ليتم التعرف على مايقيسه و بأن ذلك يتحقق من خلال مساعدة متخصصين في المناهج وخبراء في محتواها. وينوه أبو علام (١٩٨٧) ويتفق معه أبولبدة (٢٠٠٥) بأن التعرف على صدق محتوى الاختبارات التحصيلية لا يحسب له معامل صدق بل تتم مقارنة بين أسئلة الاختبار والأهداف التدريسية، ويلخص أبو لبدة خصائص الصدق وأنواعه، حيث تعتمد خصائصه على غرض الاختبار والمجموعة التي سيطبق عليها وهو محدد ونسبي ومرتبط بصدق النتائج وثباتها، وأنواعه صدق المحتوى الذي يمثل مفردات المادة وأهدافها، والصدق الظاهري الذي يتضمن أسئلة ومهارات تتعلق بما يحاول الاختبار قياسه، وصدق المفهوم الذي يعني بسمة معينة يستدل عليها من أداء الفرد في الاختبار، والصدق التجريبي نوعين الصدق التنبؤي والصدق الملازم والذي يتم فيه تحديد الارتباط أو التوافق بين درجات الاختبار و المعدل الفصلي للطالب حيث يعتبر المعدل هو المحك والوسيلة التي نتعرف من خلالها على صدق الاختبار، والصدق العاملي الذي يركز على إيجاد عامل مشترك بين عدة اختبارات تقيس نفس الشئ. ويؤثر في صدق الاختبار عدة عوامل أهمها: غموض تعليمات الإجابة و سوء طباعة الاختبار والصياغة اللغوية لفقراته

وعدم الاهتمام بخاصية نوعية ونسبية صدق الاختبار و عدم توخي الدقة في إدارة تطبيق الاختبار وتصحيحه و تهاون المفحوص في الإجابة عليه (عريفج و مصلح، ١٩٩٩).

علاقة الصدق بالثبات: يعرف باشموس وآخرون (١٩٨٥) الاختبار الثابت أنه "هو الذي يعطي نفس النتائج لنفس المجموعة من الأفراد إذا ما طبق مرة أخرى، في نفس الظروف بشرط ألا يحدث تعلم أو تدريب في الفترات بين مرات إجراء الاختبار" (ص١٤٤) وتتحدد العلاقة بين الصدق والثبات في صدق النتائج فالثبات ضروري للحصول على نتائج صادقة بينما الصدق غير لازم للحصول على نتائج ثابتة (أبوعلام، ١٩٨٧) ويحسب ثبات الاختبار بعدة طرق: طريقة إعادة الاختبار وفيها يتم تطبيق الاختبار على مجموعة من الأفراد وتسجل نتائجه، ثم يطبق نفس الاختبار مرة أخرى على نفس المجموعة بعد فترة من الزمن تكفل عدم تذكره، ثم تحسب معامل الارتباط بين نتائج الاختبارين، وكلما ارتفعت معامل الارتباط لنتائج الاختبارين زاد ثبات الاختبار، ويؤخذ على هذه الطريقة الألفة التي يكونها أفراد العينة بعد إجراء الاختبار لأول مرة وتأثر النتائج عند إجرائه للمرة الثانية مما يؤدي إلى انخفاض معامل الارتباط بين نتائج الاختبارين (الدليم، عبدالجواد، عمران، ١٩٩٧) كما يؤخذ عليه صعوبة الإجراء واستغراقه وقتاً طويلاً وجهداً كبيراً من قبل الفاحص، وهناك طريقة التجزئة النصفية والتي يطبق فيها الاختبار مرة واحدة ويتم فيها تقسيمه إلى جزأين متكافئين و تصحيح كل جزء على حدة ثم يتم حساب معامل الارتباط بين ناتج الجزأين ولكون معامل الثبات المستخرجة تتعلق بنصف الاختبار يتم تصحيحها من خلال معادلة سبيرمان براون أو معادلة رولون أو معادلة كودر ريتشاردسون، ويؤثر على ثبات الاختبار عدة عوامل منها زمن الاختبار، ودقة صياغته وتباين مستوى الصعوبة والسهولة في فقراته ونسبة التخمين التي يسمح بها الاختبار و وحالة المفحوص الصحية والنفسية (عبد الهادي، ٢٠٠٢) وفي مركز البحوث للتعلم والتدريس في جامعة ميتشغان (CRLT) في وثيقة بعنوان الاختبارات والتصحيح يشير إلى مفهومين رئيسيين في تطوير الاختبارات

وهما الصدق والثبات، فالأول يتعلق بأداة القياس لما هو المفروض قياسه والآخر ثبات النتائج بعد التكرار (Erinkson،n.d.).

علاقة الصدق بالموضوعية: يرى بلوم و هاستنجس و مادونس(١٩٨٣) انها مرتبطة بالعدل في إعطاء الدرجات بين المفحوصين وبثبات الاختبار وصدقه، فعندما يكون أي جانب من عملية التصحيح غير دقيق فهو يؤثر سلبياً على صدق المحتوى، وعندما تتدخل ذاتية المصحح في عملية التصحيح يتأثر بالتالي صدق الاختبار وثباته.

معايير تقييم الاختبار:

يضع أبوحطب و عثمان وصادق (٢٠٠٣) نموذج لتقويم الاختبارات يتألف من عدة نقاط وهي:

١. عنوان الاختبار.
٢. اسم مؤلف الاختبار ومترجمه (إن وجد).
٣. الناشر.
٤. صور الاختبار والجماعات التي يمكن أن يطبق عليها.
٥. المعالم العملية(أية خصائص عملية تتصل بالأداء فيه أو تصحيحه).
٦. نوع الاختبار(اختبار قدرات أو شخصية مثلاً).
٧. تاريخ النشر.
٨. التكاليف كراسة التعليمات و الأسئلة، ورق الإجابة.
٩. الزمن المطلوب.
١٠. الغرض الذي من أجله نحكم على الاختبار.
١١. وصف الاختبار من حيث عناصره، وتصحيحه.
١٢. غرض المؤلف وأساس اختبار عناصر الاختبار عنده.
١٣. دقة التعليمات ووضوحها، والتدريب المطلوب لمن يقوم بعملية إعطاء الاختبار، ودرجة تقنينه.
١٤. الوظائف أو السمات العقلية الممثلة في كل درجة.

١٥. تعليقات وملاحظات عن تصميم الاختبار وبنائه (ورقة الأسئلة وطريقة الإجابة).
١٦. الصدق التنبؤي (المحك، وعدد ونوع الحالات التي استخدمت في حسابه. والنتيجة).
١٧. الصدق التلازمي (المحك، وعدد ونوع الحالات المستخدمة والنتيجة).
١٨. أية أدلة تجريبية أخرى تبين ما يقيسه الاختبار (صدق التكوين الفرضي).
١٩. تعليقات وملاحظات عن صدق الاختبار بالنسبة لأغراض معينة سيستعمل من أجلها الاختبار.
٢٠. ثبات الاختبار بحساب معامل الاستقرار (الطريقة التي اتبعت للوصول إليه وعدد ونوع الحالات والنتيجة).
٢١. ثبات الاختبار بحساب معامل التكافؤ (الطريقة التي اتبعت للوصول إليه وعدد ونوع الحالات والنتيجة).
٢٢. ثبات الاختبار بحساب معاملات الاتساق الداخلي أو اتساق الأسئلة (الطريقة التي اتبعت للوصول إليها، وعدد ونوع الحالات والنتيجة).
٢٣. موضوعية الاختبار من حيث التصحيح تفسير الدرجات.
٢٤. المعايير (نوعها، وعدد ونوع الحالات).
٢٥. تعليقات وملاحظات عن دقة فئات الاختبار وصحة المعايير للغرض الذي سيستعمل الاختبار من أجله.
٢٦. تعليقات وملاحظات من قاموا باستعراض للاختبار وتقويمه ونقده.
٢٧. تقويم عام للاختبار.
٢٨. المراجع. (ص٥٠٥).

كما يضع لورنس Lawrence (١٩٩٤) مجموعة من الأسئلة تتعلق بتقويم الاختبار من حيث تغطية الاختبار للمادة واستخدامه، مناسبة العينة التي استخدمت للتعرف على صدق الاختبار، وثبات الاختبار وصدق معاييره، وصدق محتواه و فقراته وتوافقها مع أهداف الاختبار، وإدارة الاختبار وتفسير تسجيل نتائجه. ويقترح جرونلند Gronlund (٢٠٠٦) عمل قائمة لتقييم عملية تخطيط الاختبار تحتوي على الأسئلة التالية:

١. هل هدف الاختبار واضح؟
٢. هل تم تحديد وتعريف مخرجات التعلم المطلوب؟
٣. هل تم كتابة مخرجات التعلم المطلوبة سلوكياً ؟
٤. هل تم إعداد مواصفات للاختبار تشير إلى طبيعة وتوزيع الفقرات التي سيحتويها؟
٥. هل الفقرات التي تم توصيفها تمثل عينة من مهام المجال التحصيلي؟
٦. هل نوع الفقرات مناسباً لقياس مخرجات التعلم؟
٧. هل صعوبة الفقرات مناسبة لمستوى للطلاب المفحوصين وطبيعة أداة القياس؟
٨. هل عدد الفقرات مناسبا للطلاب المفحوصين وللزمن المتاح ولتفسير النتائج؟
٩. هل تحتوي خطة الاختبار على خصائص تساهم في تعزيز صدق وثبات الدرجات؟
١٠. هل تم التخطيط لترتيب فقرات الاختبار وتعليمات الإجابة والتصحيح واستخدام النتائج؟ (ص٦٦).

الاختبارات الموضوعية:

هي أداة لقياس التحصيل الدراسي. سميت بذلك لعدم تدخل عامل الذاتية عند تصحيحها، ولإمكان أي فرد تصحيحها والخروج بنفس النتائج (أبولبدة،٢٠٠٥) تصنف إلى عدة أنواع: الاختيار من متعدد، والصواب والخطأ، والمزاوجة، والتكميل، والترتيب. والتلخيص التالي يستعرض أنواع الاختبارات الموضوعية ونبذة قصيرة عن مجالات استخدامها وأبرز ما يميزها ويؤخذ عليها، وسيتم التركيز على اختباري الصواب والخطأ والاختيار من متعدد لعلاقتهما المباشرة والوثيقة بموضوع الدراسة.

اختبار المزاوجة:

يستخدم للتعرف على قدرة المفحوص على التذكر والاستدعاء والربط، ومن مميزاته سهولة الإعداد والتصحيح وانخفاض نسبة التخمين، ويؤخذ عليه عدم قياس القدرات المهمة كالتعبير والبرهنة والقياس (هاشم، ٢٠٠٦).

اختبار الترتيب:

يستخدم لقياس قدرة المفحوصين على الترتيب والتنظيم وفق معايير محددة. ويقيس قدرة المفحوص على التفكير والربط ويستخدم في المواد الاجتماعية والحساب واللغات (الزيود و عليان، ٢٠٠٢) ويوخذ عليه قياس المستويات المعرفية البسيطة.

اختبار التكميل:

يظهر بصورتين: الإجابة القصيرة و ملئ الفراغ أو التكملة، يستخدم لقياس قدرة المفحوص على تذكر المصطلحات والتعريفات والتواريخ، يتميز بخلوه من التخمين وسهولة الإعداد والتصحيح وتمثيل محتوى المادة، يؤخذ عليه بأنه أحياناً يعطي إجابات متعددة من قبل المفحوصين إذا لم تتم صياغته بصورة دقيقة مما يشكل مدخلاً للذاتية في التصحيح (يوسف و الرافعي، ٢٠٠٥).

ويطلق مركز التقويم والاختبارات بجامعة آيوا (The University of Iowa)عدداً من الأسئلة للحكم على جودة الاختبارات الموضوعية وهي: هل تعكس أسئلة الاختبار أهداف المادة؟ هل يعتبر الاختبار عادلاً في ضوء ما تم تدريسه؟ هل يتم تنفيذ الاختبار في جو يسمح لجميع الطلاب فرصة إظهار تحصيلهم؟ هل يركز الاختبار على المواضيع الهامة؟ هل طوله مناسباً للزمن المتاح؟ هل فقرات الاختبار مميزة لمستويات الطلاب؟ وهل نتائجه ثابتة بحيث أن تكون قريبة من نتائج اختبار مماثل؟ وهل هي دقيقة؟ (How to Judge the Quality of an Objective Test)، n.d.) .

اختبار الصواب والخطأ:

يعرفه الدوسري (٢٠٠١) بأنه "يتكون هذا النوع من الأسئلة غالباً من جملة إخبارية يحدد الطالب صحتها أو خطئها" (ص ٢٩٧) ويأتي بصيغ متعددة: صيغة الصواب والخطأ، وصيغة الإيجاب والنفي، وصيغة الأسئلة العنقودية، وصيغة الجمل التصحيحية. وتقتصر مجالات استخدامه على قياس المستويات المعرفية البسيطة كالمعرفة والفهم، وهو يقيس قدرة المفحوص على تذكر الحقائق البسيطة و فهم

النظريات والتعرف على صحة العلاقات السببية والتمييز بين المسلمات والفرضيات والحقائق والآراء (الزيود وعليان، ٢٠٠٥).

مزايا و عيوب اختبار الصواب والخطأ:

يحظى بشعبية كبيرة لدى المعلمين لسهولة بناء فقراته وتصحيحه وتمثيله للمادة الدراسية، إلا انه لا يحظى باحترام العاملين في مجال القياس التربوي والنفسي (أبو لبدة،٢٠٠٥) من أبرز عيوبه قياسه لأهداف تربوية دنيا وغموض فقراته عند اقتطاعها من السياق التي وردت فيه، وسهولة الإجابة عليها مما لا يوفر بيانات عن الفروق الفردية بين المفحوصين ودرجة تقديرهم الحقيقية، كما يشجع على الحفظ عوضاً عن الفهم والتفكير، ويساهم في نجاح المفحوصين بسبب التخمين أو الصدفة نظراً لارتفاع نسب التخمين عند الإجابة إلى ٥٠٪ مما يؤدي إلى انخفاض نسبة ثباته، ويعتبر أقل الاختبارات الموضوعية ثباتاً (الزيود و عليان،٢٠٠٥) وقد يشجع المتعلم لتعلم ما هو غير محمود كالغش (الكبيسي، ٢٠٠٧) ويرى بعض المختصون في التقويم التربوي بأن عيوب اختبار الصواب والخطأ لا يمكن تفاديها وينصحون باستخدام نوع آخر من الأسئلة، وعند مقارنة أسئلة الصواب والخطأ بالأسئلة الموضوعية الأخرى نجد بأنها عرضة للتخمين واحتمالات الغش وقلة التمييز بين مستويات الطلاب، مقتصرة على قياس القدرة على الحفظ والتذكر فقط (الدوسري، ٢٠٠١).

معايير تقييم فقرات الصواب والخطأ:

تتفق معظم أدبيات التقويم التربوي على قواعد بناء فقرات الصواب والخطأ والتي أيضاً تعتبر معايير لمراجعة جودة بناء هذه الأسئلة. وهي تتضمن على المفهوم الذي تحتويه الفقرة يجب أن يكون إما صحيح أو خطأ، وأن تكون الفقرة إخبارية قصيرة وواضحة وبسيطة ومفهومة خالية من الغموض ودلائل الإجابة، وأن يتساوى عدد الفقرات الصحيحة مع الخاطئة وخلطها عشوائياً (حمدان، ٢٠٠١) ويجب أن تحتوي الفقرة على فكرة واحدة فقط وأن تخلو من كلمات التعميم مثل كل وجميع و دائماً وحتماً، والتخصيص مثل في بعض الأحيان وتحت ظروف معينة وربما، والتي توحي للمفحوص

بأن الفقرة صواب، ويجب أن ينسب الرأي ووجهة النظر في الفقرة إلى مصدرها كما أفاد جرونلند Gronlund، و يجب أن يستخدم النفي بحرص ويستبعد النفي المزدوج (أبو زينة، ١٩٩٢).

وكانت الإدارة العامة للقياس والتقويم بوزارة المعارف (التربية والتعليم حالياً) بالمملكة العربية السعودية في عام ١٤٢٢هـ قد وضعت كتيب معايير مراجعة أسئلة الاختبارات التحصيلية، نختص بذكر المتعلقة باختبار الصواب والخطأ وهي كالتالي:

يختبر معلومة مهمة – مرتبط بالمادة الدراسية – ليس منقولاً حرفياً من المادة التعليمية – يتجنب التعقيد اللغوي (صعوبته القرائية واللغوية مناسبة) – غير متحيز ثقافياً – مختصر بما لا يخل بوضوح الجزء الذي يجعل العبارة صحيحة أو خطأ – عبارة السؤال صحيحة أو خطأ تماماً – يقتصر على فكرة واحدة وإذا تعددت الأفكار فجميعها صحيحة أو جميعها خاطئة – صحة أو خطأ العبارة لا يعتمد على حرف غير بارز، أو كلمة قد لا تسترعي انتباه المفحوص أو كلمة و رد في تهجئتها خطأ – (في حالة وجود صيغة النفي) تبرز أداة النفي بشكل واضح – يتجنب النفي والنفي المزدوج – يتجنب استخدام كلمات مثل غالباً، نادراً، دائماً، أبداً، في كثير من الأحيان، إلى حد ما...الخ _ (في حالة الرأي أو الاعتقاد) ينسب الرأي أو الاعتقاد إلى صاحبه (أصحابه) (ص ص ١٢، ١٣).

ويضع جرونلند Gronlund (٢٠٠٦) قائمة لتقييم فقرات الصواب والخطأ وهي كالتالي:

١. هل هذا النوع من الفقرات ملائماً لقياس نتاج التعلم المطلوب؟
٢. هل محتوى الفقرة يطابق هدف تعلمي يجب قياسه؟
٣. هل تحتوى كل فقرة على فكرة رئيسة واحدة؟
٤. هل يمكن أن يحكم على كل عبارة دون شك بأنها إما صواب أو خطأ؟
٥. هل تم صياغة الجمل باختصار وبساطة ووضوح؟
٦. هل تم استخدام النفي بحرص واستبعاد النفي المزدوج؟

٧. هل تم نسب الآراء في الجمل إلى مرجع ما؟
٨. هل تخلو الجمل من الإشارات التي تدل على الإجابة؟
٩. هل يتساوى عدد الفقرات الصواب مع عدد الفقرات الخطأ؟
١٠. هل تم ترتيب الفقرات الصواب والفقرات الخطأ عشوائياً؟ (ص ١٠٠)

اختبار الاختيار من متعدد:

يعرفه الدوسري (٢٠٠١) "يتكون سؤال الاختيار من متعدد من جزأين: جذر وبدائل. ويصاغ الجذر عموماً على شكل سؤال على المفحوص أن يحدد إجابته من بين البدائل المعطاة، أو على شكل جمل ناقصة تكملتها واحد من البدائل" (ص٢٥٣) وهو أفضل الاختبارات الموضوعية وأكثرها شيوعاً، يستخدم لقياس قدرات متعددة كالتذكر والفهم والتطبيق والتحليل وغيرهم باستثناء تنظيم الأفكار وربطها والتعبير عنها، ويتم توظيفه عندما يكون لدى المعلم المهارة لبناء فقراته، ولتمثيل مفردات المادة وللتعرف بدقة عن على قدرة المفحوص وتحصيله، وعندما يكون عدد المفحوصين كبيراً، وعندما يقوم بالتصحيح آخرين، وللاستخدام عدة مرات (الكبيسي، ٢٠٠٧) كما يستخدم لقياس قدرة المفحوص على التعرف على العلاقات السببية والتزامن واستخراج الخطأ ومعرفة التشابه والفروق والقدرة على التقييم وبيان الأثر (غانم، ١٩٩٧) وتتعدد الصور التي يستخدم فيها هذا الاختبار فقد يظهر الجذر كجملة ناقصة أو على شكل سؤال أو قد تحتوي البدائل على أكثر من إجابة ويطلب من المفحوص استخراج أفضلها أو ترتيب محتوى بدائلها تبعاً لما طلب في الجذر (أبو علام، ١٩٨٧) ويتفق معه الدوسري ويضيف (٢٠٠١) فئة الأسئلة التعديلية وفئة الأسئلة المركبة وفئة الاختيار المتعدد كصور أخرى يظهر بها هذا الاختبار، ويشير إلى تفضيل الطلاب والمعلمين لهذا الاختبار وإمكانية التحكم بمستوى الصعوبة في فقراته.

مزايا و عيوب اختبار الاختيار من متعدد:

يتميز بانخفاض نسبة التخمين مقارنة باختبار الصواب والخطأ وقياس القدرات العقلية العليا وسهولة التصحيح (الدليم وآخرون، ١٩٩٧) كما يمتاز بتمثيله لمفردات

المادة وتوفر شروط الصدق والثبات وخلوها من الذاتية عند التصحيح (سيد، سالم، ٢٠٠٤) أما عيوبه يشير ماكملان McMillan (١٩٩٧) إلى طول المدة الزمنية للإجابة على فقراته مقارنة بالاختبارات الموضوعية الأخرى، و صعوبة كتابة فقراته خاصة الموهات أو البدائل حيث أن الكثير من المعلمين لا يجدون صعوبة في كتابة بديل أو اثنين ولكن الثالث والرابع يكونا ضعيفين مما يسهل على الطلاب استبعادهم وبالتالي لا يجدون صعوبة في تخمين الإجابة الصحيحة، كما يقوم الطلاب بالاعداد لاختبار الاختيار من متعدد عن طريق قراءة مادة الاختبار عدة مرات والتركيز للتعرف عليها، وبذلك يبذل جهد قليل لتذكر المعلومات، وهذه الطريقة للدراسة لا تلائم النظريات المعاصرة للتعلم. وعلى الرغم من صعوبة بناء اختبار الاختيار من متعدد إلا أنه اكتسب شهرة واسعة وإقبالاً كبيراً في الأوساط التعليمية وذلك لموضوعيته وسهولة تصحيحه. ويشير باركس Parkes (د.ت) إلى ملاحظة دوير Dwyer بأن التسليم بموضوعية اختبار الاختيار من متعدد ما هو إلا أسطورة، فقد يصاغ بصورة سيئة تجعل منه أقل موضوعية من الاختبار المقالي. ويرى الزيود وعليان (٢٠٠٥) أن من أبرز عيوب هذا الاختبار انه يتطلب " مهارة عالية وجهداً فنياً متميزاً وإلماماً كبيراً بتفاصيل محتوى المادة الدراسية المراد قياسها، وقدرة لغوية عالية" ص ١١٠. ويتفق معه الظاهر وآخرون (١٩٩١) بأن صياغة فقرات الاختيار من متعدد تحتاج إلى مهارة ودقة عاليين من قبل المعلم وبأنه مكلف مادياً حيث يتطلب ورقاً كثيراً، ويعيب عليه سهولة الغش مقارنة بالاختبارات المقالية ولوجود التخمين والصدفة دوراً فيه.

معايير تقييم فقرات الاختيار من متعدد:

إن معايير تقييم هذا الاختبار هي نفس الشروط التي يجب مراعاتها عند صياغة فقراته وهي: أن يحتوي جذر السؤال على مشكلة واضحة ومحددة ويخلو من الجمل المعترضة ودلائل تشير إلى الإجابة الصحيحة والنفي، أن تنسجم البدائل مع صياغة الجذر وترتب الإجابات عشوائياً وأن لا يربط فقرتين ببعضهما البعض بحيث لو أخطأ المفحوص بالأولى يخسر الثانية تلقائياً، أن يقلل من البدائل والموهات التي تحتوي على

عبارة كل ما ذكر صواب أو كل ما ذكر خطأ، ويقلل من جعل الإجابة الصحيحة من بديلين و أن يكون عدد البدائل بين ثلاثة وخمسة والتأكد من عدم مرادفة بديلين، وترتيب البدائل ترتيباً منطقياً (الشيخ، عبدالرحمن، عبدالحميد، ٢٠٠٤) ويلخص هوجان Hogan (٢٠٠٧) رأي البروفسور هالادينا وزملاؤه Haladyna في تعليمات بناء فقرات الاختيار من متعدد من حيث المحتوى والأشكال التي يظهر فيها الاختبار وتصميمه وكتابة فقراته، فمن ناحية المحتوى يجب أن تعكس كل فقرة محتوى محدداً يحتوي على عملية عقلية واحدة، وأن يكون محتوى الفقرة هاماً كما يجب تجنب الفقرات اللغزية، وقياس مستويات التعلم العليا، وعدم ربط فقرتين ببعضهما البعض، و تجنب التفصيل الدقيق والتعميمات، وكذلك تجنب الفقرات التي تحتوي على آراء، وجعل المفردات المستخدمة في صياغة الفقرات بسيطة، وفيما يتعلق بأشكال الاختبار يرى هالادينا استخدم السؤال المباشر وتكملة الجملة و اختيار الإجابة الأفضل وتنظيمها رأسياً عوضاً عن أفقياً، أما تصميم الفقرات فيجب مراجعتها وتصحيحها، ويجب استخدام جمل سليمة نحوياً وإملائياً وتقليل كمية القراءة التي يقرأها المفحوص، وعند كتابة الفقرات يجب أن يكون الجذر واضحاً محتوياً على الفكرة الرئيسة متجنباً الإطالة في الكلمات و استخدام العبارات المنفية، وفيما يتعلق بالبدائل يرى هالادينا بأن يتم عمل العديد من البدائل ثم استخدام ثلاثة فقط والتأكد من أن بديلاً واحداً فقط يمثل الإجابة الصحيحة، وينوه بأهمية تغيير مكان الإجابة الصحيحة في كل فقرة ووضع البدائل بتسلسل منطقي أو رقمي متجانس في المحتوى متقارب في الطول، كما ينوه إلى الحرص عند استخدام عبارة كل ما ذكر خطأ وتجنب عبارة كل ما ذكر صحيح كبديل وتجنب استخدام النفي فيها أو إشارات تدل على الإجابة مثل دائماً و أبداً وبالتأكيد، أو صياغة البديل كتكملة للجذر أو وجود بديل واضح بأنه الإجابة الصحيحة أو وجود إشارات في بديلين أو أكثر تدل على الإجابة أو وضع بدائل سيئة وغير منطقية، ويرى وضع بدائل منطقية تستحق التصديق و استخدام أخطاء الطلاب المتكررة كبدائل.

ويشتمل كتيب معايير مراجعة أسئلة الاختبارات التحصيلية الذي أعدته وزارة التربية والتعليم (١٤٢٢هـ) على معايير تقييم فقرات اختبار الاختيار من متعدد وهي كالتالي:

السؤال بشكل عام:

يختبر معلومة / هامة - مرتبط بالمادة الدراسية – ليس منقول حرفياً من المادة التعليمية – يتجنب التعقيد اللغوي(صعوبته القرائية واللغوية مناسبة) – غير متحيز ثقافياً – الفراغات(المسافات) بين المقدمة والبدائل نفسها مناسبة – (في مجموعة الأسئلة المبنية على مادة تعليمية مشتركة) مستقل عن غيره من أسئلة المجموعة – (في مجموعة الأسئلة المبنية على مادة تعليمية مشتركة) ترتيبه يتناسب مع موقع بياناته في المادة التعليمية المشتركة.

مقدمة السؤال:

تحدد المطلوب بوضوح – مختصرة بما لا يخل بوضوح المطلوب – تتلافى تكرار الكلمة / الكلمات في البدائل – تستخدم صيغة النفي فقط عندما تكون هي الأنسب – (في حالة وجود صيغة النفي) تبرز أداة النفي بشكل واضح.

بدائل السؤال بشكل عام:

جميعها إجابات محتملة للسؤال- تتفق نحوياً مع المقدمة- متشابهة في الطول- متشابهة في التركيب- (الأعداد، التواريخ، ...) فيها مرتبة تصاعدياً/تنازلياً_ لا تحتوي على إيحاءات(مؤشرات) للحل_ لا تحتوي على كلمات مكررة (فالكلمات المشتركة بين جميع البدائل يمكن وضعها في المقدمة)- عددها أربعة أو خمسة- مرتبة عمودياً.

الإجابة الصحيحة للسؤال:

متفق على أنها الإجابة الصحيحة الوحيدة أو الإجابة الأنسب- متجانسة مع البدائل الأخرى.

المموهات:

تعتمد على أخطاء شائعة- لا يمكن تفسيرها إجابات صحيحة (الإدارة العامة للقياس والتقويم، ١٤٢٢، ص٨ - ١١)

ويضع جرونلند Gronlund (٢٠٠٦) قائمة لتقييم فقرات الاختيار من متعدد وهي كالتالي:

١. هل هذا النوع من الفقرات ملائماً لقياس نتاج التعلم المطلوب؟
٢. هل محتوى الفقرة يطابق هدفا تعلمي يجب قياسه؟
٣. هل يمثل الجذر مشكلة واحدة بوضوح؟
٤. هل تم صياغة الجذر بطريقة بسيطة وواضحة؟
٥. هل تم صياغة الجذر بطريقة لا تتطلب إعادة كلماته في البدائل؟
٦. هل تم كتابة الجذر كعبارة إخبارية إذا أمكن؟
٧. إذا استخدمت كلمات منفية في الجذر هل تم إبرازها (بوضع خط تحتها أو تكبير حروفها)؟
٨. هل الإجابة المطلوبة بين البدائل هي الوحيدة أم أفضلها؟
٩. هل كل البدائل متجانسة نحويا وصيا غيا مع الجذر؟
١٠. هل تخلو البدائل من إشارات تدل على الإجابة؟
١١. هل المموهات تتسم بالتصديق لمن لم يدرس؟
١٢. لتجنب طول عبارة البديل التي توحي بالإجابة هل طول عبارة الإجابة الصحيحة مختلفة عن باقي الاختيارات؟
١٣. هل تم تجنب وضع "كل ما ذكر أعلاه" كبديل واستخدام " لاشيء مما ذكر أعلاه" عند الحاجة؟
١٤. هل تم تغيير مكان الإجابة الصحيحة بين الاختيارات في الاختبار؟
١٥. هل طريقة تنظيم الفقرات و صياغتها نحوياً يؤهل الاختبار للاستخدام بفعالية؟ (ص٩١).

وفي هذا الشأن طور مركز العلوم والرياضيات التعليمي التابع لكلية التربية بجامعة تكساس فرع أوستن عبر المشروع التعاوني الإقليمي بولاية تكساس Texas Regional Collaboratives (د.ت) قائمة لفحص فقرات اختبار الاختيار من متعدد تركز على أربعة مجالات رئيسة وهي: الارتباط ويقصد به ارتباط الفقرة بحادثة ما واتصالها بالعالم الواقعي ومناسبتها للمرحلة الدراسية، فحص جذر الفقرة وعلاقته المباشرة بالهدف وأهميته وقياسه لمستويات معرفية عليا ووضوحه وسلامته من الأخطاء النحوية والإملائية وخلوه من النفي والتحيز لجنس أو ثقافة معينة، فحص البدائل و يتعلق ذلك بوجود إجابة صحيحة واحدة وبدائل قابلة للتصديق مع وضوحها وتساوي طولها وصياغتها المركبة والنحوية وترتيبها منطقياً، وخلوها من الخدعة وتكرار المعلومات وخلوها من الإشارة للجذر أو وضع بديل عكس بديل آخر، وفحص المهارة والتكوين أو التصميم وفي ضوئها يتم فحص: الفائدة التي يحققها الطالب، وعدم دلالة الفقرات على الإجابة والدقة و مناسبة العناوين وسلامة تصنيفها وتناسق تكوين الفقرات، وفحص أخير للفقرات يحتوي على نقطتين: وهما مراجعة الفقرات بعد أكثر من يومين من المراجعة الأولى، ووجود أكثر من مراجع.

الدراسات السابقة:

فيما يلي استعراض لبعض الدراسات السابقة العربية والأجنبية في مجال تحليل الاختبارات التحصيلية ثم التعليق عليها وصياغة فروض الدراسة.

١. دراسة تجريبية لوليامز Williams (٢٠٠٦) بكلية الأعمال للدراسات العليا بجامعة كوينزلاند للتقنية في استراليا Queensland University of Technology (QUT) قام فريق من الباحثين بتجربة تهدف للكشف عن مدى قوة أختبار يقيس العمليات العقلية العليا من نوع الاختيار من متعدد يختلف عن اختبار الاختيار من متعدد التقليدي (Assertion Reasoning Questions) (ARQ) بحيث يحتوي السؤال على عبارتين تمثلان مفهومين قد يكونا مرتبطتين منطقياً أو غير

مرتبطتين ويكون عدد بدائله خمس يقوم المفحوص باختيار أحدهما، والبدائل في هذا الاختبار موحدة لجميع الأسئلة ويكون شكلها وتفسيرها كالتالي:

أ . صواب: صواب؛ التفسير: المفهومين صواب والربط المنطقي صحيح.

ب. صواب: صواب؛ التفسير: المفهومين صواب والربط المنطقي خطأ.

ج . صواب: خطأ؛ التفسير: المفهوم الأول صواب والثاني خطأ ولا ينتج عن ذلك ربط منطقي.

د . خطأ : صواب؛ التفسير: المفهوم الأول خطأ والثاني صواب ولا ينتج عن ذلك ربط منطقي.

هـ. خطأ: خطأ؛ التفسير: المفهومين خطأ ولا ينتج عن ذلك ربط منطقي.

ويهدف هذا الاختبار لمعرفة إمكانية بناء أسئلة تقيس عمليات عقليا عليا تعتبر عادة مطلب أساسي من طلاب الدراسات العليا، وتم تنفيذ التجربة من خلال الانترنت على طلاب في مادتين بقسم إدارة الأعمال لمرحلة الماجستير - معظمهم من الطلاب الأجانب - تم استفتائهم لاحقاً عن مدى قياس الاختبار للعمليات العقلية العليا، وأفرزت التجربة أن الاختبار يعتبر بديلاً جيداً لاختبار الاختيار من متعدد التقليدي واختبار الإجابات القصيرة، ومؤشراً جيداً لأداء الطلاب في الاختبار المقالي ولكنها أخفقت في الوصول إلى نتائج تخدم الهدف الرئيس للاختبار إلا وهو قياس التفكير العميق والعمليات العقلية العليا حيث من وجهة نظر المفحوصين يقيس الاختبار مهارتهم وتمكنهم من اللغة الانجليزية.

٢. دراسة عام ٢٠٠٦ لواجنر وهارفي Wagner & Harvey استخدم الباحثان نظرية بند الاستجابة (Item Response Theory) للتعرف على درجة التخمين في اختبار واتسون وجلاسر للتفكير الناقد (WGCTA) وتطوير اختبار (Wagner Assessment Test)يقيس التفكير الناقد ويركز على التقليل من نسبة التخمين الناجح لذوي التحصيل المتوسط والمتدني وتطبيق الاختبارين ومقارنة تحليل النتائج بينهما وكانت النتيجة لصالح الاختبار المطور الذي راعى زيادة بدائل الإجابة.

٣. دراسة الزهراني (١٤٢٤هـ) بعنوان "مدى توفر صدق المحتوى وشروط الصياغة والإخراج الجيدة لأسئلة اختبارات وزارة التربية والتعليم لمادة الرياضيات للثانوية العامة لطلاب (العلوم الطبيعية) للأعوام الدراسية ١٤١٦هـ- ١٤٢٢هـ دراسة تقويمية" أستهدف الباحث تصميم معيار وتوظيفه لتحليل أسئلة الاختبارات والتعرف على مدى تحقيق الاختبارات لشروط صدق المحتوى في ضوء المستويات المعرفية والأوزان النسبية لموضوعات المقرر الدراسي والمحتوى الرياضي والفروق بينها، والتعرف على مستوى صياغة الأهداف في ضوء شروط الصياغة من قبل المعلمين والمشرفين التربويين، وتقديم نموذج مقترح لاختبار مادة الرياضيات يتوفر فيه صدق المحتوى والصياغة والإخراج الجيد لورقة الاختبار. ومن أهم النتائج التي توصل لها الباحث انخفاض صدق محتوى الاختبارات وإيجاد فروق ذات دلالة إحصائية في صياغة الاختبار وإخراجه لصالح الصياغة لاختلاف الأعوام التي تشملها الدراسة، وإيجاد فروق ذات دلالة إحصائية بين المعلمين والمشرفين التربويين في إخراج أسئلة الاختبار فقط لصالح المعلمين ولم يجد فروق ذات دلالة إحصائية بينهما في صياغة الاختبار، وخرجت الدراسة بعدة توصيات من ضمنها نشر ثقافة إعداد الاختبارات التحصيلية بين العاملين في هذا المجال عبر الدورات وورش العمل والاستعانة بأكاديميين في الرياضيات وطرق التدريس والقياس والتقويم لوضع أسئلة وزارة التربية والتعليم.

٤. دراسة تحليلية تقويمية لأسئلة اختبارات الشهادة الثانوية العامة لمقررات اللغة العربية بقسميها العلمي والأدبي خلال الفترة ١٩٩٧- ٢٠٠١ بدولة قطر أجراها السليطي و تايه (٢٠٠٤) للتعرف على المستويات المعرفية - حسب تصنيف بلوم- ومستويات الفهم التي تقيسها أسئلة الاختبار، وتنوعها مابين الإنشائي والموضوعي، والفروق بين أسئلة اختبارات قسمي العلمي والأدبي، وقد أظهرت الدراسة أن الأسئلة التي تقيس التذكر كان لها الحظ الأكبر من ضمن المستويات المعرفية حيث بلغت نسبتها ٤٢و٩ ٪ في القسم العلمي و ٨، ٤٤٪ من عينة

القسم الأدبي، بينما بلغت نسبة الأسئلة التي تقيس مستويات الفهم ٨،٣٢ ٪ في القسم العلمي و ٢١و٢٥ ٪ في القسم الأدبي، كما أظهرت الدراسة قلة الأسئلة التي تعكس المجال الوجداني، والتركيز على الاختبارات المقالية وعدم الدقة في صياغة الأسئلة، واختتمت الدراسة بعدة توصيات من ضمنها تدريب المعلمين والمعلمات والموجهين والموجهات على كيفية إعداد الاختبارات التحصيلية.

٥. دراسة قدمها عمر وخريوش عام ٢٠٠٤ في مؤتمر النوعية في التعليم الجامعي الفلسطيني في جامعة القدس المفتوحة حول أثر الأسئلة الموضوعية في نجاح طلاب جامعة القدس المفتوحة، أظهرت الدراسة أن للأسئلة الموضوعية أثرا في نجاح الطلاب، وأن نسبة النجاح في الأسئلة الموضوعية أعلى منه في الأسئلة المقالية، كما أظهرت ارتفاع نسبة النجاح عندما يحتوي الاختبار على أكثر من ٢٠٪ من الأسئلة الموضوعية وتزداد نسبة النجاح بازدياد العلامات المخصصة للأسئلة الموضوعية.

٦. دراسة الدويش (٢٠٠٤) بعنوان "دراسة تحليلية لتطوير أسئلة اختبارات مادة الرياضيات في المرحلة المتوسطة في منطقة الرياض التعليمية في ضوء أهداف المنهج التعليمية" استهدف الباحث مدى تمثيل أسئلة الاختبار لأهداف المنهج التعليمية في المرحلة المتوسطة والأوزان النسبية للمستويات المعرفية التي تقيسها معتمداً على قائمة (ديلتز) في عملية التحليل ثم في إجراءات التطوير لاحقاً. وتوصل الباحث أن نسبة تمثيل أسئلة الاختبارات لأهداف المنهج التعليمية بصورة عامة ١٦،٣٨٪ وبلغت نسبة المستويات المعرفية الدنيا التي تقيسها أسئلة الاختبارات ٦٨،٨٤٪ بينما وصلت نسبة المستويات المعرفية المعقدة التي تقيسها أسئلة الاختبارات ٣٢،١٥٪ كما وصلت نسبة تحقيق أسئلة الاختبارات للأداءات المحددة في عينة الدراسة ١٠٠٪، وفي ضوء هذه النتائج وضع الباحث نموذجاً لكيفية وضع أسئلة اختبارات الرياضيات مستخدماً قائمة (ديلتز) المعرفية كمعيار بحيث يتوفر فيه تمثيل

المستويات المعرفية وتحديد ثقلها المئوي وتستخدم كدليل لمعلمي الرياضيات في المرحلة المتوسطة.

٧. دراسة الخضيري في عام ١٤٢٤هـ (٢٠٠٤) بعنوان "دراسة تقويمية ومقارنة لأسئلة اختبارات العلوم للصف الثاني المتوسط وأسئلة الكتاب المقرر في ضوء تصنيف بلوم للأهداف المعرفية في منطقة القصيم" استهدف الباحث تحديد نسبة المستويات المعرفية وعددها في أسئلة الاختبار وأسئلة الكتاب المقرر ونوعية الأسئلة في كلاهما ثم إجراء المقارنة بينهما. توصل الباحث أن تمثيل المستويات المعرفية في أسئلة الاختبار غير متناسباً حيث بلغت نسبة الأسئلة التي تقيس التذكر ٦١٪ والفهم ٢٥٪ بينما بلغ تمثيل المستويات العليا ١٤٪، وغير متوافقاً مع أسئلة الكتاب حيث بلغت نسبة التذكر ٤١٪ والفهم ٣٥٪ والمستويات العليا ٢٤٪ والتي قام الباحث بتحليلها لإجراء المقارنة، أما نوعية الأسئلة فقد أفرزت النتائج أن هناك فروق ذات دلالة إحصائية بين نوعية الأسئلة في الاختبار وأسئلة الكتاب حيث بلغت نسبة الأسئلة الموضوعية في الاختبار ٧٣٪ والمقالية ٢٧٪ بينما بلغت النسب في أسئلة الكتاب الموضوعية ٢٤٪ والمقالية ٧٦٪، ويوصي الباحث بإعداد قوائم تحتوي على الأهداف التربوية ونشرها بين المعلمين وتدريبهم على كيفية صياغة الأهداف وتحليل المحتوى وبناء جدول مواصفات الاختبارات، وإعداد جداول مواصفات لكل مادة وتزويد المعلمين به.

٨. دراسة العتيبي عام ١٤٢١هـ بعنوان "دراسة تقويمية ومقارنة بين أسئلة اختبار الفصل الأول والثاني لمادة الرياضيات للمرحلة الثانوية وفق تصنيف بلوم وأصناف المحتوى الرياضي للأعوام ١٤١٥هـ- ١٤١٩هـ بمدينة مكة المكرمة" استهدف الباحث الكشف عن ما تتضمنه أسئلة اختبارات معلمي الرياضيات و أسئلة اختبارات وزارة المعارف من محتوى رياضي ومستويات معرفية، وإيجاد الفروق بين ما تتضمنه أسئلة اختبارات الفصل الأول والفصل الثاني، وأفرزت الدراسة عدة نتائج من ضمنها اختلاف تمثيل المحتوى والمستويات المعرفية بين معلمي الرياضيات

والمشرفين التربويين في الوزارة، ووجود فروق ذات دلالة إحصائية بين أسئلة اختبارات الفصل الأول وأسئلة اختبارات الفصل الثاني تتعلق بأصناف المحتوى وبعض المستويات المعرفية، ويوصي الباحث بإجراء المزيد من الدراسات في هذا المجال وتدريب المعلمين على كيفية تطبيق مفاهيم مادة الرياضيات.

٩. دراسة الذوييي في عام ١٤٢٠هـ بعنوان "اتجاهات طلاب جامعة أم القرى نحو الاختبارات المقالية والاختبارات الموضوعية" استفتى الباحث اتجاه عينة من طلاب جامعة أم القرى لمرحلة البكالوريوس باستخدام مقياس الاتجاه للتعرف على اتجاهاتهم نحو الاختبارات المقالية والموضوعية باختلاف مستواهم الدراسي وتخصصاتهم وأظهرت النتائج فروق ذات دلالة إحصائية بين كل من:متوسط اتجاه الطلاب نحو الاختبار المقالي ومتوسط اتجاههم نحو الاختبار الموضوعي، وبين متوسط بعض البنود في مقياس الاتجاه نحو الاختبار المقالي وبين متوسط البند الذي يقابله في مقياس الاتجاه نحو الاختبار الموضوعي، وفي متوسط الاتجاه نحو الاختبار المقالي بين التخصص الأدبي وكلا التخصصين الشرعي والعلمي، كما لم توجد فروق ذات دلالة إحصائية في متوسط الاتجاه نحو الاختبارين بين طلاب المستويات الدنيا والعليا، وخرج الباحث بعدة توصيات من ضمنها حث العاملين في مجال الاختبارات بالجمع بين الاختبارات المقالية والموضوعية حسب ما تقتضيه طبيعة المادة الدراسية ورفع مستوى الاختبارات.

١٠. دراسة البكر عام ١٤١٨هـ استهدفت بناء قائمة معايير لأسئلة الاختبارات التحصيلية ثم توظيفها لتحليل أسئلة اختبارات مادة الفيزياء لطلاب وطالبات الصف الثالث ثانوي في المملكة العربية السعودية أظهرت النتائج تركيز الاختبار على الأسئلة المقالية المقيدة وإهمال الأسئلة الموضوعية، قلة درجة الارتباط بين نتائج الفصلين الأول والثاني، عدم وضوح أسلوب التقييم كما اهتمت الأسئلة بالمستويات المعرفية التذكر والفهم والتطبيق والتحليل وعمليات العلم الأساسية والتكاملية، كما أظهرت النتائج مراعاة الأسئلة للوزن النسبي للمواضيع

الدراسية ومناسبة وزن السؤال لزمن الإجابة عليه، واحتوائها على الظواهر الفيزيائية وتوظيف المعلومات في حل المشكلات ومناسبة صياغة الأسئلة ومراعاة المعايير الفنية للاختبار.

١١. دراسة محمود التحليلية عام ١٤١٦هـ لأسئلة الثانوية العامة في مادتي الحديث والثقافة الإسلامية والتوحيد في المملكة العربية السعودية في ضوء المستويات المعرفية وفق تصنيف بلوم من عام ١٤٠٥- ١٤١٠هـ أظهرت التركيز على قياس مستوى التذكر يليه مستوى الفهم وإهمال المستويات المعرفية العليا من تطبيق وتحليل وتركيب وتقويم ووصت بضرورة تدريب معلمي علوم التربية الإسلامية بكيفية إعداد الاختبارات وطرق التقويم.

١٢. دراسة الحميضي عام ١٤١٣هـ استهدفت تقويم الكفاءات التعليمية في إعداد وتصحيح وتحليل الاختبارات لدى عينة عشوائية من معلمات الدراسات الاجتماعية في المرحلة الثانوية بمدينة الرياض في المملكة العربية السعودية في ضوء المؤهل العلمي والتخصص وسنوات الخبرة التعليمية أظهرت النتائج قلة معرفة معلمات الدراسات الاجتماعية حول إعداد وتصحيح وتحليل الاختبارات ووجود فروق دالة إحصائياً لصالح المعلمات ذو بكالوريوس التربية في مدى المعرفة وإعداد وتصحيح وتحليل الاختبارات كما أظهرت النتائج وجود فروق دالة إحصائياً لصالح المعلمات ذو الخبرة في التدريس ١١ سنة فأكثر في صياغة الأهداف السلوكية وتصحيح وتصحيح جزئيات الاختبار وسرعة إعادة الأوراق و التعرف على ميول الطالبات من خلال نتائج الاختبار واستخدام الاختبار المقالي، ولم تظهر الدراسة أية فروق دالة إحصائياً للتخصص بين فئتي المعلمات، وخرجت الدراسة بعدة توصيات كان من أبرزها تدريب المعلمات على كيفية إعداد الاختبارات التحصيلية التي تتسم بصدق المحتوى وثبات النتائج.

تعليق على الدراسات السابقة:

ركزت دراسات الزهراني والسليطي وتايه و الدويش والخضيري والعتيبي ومحمود على تحليل محتوى الاختبارات في ضوء المستويات المعرفية حسب تصنيف بلوم وأشارت في توصياتها إلى ضرورة رفع كفاءة الاختبارات التحصيلية من خلال تدريب العاملين عليها من معلمين ومعلمات وانخراطهم في الدورات التدريبية وورش العمل، أما تمثيل محتوى الاختبارات للأهداف الدراسية الذي يشكل محورا رئيساً في الدراسة الحالية يظهر منخفضاً في نتائج دراستي الزهراني و الدويش، ولا يوجد دراسات عربية سابقة – حسب جهد الباحثة- تتعلق بتحليل أسئلة الاختبارات في المرحلة الجامعية أو تحليل أسئلة الصواب والخطأ والاختيار من متعدد، وعلى الرغم من أن دراستي ويليامز (Williams) و واجنر وهاري (Wagner & Harvey) تشيران إلى محاولات تجريبية للإرتقاء باختبار الاختيار من متعدد التقليدي لم تتوفر في الدراسات العربية إلا إنها تتفق معها في الحاجة لتطوير الاختبارات عامة و قياس عمليات عقلية عليا. وفي ضوء الإطار النظري والدراسات السابقة وما ورد من تعليق عليها تم صياغة فرضيات الدراسة كما يلي:

١. تباين في تمثيل الاختبارات الموضوعية للأهداف الدراسية على مستوى القسم.
٢. تباين في توفر شروط صياغة أسئلة الصواب والخطأ على مستوى الاختبار في القسم.
٣. تباين في توفر شروط صياغة أسئلة الاختيار من متعدد على مستوى الاختبار في القسم.
٤. يوجد فروق ذات دلالة إحصائية في نتائج تحليل الاختبارات الموضوعية بين الأقسام التعليمية.

منهجية وإجراءات الدراسة:

منهج الدراسة:

استخدمت الباحثة المنهج التحليلي الوصفي الذي يعتمد على الأسلوبين الكمي والكيفي في رصد وتحليل مشكلة الدراسة. ويشير عبيدات وعدس وعبد الحق (١٩٩٣) أن من المجالات التي حددها بيرلسون (Perelson) لاستخدام تحليل المحتوى "مطابقة المحتوى العلمي في الكتب الدراسية للأهداف التربوية المتوخاة، فمن خلال عملية تحليل المحتوى في ضوء الأهداف التربوية المختلفة يمكن أن يتضح أي الأهداف جرى التركيز عليها بصورة مقبولة وأيها لم يحظ بمثل هذا النوع من التركيز" (ص ١٧١) ومن خلال تحليل أسئلة الاختبارات الموضوعية بكلية التربية لإعداد معلمات الابتدائي بالرياض يمكن التعرف على مدى صلاحيتها و الوصول إلى التوصيات والمقترحات التي قد تساهم في تطويرها في المستقبل.

مجتمع الدراسة وأسلوب اختيار العينة:

تكون مجتمع الدراسة من الاختبارات الموضوعية في امتحانات الفصل الدراسي الأول من العام الجامعي ١٤٢٧- ١٤٢٨هـ لمرحلة البكالوريوس في كلية التربية لإعداد معلمات الابتدائي بمدينة الرياض. وللتعرف على مستوى تمثيل الاختبارات الموضوعية للأهداف التعليمية تم اختيار عينة الدراسة بالطريقة العشوائية الطبقية لدقة تمثيلها للأقسام التعليمية بمراحلها الدراسية الأربعة، وتشكلت العينة من ٣٢ اختبار من مجموع ١٢٧، لكل قسم ٨ اختبارات ولكل مرحلة دراسية اختبارين، وقد بلغ عدد أسئلة جميع الاختبارات ١٩٦٠ سؤال بينما بلغ عدد الأهداف التعليمية وفق الخطط الدراسية ١٩٥ هدف. ولتقييم الأسئلة الموضوعية تم أخذ عينة عشوائية طبقية من الأسئلة لنفس الأسباب المذكورة مسبقاً تشكلت من ٦٤٠ سؤال من أصل ١٩٦٠ موزعة بالتساوي بين نوعي أسئلة الصواب والخطأ والاختيار من متعدد.

أدوات الدراسة:

استخدمت الدراسة الأدوات التالية:

١. الاختبارات الموضوعية في الفصل الدراسي الأول من العام الجامعي ١٤٢٧/١٤٢٨هـ لجميع الفرق والأقسام بكلية التربية لإعداد معلمات الابتدائي بالرياض.

٢. الخطط الدراسية لمقررات الفصل الدراسي الأول التي استخدمت فيها الاختبارات الموضوعية.

٣. استمارات تحليل محتوى الاختبارات الموضوعية قامت الباحثة بإعدادها، وعددها ثلاثة وهي كالتالي:

استمارة (١): تمثل مقياس تقدير استخدم كمعيار لتحليل الأسئلة يصاحبه تفسير لقواعد التحليل، يحتوي على قائمة بالأهداف التعليمية للمقرر الدراسي يقابلها سلم تقديرخماسي تصاعدي، يهدف لتقدير مستوى صدق محتوى الاختبار ودرجة تمثيله لأهداف الخطة الدراسية عبر سلم التقدير اللفظي: منخفض – دون المتوسط- متوسط – فوق المتوسط– عالي. وتختلف قائمة الأهداف في الاستمارة تبعاً لمسمى المقرر الدراسي.

استمارة (٢): قائمة مراجعة أسئلة الصواب والخطأ موضحة في ملحق (أ) تحتوي على قائمة معايير لمراجعة أسئلة الصواب والخطأ أعدت مسبقا من قبل الإدارة العامة للقياس والتقويم (١٤٢٢هـ) تم اختيارها من بين المعايير التي ذكرت في الإطار النظري من الدراسة لشمولها لكافة المعايير وخضوعها لمراجعة إدارة القياس والتقويم بالوزارة، إضافة إلى عدم احتياجها للترجمة. وقد قامت الباحثة بوضعها في الإطار الذي يخدم الدراسة من خلال إعداد جدول ذو بعدين أفقي ورأسي، البعد الأول يعرض أرقام متسلسلة ترمز إلى فقرات عشوائية من الاختبار والبعد الآخر يحتوي على شروط بناء فقرات اختبار الصواب والخطأ ثم يتم مراجعة كل فقرة أو سؤال ليتم التأكد من توفر شروط كتابة الاختبار من عدمه. وقد قامت الباحثة بتعديل المعايير بما يخدم الدراسة وذلك بحذف المعايير الثلاثة الأولى من قائمة المراجعة وهم: "يختبر معلومة/

مهارة هامة" و "مرتبط بالمادة الدراسية" و "ليس منقولاً حرفياً من المادة التعليمية" وذلك لعدم اتصالها المباشر بصياغة السؤال فنياً والالتزام بشروط كتابته، أما المعيار الثالث فقد حذف لتعدد المراجع التي يستخدمها المدرس الجامعي مما يجعل التعرف على عبارة منقولة حرفياً أمراً صعباً.

استمارة (٣): قائمة مراجعة أسئلة الاختيار من متعدد موضحة في ملحق (ب) أعدت معاييرها مسبقاً من قبل الإدارة العامة للقياس والتقويم (١٤٢٢هـ) وهي مماثلة للاستمارة السابقة بفرق احتواء البعد الرأسي على شروط وقواعد بناء فقرات الاختيار من متعدد لرصد مدى توفر هذه الشروط من عدمه. وقد قامت الباحثة بتعديل القائمة من خلال استبعاد المعايير التي تتعلق بمجموعات الأسئلة المبنية على مادة تعليمية مشتركة لعدم وجود هذا النوع من الأسئلة، كما تم إضافة معيارين وهما "تتجنب وضع عبارة كل ما ذكر صواب / خطأ " و "تتجنب وضع عبارة تجمع بين بديلين".

وقد تم عرض الاستمارتين الثانية والثالثة على إدارة الجودة بالإدارة العامة للتقويم والجودة بوزارة التربية والتعليم (الإدارة العامة للقياس والتقويم سابقاُ) وأُخذت الموافقة على استخدام المعايير المعدة من قبلهم في الإطار الذي وضعته الباحثة.

صدق الاستمارات:

تم عرض الاستمارات على تسعة من المحكمين والمختصين في مجالي المناهج والتقويم التربوي للتأكد من الصدق الظاهري لأداة الدراسة والتعرف على مدى صلاحيتها لما أعدت من أجله، وقد اتفقوا على صلاحيتها للاستخدام، وقد كان لبعضهم ملاحظات فردية تتعلق بصياغة بعض الكلمات في قائمة مراجعة أسئلة اختباري الصواب والخطأ والاختيار من متعدد تم معالجتها وأخذها في الاعتبار. وقد بلغت نسبة الاتفاق بين المحكمين على صلاحية استخدام استمارة (١) و (٣) ٩٧٪ واستمارة (٢) ٩٨٪ وهي نسب مرتفعة دالة على صدق أدوات الدراسة.

إجراءات الدراسة:

تمت إجراءات الدراسة بالخطوات التالية:

١. التفاهم مع عميدة كلية التربية لإعداد معلمات الابتدائي مجتمع الدراسة بشأن التعاون وتوفير ما تتطلبه الدراسة من وثائق ومعلومات.
٢. جمع الخطط الدراسية و الاختبارات الموضوعية في امتحان الفصل الدراسي الأول وتصنيفها تبعاً للمراحل الدراسية والأقسام التعليمية.
٣. تصميم استمارات جمع البيانات وتحكيمها، تم توضيحه في أدوات الدراسة.
٤. الاستعانة بأعضاء هيئة تدريس في التخصصات العلمية: الدراسات الإسلامية واللغة العربية والعلوم الاجتماعية والعلوم والرياضيات والتربية وعلم النفس من مختلف الجامعات السعودية للمشاركة في تحليل الاختبارات الموضوعية في هذه المجالات والاتفاق معهم على طريقة التواصل وقواعد التحليل والإطار الزمني للدراسة، وقد بلغ عددهم ٢٩ عضو. وفي هذا الشأن ترى الجلبي (٢٠٠٥) أن يقوم الباحث بتوكيد مهارته في تحليل المحتوى بعرض مجموعة من العينات على ذوو الخبرة للتأكد من سلامتها، كما ترى استعانة الباحث بمن لهم خبرة في مجال التحليل.
٥. تحليل كل اختبار من الاختبارات الموضوعية ثلاث مرات، التحليل الأول من قبل الباحثة والتحليلين الآخرين من قبل محللين آخرين كل على حدة. ولزم التنويه بأن هناك من قام بتحليل أكثر من اختبار وذلك حسب التخصص.
٦. التحليل الإحصائي للبيانات واستخراج متوسط تقييم المحللين.

كيفية تحليل الاختبارات وحساب ثباتها:

تتدرج عملية التحليل بعدة خطوات، وقد أعدت الباحثة أرضية مشتركة لتنظيم وتوحيد نقاط التحليل موضحة في الخطوات ١ و ٢ و ٤، بينما توضح الخطوات ٣ و ٥ دور المحللين المشاركين إلى جانب الباحثة في عملية التحليل وهي كما يلي:

١. تحديد قواعد التحليل المتعلقة بتقدير درجة تمثيل الاختبارات الموضوعية للأهداف التعليمية للمقرر الدراسي.

٢. كتابة الأهداف الدراسية المدونة في الخطط الدراسية في استمارة (١) لجميع اختبارات عينة الدراسة التي بلغ عددها ٣٢ اختبار موضوعي لتوحيد الأهداف التعليمية المراد تقدير تمثيلها في الاختبارات.

٣. قراءة أسئلة الاختبارات بتمعن وتأني ومقارنتها بالأهداف الموثقة في الخطط الدراسية للتعرف على عدد الأهداف التي تمثلها، ووفق قواعد التحليل تقدير درجة تمثيلها عبر سلم التقدير الخماسي في استمارة (١)

٤. اختيار ٢٠ فقرة عشوائية من أسئلة الصواب والخطأ والاختيار من متعدد في كل اختبار وتظليلها باللون البرتقالي ليتم التعرف عليه بسهولة من قبل المحللين الآخرين.

٥. قراءة أسئلة الاختبارات التي تم اختيارها قراءة فاحصة ودقيقة ثم رصد مدى توفر شروط كتابتها في الاستمارات (٢) و (٣) بوضع إشارة √ عند توفرها أو إشارة X عند عدمه في الخانات التي تحمل أرقام الأسئلة.

٦. تم حساب معامل ثبات التحليل على عينة مقدارها ٢٥٪ من النتائج وفق معادلة هولستي (Holsti، 1969، p140)حيث كانت نسبة الاتفاق بين الباحثة والمحللين الآخرين ٦١٪ في تمثيل الاختبارات للأهداف السلوكية في الخطط الدراسية و٨١٪ في التقييم العام لأسئلة الصواب والخطأ و ٨٢٪ في التقييم العام لأسئلة الاختيار من متعدد وهي نسب تدل على ثبات التحليل.

الأساليب الإحصائية للدراسة:

استخدمت الباحثة المتوسطات والنسب المئوية، وتحليل التباين الأحادي (One Way ANOVA) لعقد المقارنات البعدية بين تحليل أسئلة الاختبارات الموضوعية في جميع الأقسام.

نتائج الدراسة:

أولاً: ما مستوى تمثيل أسئلة الاختبارات الموضوعية لمحتوى المقررات في ضوء الأهداف التدريسية الموثقة في الخطط الدراسية وفق آراء أصحاب الاختصاص؟

يوضح جدول (٢) متوسط درجة تمثيل أسئلة اختباري الصواب والخطأ والاختيار من متعدد لأهداف المقررات الموضحة في الخطط الدراسية لجميع المقررات التي شملتها عينة الدراسة في كل أقسام الكلية وفق سلم التقدير الخماسي الذي ورد في استمارة (١) مع العلم أن تفسير درجة تمثيل أسئلة الاختبار لأهداف المقرر هي كالتالي:

١,٠٠ منخفض- ٢,٠٠ دون المتوسط – ٣,٠٠ متوسط - ٤,٠٠ فوق المتوسط - ٥,٠٠ عالي

جدول (٢)

متوسط درجة تمثيل أسئلة الصواب والخطأ والاختيار من متعدد لأهداف المقررات في جميع الأقسام

القسم	اسم المقرر	متوسط درجة تمثيل الأسئلة لأهداف المقرر❖
القرآن الكريم والدراسات الإسلامية	التفسير	٣,٥٩
	الفقه	٣,٣٠
	مناهج المدرسة الابتدائية	٣,٤٣
	الحديث	١,٩٢
	علوم الحديث ونصوص منه	٢,٩٠
	النحو والصرف	٣,٦٧
	أصول الفقه	٣,٣٦
	علم نفس النمو	٣,٠٧
جميع مواد القسم		٣,١٥
العلوم والرياضيات	أساسيات علم التصنيف	٢,٦٥
	الحضارة الإسلامية	٣,٨٨
	حاضر العالم الإسلامي	٣,٦٣
	طرق تدريس خاصة	٣,٣٣
	فسيولوجيا عامة	٣,٢٨

تابع جدول (٢)

	حيوان عام	٣,٠٥
	ميكروبلوجيا عامة	٣,١٣
	كيمياء عامة	٣,٣٤
جميع مواد القسم		٣,٢٩
اللغة العربية والعلوم الاجتماعية	قاعة البحث والمكتبة العربية	٣,١١
	علم الأصوات	٤,٦٧
	نصوص أدبية من العصر الأندلسي	٤,٢٠
	جغرافية العالم الإسلامي	٣,٠٠
	مبادئ جغرافية اقتصادية	٣,٠٧
	جغرافية طبيعية	٣,٢٥
	مبادئ الإحصاء	٣,١٠
	النحو والصرف	٣,٧٢
جميع مواد القسم		٣,٥١
رياض الأطفال	تنمية مهارات الاتصال..	٢,٦٠
	تنمية المهارات اللغوية....	٣,٧٧
	التربية الحركية....	٣,٥٦
	القياس والتقويم....	٣,٢٢
	سيكولوجية رسوم الأطفال....	٤,٣٣
	مدخل إلى رياض الأطفال	٢,٢٤
	علم البيئة والتلوث	٢,٦٢
	سيكولوجية الابتكار....	٣,٠٠
جميع مواد القسم		٣,١٧

❖ المتوسط من ٥

أبرز ما يتضح من الجدول السابق ما يلي:

١. في قسم القرآن الكريم والدراسات الإسلامية بلغ متوسط درجة تمثيل أسئلة جميع اختبارات المقررات لأهدافها المدونة في الخطط الدراسية ٣,١٥ بتقدير "متوسط" وفق سلم التقدير الخماسي الموضح أعلاه. وبلغ متوسط تمثيل اختبار مقرر "الحديث" (١,٩٢) بتقدير "منخفض" وهي أدنى درجة لتمثيل أهداف مقرر في جميع الأقسام، بينما وصل متوسط مستوى تمثيل الاختبار لأهدافه في اختبار " علوم الحديث ونصوص منه" (٢,٩٠) أي بتقدير "دون المتوسط". واشتركت اختبارات مقررات القسم الباقية في تمثيل متوسط لأهدافها المدونة في الخطط الدراسية حيث تراوحت بين ٣,٠٧ – ٣,٦٧ بدرجة تمثيل متوسطة وفق سلم التقدير.

٢. في قسم العلوم والرياضيات بلغ متوسط تمثيل جميع اختبارات المقررات ٣,٢٩ بتقدير "متوسط"، وباستثناء اختبار "أساسيات علم التصنيف" الذي حصل على متوسط ٢,٦٥ وتقدير "دون المتوسط" في تمثيل الأهداف تراوح متوسط تمثيل باقي الاختبارات لأهدافها بين ٣,٠٥- ٣,٨٨ بتقدير "متوسط" في مستوى التمثيل. وتجدر الإشارة أن بعض الاختبارات في مواد التخصص احتوت على أسئلة موضوعية أخرى مثل المزاوجة والتكميلي وفقرات مقالية تم اعتبارها عند تقدير درجة التمثيل. كما يجب التنويه أن اختبارات الجزء العملي في مقررات التخصص فقط وهي كيمياء عامة وحيوان عام وأساسيات علم التصنيف وفسيولوجيا عامة وميكروبولوجيا عامة لم يتم تحليلها وبالتالي فإن التقدير الذي حصلت عليه هذه المقررات في درجة تمثيل أسئلة الاختبارات لأهداف الخطط الدراسية قد لا يمثل حقيقة واقعها.

٣. في قسم اللغة العربية والعلوم الاجتماعية بلغ متوسط تمثيل أسئلة جميع اختبارات المقررات لأهدافها المذكورة في الخطط الدراسية ٣,٥١ بتقدير "متوسط" وفق سلم التقدير. وحصل اختبار "علم الأصوات" على متوسط (٤,٦٧) واختبار "نصوص أدبية من العصر الأندل الخماسي"، توسط (٤,٢٠) مما يعني تقدير فوق المتوسط في

السلم الخماسي، بينما حصلت باقي الاختبارات على تمثيل بدرجة متوسطة لأهداف مقرراتها.

٤. في قسم رياض الأطفال بلغ المتوسط العام لتمثيل جميع الاختبارات لأهدافها ٣,١٧ بتقدير "متوسط" في مستوى التمثيل. وبلغ متوسط تمثيل اختبار المتوسط"، رسوم الأطفال" ٤,٣٣ مما يعني تقدير "فوق المتوسط"، كما بلغ متوسط تقدير الاختبارات " تنمية المهارات اللغوية.." و "التربية الحركية.." و "القياس والتقويم.." وسيكولوجية الابتكار.." بمستوى "متوسط" في سلم التقدير الخماسي، بينما بلغ متوسط تمثيل الأهداف لاختبار "تنمية مهارات الاتصال.."درجة ٢,٦٠ واختبار" علم البيئة والتلوث" درجة ٢,٦٢ مما يجعلهما أدنى الاختبارات في القسم تمثيلاً لأهداف مقرراتها بمستوى تقدير "دون المتوسط"

ويلخص الجدول التالي رقم (٣) متوسط ومستوى تمثيل أسئلة الاختبارات لأهداف المقررات الدراسية في جميع الأقسام، وتظهر جميعها متقاربة في درجة التمثيل و تشابهه في مستواه.

جدول (٣)

المتوسط الحسابي ❖ ومستوى تمثيل أسئلة اختباري الصواب والخطأ والاختيار من متعدد لأهداف المقررات الدراسية

القرآن الكريم والدراسات الإسلامية		العلوم والرياضيات		اللغة العربية والعلوم الاجتماعية		رياض الأطفال	
المتوسط	مستوى التمثيل	المتوسط	مستوى التمثيل	المتوسط	مستوى التمثيل	المتوسط	مستوى التمثيل
٣,١٥	متوسط	٣,٢٩	متوسط	٣,٥١	متوسط	٣,١٧	متوسط

❖المتوسط من ٥

ثانياً: ما مدى الالتزام بشروط بناء فقرات الصواب والخطأ عند إعداد الاختبارات وفق معايير تقييم أسئلة الصواب والخطأ في أدبيات التقويم التربوي وبشكل خاص المعايير المعدة من قبل وزارة التربية والتعليم؟

يوضح جدول (٤) المتوسط لمدى الالتزام بشروط بناء أسئلة الصواب والخطأ في جميع الأقسام تبعاً لمعايير مراجعة أسئلة الصواب والخطأ التي وردت في استمارة (٢)

جدول (٤)

متوسط مدى الالتزام بشروط بناء أسئلة الصواب والخطأ

في اختبارات جميع الأقسام

معايير مراجعة أسئلة الصواب والخطأ	القرآن الكريم و الدراسات الإسلامية	العلوم والرياضيات	اللغة العربية والعلوم الاجتماعية	رياض الأطفال	جميع المقررات
مختصر بما لا يخل بوضوح الجزء الذي يجعل العبارة صحيحة أو خطأ	٠,٩٠	٠,٩٣	٠,٨٦	٠,٩١	٠,٩٠
يتجنب التعقيد اللغوي	٠,٩٢	٠,٩١	٠,٩٤	٠,٩٥	٠,٩٣
عبارة السؤال صحيحة تماماً أو خاطئة تماماً	٠,٩٢	٠,٩٨	٠,٩٣	٠,٩٧	٠,٩٥
يقتصر على فكرة واحدة- وإذا تعددت الأفكار فجميعها صحيحة أو خطأ تماماً	٠,٩٣	٠,٩٥	٠,٩٥	٠,٩٧	٠,٩٥
صحة أو خطأ العبارة لا يعتمد على حرف أو كلمة قد لا تسترعي انتباه المفحوص	٠,٩٢	٠,٩٦	٠,٩٣	٠,٩٨	٠,٩٥

تابع جدول رقم (٤)

معايير مراجعة أسئلة الصواب والخطأ	القرآن الكريم و الدراسات الإسلامية	العلوم والرياضيات	اللغة العربية والعلوم الاجتماعية	رياض الأطفال	جميع المقررات
في حالة وجود صيغة النفي تبرز أداة النفي بشكل واضح	٠,٦٢	٠,٤٩	٠,٨٣	٠,٤٤	٠,٥٦
يتجنب النفي والنفي المزدوج	٠,٨٩	٠,٩٨	٠,٩٠	٠,٩١	٠,٩٢
يتجنب استخدام كلمات مثل غالباً، نادراً، دائماً، في كثير من الأحيان، إلى حد ما	٠,٩٥	٠,٩٧	٠,٩٦	٠,٩٤	٠,٩٦
في حالة الرأي أو الاعتقاد ينسب الرأي أو الاعتقاد إلى صاحبه (أصحابه)	٠,٨٠	٠,٦٧	٠,٦٧	٠,٤٠	٠,٦٣
غير متحيز ثقافياً	٠,٩٢	١,٠٠	١,٠٠	١,٠٠	٠,٩٨
التقييم العام لأسئلة الصواب والخطأ	٠,٨٩	٠,٩١	٠,٩١	٠,٨٥	٠,٨٩

من أبرز ما يدل عليه الجدول السابق ما يلي:

١. في قسم القرآن الكريم و الدراسات الإسلامية توفرت شروط صياغة فقرات الصواب والخطأ بمتوسط عام بلغ ٠,٨٩، حيث لم تتوفر بعض الشروط في بعض أسئلة اختبارات القسم مثل " في حالة وجود صيغة النفي تبرز أداة النفي بشكل واضح" حيث بلغ متوسط توفر هذا الشرط ٠,٦٢ و " في حالة الرأي أو الاعتقاد ينسب الرأي أو الاعتقاد إلى صاحبه (أصحابه) بلغ متوسط توفره ٠,٨٠ و "تجنب

النفي والنفي المزدوج" بلغ متوسط توفره ٠,٨٩ و انفرد هذا القسم عن باقي الأقسام بعدم خلو أسئلته من التحيز الثقافي حيث بلغ شرط " غير متحيز ثقافياً" متوسط ٠,٩٢ بينما خلت باقي الأقسام من هذا الشرط بمتوسط ١,٠٠ وتوفرت باقي الشروط بمتوسطات متقاربة بين ٠,٩٠ – ٠,٩٥

٢. في قسم العلوم والرياضيات توفرت شروط بناء فقرات الصواب والخطأ بمتوسط ٠,٩١ بشكل عام، وبشكل خاص بلغ متوسط توفر شرط " في حالة وجود صيغة النفي تبرز أداة النفي بشكل واضح" ٠,٤٩ بينما بلغ متوسط توفر شرط " في حالة الاعتقاد ينسب الرأي أو الاعتقاد إلى صاحبه ٠,٦٧ وتوفرت باقي الشروط بمتوسطات متفاوتة بين ٠,٩١- ١,٠٠

٣. في قسم اللغة العربية والعلوم الاجتماعية توفرت شروط الصياغة بشكل عام بمتوسط٠,٩١ وبشكل خاص وصل المتوسط ٠,٨٦ للشرط أو المعيار " مختصر بما لا يخل بوضوح الجزء الذي يجعل العبارة صحيحة أو خطأ" بينما توفر بند إظهار النفي بمتوسط ٠,٨٣ و بند إسناد الرأي أو الاعتقاد إلى صاحبه بمتوسط ٠,٦٧ كما بلغ في قسم العلوم والرياضيات. وتوفرت باقي الشروط بمتوسطات متفاوته بين ٠,٩٠ – ١,٠٠

٤. في قسم رياض الأطفال توفرت شروط صياغة أسئلة الصواب والخطأ بصورة عامة بمتوسط ٠,٨٥ وهو أدنى متوسط بين الأقسام، وبصورة خاصة بلغ متوسط توفر شرط إبراز النفي ٠,٤٤ بينما بلغ متوسط شرط إسناد الرأي لصاحبه ٠,٤٠ والمتوسطين الأخيرين يشكلان أدنى متوسط بين الأقسام لهذين الشرطين. وتوفرت باقي الشروط بمتوسطات متفاوتة بين ٠,٩١ – ١,٠٠

ثالثاً: ما مدى الالتزام بشروط بناء فقرات الاختيار من متعدد عند إعداد الاختبارات وفق معايير تقييم أسئلة الاختيار من متعدد في أدبيات التقويم التربوي وبشكل خاص المعدة من قبل وزارة التربية والتعليم؟

يوضح جدول (٥) متوسط مدى التزام أقسام الكلية بشروط بناء فقرات الاختيار من متعدد وفق معايير مراجعة أسئلة الاختيار من متعدد التي وردت في استمارة (٣)

جدول رقم (٥)

متوسط مدى توفر شروط بناء أسئلة الاختيار من متعدد في اختبارات جميع الأقسام

معايير مراجعة أسئلة الاختيار من متعدد		القسم				
		القرآن الكريم والدراسات الإسلامية	العلوم والرياضيات	اللغة العربية والعلوم الاجتماعية	رياض الأطفال	جميع المقررات
السؤال بشكل عام	يتجنب التعقيد اللغوي	٠,٩٣	٠,٩٥	٠,٩٨	٠,٩٧	٠,٩٦
	غير متحيز ثقافياً	٠,٩٥	١,٠٠	١,٠٠	١,٠٠	٠,٩٩
مقدمة السؤال	تحدد المطلوب بوضوح	٠,٩٢	٠,٩٦	٠,٩٢	٠,٩٦	٠,٩٤
	مختصرة بما لا يخل بوضوح المطلوب	٠,٩٣	٠,٩٠	٠,٩٥	٠,٩٥	٠,٩٣
	تتلافى تكرار الكلمة/الكلمات في البدائل	٠,٩٢	٠,٩٢	٠,٩٣	٠,٩٦	٠,٩٣
	تستخدم صيغة النفي فقط عندما تكون هي الأنسب	٠,٦٩	٠,٦٣	٠,٧٢	٠,٤٦	٠,٦٠
	في حالة وجود صيغة النفي تبرز أداة النفي بشكل واضح	٠,٦٧	٠,٦٣	٠,٤٣	٠,٢٧	٠,٤٥
بدائل السؤال بشكل عام	جميعها إجابات محتملة للسؤال	٠,٨٢	٠,٩٦	٠,٧٧	٠,٩٨	٠,٨٨
	تتفق نحوياً مع المقدمة	٠,٩٥	٠,٩٦	٠,٩٨	٠,٩٧	٠,٩٧
	متشابهة في الطول بقدر الإمكان	٠,٩٧	٠,٩٨	٠,٩٩	٠,٩٩	٠,٩٨
	متشابهة في التركيب	٠,٩٦	٠,٩٨	٠,٩٩	٠,٩٩	٠,٩٨

تابع الجدول رقم (٥)

معايير مراجعة أسئلة الاختيار من متعدد		القسم				
		القرآن الكريم والدراسات الإسلامية	العلوم والرياضيات	اللغة العربية والعلوم الاجتماعية	رياض الأطفال	جميع المقررات
	الأعداد، التواريخ، وما شابها مرتبة تصاعدياً/تنازلياً	٠,٧٥	٠,٣٣	٠,٧٠	٠,١٧	٠,٤٤
	لا تحتوي على إيحاءات (مؤشرات) للحل	٠,٩٤	٠,٩٥	٠,٩٥	٠,٩٦	٠,٩٥
	لا تحتوي على كلمات مكررة (مشتركة) بين جميع البدائل	٠,٨٣	٠,٨٩	٠,٩٠	٠,٨٩	٠,٨٧
	عددها أربعة أو خمسة	٠,٢٥	٠,٧١	٠,٠٠	٠,٢٧	٠,٢٩
	تتجنب وضع عبارة (كل ما ذكر صواب / خطأ) كبديل	٠,٩٢	٠,٩٥	٠,٩٢	٠,٧٥	٠,٨٩
	تتجنب وضع عبارة تجمع بين بديلين	٠,٨٧	٠,٩٤	٠,٩٥	٠,٩٦	٠,٩٣
الإجابة الصحيحة للسؤال	متفق على أنها الإجابة الصحيحة الوحيدة أو الأنسب	٠,٨٩	٠,٩٨	٠,٩٤	١,٠٠	٠,٩٥
	متجانسة مع البدائل الأخرى	٠,٩٤	٠,٩٩	٠,٩٩	٠,٩٩	٠,٩٨
الموهمات	تعتمد على أخطاء شائعة	٠,٥٨	٠,٨١	٠,٦٥	٠,٠٢	٠,٥١
	لا يمكن اعتبارها إجابات صحيحة من قبل المفحوص	٠,٨١	٠,٩٣	٠,٨٥	٠,٩٩	٠,٨٩

يتضح من الجدول السابق ما يلي:

١. في قسم القرآن الكريم والدراسات الإسلامية بلغ متوسط التقييم العام لأسئلة الاختيار من متعدد ٠,٨٤ بينما كان متوسط بند "عدد البدائل أربعة أو خمسة" منخفضاً بدرجة ٠,٢٥ كما بلغ متوسط البند "تعتمد –أي الموهات- على أخطاء شائعة" ٠,٥٨ بينما بلغ متوسط البندين "تستخدم صيغة النفي فقط عندما تكون هي الأنسب" و "في حالة وجود صيغة النفي تبرز أداة النفي بشكل واضح" ٠,٦٩ و ٠,٦٧ على التوالي. كما بلغ متوسط بند "ترتيب الأعداد و التواريخ وما شابه ترتيباً تصاعدياً أو تنازلياً " ٠,٧٥ كما يتضح أن خمس من الشروط أو البنود حصلت على متوسطات مابين ٠,٨١ و٠,٨٩ وهي شرط يتعلق بالموهات" لا يمكن اعتبارها إجابات صحيحة من قبل المفحوص" و شرطين مرتبطين بالبدائل "جميعها إجابات محتملة للسؤال" و " لا تحتوي على كلمات مكررة (مشتركة) بين جميع البدائل" و البند المرتبط بالإجابة الصحيحة "متفق على أنها الإجابة الصحيحة الوحيدة أو الأنسب " وحصلت باقي الشروط على متوسطات مرتفعة مابين ٠,٩٢ – ٠,٩٧

٢. في قسم العلوم والرياضيات ظهر متوسط التقييم العام لأسئلة الاختيار من متعدد مرتفعاً حيث بلغ ٠,٩٠. وظهر منخفضاً متوسط توفر الشرط المتعلق بترتيب الأعداد والتواريخ ترتيباً تصاعدياً أو تنازلياً حيث بلغ ٠,٣٣ كما بلغ متوسط توفر الشرطين المتعلقين باستخدام صيغة النفي عندما تكون الأفضل للاستخدام و إبرازها مماثلاً وهو ٠,٦٣ وبلغ متوسط توفر البند المرتبط بعدد البدائل أربعة أو خمسة ٠,٧١ وتوفر الشرطين المرتبطين باعتماد الموهات على أخطاء شائعة وباحتواء كلمات مكررة بين جميع البدائل بمتوسطين ٠,٨١ و ٠,٨٩ بالترتيب. أما باقي الشروط وعددها أربعة عشر اتسمت متوسطاتها بالارتفاع بين ٠,٩٠ – ١,٠٠.

٣. في قسم اللغة العربية والعلوم الاجتماعية بلغ متوسط التقييم العام لأسئلة الاختيار من متعدد ٠,٨٦ وقد توفر البند أو المعيار المتعلق بعدد البدائل في السؤال على متوسط ٠,٠٠ وبلغ متوسط إبراز أداة النفي بشكل واضح ٠,٤٣ واستخدامه –أي

النفي - عندما يكون الأنسب ٠,٧٢ كما بلغ متوسط اعتماد المموهات في السؤال على أخطاء شائعة ٠,٦٥ كما حصل كل من ترتيب البدائل ترتيباً منطقياً من حيث العدد والتواريخ وما شابه على ٠,٧٠ و كونها جميعها _ أي البدائل- إجابات محتملة على متوسط بلغ ٠,٨٥ باستثناء الشرط أو المعيار المتعلق بعدم اعتبار المموهات إجابات صحيحة من قبل المختبر والذي بلغ متوسطه ٠,٨٥ وتوفرت باقي الشروط بمتوسطات مرتفعة تتراوح بين ٠,٩٠ – ١,٠٠.

٤. في قسم رياض الأطفال بلغ متوسط التقييم العام لأسئلة الاختيار من متعدد ٠,٧٩ وهي أدنى درجة بين الأقسام. وبلغ متوسط اعتماد المموهات على أخطاء شائعة منخفض جداً فقد توفر بدرجة ٠,٠٢ أما ترتيب الأعداد والتواريخ وما شابه ترتيباً منطقياً فقد بدا منخفضاً أيضا بمتوسط ٠,١٧ بينما بلغ متوسط الشرطين إبراز أداة النفي في الفقرة و وصياغة أربعة أو خمس بدائل للسؤال منخفضاً كذلك حيث كان ٠,٢٧ كما بلغ متوسط استخدام صيغة النفي عندما يكون الأفضل ٠,٤٦ و تجنب وضع عبارة كل ما ذكر صواب/خطأ كبديل أو اختيار ٠,٧٥ وباستثناء البند " لا تحتوي على كلمات مكررة (مشتركة) بين جميع البدائل" الذي توفر بمتوسط ٠,٨٩ توفرت باقي الشروط بمتوسطات مرتفعة تتراوح بين ٠,٩٤ – ١,٠٠.

ويلخص الجدول (٦) التقييم العام لأسئلة الاختيار من متعدد ومقدمتها وبدائلها وإجاباتها الصحيحة ومموهاتها في جميع أقسام الكلية.

جدول (٦)

ملخص تقييم أسئلة الاختيار من متعدد في جميع الأقسام

التقييم العام	القرآن الكريم والدراسات الإسلامية	العلوم و الرياضيات	اللغة العربية والعلوم الاجتماعية	رياض الأطفال
السؤال بشكل عام	٠,٩٤	٠,٩٨	٠,٩٩	٠,٩٨
مقدمة السؤال	٠,٨٥	٠,٩١	٠,٨٩	٠,٧٢
بدائل السؤال	٠,٨٣	٠,٨٧	٠,٨٢	٠,٧٩
الإجابة الصحيحة	٠,٩١	٠,٩٨	٠,٩٦	٠,٩٩
المموهات	٠,٦٩	٠,٨٧	٠,٧٨	٠,٥٠

يبين الجدول السابق أن جميع الأقسام تحظى بدرجات مرتفعة فيما يتعلق بالسؤال بشكل عام والإجابة الصحيحة، أما مقدمة السؤال وبدائله والمموهات ينفرد قسم العلوم والرياضيات بأعلى درجة بينها. قسم رياض الأطفال يعتبر أقل الأقسام حظاً في توفر شروط بناء أسئلة الاختيار من متعدد، فعلى الرغم من حصوله على متوسط ٠,٩٩ بالشروط المتعلقة بالإجابة الصحيحة و ٠,٩٨ في السؤال بشكل عام إلا أنه يتخلف عن جميع الأقسام.

رابعاً: هل توجد فروق ذات دلالة إحصائية بين الأقسام التعليمية عند تحليل أسئلة الاختبارات الموضوعية؟

أ- إيجاد فروق ذات دلالة إحصائية بين الأقسام لدرجة تمثيل اختباري الصواب والخطأ والاختيار من متعدد لأهداف المقررات الدراسية.

جدول (٧)

تحليل التباين الأحادي لدلالة الفروق في درجة تمثيل الأسئلة لأهداف المقرر باختلاف القسم

مصدر التباين	مجموع المربعات	درجات الحرية	متوسط المربعات	قيمة ف	مستوى الدلالة	التعليق
بين المجموعات	٠,٦٨	٣	٠,٢٣	٠,٦٨	٠,٥٧١	غير دالة
داخل المجموعات	٩,٢٤	٢٨	٠,٣٣			

يتضح من الجدول (٧) أن قيمة (ف) غير دالة، مما يشير إلى عدم وجود فروق ذات دلالة إحصائية في درجة تمثيل الأسئلة لأهداف المقرر باختلاف القسم.

ب- إيجاد فروق ذات دلالة إحصائية بين الأقسام في التقييم العام لمعايير أسئلة الصواب والخطأ.

جدول (٨)

تحليل التباين الأحادي لدلالة الفروق في درجة التقييم العام لمعايير أسئلة الصواب والخطأ باختلاف القسم

مصدر التباين	مجموع المربعات	درجات الحرية	متوسط المربعات	قيمة ف	مستوى الدلالة	التعليق
بين المجموعات	٠,٠٢	٣	٠,٠١	١,١١	٠,٣٦١	غير دالة
داخل المجموعات	٠,١٦	٢٨	٠,٠١			

يتضح من الجدول (٨) أن قيمة (ف) غير دالة، مما يشير إلى عدم وجود فروق ذات دلالة إحصائية في درجة التقييم العام لمعايير أسئلة الصواب والخطأ باختلاف القسم.

جـ- إيجاد فروق ذات دلالة إحصائية بين الأقسام في درجة التقييم العام لأسئلة الاختيار من متعدد.

جدول (٩)

تحليل التباين الأحادي لدلالة الفروق في درجة التقييم العام لأسئلة الاختيار من متعدد باختلاف القسم

مصدر التباين	مجموع المربعات	درجات الحرية	متوسط المربعات	قيمة ف	مستوى الدلالة	التعليق
بين المجموعات	٠,٠٥	٣	٠,٠٢	٤,٨٧	٠,٠٠٨	دالة عند مستوى ٠,٠١
داخل المجموعات	٠,٠٩	٢٦	٠,٠٠			

يتضح من الجدول (٩) أن قيمة (ف) دالة عند مستوى ٠,٠١، مما يشير إلى وجود فروق ذات دلالة إحصائية في درجة التقييم العام لأسئلة الاختيار من متعدد باختلاف القسم. وباستخدام اختبار توكي للكشف عن مصدر تلك الفروق جدول (١٠) يبين أن هناك فروق دالة عند مستوى ٠,٠٥ في درجة التقييم العام لأسئلة الاختيار من متعدد بين قسم العلوم والرياضيات وقسم رياض الأطفال، وذلك لصالح قسم العلوم والرياضيات.

جدول (١٠)

اختبار توكي لتوضيح مصدر الفروق في درجة التقييم العام لأسئلة الاختيار من متعدد باختلاف القسم

القسم	المتوسط الحسابي	القرآن الكريم والدراسات الإسلامية	العلوم والرياضيات	اللغة العربية والعلوم الاجتماعية	رياض الأطفال	الفرق لصالح
القرآن الكريم والدراسات الإسلامية	٠,٨٤					
العلوم والرياضيات	٠,٩٠				❖	العلوم والرياضيات
اللغة العربية والعلوم الاجتماعية	٠,٨٦					
رياض الأطفال	٠,٧٩					

❖ تعني وجود فروق دالة عند مستوى ٠,٠٥

مناقشة النتائج:

أولاً: ما مستوى تمثيل أسئلة الاختبارات الموضوعية لمحتوى المقررات في ضوء الأهداف التدريسية الموثقة في الخطط الدراسية وفق آراء أصحاب الاختصاص ؟

وانبثقت منه الفرضية: تباين في درجة تمثيل الاختبارات الموضوعية للأهداف الدراسية على مستوى القسم.

تشير نتائج التحليل أن متوسط تمثيل الاختبارات الموضوعية للأهداف التعليمية الموثقة في الخطط الدراسية ظهر متبايناً على مستوى القسم مما يشير إلى تباين قدرات ومهارات واضعي الاختبارات في كل قسم، بينما ظهر متشابهاً في جميع الأقسامً

بتقدير "متوسط". ولا شك أن مفردات هذه الاختبارات لم تعكس تمثيلاً شاملاً لأهداف المقررات في الخطط الدراسية وبالتالي انعكس ذلك على صدق الاختبارات ودرجة قياسها لما وضعت لقياسه. وعلى الرغم من ما تتميز به الاختبارات الموضوعية من الشمول في أدبيات التقويم التربوي - مقارنة بأنواع أخرى من الاختبارات- إلا أنها في هذه الحالة لم تستثمر في تمثيل مفردات المنهج. كذلك إغفال أهمية الأعداد والتخطيط للاختبار من خلال تحليل محتوى المواضيع المقاسة يشكل عامل رئيس في قصور تمثيل الاختبارات لها كما يرى الدوسري (٢٠٠١) وأبولبدة (٢٠٠٥) والصراف (٢٠٠٢) وغيرهم.

وعلى الرغم من أن اختبارات منتصف الفصل الدراسي ومتطلبات دراسية أخرى كالبحوث والواجبات المنزلية والنشاطات الإضافية تقيس جزءًا من الأهداف إلا أن ذلك لا يشكل أكثر من ٣٠٪ من درجة الطالبة ويبقي ٧٠٪ للاختبار النهائي، يستثنى من ذلك المقررات التخصصية بقسم العلوم والرياضيات التي تسخر ٢٠٪ من درجة الطالبة للاختبارات العملية وتبقي ٥٠٪ من الدرجة للاختبارات النهائية وفق النظام المعمول به في الكلية مجتمع الدراسة. وربما يكون تواضع مستوى تمثيل الاختبارات للأهداف ناتجاً عن عدم مناسبة أساليب القياس المستخدمة لنتاج التحصيل. ويطرح جرونلند Gronlund (٢٠٠٦) سؤالاً هاما عن مدى مناسبة استخدام هذا النوع من الأسئلة لقياس نتاج التعلم. ويؤكد عودة (٢٠٠٢) أهمية اختيار الأسلوب الملائم لقياس الأهداف السلوكية، كما رأى الظاهر وآخرون (١٩٩١) أن طبيعة المحتوى المراد قياسه تؤثر في شكل الاختبار والتعبير عن واقع التحصيل. والسؤال الذي يطرح ذاته في هذه الدراسة هل استخدام اختباري الصواب والخطأ والاختيار من متعدد كان مناسباً لقياس نتاج التعلم في ٨٨٪ من المقررات الدراسية بالكلية؟ ربما لم يكونا الوسيلة الأنسب لقياس التحصيل في بعض أجزاء المقررات الدينية حيث لا يمكن خلوهما من التحيز الثقافي الذي يعتبر شرطاً هاماً من شروط هذين الاختبارين. وربما لم يكونا مناسبين لقياس نتاج التحصيل في أجزاء بعض المقررات التي تركز على التطبيقات العملية مثل مقررات علم

الأصوات وتنمية المهارات اللغوية والتربية الحركية أو بعض أجزاء المقررات التي تتطلب عرض الأفكار وتسلسلها والدفاع عنها مثل مقررات سيكولوجية الابتكار ونصوص أدبية من العصر الأندلسي وحاضر العالم الإسلامي. وقد يكون تواضع مستوى التمثيل ناتجاً عن عدم قياس الأهداف السلوكية المتعلقة بالعمليات العقلية العليا كالمقارنة والتحليل والتقييم لما تتطلبه من تدريب وخبرة بالمستويات المعرفية العليا وطرق قياسها مما أدى إلى إسقاط هذه الأهداف أو استبدالها بأهداف أخرى تعالج عمليات عقلية بسيطة. وقد ورد في الدراسات السابقة من هذه الدراسة نتائج عدة دراسات تم التركيز فيها على قياس العمليات العقلية الدنيا مثل دراسة محمود، والسليطي وتايه، والدويش، والخضيري وما يخلفه من آثار سلبية على العملية التعليمية. وقد يكون تواضع مستوى تمثيل الأهداف السلوكية في الاختبارات ناتجاً عن عدم اتسامها بالدقة والوضوح والمباشرة والشمول في الخطط الدراسية مما أدى إلى عدم دلالتها لنتاج التعلم وتعثر قابليتها للقياس كما يرى عريفج ومصلح (١٩٩٩) سمات الهدف السلوكي.

ثانياً: ما مدى الالتزام بشروط بناء فقرات الصواب والخطأ عند إعداد الاختبارات وفق معايير تقييم أسئلة الصواب والخطأ في أدبيات التقويم التربوي وبشكل خاص المعدة من قبل وزارة التربية والتعليم؟

وانبثقت منه الفرضية: تباين في توفر شروط صياغة أسئلة الصواب والخطأ على مستوى الاختبار في القسم.

أظهرت نتائج الدراسة تباين في توفر شروط صياغة أسئلة الصواب والخطأ على مستوى الاختبار في القسم، فهناك شروط تم أخذها بالاعتبار وشروط تم إغفالها، كما أظهرت النتائج تشابهاً بين الأقسام التعليمية في متوسطات التزامها بشروط الصياغة، وقد بلغ متوسط التقييم العام لجميع الأقسام نسبة ٨٩٪ وهي نسبة جيدة دالة على القدرة الفنية في إخراج هذا النوع من الأسئلة، ولكن هذا النوع من الاختبارات أقل الاختبارات الموضوعية ثباتاً، أي لا يمكن الوثوق بنتائجه خاصة وانه لم تطبق عليه معادلة إزالة أثر التخمين عند إجراء التصحيح في مجتمع الدراسة. وينوه أبو

لبدة (٢٠٠٥) بأنه يساعد على نجاح المفحوصين بسبب التخمين حيث تبلغ نسبة التخمين فيه ٥٠٪. كما يعتبره الدوسري (٢٠٠١) قاصراً على قياس القدرة على الحفظ والتذكر وعرضة لاحتمالات الغش بين الطلاب وقلة التمييز بين مستوياتهم.

ثالثاً: ما مدى الالتزام بشروط بناء فقرات الاختيار من متعدد عند إعداد الاختبارات وفق معايير تقييم أسئلة الاختيار من متعدد في أدبيات التقويم التربوي وبشكل خاص المعدة من قبل وزارة التربية والتعليم؟

وانبثقت منه الفرضية: تباين في توفر شروط صياغة أسئلة الاختيار من متعدد على مستوى الاختبار في القسم.

أظهرت نتائج الدراسة تباين في توفر شروط صياغة أسئلة الاختيار من متعدد على مستوى الاختبار في القسم الواحد، فهناك شروط تم إهمالها من قبل واضعي الاختبارات مما يشير إلى اختلاف مهاراتهم ومعارفهم عند بناء هذا النوع من الأسئلة. وينوه دوير Dwyer بأن صياغة أسئلة الاختيار من متعدد بصورة سيئة يجعلها أقل موضوعية من الاختبارات المقالية (Parkes, n.d.). ويتطلب إعداد هذا النوع من الأسئلة "مهارة عالية وجهداً فنياً متميزاً وإلماماً كبيراً بتفاصيل محتوى المادة الدراسية المراد قياسها، وقدرة لغوية عالية.." كما يرى الزيود وعليان (ص١١٠، ٢٠٠٥) على الرغم من توفر معظم شروط بناء فقرات الاختيار من متعدد التي تزيد من مؤشرات جودة الإعداد في جميع الأقسام باستثناء قسم رياض الأطفال إلا إنه تم إغفال بعض الشروط الهامة مثل إبراز أداة النفي واستخدامها عند الضرورة، و بناء مموهات قابلة للتصديق مما يؤدي إلى ارتفاع نسبة التخمين عند الإجابة وبالتالي التأثير سلباً على صدق النتائج. ويضع جرونلند Gronlund (٢٠٠٦) هذه المعايير في قائمة مراجعة أسئلة الاختيار من متعدد كمرجع للتأكد من جودة بناء مكوناته. ويظهر أن الهيئة التعليمية في الكلية متمثلة بأقسامها تضع فقرات موضوعية جيدة وربما يعود ذلك لتلقيها دورات في إعداد إختباري الصواب والخطأ والاختيار من متعدد عندما تم تفعيل برنامج الطالبات المنتسبات بالكلية الذي تطلب إعداد إختبارات موضوعية يمكن تصحيحها آلياً.

رابعاً: هل توجد فروق ذات دلالة إحصائية بين الأقسام التعليمية عند تحليل أسئلة الاختبارات الموضوعية؟

انبثقت منه الفرضية: يوجد فروق ذات دلالة إحصائية في نتائج تحليل الاختبارات الموضوعية بين الأقسام التعليمية.

على الرغم من تقارب الأقسام في معظم نتائج الدراسة إلا إنه وجدت فروق ذات دلالة إحصائية في التقييم العام لبناء أسئلة الاختيار من متعدد بين قسمي العلوم والرياضيات و رياض الأطفال لصالح الأول. وقد تميزت اختبارات قسم العلوم والرياضيات بأعلى نسبة في الالتزام بشروط بناء أسئلة الاختيار من متعدد حيث بلغت نسبة التقييم العام ٩٠٪ وقد يعود ذلك لطبيعة مقرراته التخصصية واحتوائها على مفردات وجزئيات تشكل حقلاً غنياً بالفقرات والبدائل إضافة إلى مهارة وخبرة واضعي أسئلة الاختبارات التي بدورها تشكل عاملاً هاماً في سلامة إعداد هذا النوع من الأسئلة. وانفرد قسم رياض الأطفال بأدنى نسبة في الالتزام بشروط أسئلة الاختيار من متعدد حيث بلغت ٧٩٪ وربما يعود ذلك لقلة مهارة واضعي الاختبارات أو لضيق الوقت الذي خُصص لبناء الأسئلة أو كلاهما.

التوصيات:

في ضوء النتائج السابقة توصي الباحثة بما يلي:

١. التأكيد على أهمية تحليل محتوى المقررات التي تدرس لضمان تمثيل مفردات الاختبارات للمحتوى المقاس.
٢. تطوير الخطط الدراسية للمقررات والتركيز على الأهداف السلوكية المشتقة من الأهداف العامة للمقرر ليتمكن أعضاء هيئة التدريس من قياسها بالأسلوب الأكثر مناسبة وفعالية.
٣. استخدام الأسئلة الموضوعية عندما تكون هي الأكثر مناسبة لطبيعة المحتوى المقاس.

٤. تعاون أعضاء هيئة التدريس ذو المهارة في كتابة الاختبارات الموضوعية مع آخرون يرغبون في تطوير مهاراتهم في هذا المجال.
٥. استخدام قوائم المراجعة (check-list) المتوفرة في أدبيات الاختبارات الموضوعية عند كتابة أسئلة الاختبار.
٦. حث الهيئة التعليمية على بناء أسئلة الاختبارات الموضوعية مبكراً، ربما بشهرين قبل بدء الاختبارات النهائية لما تتطلبه من مهارة ووقت وجهد في البناء والمراجعة.
٧. تحليل نتائج الاختبارات عبر البرامج المتوفرة مثل (Mega Start)(MINITAB) (SPSS) (Excel) ومناقشة إيجابياتها وسلبياتها في الأقسام التعليمية لتعميم الفائدة.
٨. إشراف أعضاء هيئة التدريس على الاختبارات الموضوعية التي توضع من قبل المعيدات لحداثة خبرتهن وقلة مهارتهن في هذا المجال.
٩. رفع كفاءة أعضاء هيئة التدريس في إعداد الاختبارات التحصيلية عامة والاختبارات الموضوعية خاصة من خلال الدورات وورش العمل.
١٠. إجراء دراسة تحليلية إحصائية لنتائج الاختبارات الموضوعية بقسم رياض الأطفال تسلط الضوء على معامل ثبات وصعوبة وتمييز الاختبارات الموضوعية لتوضيح الصورة عن واقع الاختبارات الموضوعية بهذا القسم.

المراجع:

المراجع العربية:

١. أبو حطب، فؤاد؛ عثمان، سيد؛ صادق، أمال. (٢٠٠٣). التقويم النفسي. (الطبعة الثالثة).
٢. القاهرة: مكتبة الأنجلو المصرية.
٣. أبوزينة، فريد.(١٩٩٢). أساسيات القياس والتقويم في التربية.الكويت:مكتبة الفلاح.
٤. أبوعلام، رجاء.(١٩٨٧). قياس وتقويم التحصيل الدراسي.الكويت:دار القلم للنشر والتوزيع.
٥. أبولبدة، سبع.(٢٠٠٥). مبادئ القياس النفسي والتقييم التربوي.(الطبعة الثالثة).عمان: جمعية
٦. عمال المطابع التعاونية.
٧. إحصائية الفصل الدراسي الأول للعام الجامعي ١٤٢٨/١٤٢٧هـ. (١٤٢٨). الرياض: كلية التربية لإعداد معلمات الابتدائي.
٨. الإدارة العامة للقياس والتقويم.(١٤٢٢هـ). معايير مراجعة أسئلة الاختبارات التحصيلية. الرياض: وزارة المعارف.(بدون رقم نشر)
٩. باشموس،سعيد؛ و خيري، السيد؛ و مهني، يحي.(١٩٨٥). التقويم التربوي.(الطبعة
١٠. الثانية). الرياض: مطابع البلاد.
١١. البكر، مشاعل.(١٤١٨). تقويم أسئلة اختبارات الفيزياء لشهادة الثانوية العامة في المملكة العربية السعودية دراسة تحليلية مقارنة لمضمون الأسئلة بنين وبنات. رسالة ماجستير غير منشورة. قسم
١٢. المناهج وطرق التدريس، كلية التربية، جامعة الملك سعود: الرياض.
١٣. بلوم، بنيامن؛ و هاستنجس، توماس؛ و مادوس، جورج.(١٩٨٣). تقييم تعلم الطالب التكويني
١٤. والتجميعي. (ترجمة المفتي، محمد؛ و النجار، زينب؛ و شلبي، أحمد).القاهرة:المركز الدولي للترجمة والنشر.
١٥. الجلبي، سوسن. (٢٠٠٥). أساسيات بناء الاختبارات والمقاييس النفسية والتربوية. دمشق: مؤسسة علاء الدين للنشر والتوزيع.
١٦. الحلوة، مصطفى. (١٩٩٠). التقويم في مراحل التعليم العام. الكويت: دار القلم للنشر والتوزيع.
١٧. حمدان، محمد. (٢٠٠١). تقييم التعلم والتحصيل كتاب يدوي للقياس والتقويم التربوي في سلسلة التربية الحديثة الرابعة والعشرون. الفيحاء:دار التربية الحديثة.
١٨. الحميضي، فاطمة. (١٤١٣). تقويم الكفاءات التعليمية في إعداد وتصحيح وتحليل الاختبارات لدى معلمات الدراسات الاجتماعية في المرحلة الثانوية بمدينة الرياض في المملكة العربية

السعودية. رسالة ماجستير غير منشورة. قسم المناهج وطرق التدريس، كلية التربية، جامعة الملك سعود: الرياض.

١٩. خضر، فخري. (١٩٨٧). التقويم التربوي. دبي:دار القلم للنشر والتوزيع.

٢٠. الخضيري، إبراهيم.(١٤٢٤). دراسة تقويمية ومقارنة لأسئلة اختبارات العلوم للصف الثاني المتوسط وأسئلة الكتاب المقرر في ضوء تصنيف بلوم للأهداف المعرفية في منطقة القصيم التعليمية. رسالة ماجستير غير منشورة. قسم علم النفس، كلية التربية، جامعة أم القرى: مكة المكرمة.

٢١. الدليم، فهد؛ وعبدالجواد، عبدالله؛ و عمران، محمد.(١٩٩٧) أسس ومفاهيم القياس والتقويم في مجال التربية.(الطبعة الثانية). الرياض: مطابع المجد.

٢٢. الدوسري، إبراهيم.(٢٠٠١). إطار مرجعي للتقويم التربوي.(الطبعة الثالثة).الرياض:مكتب التربية العربي لدول الخليج.

٢٣. الدويش، سليمان.(٢٠٠٤). دراسة تحليلية لتطوير أسئلة اختبارات مادة الرياضيات في المرحلة المتوسطة في منطقة الرياض التعليمية في ضوء أهداف المنهج التعليمية. مجلة كليات المعلمين. مجلد ٤(١). ١٤٠- ١٤٨.

٢٤. الذويبي، عبد الله. (١٤٢٠). اتجاهات طلاب جامعة أم القرى نحو الاختبارات المقالية والاختبارات الموضوعية (دراسة مقارنة). رسالة ماجستير غير منشورة. قسم علم النفس، كلية التربية، جامعة أم القرى: مكة المكرمة.

٢٥. الزيود، نادر؛ وعليان، هشام. (٢٠٠٥). مبادئ القياس والتقويم في التربية. (الطبعة الثالثة). عمّان: دار الفكر للطباعة والنشر والتوزيع.

٢٦. الزهراني، محمد.(١٤٢٤). مدى توفر صدق المحتوى وشروط الصياغة والإخراج الجيدة لأسئلة

٢٧. اختبارات وزارة التربية والتعليم لمادة الرياضيات للثانوية العامة لطلاب (العلوم الطبيعية)

٢٨. للأعوام الدراسية ١٤١٦هـ- ١٤٢٢هـ - دراسة تقويمية- . رسالة ماجستير غير منشورة.

٢٩. قسم علم النفس،كلية التربية، جامعة أم القرى: مكة المكرمة.

٣٠. السكران، راشد.(٢٠٠٢، ٢٩ ديسمبر). الرشودي:قبول ١١١ ألف طالبة انتظاماً وانتساباً والدراسة عن بعد ملائمة للطالبات.الرياض. تم استرجاعه في ٢٠٠٧/١٧/٢ على الرابط http://www.alriyadh.com/Contents//29-12-2002/Mainpage/LOCAL_4490.php

٣١. السليطي، حمده ؛ تايه، خضر.(٢٠٠٤). دراسة تحليلية تقويمية لأسئلة اختبارات الشهادة الثانوية العامة لمقررات اللغة العربية (القسمان:العلمي والأدبي) من عام ١٩٩٧- ٢٠٠١ بدولة قطر. رسالة الخليج العربي.٩٣، ٩٧- ١٣٣.

٣٢. سيد،علي؛ وسالم، أحمد.(٢٠٠٤). التقويم في المنظومة التربوية.(الطبعة الثانية). الرياض: مكتبة الرشد.

٣٣. شعلة، الجميل. (٢٠٠٠). التقويم التربوي للمنظومة التعليمية اتجاهات وتطلعات.القاهرة:دار الفكر العربي.

٣٤. الشيخ، تاج السر؛ و عبدالرحمن، نائل؛ و عبد الحميد، بثينة.(٢٠٠٤). القياس والتقويم التربوي. الرياض: مكتبة الرشد.

٣٥. الصراف، قاسم.(٢٠٠٢). القياس والتقويم في التربية والتعليم. الكويت: دار الكتاب الحديث.

٣٦. الظاهر، زكريا؛ و تمرجيان، جاكلين؛ و عبد الهادي، عزت.(محرر) (١٩٩١). مبادئ القياس والتقويم في التربية. عمان: مكتبة دار الثقافة للنشر والتوزيع.

٣٧. عبد السلام، فاروق؛ وطاهر، ميسرة؛ ومهني، يحي.(١٩٩٢). مدخل إلى القياس التربوي والنفسي. (الطبعة الثانية). مكة المكرمة: دار القافلة.

٣٨. عبد الهادي، نبيل.(٢٠٠٢). مدخل إلى القياس والتقويم التربوي.(الطبعة الثانية). عمان: دار وائل

٣٩. للنشر.

٤٠. عبيدات، زوقان؛ وعدس، عبدالرحمن؛ عبدالحق، كايد. (١٩٩٣). البحث العلمي مفهومه أدواته أساليبه. الرياض: دار أسامة للنشر والتوزيع.

٤١. العبيدي، غانم؛ والجبوري، حنان.(١٩٨١). أساسيات القياس والتقويم في التربية والتعليم. الرياض: دار العلوم للطباعة والنشر.

٤٢. العتيبي، سعود. (١٤٢١). دراسة تقويمية ومقارنة بين أسئلة اختبار الفصل الأول والثاني لمادة الرياضيات للمرحلة الثانوية وفق تصنيف بلوم وأصناف المحتوى الرياضي للأعوام١٤١٥هـ-١٤١٩هـ بمدينة مكة المكرمة. رسالة ماجستير غير منشورة.قسم علم النفس،

٤٣. كلية التربية، جامعة أم القرى: مكة المكرمة.

٤٤. عريفج، سامي؛ و مصلح، خالد.(١٩٩٩). في القياس والتقويم.(الطبعة الرابعة). عمان: دار مجدلاوي للنشر.

٤٥. علام، صلاح الدين. (٢٠٠٧). القياس والتقويم التربوي في العملية التدريسية. عمان: دار المسيرة للنشر والتوزيع.

٤٦. عمادة القبول والتسجيل. (١٤٢٧) بيان بأعداد الطالبات الموجودات على مقاعد الدراسة لعام ١٤٢٧/١٤٢٨هـ. الرياض: كليات البنات.

٤٧. عمر، فيصل ؛ خريوش، عبد الرءوف.(٢٠٠٤، يوليو). الأسئلة الموضوعية في الامتحانات في جامعة القدس المفتوحة وسبل تطويرها. بحث مقدم إلى مؤتمر التوعية في التعليم الجامعي الفلسطيني. رام الله. تم استرجاعه في ٢٠٠٧/٢٢/٢ على الرابط
http://www.qou.edu/homepage/arabic/qulitydepartment/quilityconference/papers/session5

٤٨. عودة، أحمد.(٢٠٠٢). القياس والتقويم في العملية التدريسية.(الطبعة الخامسة). إربد: دار الأمل.

٤٩. غانم، محمود.(١٩٩٧). القياس والتقويم. عمان: دار الأندلس للنشر والتوزيع.

٥٠. الغريب، رمزية.(١٩٨٨). التقويم والقياس النفسي والتربوي. القاهرة: مكتبة الأنجلو المصرية.

٥١. فتح الله، مندور.(٢٠٠٠). التقويم التربوي. الرياض: دار النشر الدولي.

٥٢. القحطاني،سالم؛ و العامري،أحمد؛ و آل مذهب، معدي؛ والعمر،دران. (٢٠٠٤). منهج البحث في العلوم السلوكية.(الطبعة الثانية).الرياض: العبيكان.

٥٣. القرني، ناصر.(محرر).(١٤٢٠هـ). دليل المعلمين والمعلمات في تقويم الاختبارات التحصيلية.

٥٤. الرياض.

٥٥. الكبيسي، عبد الواحد.(٢٠٠٧). القياس والتقويم تجديدات ومناقشات. الرياض:دار جرير للنشر والتوزيع.

٥٦. محمود، عبد الله.(١٤١٦). دراسة تحليلية لأسئلة الثانوية العامة في مادتي(الحديث والثقافة الإسلامية و(التوحيد) بالمملكة العربية السعودية. مجلة الجامعة الإسلامية بالمدينة المنورة.١٠١- ١٠٢،

٥٧. ٤٥٥- ٤٨٤. تم استرجاعه في ٢٠٠٧/١١/٢٧ على الرابط
http://www.edu.sa/magazine/101-102/6.htm

٥٨. ملحم،سامي. (٢٠٠٥). القياس والتقويم في التربية وعلم النفس.(الطبعة الثالثة). عمان: دار المسيرة للنشر والتوزيع والطباعة.

٥٩. منسي، حسن. (٢٠٠٢). التقويم التربوي. إربد: دار الكندي للنشر.

٦٠. منسي، محمود. (١٩٩٨). التقويم التربوي. الإسكندرية: دار المعرفة الجامعية.

٦١. منصور، عبد المجيد؛ و الشربيني، زكريا؛ و الحشاش، عبد اللطيف.(١٩٩٦). التقويم التربوي (الأسس والتطبيقات). القاهرة: دار الأمين للنشر والتوزيع.

٦٢. هاشم،كمال. (٢٠٠٦). التقويم التربوي مفهومه- أساليبه- مجالاته- توجهاته ألحديثه.الرياض: مكتبة الرشد.

٦٣. يوسف، ماهر؛ والرافعي، محب.(٢٠٠٥). التقويم التربوي أسسه وإجراءاته. (الطبعة الرابعة).

٦٤. الرياض: مكتبة الرشد.

٦٥. وكالة كليات البنات.(١٤٢٥هـ). دليل موجز لكليات البنات.الرياض:وزارة التربية والتعليم. (بدون رقم نشر).

المراجع الأجنبية:

66. Devine، M & Yaghlian، N. (n.d.) Construction of Objective Tests.
67. [Available online]. Retrieved Feb 4، 2007 from http://www.clt.cornell.edu/campus/teach/faculty/materials/ testconstruction/manual.pdf
68. Erinkson، S. (n.d.). Testing and Grading.[Available on line]. Retrieved Feb 3، 2007 from http://www.crlt.umich.edu/publinks/occ2.html
69. Evaluation and Examination Service.(n.d.). How to Judge the Quality of an Objective Test. The University of Iowa. [Available online]. Retrieved Feb 4، 2007 from http://www.uiowa.edu/~examserv/level_2/resources/technical%20 bulletins/tech%20bulletin%2006.pdf
70. Gronlund، N. (2006). Assessment of Student Achievement. 8th Ed. Boston: Pearson.
71. Guba،E & Lincoln، Y. (1992). Effective Evaluation:Improving the
72. Usefulness of Evaluation Results Through Responsive and
73. Naturalistic Approaches. Reprint Ed. San Francisco: Jossey Bass.
74. Hogan،T. (2007). Educational Assessment A Practical Introduction. NJ:
75. Wiley Jossey-Bass Education.
76. Holsti،O.R. (1969). Content Analysis For The Social Sciences Ana Humanities. MA: Addison-Wesley Publishing Company.
77. Lawrence، R. (1994). Questions To Ask When Evaluating Tests.
78. [Available on line]. Retrieved Nov 21، 2007 from
79. http://pareonline.net/getvn.asp?v=4&n=2
80. McMillan، J. (1997). Classroom Assessment:Principles and Practice for
81. Effective Instruction. Boston: Allyn and Bacon.
82. Nunnaly، J. (1972). Educational Measurement and Evaluation. 2nd Ed.
83. NY:McGraw-Hill.

84. Parkes، J. (n.d.) Multiple Choice Test. [Available on line]. Retrieved
85. Feb 4، 2007 from
86. http://www.flaguide.org/cat/mutiplechoicetest/multiple-choice-test7.php

87. Texas Regional Collaboratives.(n.d). Checklist for Reviewing Multiple
88. Choice Items. The University of Texas at Austin.[Available online] http://thetrc.org/download/tsds/checklist_for_reviewing%20items.doc
89. Wagner، T & Harvey، R. (2006). Development of a New Critical Thinking
90. Test Using Item Response Theory. Psychological Assessment، 18 (1)،100-105.
91. Williams، J. (2006). Assertion-reason multiple-choice testing as a tool for deep Learning: a qualitative analysis. Assessment & Evaluation in Higher Education، 31 (3)، 287-301.

الملاحق:

ملحق (أ)

استمارة (٢)

قائمة مراجعة أسئلة الصواب والخطأ

المقرر: الاسم:

الفرقة: التخصص:

القسم: الهاتف:

الدرجة العلمية:

تعليمات تعبئة الاستمارة:

مرفق اختبار المقرر، الرجاء مراجعة الأسئلة التي تم تظليلها باللون البرتقالي فقط، و فضلاً - في الاستمارة - وضع إشارة √ عند توفر بنود قائمة المراجعة، أو وضع إشارة X عند عدم توفرها، أو تركها فارغة عند عدم وجود هذا النوع من الأسئلة وذلك لكل سؤال في الخانة التي تحمل رقم السؤال. ثم تعبئة المعلومات الخاصة بك في أعلى يسار الاستمارة في حال الرجوع إليكم.

العدد	قائمة مراجعة أسئلة الصواب والخطأ	أرقام الأسئلة									
		١	٢	٣	٤	٥	٦	٧	٨	٩	١٠
١	مختصر بما لا يخل بوضوح الجزء الذي يجعل العبارة صحيحة أو خطأ										
٢	يتجنب التعقيد اللغوي										
٣	عبارة السؤال صحيحة تماماً أو خاطئة تماماً										

العدد	قائمة مراجعة أسئلة الصواب والخطأ	أرقام الأسئلة									
		١	٢	٣	٤	٥	٦	٧	٨	٩	١٠
٤	يقتصر على فكرة واحدة- وإذا تعددت الأفكار فجميعها صحيحة أو خطأ تماماً										
٥	صحة أو خطأ العبارة لا يعتمد على حرف أو كلمة قد لا تسترعي انتباه المفحوص										
٦	في حالة وجود صيغة النفي تبرز أداة النفي بشكل واضح										
٧	يتجنب النفي والنفي المزدوج										
٨	يتجنب استخدام كلمات مثل غالباً، نادراً، دائماً، في كثير من الأحيان، إلى حد ما										
٩	في حالة الرأي أو الاعتقاد ينسب الرأي أو الاعتقاد إلى صاحبه (أصحابه)										
١٠	غير متحيز ثقافياً										

مصدر القائمة: معايير مراجعة أسئلة الاختبارات التحصيلية، إعداد/الإدارة العامة للقياس والتقويم بوزارة المعارف (التربية والتعليم حاليا)عام ١٤٢٢هـ.

ملحق (جـ)

استمارة (٣)

قائمة مراجعة أسئلة الاختيار من متعدد

المقرر: الاسم:

الفرقة: التخصص:

القسم: الهاتف:

الدرجة العلمية:

تعليمات تعبئة الاستمارة:

مرفق اختبار المقرر، الرجاء مراجعة الأسئلة التي تم تظليلها باللون البرتقالي فقط، و فضلاً - في الاستمارة- وضع إشارة √ عند توفر بنود قائمة المراجعة، أو وضع إشارة X عند عدم توفرها، أو تركها فارغة عند عدم وجود هذا النوع من الأسئلة وذلك لكل سؤال في الخانة التي تحمل رقم السؤال. ثم تعبئة المعلومات الخاصة بك في أعلى يسار الاستمارة في حال الرجوع إليكم.

	العدد	قائمة مراجعة أسئلة الاختيار من متعدد	أرقام الأسئلة									
			١	٢	٣	٤	٥	٦	٧	٨	٩	١٠
أولاً: السؤال بشكل عام	١	يتجنب التعقيد اللغوي.										
	٢	غير متحيز ثقافياً.										
ثانياً: مقدمة السؤال	٣	تحدد المطلوب بوضوح.										
	٤	مختصرة بما لا يخل بوضوح المطلوب.										

	العدد	قائمة مراجعة أسئلة الاختيار من متعدد	أرقام الأسئلة									
			١	٢	٣	٤	٥	٦	٧	٨	٩	١٠
	٥	تتلافى تكرار الكلمة/الكلمات في البدائل.										
	٦	تستخدم صيغة النفي فقط عندما تكون هي الأنسب.										
	٧	في حالة وجود صيغة النفي تبرز أداة النفي بشكل واضح.										
ثالثا: بدائل السؤال بشكل عام	٨	جميعها إجابات محتملة للسؤال.										
	٩	تتفق نحوياً مع المقدمة.										
	١٠	متشابهة في الطول بقدر الإمكان.										
	١١	متشابهة في التركيب.										
	١٢	الأعداد، التواريخ، وما شابها مرتبة تصاعدياً/تنازلياً.										
	١٣	لا تحتوي على إيحاءات (مؤشرات) للحل.										
	١٤	لاحتوي على كلمات مكررة(مشتركة) بين جميع البدائل.										
	١٥	عددها أربعة أو خمسة.										
	١٦	تتجنب وضع عبارة (كل ما ذكر صواب / خطأ) كبديل										

	العدد	قائمة مراجعة أسئلة الاختيار من متعدد	أرقام الأسئلة									
			١	٢	٣	٤	٥	٦	٧	٨	٩	١٠
	١٧	تتجنب وضع عبارة تجمع بين بديلين.										
رابعا: الإجابة الصحيحة للسؤال	١٨	متفق على أنها الإجابة الصحيحة الوحيدة أو الأنسب.										
	١٩	متجانسة مع البدائل الأخرى.										
خامسا: الموهات	٢٠	تعتمد على أخطاء شائعة.										
	٢١	لا يمكن اعتبارها إجابات صحيحة من قبل المفحوص.										

مصدر القائمة: الإدارة العامة للقياس والتقويم. (١٤٢٢). معايير مراجعة أسئلة الاختبارات التحصيلية.

الرياض: وزارة المعارف (التربية والتعليم حاليا). (بدون رقم نشر)

Objective Tests of Some Undergraduate Courses at the College of Education for Elementary Teachers Preparation: An Analytical Study

Sana A. Almansour

College of Education, Princess Noura Bint Abdelrahman University (PNU), Riyadh, Saudi Arabia

Abstract:

This study was aimed at analyzing objective test items that was employed as a testing measure in the final examination of undergraduate courses in the College of Education for Elementary Teachers Preparation in Riyadh. The study focused on three major areas، first: the degree of representation – in light of course syllabus – to course targets based on the opinion of experts in the subject matter،second: the extent of compliance to construction terms of true-false and multiple choice test items in accordance with the criteria in the educational evaluation literature with special emphasis on criteria of test review drafted by the Ministry of Education in Saudi Arabia، third: to find out –if there is any- statistical differences regarding analysis results between departments' tests. Sample of the study constituted of 32 objective tests out of 127 representing college departments and four levels of schooling. The researcher used 3 forms to place test analysis، 29 faculties from different disciplines participated in the process، each test in the sample was analyzed separately by the researcher and two others specialized in the subject matter of the test. The most significant results of the study have shown that the degree of representation of tests items to course objectives was not inclusive and has gotten a "medium" rating. Study results have also shown statistical differences in compliance to multiple choice items construction terms between Science and Mathematics department and Kindergarten department for the benefit of the first one.